Causal Inference and Omitted Variable Bias in Financial Aid Research: Assessing Solutions

Stephanie Riegg Cellini

In the United States, the real cost of a college education has climbed almost 30% in the past 10 years and shows no sign of stabilizing in the near future (U.S. Department of Education, 2006a, Table 312). In this environment, student financial aid is perhaps the most important policy tool for maintaining and increasing access to postsecondary education among low- and middle-income students. With more than $13 billion spent on the federal Pell Grant program alone in 2004–2005 (U.S. Department of Education, 2006b)—not to mention the many other federal, state, local, and private aid programs—an accurate assessment of the causal impact of financial aid on college-going is a crucial consideration for college administrators and policymakers at all levels of government.

Researchers in the social sciences have analyzed the price sensitivity of college students from many different angles using a wide variety of methods. This article addresses an important challenge in the quantitative literature on

Acknowledgements: I gratefully acknowledge comments and advice from Latika Chaudhary, Dylan Conger, Amaury Nora, Edward St. John, two anonymous referees, and participants in the Lumina Conference on Barriers to Financial Aid and Access. I thank Phillip Stegner for excellent research assistance.
this topic—determining the causal effect of financial aid on student access to college. While simple ordinary least squares estimates of the impact of aid on college-going can reveal a correlation between financial aid policies and enrollment, these estimates are likely to suffer from omitted variable bias due to self-selection, potentially overestimating or underestimating the causal impact of these policies on enrollment.

I draw on recent quantitative studies of the impact of student financial aid on college-going to illustrate the problem of omitted variable bias and highlight the most promising solutions. In so doing, I seek to bridge the economics and education literature. While economists have spent several decades developing and implementing solutions to the problems of omitted variable bias in education, researchers in the field of higher education have only more recently begun to acknowledge the problem, and few have adopted convincing alternate estimation strategies to address it.

In this article, I first discuss the problem of omitted variables and the bias they create. I then explain and assess the most common techniques used to control for this problem in the economics and education literature on financial aid: random assignment, multivariate regression, proxy variables, fixed effects, difference-in-differences, regression discontinuity, and instrumental variables. In contrast to traditional econometrics texts, my focus is on the intuition, assumptions, and applications of each method. The approach is novel in that I examine each method in the context of the same question: Does financial aid increase college enrollment? While this strategy allows for straightforward comparisons between methods, it is worth noting that the same methods may be applied to any number of other research questions that involve similar problems of self-selection and omitted variable bias. I conclude with recommendations for future research, highlighting the methodological approaches that hold the most promise in overcoming omitted variable bias to more accurately measure the causal impact of student financial aid on access to education.

---

1For reviews of this literature, see Angrist & Kreuger, 1999; Card, 1999.
2See, for example, Alon, 2005; Becker, 2003; Curs & Singell, 2002; DesJardins, Alburg, & McCall, 2006; Dowd & Coury, 2006; Dowd, forthcoming.
3Though I draw on many studies of financial aid policy in both the economics and higher education literature, this article is not intended to be a comprehensive review of the literature. Rather, the goal is to use selected articles to illustrate salient features of the methods used to address omitted variable bias. For a more detailed review of this literature, see Dowd (forthcoming).
4For more detailed explanations of these methods, see, for example, Angrist & Krueger, 1999; Becker, 2003; Greene, 2000; Meyer, 1995; Shadish, Cook, & Campbell, 2002; Wooldridge, 2002.
Despite an extensive literature on the impact of financial aid and tuition on college enrollment, few studies can truly estimate the causal effect of a particular policy. The reason is what economists generally refer to as endogeneity: the idea that change in the variable of interest comes from within the system or model under study, rather than from outside factors. Two potential sources of endogeneity create bias in traditional linear ordinary least squares estimates of the causal impact of financial aid policy—namely, simultaneity and omitted variable bias. Simultaneity, or reverse causality, occurs when the independent variable of interest and the outcome are determined jointly or at the same time, making the direction of causality unclear. DesJardins, Ahlborg, and McCall (2006) and Curs and Singell (2002) provide excellent discussions of the problem of simultaneity in application, enrollment, and financial aid receipt. This article addresses the second and equally important source of endogeneity in financial aid research—omitted variable bias.

The problem of omitted variable bias arises because states that implement certain financial aid policies and students who receive financial aid are typically self-selected. The idea of self-selection is that individuals, states, or other entities make conscious choices about whether to adopt a policy, apply for aid, go to college, and so forth. The basis for these decisions may be related to characteristics that we can observe, but they may also be based on characteristics that we cannot observe—a problem economists refer to as unobserved heterogeneity.

In studies that use states as the unit of observation, whether a state adopts a particular financial aid policy is likely to be related to other policies and characteristics of that state, for example, how many low-income students are in the state or how many colleges are within its borders. Similarly, in studies that focus on students, variation in the receipt of financial aid will undoubtedly be driven by factors such as financial need, academic achievement, and the student’s knowledge of aid programs. In both cases, financial aid policy and receipt are considered endogenous variables because changes in these variables are driven by some of the same factors that also influence enrollment—the outcome of interest in the model. In both cases, if any factors that are correlated with both financial aid and enrollment are left out of a model, simple cross-sectional ordinary least squares (OLS) estimates will be biased.

To look more closely at the first situation—where we take states as the unit of analysis—consider the following equation of interest:

\[
Enroll_s = \alpha + \beta F_s + \epsilon_s
\]  

Where \( Enroll_s \) is the number of students enrolled in four-year colleges in state \( s \), \( F_s \) is a binary variable (also known as an indicator or dummy vari-
able) that equals 1 if the state has a particular financial aid policy in place (e.g., a merit aid program), and 0 otherwise. For simplicity and without loss of generality, I use financial aid as the endogenous regressor (typically denoted $X$ in textbooks) and enrollment as the outcome of interest (typically denoted $Y$) throughout.

If one were to perform univariate OLS regression analysis (or simply calculate a correlation coefficient), our estimate of the impact of financial aid on enrollment would be determined as follows:

$$
\hat{\beta} = \frac{\text{cov}(F, \text{Enroll})}{\text{var}(F)}.
$$

(2)

Substituting in for $\text{Enroll}$ from equation (1) and simplifying, the equation becomes:

$$
\hat{\beta} = \frac{\text{cov}(F, (\alpha + \beta F + \epsilon))}{\text{var}(F)} = \beta \frac{\text{var}(F)}{\text{var}(F)} + \frac{\text{cov}(\epsilon, F)}{\text{var}(F)} = \beta + \frac{\text{cov}(\epsilon, F)}{\text{var}(F)}.
$$

(3)

As equation (3) reveals, the estimated $\hat{\beta}$, equals the “true” effect ($\beta$) plus a potential bias term ($\text{cov}(\epsilon, F)/\text{var}(F)$). If this last term is not equal to 0 (that is, if there is some correlation between the error term and the policy), the estimator will be biased. We will either over- or underestimate the effect of the policy, depending on the sign of the final term. Intuitively, if there are any other factors left out of this model that influence both the adoption of the financial aid policy and enrollment in the state, our estimate of the impact of the policy will be biased.

If, for example, states with low per-capita income have both low enrollment rates and strong financial aid policies, the omission of the per-capita income variable will cause ($\text{cov}(\epsilon, F)$) to be negative and $\hat{\beta}$ will underestimate the true causal impact of financial aid. Similarly, states with fewer community colleges may have higher four-year college enrollment and generous financial aid policies, making ($\text{cov}(\epsilon, F)$) positive, and causing $\hat{\beta}$ to overestimate the causal impact of financial aid on enrollment.

An analogous argument can be made when studying individuals and their probability of enrolling, when facing a certain financial aid offer. To see the probability of enrolling in the student-level model, we consider the following linear probability specification.6

5In this oversimplified model, the error term $\epsilon$ includes every other observable and unobservable factor that affects enrollment in the state as well as a component of random noise. Mathematically, the error term could be described as $\epsilon = \lambda X + \eta$, where $X$ is a vector of all state characteristics affecting enrollment and $\eta$ is noise. However, for simplicity, I consider $\epsilon$ alone in the examples that follow.

6 Probit and logit models are more prevalent in the literature, but I use a linear probability model for simplicity. See Becker (2003) for an excellent description of the implications of omitted variable bias in linear vs. non-linear models.
\[ Enroll_i = \alpha + \beta R_i + \epsilon_i, \quad (4) \]

where \( i \) indexes individual students and \( Enroll \) now represents the student’s binary choice to enroll in college. \( (Enroll_i \) equals 0 if the student does not enroll and 1 if she does). \( R \) equals 1 if the student received a particular form of financial aid (e.g., a Pell Grant) and 0 otherwise. As in the state-level case, we can show that \( \hat{\beta} = \beta + \frac{\text{cov}(\epsilon_i, F_i)}{\text{var}(R_i)}. \)

In this case, the factors in \( \epsilon_i \) that influence both a student’s enrollment decision and her financial aid award, such as parental income, academic achievement in high school, and knowledge of financial aid or postsecondary options, will cause estimates of \( \hat{\beta} \) to be biased. To take a typical example, if students who receive need-based financial aid are less likely to enroll in college for any number of reasons that are not captured in this model (for example, lower than average income, low levels of parental support, or little knowledge of postsecondary options), estimates of \( \hat{\beta} \) will underestimate the impact of financial aid on enrollment—potentially attributing false negative effects of the aid program on enrollment. A similar, but opposite bias is likely to occur in the case of merit-based scholarships. If students who receive merit-based scholarships also tend to be more likely to enroll in college in the absence of the aid (due to higher academic achievement, higher income, or higher family expectations) estimates of \( \hat{\beta} \) will overestimate the impact of the aid on enrollment—potentially finding an overly positive impact of aid—as these students would have been more likely to attend college even without it.

Though the examples are oversimplified—univariate OLS is rarely if ever used to draw causal inferences—they illustrate in the simplest terms the problem of omitted variable bias. In the sections that follow, I focus my attention on possible solutions to this problem.

**Solutions**

**Random Assignment**

In the social sciences, random assignment has become the gold standard for proving causality. In this approach, observations (in this case, states or students) are randomly assigned (a process akin to a coin toss) to either the “treatment” or “control” groups. In the case of students, each member of the treatment group might receive a grant to attend the college of his or her choice, while those in the control group would not. Since financial aid receipt is completely unrelated to any characteristics of the student (including characteristics related to his or her enrollment decision), then \( \text{cov}(\epsilon_i, R_i) = 0 \) and the bias is completely eliminated from univariate OLS estimates. One
must compare only the mean enrollment rates of the treatment and control groups using t-tests to obtain unbiased estimates of the grant’s impact. The main identifying assumption of this approach is that, given a large enough sample size, there should be no mean differences in characteristics between the group that was selected to receive the grant and the group that was not. As a robustness check, one can also use t-tests to compare the means of all observable characteristics of the two groups before treatment to ensure that the identifying assumption holds.

This last point suggests that random assignment may not be feasible when looking at state-level financial aid policy. With just 50 states—25 in each experimental group—the sample size would likely be too small to assure that there were no mean differences between the two groups. Random assignment may be possible with smaller geographic areas, if, for example, certain counties were assigned to a particular financial aid policy while others were not. Random assignment of colleges would be another potentially fruitful approach for financial aid research. If half of the colleges in a state were randomly assigned to a new financial aid policy, while others continued with the status quo, one would in theory be able to judge differences in student enrollment and persistence under the two schemes. Though the exact number of observations needed to successfully carry out a random assignment experiment depends on many factors, such as the level of randomization (e.g., schools versus students) and the outcome under study, in general, the more observations one has, the more power one has in detecting significant differences between treatment and control groups.\(^7\)

Despite the simplicity of the approach and the accuracy of the estimates derived from random assignment experiments, the approach has some significant drawbacks. First, it is costly in terms of time and money. A research team must establish a close relationship with policymakers and program administrators to be allowed to carry out the experiment. They must also identify and solicit potential participants, design a lottery to assign participants to treatment and control groups, implement the program for the treatment group, and track both groups’ outcomes over time. Despite these substantial short-term costs, Cook (2002) argues that the costs of undertaking random assignment experiments are likely to be lower than other research designs in the long run. He reasons that, because fewer experiments are needed to achieve the same level of confidence in causal estimates, random assignment experiments are more cost-effective in answering certain research questions. Moreover, the increased likelihood of obtaining accurate estimates of program effectiveness under random assignment will save money and

\(^{7}\)For more details on sample size requirements, see Raudenbush, 1997; Raudenbush, Martinez, & Spybrook, 2007; Shadish, Cook, & Campbell, 2002.
time overall since policymakers will be more likely to implement the most appropriate and effective policies.

A second important practical challenge in conducting random assignment experiments is in implementing the treatment and measuring outcomes. Manipulating and isolating the effect of the treatment can be particularly difficult in educational settings. Incomplete compliance may occur if some members of the assigned treatment group refuse treatment—that is, don’t avail themselves of the financial aid that is offered to them. Similarly, some control group members may seek treatment outside of the experiment by obtaining a different source of financial aid. Similar difficulties arise in measuring outcomes under differential attrition, where participants drop out of the experiment for potentially important unobservable reasons. Cook (2002) suggests that these types of problems can be minimized with proper monitoring and incentives for participation, and he stresses the need to study implementation quality in its own right. As discussed in more detail below, new methods, such as instrumental variables techniques, also offer promise in addressing some of these problems.

Third, random assignment introduces issues of equity that can be difficult to overcome. This is a particularly important consideration in the case of financial aid research. In the student aid example above, can random assignment be justified if it means that some students with high incomes receive grants as part of the treatment group while some students with low incomes are left without grants in the control group? From an ethical perspective, the situation is troublesome, but there are ways to ameliorate the situation. In particular, one can pick a treatment and control group from a more equal pool of students. If all students are equally eligible for the aid—for example, all are in households with incomes barely below the poverty line—then random assignment is a bit less troubling on this dimension. Along the same lines, programs that are oversubscribed are good candidates for random assignment. If, for example, there are a finite number of grants available and a large number of equally qualified and eligible students, one could enter these students into a lottery for the grants. In some cases, this process might be more justifiable than selecting recipients based on other arguably arbitrary measures—such as a student’s state of residence or gender.

One final drawback of random assignment experiments is their lack of external validity or generalizability. Every random assignment experiment is done in the context of a specific treatment in a specific location, making inferences to other contexts difficult. For example, a random assignment evaluation of a financial aid policy in San Francisco may show that every $1,000 spent on a program induces 20 more students to enroll. However, the

---

8For more details on measurement and compliance, see Angrist & Krueger, 1991; Chatterji, 2005; Cook, 2002; Cronbach et al., 1980; Shadish, Cook, & Campbell, 2002.
same policy randomly evaluated in Milwaukee might show negligible effects because the students of Milwaukee face a very different set of circumstances than those in San Francisco. In this case random assignment can control for individual differences in students—it is internally valid—but it cannot account for the context of the experiment. Still, with many small-scale context-specific experiments, patterns may emerge that lend themselves to generalization. Moreover, there has been renewed attention in recent years to the importance of exploring and understanding the specific context of each experiment, and employing multiple research methods such as ethnography and descriptive quantitative analysis to describe, understand, and evaluate the context of a particular random assignment experiment before implementation.\(^9\)

Despite its limitations, random assignment is becoming increasingly popular in education research; and the Department of Education’s Institute for Education Sciences has made a concerted effort to fund studies of this type in the past few years (Myers & Dynarski, 2003; U.S. Department of Education, 2003). As such, a substantial literature has grown in detailing and overcoming the challenges of implementing random assignment experiments in education.\(^10\) At the same time, a literature has grown debating its role, and that of the federal government, in shaping the education research agenda (Berliner, 2002; Chatterji, 2005; Dowd, forthcoming; Dowd & Tong, 2007; Erickson & Gutierrez, 2002; Feuer, Towne, & Shavelson, 2002a, 200b; St. Pierre, 2002).

Some of the most notable random assignment experiments in education are based on oversubscribed elementary school voucher programs in cities like Milwaukee (Rouse, 1998) and New York (Myers & Mayer, 2003; Mayer et al., 2002; Krueger & Zhu, 2004), among others. Other large-scale random assignment projects include an evaluation of the effects of class size in the Tennessee Student/Teacher Achievement Ratio (STAR) Experiment (Krueger, 1999), Upward Bound college readiness program (Myers & Schirm, 1999; Myers et al., 2004), the Quantum Opportunities After-School Program (Hahn, Leavitt, & Aron, 1994), and the Opening Doors Community College Program (Purnell & Blank, 2004). The latter program randomly assigned low-income community college students to the Opening Doors Program, which offered enhanced student services, curricular and instructional reform, and financial aid in the form of vouchers for transportation, childcare,

---

\(^9\)See, for example, Chatterji, 2005; Cohen, Raudenbush, & Ball, 2003; Dowd & Tong, 2007; Erickson & Gutierrez, 2002; Raudenbush, 2005.

\(^10\)For more details on the implementation of random assignment experiments, see Chatterji, 2005; Cohen, Raudenbush, & Ball, 2003; Cook, 2002; Cronbach & Associates, 1980; Raudenbush, 1997; Raudenbush, Martinez, & Spybrook, 2007; Shadish, Cook, & Campbell, 2002).
and books. This study is among the first to use random assignment to study financial aid in postsecondary education.

**Multivariate Regression**

Multivariate regression—including linear and nonlinear estimation—is still the most often-used method for addressing omitted variable bias in the financial aid literature. With this approach, one can control for all observed and measurable variables that were previously omitted from the univariate model. Adding a vector, $X$, that includes every exogenously determined, state-level variable that might be related to enrollment to equation (1), will remove these factors from the error term and reduce the bias of the univariate OLS estimator. Again, using the linear specification for simplicity, the equation is now:

$$ Enroll_s = a + \beta F_s + \delta X_s + \epsilon_s $$

In our case, economic theory suggests that $X_s$ would include variables such as population, unemployment rate, poverty rate, percent minority—all of the measurable variables that might make enrollments in one state different from another. However, it is worth emphasizing that the control variables should be determined exogenously (outside the model), irrespective of the state’s enrollment or financial aid policies, or they will introduce further bias.

In the student-level analysis the equation looks similar:

$$ Enroll_i = a + \beta R_i + \delta X_i + \epsilon_i. $$

In this case, $X_i$ represents student-level characteristics that might include parental income or education level, academic achievement in high school, the type of high school attended, and the student’s race, ethnicity, gender, and age. These variables are generally agreed to be important determinants of enrollment in both the economics and education literature, though perhaps for different reasons. In economics, these variables are typically derived from models of human capital investment and expected lifetime utility maximization developed by Mincer (1958), G. S. Becker (1964), Kohn, Manski, and Mundel (1976), and Fuller, Manski, and Wise (1982). In education, the choice of variables has been influenced by the work of Hossler, Braxton, and Coopermith (1989), Hossler and Gallagher (1987), Weiler (1990), Welki and Navratil (1987), among others. While both traditions emphasize the role of student and institutional characteristics—including financial aid—the

---

11 See, for example, Curs & Singell, 2002; DesJardins et al., 2006; Dowd, 2004; Dowd & Coury 2006; Fuller, Manski and Wise, 1982; Jackson 1978, 1990; Moore, Stuedenmund, & Slobko, 1991; Seneca & Taussig, 1987; St. John, Musoba, & Simmons, 2003; St. John 1990, 1999.
education research has gone further in integrating the role of contextual and sociocultural factors such as parental support, peers’ college plans, and student perceptions.

While multivariate regression analysis is relatively easy to understand and carry out, the only way this method can fully eliminate omitted variable bias is if all possible omitted variables that might be correlated with enrollment are included in the model, mathematically: $\frac{\text{cov}(F_s, \varepsilon_s | X_s)}{\text{var}(F_s | X_s)} = 0$. This situation might occur if the criteria for selection are known precisely, and a study of the Vietnam-era draft by Angrist (1998) comes close. However, most often economists argue that unobservable heterogeneity remains—for example, the preferences of voters, the public perception of community colleges, and other unobservable or unmeasurable characteristics that make Wisconsin different from California. In the student financial aid case, innate ability, a student’s knowledge about college and financial aid programs, and sociocultural factors such as a family’s expectations of college enrollment, are unobservable, rendering straightforward multivariate regression unconvincing. Moreover, even if some of these variables can be measured and included in multivariate models, they can introduce additional bias if they are endogenously determined.

As Angrist and Krueger (2001) point out, it is rare that a theory specifies all of the variables that must be controlled for in a given relationship; and even if it did, it would be nearly impossible to observe and measure them all. While rich datasets with extensive background characteristics are helpful in this regard, omitted variable bias remains problematic in multivariate regression analysis. In the following sections, I discuss several methods that can be used to ameliorate this problem.

**Proxy Variables**

One of the simplest methods to mitigate bias from unobservable omitted variables is the use of proxy variables. This approach involves finding suitable observable variables that can “stand in” for each unobservable. Adding a proxy, denoted $P_s$, to the model in equation (5) above, we have:

$$Enroll_s = \alpha + \beta F_s + \delta X_s + \lambda P_s \varepsilon_s.$$  

(7)

There are two key conditions that must be met by a potential proxy. First, the proxy variable must be ignorable, or redundant, in the equation. Mathematically, $E(Enroll_s | F_s, X_s, P_s) = E(Enroll_s | F_s, X_s)$. Intuitively, the proxy must be irrelevant in explaining enrollment once $F_s$ and $X_s$ have been controlled for.

Assume for a moment that there is only one omitted variable in the student-level analysis—a student’s unobservable innate ability. (In fact, ability
is the classic omitted variable in the economics of education literature. See Angrist & Krueger, 1999; Card, 1999.) If we believe that a student’s performance on an IQ test is a good measure of his or her innate ability, the IQ score could serve as a proxy for ability, since the redundancy condition is met. That is, if we knew a student’s true innate ability, we would not need to include IQ in the equation (Wooldridge, 2002). The second condition that must be met for a good proxy, is that it must be sufficiently closely related to the omitted variable, so that once \( P \) is included in the equation, there is no remaining variation in the unobservable variable that is correlated with \( F \) or \( X \). If this condition does not hold—and we can never know with certainty if it does—then omitted variable bias may remain. Still, as Wooldridge points out, if it does not hold, a reasonable “imperfect proxy” may still reduce omitted variable bias.

When using proxy variables, as in the case of more straightforward multivariate regression analysis, it is a good idea to add proxies sequentially and compare coefficients on the variable of interest. Since reasonable proxies reduce omitted variable bias, if estimates do not change with the addition of these variables, it is plausible to believe that any remaining unobservables also have little effect on the estimates. Neumark and Rothstein’s (2003) research on school-to-work programs has made extensive use of this method; and in financial aid research, St. John (1990) includes test scores as proxies for ability and postsecondary plans as proxies for aspirations. In another example from the higher education literature, Seneca and Taussig (1987) use faculty compensation as a proxy for the academic prestige of a university.

The main constraint with this method is finding datasets that contain a rich array of potential proxies and convincing readers that the proxied variables are the only unobservables of concern. Moreover, exogeneity of the proxy variable is again important to consider, as the proxies must be determined by factors unrelated to the student’s enrollment and financial aid choices to avoid introducing additional bias. Often this can be accomplished by using variables measured in years prior to the choice, such as test scores or aspirations reported by students in middle or high school (as in St. John, 1990), before a student has the option of enrolling in college.\(^{12}\)

**Fixed Effects**

Fixed effects methods hold considerable promise in research on financial aid, and many recent studies in both the education and economics literature have implemented this approach (Ehrenberg, Zhang, & Levin, 2006; Card & Lemieux, 2000; Heller, 1999; Kane, 1994, 1995, 2004). The essential element for fixed effects and difference-in-differences estimates (described in the next

---

\(^{12}\)For more details on proxy variables, see Stahlecker & Trenkler, 1993; Wooldridge, 2002.
section) is access to longitudinal or panel data—or multiple observations for each unit of analysis. Most often, this type of data tracks cross-sectional units (in our case, states, students, or institutions) over time. Examples include High School and Beyond, Beginning Postsecondary Students, the Integrated Postsecondary Education System, and the National Longitudinal Survey of Youth. But fixed effects methods can also be employed as long as there are multiple observations within larger units such as states, colleges, or families—making this method feasible even with cross-sectional datasets such as the National Postsecondary Student Aid Study or the Current Population Survey.

While basic OLS estimation compares different states or students to each other, the fixed effects approach essentially compares a unit of analysis to itself. The model is identified by changes or differences within units. For this reason, the fixed effects estimator is sometimes referred to as a “within estimator.” Because the estimates are derived from within units, the approach eliminates the bias from any characteristics—either observed or unobserved—that are common to units.

In the most typical example, if we have data on each state for several years, fixed effects estimation will control for all time-invariant characteristics of the state—both unobservable and observable—that might bias cross-sectional OLS estimates. These include many, but not all of the characteristics that make Wisconsin different from California, such as the size and structure of the public university system or state high school graduation requirements. However, time-varying unobservables may remain, an issue I return to in the following sections.

To implement the fixed effects approach, one simply adds dummy variables for each state (or individual) to the model as follows:

\[
Enroll_{st} = a + \beta F_{st} + \delta X_{st} + d_s + \epsilon_{st}, \tag{8}
\]

where the vector of dummy variables is denoted \(d_s\). In this case, one can also add time fixed effects, or dummy variables for each year, denoted \(d_t\), as in equation (9) below:

\[
Enroll_{st} = a + \beta F_{st} + \delta X_{st} + d_s + d_t + \epsilon_{st} \tag{9}
\]

These time or year fixed effects will absorb any time trends that are common to all states—for example, inflation or changes in federal laws that affect all states in the same way.

In a good example of this approach, Heller (1999) implements a fixed effects model to look at the impact of tuition and state need-based grants on enrollment rates. In addition to state-level fixed effects, his analysis also
includes year fixed effects that interacted with regional fixed effects. That is, he creates a dummy variable for each region of the country (e.g., South, West) and multiplies that by an indicator for the year to create new dummy variables. This approach allows any trends over time to vary by region.

Because fixed effects estimates rely on variation within units, this approach uses only units that experience changes or differences in the variable of interest; all others are dropped from the analysis. If, for example, $F_{st}$ represents the value of state need-based grants to students in a particular state and year, as in Heller’s work, states that offer students the same amount every year will not be included in the analysis, possibly reducing the sample size. $F_{st}$ must vary over time within states to be identified. Moreover, any control variables in $X_{st}$ that do not vary over time within units will also be unidentified; by eliminating time-invariant unobservables, fixed effects also eliminate the time-invariant observables. This is one reason why Heller includes very few control variables in his model. In fact, the only variable he adds is the unemployment rate, since it is one of the few that varies over time within a state.

Again, in order to prove causality, it is important to ensure that the variation over time in the variable of interest is exogenous. That is, fluctuations in the amount of financial aid offered to students each year could be determined by the idiosyncratic state budgetary process or by a formula based on national data, but they should not be determined by factors correlated with enrollment in the state.

Finally, it is worth returning to an earlier point—that fixed effects can also be used at other levels of aggregation, depending on data availability. For example, rather than using state or individual fixed effects, Ehrenberg, Zhang, and Levin (2006) use college-level fixed effects to look at the impact of institution-level National Merit Scholarships on the enrollment of low-income students. Generally, the more precise or homogeneous the groups are in fixed effects estimation, the less bias will remain in the estimates. Using the state policy example, state-level fixed effects are more meaningful than region-fixed effects since states in the same region of the country can have very different systems of higher education. Similarly, in studies of individuals, family-level fixed effects are more effective at reducing omitted variable bias than institution-level fixed effects since children in the same families share more characteristics than students in the same school.\(^{13}\) Individual-level fixed effects would theoretically provide the best control for omitted variable bias, but this is rarely possible since one would need to observe the same individual’s enrollment response both with and without an offer

---

\(^{13}\) For examples of family fixed effects strategies in education, see Cellini, 2006; Currie & Thomas, 1995.
of financial aid to be able to identify effects. Because of this requirement, fixed effects methods are typically considered only a partial correction for omitted variable bias.\textsuperscript{14}

**Difference-in-Differences**

The difference-in-differences approach goes one step further than the fixed effects approach by adding an additional level of variation. In financial aid research, this approach typically uses a binary variable to represent a policy change. The variable $F_s$ again takes on a value of 0 for states that do not implement a given policy and 1 if they do, but it is now multiplied by $T_t$, an indicator for the time period that takes on value of 0 in years before the policy change and a 1 after. The model is estimated as follows:

$$
Enroll_{st} = \alpha + \beta (F_s * T_t) + \delta X_{st} + d_s + d_t + \varepsilon_{st},
$$

(10)

This approach introduces two levels of variation. The “first difference” is the variation across time. The “second difference” is the adoption of the policy. That is, the model compares the within-state changes over time between states that adopted the policy and those that did not.

It is worth noting that the difference-in-differences approach and the regression discontinuity approach (described in the next section) rely on “natural experiments” or “quasi-experiments” for identification. In the absence of a true random assignment experiment, these two approaches exploit seemingly random variation that occurs in nonexperimental (or observational) settings, such as in nature, government policy, or institutional design. In the case of difference-in-differences, the model above mimics random assignment by comparing enrollment in “treatment” states before and after the policy change to changes in enrollment over the same time period in the “control” states that did not implement the policy. The key identifying assumption is that the policy change is exogenous—that is, passage of the policy must not be related to enrollment in the state. Typically, this claim cannot be proven directly but must be argued by the researcher. If the policy or law is determined at the federal level, this assumption is typically uncontroversial, but it becomes less plausible at the local level.

Another important consideration is that the difference-in-differences approach introduces serial correlation. That is, the observations for each person or state are correlated with the other observations for that state in the previous year. To take an example from Cornwell, Mustard, and Sridhar (2006), the state of Georgia implemented a merit-based financial aid program in 1993. It would therefore be assigned a 0 in 1990, 1991, and

\textsuperscript{14}For more technical details on fixed effects methods, see Chamberlain, 1980, 1984; Wooldridge, 2002, 2005.
1992 before introducing the program, and a 1 in 1993, 1994, and 1995. If, as we suspect, enrollment goes up in 1993 and stays high in the following years in response, the errors will be correlated over time. Fortunately, there are several solutions to this problem, the simplest of which is clustering standard errors.\textsuperscript{15}

Research on financial aid has made extensive use of the difference-in-differences approach (Conley & Taber 2005; Cornwell, Mustard, Sridhar 2006, Dynarski, 2000, 2003; Kane 2003; Lisenmeier, Rosen, & Rouse, 2003); and there are likely to be many more creative applications of this strategy in coming years. A promising new direction for difference-in-differences is to add additional levels of variation for a difference-in-difference-in-differences (or more) estimator. This could be achieved in financial aid research by identifying groups of states or students who might be more strongly affected by a policy than others, such as low-income students or states with high poverty rates. The difference-in-differences and triple differences approaches are fairly straightforward to implement and, as in the case of fixed effects, can control for all time-invariant unobservables, though time-invariant unobservables may remain.

**Regression Discontinuity**

In this quasi-experimental approach, identification rests on an exogenous discontinuity, or cutoff. The key assumption is that students’ or states’ unobservable characteristics vary smoothly across this cutoff, while the cutoff creates a sharp difference in one variable that can be used to identify the causal effect of a program or policy. In the discussion that follows, I focus on “sharp” regression discontinuity designs where the cutoff is deterministic in nature. However, so-called “fuzzy” regression discontinuity designs are also feasible in cases of incomplete compliance. Interestingly, this approach was first developed for use in models of financial aid and college access by Thistlewaite and Campbell (1960) but has only more recently become widespread. Kane (2003), Van der Klaauw (2002), and Bettinger (2004) all draw on this strategy to identify the effects of financial aid policies.

To take an example from Kane (2003), California’s CalGrant financial aid program used a GPA cutoff of 3.15 to determine which students would be eligible for grants. Students just below the cutoff did not receive grants, but students just above it did. Kane argues that, within a narrow range, these students are likely to be quite similar—that is, there is no reason to believe that students with 3.14 GPAs are systematically different (on observable or unobservable dimensions) than those with 3.15 GPAs who received grants.

\textsuperscript{15}For more details on difference-in-differences methods and solutions to this problem, see Athey & Imbens, 2006; Bertrand, Duflo, & Mullainathan, 2004; Conley & Taber, 2005; Lee & Kang, 2006; Meyer, 1995.
The only mean difference should be their grant receipt. The causal impact of the CalGrant program can then be identified by the difference in the outcomes of those students who just barely made the cutoff versus those who did not.

Institutionally derived, arbitrary, or formulaic cutoffs are good candidates for regression discontinuity research designs, as are physical or geographic borders or boundaries. While physical boundaries have not yet been used in financial aid policy research, elementary school district boundaries have been widely used to study the effects of school quality on housing prices (see, e.g., Black, 2002; Kane, Riegg, & Staiger, 2006; Rothstein, 2006).

There are two ways to implement the regression discontinuity approach. The first is simply to limit one’s sample to observations very close to the discontinuity on either side (e.g., if the cutoff is 3.15, then limit the sample to students with GPAs between 3.12 and 3.17). The equation would look like equation (6), where $R$ represents a dummy variable that equals 1 if the student made the GPA cutoff and therefore received the grant.

Alternatively, if the sample size is relatively small, one can use the entire sample of students, but then the specification must include a polynomial in GPA to account for the smoothness of the underlying function, as follows:

$$Enroll_i = a + \beta R_i + \lambda (GPA_i)^2 + \delta X_i + \epsilon_i$$  \hspace{1cm} (11)

Note that the polynomial is usually a quadratic or cubic term; anything higher may absorb all variation. Also, whether one uses the specification in equation (11) or uses the limited sample approach, the coefficient on the variable of interest should be the same in both cases. However, the standard errors may be smaller in the polynomial approach since the sample size is larger.

There are two important considerations to keep in mind when implementing a regression discontinuity design. First, the cutoff must be exogenously determined. In Kane’s case, the GPA cutoff for the grant changed every year based on the number of applicants—not based on characteristics of the applicants themselves.

A second but related point is that only the variable of interest should change at the cutoff. In Kane’s case, the cutoff was not known a priori by students, which meant that motivated students could not work extra hard to push themselves over the threshold. If this had been the case, then estimates of the causal effect of the grants would have been biased, as this unobservable motivation would have varied along the same cutoff. Graphical analyses can be particularly useful in showing that only the treatment varies at the cutoff. Providing a visual comparison of the existence of the discontinuity in the variable of interest to the lack of a discontinuity in other observable
characteristics lends credibility to the case that the discontinuity can isolate causal effects (Imbens & Lemieux, 2007; Lee & Card, 2006; Shadish, Cook, & Campbell, 2002).

**Instrumental Variables**

Though widely used in assessing the returns to schooling (Angrist & Krueger, 1991; Card, 1995; Kane & Rouse, 1993, Staiger & Stock, 1997), instrumental variables (IV) methods have rarely been applied to studies of financial aid and student access (Alon, 2005; Seneca & Taussig, 1987). The goal of the IV approach is to find a convincing variable that can be used as an “instrument” for the endogenous variable of concern—in this case, financial aid policy. The instrument allows one to identify and isolate a source of variation in the endogenous regressor that is not affected by omitted variables. To borrow an excellent intuitive explanation from Angrist and Krueger (2001, p. 73), IV estimation solves the omitted variable problem by using only part of the variability in the endogenous variable—a part that is uncorrelated with the omitted variables—to estimate the relationship between the endogenous regressor and the dependent variable.

Two conditions must be met to implement IV estimation successfully. First, the instrument must be correlated (after controlling for exogenous variables) with the endogenous variable. This is known as the strength of the instrument and is denoted mathematically as \( E(F, Z | X) \neq 0 \), where \( Z \) is the instrumental variable and \( F \) is the financial aid policy variable.

Second, the instrument must be exogenous to all other factors that might affect enrollment. That is, it must be uncorrelated with the error term \( \epsilon \), mathematically, \( E(\epsilon, Z) = 0 \). This is referred to as the validity of the instrument and is extremely important for causal inference. To see this, consider the equation for the IV estimator, where \( Z \) is an instrument for financial aid:

\[
\hat{\beta}_{IV} = \frac{\text{cov}(Z, Enroll)}{\text{cov}(F, Z)}
\]

(12)

As in the case of the OLS estimator in equation (3), substituting for \( Enroll \) and simplifying yields the “true” \( \beta \) plus a bias term:

\[
\hat{\beta}_{IV} = \frac{\text{cov}(Z, (a + \beta F + \epsilon))}{\text{cov}(F, Z)} = \beta \frac{\text{cov}(F, Z)}{\text{cov}(F, Z)} + \frac{\text{cov}(\epsilon, Z)}{\text{cov}(F, Z)} \beta + \frac{\text{cov}(\epsilon, Z)}{\text{cov}(F, Z)}
\]

(13)

If the instrument is valid, then \( \text{cov}(\epsilon, Z) = 0 \), and \( \hat{\beta}_{IV} \) will produce unbiased estimates of \( \beta \).

Putting the strength and validity conditions together in the context of our financial aid examples, the instrument must affect a state’s adoption of a financial aid policy (or a student’s ability to get it), but have no independent affect on enrollment. Put another way, the instrument must impact
enrollment only through the state’s adoption of a financial aid policy. For example, the instrument cannot also affect the adoption of higher tuition or an increase in the number of colleges, since it would also impact enrollment through those variables.

Once a potential instrument has been found, IV estimation can be implemented using substitution or two-stage least squares (2SLS). Using 2SLS as an example, in the first stage, one estimates a linear projection of the endogenous variable $F_s$:

$$F_s = \gamma + \lambda Z_s + \theta X_s + \eta_s, \quad (14)$$

where again, $Z_s$ is an instrument for the omitted variable. From this equation, one calculates predicted values, $\hat{F}_s$. These are used in the second stage regression equation to estimate:

$$\text{Enroll}_s = \alpha + \beta \hat{F}_s + \delta X_s + \varepsilon_s. \quad (15)$$

Using the predicted values of $F_s$ essentially leaves behind the residuals from the first stage, thereby eliminating the part of the variation in $F_s$ that is correlated with $\text{Enroll}_s$.

While many statistical software packages make IV estimation straightforward to carry out, the most significant challenge of this approach is finding convincing instruments. One can test the strength of the instrument using the first-stage estimation. If equation (14) produces a significant coefficient on $Z_s$ (in this case, $\hat{\lambda}$) then the instrument is considered strong. If the instrument is only weakly correlated with the endogenous variable, however, then additional problems of bias and inconsistency may be introduced.\(^{16}\)

In contrast to an instrument’s strength, the validity of an instrument cannot be tested directly and must be argued by the researcher. As such, finding plausibly valid instruments is difficult. In an effort to control for the unobservable innate ability, researchers estimating the returns to schooling have used instruments derived from natural experiments, such as quarter of birth in conjunction with compulsory schooling laws (Angrist & Krueger, 1991), and proximity to college (Card, 1995) but even these instruments are not completely free of criticism on the grounds of validity. In financial aid research, finding convincing instruments may be just as challenging.

Still, as pointed out by Angrist and Krueger (2001), a promising new direction for education research is the use of instrumental variables in random assignment experiments with incomplete compliance. As described above, compliance issues can compromise causal inference in pure random assignment evaluations if some individuals assigned to the treatment group

\(^{16}\)For more on this issue, see Bound, Jaeger, & Baker, 1995.
refuse treatment or if some assigned to the control group receive it. However, using IV estimation, one can still effectively estimate the causal effects of the treatment in question. The individual’s actual treatment status would be considered an endogenous regressor and the assigned group (treatment or control) would serve as the instrument. The instrument would be strong, because assignment status is likely to be highly correlated with actual treatment, and it would be valid since it was assigned at random. Instrumental variables estimation would then yield an accurate estimate of the causal effect of the treatment on population that complied with the random assignment. This approach has been used in the education literature assessing the impact of class size (Kreuger, 1999), hours of study (Powers & Swinton, 1984), and voucher programs (Howell, Wolf, Campbell, & Peterson, 2002) on students’ achievement test scores. To date, however, it has not been used in financial aid research.17

**Discussion and Conclusions**

The discussion above has outlined several methods for addressing the problem of omitted variable bias in financial aid research. While random assignment is undoubtedly the most methodologically sound and mathematically simple approach to reducing this bias, the ethical considerations and resources needed to carry out such evaluations make this method impractical in many situations. On the other hand, the Department of Education has made a strong effort to support large-scale random assignment evaluations in education in the past few years, making this approach increasingly feasible in education research.

For smaller-scale projects, proxy variable, fixed effects, and difference-in-differences approaches are becoming quite common. Indeed, these approaches have replaced basic multivariate regression as the new standard for education research in the economics literature. New applications and extensions of these approaches—such as triple differences—may yield high returns in financial aid research. The use of instrumental variables techniques, on the other hand, has declined in recent years, since finding strong and valid instruments has proven to be an almost insurmountable obstacle. On the other hand, the approach has proved useful in the context of random assignment evaluations with incomplete compliance. Finally, regression discontinuity holds great promise for the future of nonexperimental financial aid research, as the identifying assumptions are straightforward and the approach is relatively simple to implement.

17For more details on IV methods, see Angrist & Kreuger, 2001; Angrist, Imbens, & Rubin, 1996; Meyer, 1995; Staiger & Stock, 1997; Wooldridge, 2002.
Future research on the causal effects of financial aid should go further in applying these new methods. Often, the variables needed to apply these techniques already exist in commonly used and publicly available datasets. In other cases, a bit of resourcefulness in data collection or manipulation is all that is required to identify and exploit a suitable proxy, discontinuity, or instrument. Researchers should also consider employing and comparing the results of multiple methods to assess the magnitude of omitted variable bias and the extent to which various methods can control for it. For example, comparing multivariate regression estimates with fixed effects estimates can identify the magnitude of omitted variable bias from factors that are constant within groups. Similarly, comparing fixed effects estimates with those derived from regression discontinuity methods can indicate the size of the bias from remaining unobservables that vary within groups.

It is worth noting, however, that quantitative causal inference is by no means the only valid method of inquiry in financial aid or educational research generally. Indeed, descriptive quantitative and qualitative analyses are equally important in establishing patterns of correlation, developing theory, and directing our attention to areas where further research is warranted. They are also vitally important in understanding the nature of the forces that drive the causal effects of financial aid on college-going. These forces—whether they are economic, familial, psychological, social, cultural, or political—are essential determinants of college-going and financial aid receipt in their own right. In fact, these are the unobservable factors that create the problem of self-selection and omitted variable bias in the first place.

In this article I am concerned only with suggesting ways to mitigate or eliminate the effects of some or all of these forces—whether we can identify them or not—in order to make inferences about the independent effects of financial aid on enrollment. I argue simply that omitted variable bias must be explored and addressed before we make claims about whether policies, programs, or treatment interventions cause an effect or are simply correlated with it. The distinction is paramount.

I assess and explore the strengths and weaknesses of some of the most common methodological solutions to the problem of omitted variable bias and provide practical guidance for researchers interested in implementing these approaches. Each question we ask and each dataset we employ will present unique opportunities and challenges, requiring careful thought and a nuanced understanding of the issues before determining the most appropriate research strategy. And while the methods presented above all have disadvantages, each also holds promise for the future of causal inference in financial aid research. The possibilities are limited only by the creativity of the researcher.
REFERENCES


