6. Instrumental Variable Estimation

Summary

Instrumental variable (IV) methods allow for endogeneity in individual participation, program placement, or both. With panel data, IV methods can allow for time-varying selection bias. Measurement error that results in attenuation bias can also be resolved through this procedure. The IV approach involves finding a variable (or instrument) that is highly correlated with program placement or participation but that is not correlated with unobserved characteristics affecting outcomes. Instruments can be constructed from program design (for example, if the program of interest was randomized or if exogenous rules were used in determining eligibility for the program).

Instruments should be selected carefully. Weak instruments can potentially worsen the bias even more than when estimated by ordinary least squares (OLS) if those instruments are correlated with unobserved characteristics or omitted variables affecting the outcome. Testing for weak instruments can help avoid this problem. Another problem can arise if the instrument still correlates with unobserved anticipated gains from the program that affect participation; local average treatment effects (LATEs) based on the instruments can help address this issue.

Learning Objectives

After completing this chapter, the reader will be able to discuss

- How instrumental variables can resolve selection bias in participation, program placement, or both
- How the IV approach differs in assumptions from propensity score matching (PSM) and double-difference (DD) methods
- What sources are available for finding good instruments
- How to test for weak instruments
- What the difference is between standard IV methods and the LATE

Introduction

This handbook now turns to methods that relax the exogeneity assumption of OLS or PSM and that are also robust to time-varying selection bias, unlike DD. Remember that for DD methods one cannot control for selection bias that changes over time (chapter 5). By relaxing the exogeneity assumption, the IV method makes different identifying assumptions from the previous methods—although assumptions underlying IV may not apply in all contexts.

Recall the setup discussed in chapter 2 of an estimating equation that compares outcomes of treated and nontreated groups:

$$Y_i = \alpha X_i + \beta T_i + \varepsilon_i \tag{6.1}$$

If treatment assignment T is random in equation 6.1, selection bias is not a problem at the level of randomization (see chapter 3). However, treatment assignment may not be random because of two broad factors. First, *endogeneity* may exist in program targeting or placement—that is, programs are placed deliberately in areas that have specific characteristics (such as earnings opportunities or social norms) that may or may not be observed and that are also correlated with outcomes Y. Second, *unobserved individual heterogeneity* stemming from individual beneficiaries' self-selection into the program also confounds an experimental setup. As discussed in chapter 2, selection bias may result from both of these factors because unobserved characteristics in the error term will contain variables that also correlate with the treatment dummy T. That is, $\text{cov}(T, \varepsilon) \neq 0$ implies violation of one of the key assumptions of OLS in obtaining unbiased estimates: independence of regressors from the disturbance term ε. The correlation between T and ε naturally biases the other estimates in the equation, including the estimate of the program effect β.

Equation 6.1, as well as the corresponding concerns about endogeneity, can be generalized to a panel setting. In this case, unobserved characteristics over time may be correlated with the program as well as other with observed covariates. To an extent, this issue was discussed in chapter 5. DD methods resolved the issue by assuming that unobserved characteristics of targeted and nontargeted units were time invariant and then by differencing out the heterogeneity. When panel data are available, IV methods permit a more nuanced view of unobserved heterogeneity, allowing for these factors to change over time (such as unobserved entrepreneurial talent of targeted subjects, ability to maintain social ties and networks, and so on, all of which may vary with the duration of the program).

The IV aims to clean up the correlation between T and ε. That is, the variation in T that is uncorrelated with ε needs to be isolated. To do so, one needs to find an instrumental variable, denoted Z, that satisfies the following conditions:

1. Correlated with T: $\text{cov}(Z,T) \neq 0$
2. Uncorrelated with ε: $\text{cov}(Z,\varepsilon) = 0$

Thus, instrument Z affects selection into the program but is not correlated with factors affecting the outcomes (also known as an *exclusion restriction*).

A related issue is that measurement error in observed participation may underestimate or overestimate the program's impact. As discussed in chapter 3, an IV can be introduced to resolve this attenuation bias by calculating an intention-to-treat (ITT) estimate of the program. This estimate would account for actual participation being different from intended participation because of targeting and eligibility rules.

Khandker (2006) provides an example of how concerns regarding exogeneity and attenuation bias can be addressed. In this study, the impact of microfinance expansion on consumption expenditure and poverty is estimated using panel data from Bangladesh, spanning household surveys for 1991–92 and 1998–99.[1] The study intended to test the sensitivity of findings in Pitt and Khandker (1998) using the 1991–92 data set. Households were sampled in villages with and without a program; both eligible and ineligible households were sampled in both types of villages, and both program participants and nonparticipants were sampled among the eligible households in villages with microfinance programs. The two central underlying conditions for identifying the program's impact were (a) the program's eligibility restriction (any household with a landholding of less than half an acre was eligible to participate in microfinance programs) and (b) its gender-targeted program design (men could join only groups with other men, and women could join only groups with other women). A gender-based restriction is easily enforceable and thus observable, whereas a land-based identification restriction, for various reasons, may not be (see Morduch 1998). Thus, if the land-based restriction is not observable, using the gender-based program design to identify the program effect by gender of participation is far more efficient.

A village-level fixed-effect DD method might be used to resolve unobserved heterogeneity in this example, given the existence of panel data. However, the assumption of time-invariant unobserved heterogeneity might be violated. For example, unobserved household income, which may condition credit demand, may increase temporarily from the program so that with a larger cushion against risk, households may be willing to assume more loans. Similarly, unobserved local market conditions that influence a household's demand for credit may change over time, exerting a more favorable effect on credit demand. Also, the unmeasured determinants of credit at both the household and the village levels may vary over time, and if credit is measured with errors (which is likely), the error is amplified when differencing over time, especially with only two time periods. This measurement error will impart attenuation bias to the credit impact coefficients, biasing the impact estimates toward zero. A standard correction for both types of bias (one attributable to measurement error and one to time-varying heterogeneity in credit demand) is IV estimation. This approach is discussed further later in the chapter.

Two-Stage Least Squares Approach to IVs

To isolate the part of the treatment variable that is independent of other unobserved characteristics affecting the outcome, one first regresses the treatment on the instrument

Z, the other covariates in equation 6.1, and a disturbance, u_i. This process is known as the *first-stage regression*:

$$T_i = \gamma Z_i + \phi X_i + u_i. \tag{6.2}$$

The predicted treatment from this regression, $\hat{T}$, therefore reflects the part of the treatment affected only by Z and thus embodies only exogenous variation in the treatment. $\hat{T}$ is then substituted for treatment in equation 6.1 to create the following reduced-form outcome regression:

$$Y_i = \alpha X_i + \beta(\hat{\gamma} Z_i + \hat{\phi} X_i + u_i) + \varepsilon_i. \tag{6.3}$$

The IV (also known as *two-stage least squares*, or 2SLS) estimate of the program impact is then $\hat{\beta}_{IV}$. Specifically, looking at $Y_i = \beta T_i + \varepsilon_i$, a simplified version of equation 6.1, and knowing that by assumption cov(Z, ε) = 0, one can also write the treatment effect under IV (β) as cov(Y, Z)/cov(T, Z):

$$\text{cov}(Y_i, Z_i) = \text{cov}[(\beta T_i + \varepsilon_i), Z_i] = \beta\text{cov}(T_i, Z_i) \tag{6.4}$$

$$\Rightarrow \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(T_i, Z_i)} = \beta. \tag{6.5}$$

This derivation becomes important when examining the effects of instrument quality on the estimated program impact under IV (see the next section 6.3).

Through instrumenting, therefore, T is cleaned of its correlation with the error term. If the assumptions cov(T, Z) $\neq$ 0 and cov(Z, ε) = 0 hold, then IV consistently identifies the mean impact of the program attributable to the instrument. Specifically, it can be shown that $\hat{\beta}_{IV} = \beta + \text{cov}(Z, \varepsilon)/\text{cov}(Z, T)$. This idea is also discussed further in the next section.

Although detailed information on program implementation and participation can directly reveal the presence of selection bias, endogeneity of treatment can also be assessed using the Wu-Hausman test, which in the following example uses a regression-based method:

1. First, regress T on Z and the other exogenous covariates X, and obtain the residuals $\hat{u}_i$. These residuals reflect all unobserved heterogeneity affecting treatment not captured by the instruments and exogenous variables in the model.
2. Regress Y on X, Z, and $\hat{u}_i$. If the coefficient on $\hat{u}_i$ is statistically different from zero, unobserved characteristics jointly affecting the treatment T and outcomes Y are significant, and the null that T is exogenous is rejected.

The IV model has some variations. For example, one could rewrite the instrument equation as a nonlinear binary response model (such as a probit or logit) and use the predicted propensity score as the IV for program placement. Also, if panel data exist, IV can be combined with a panel fixed-effects approach as follows (see Semykina and Wooldridge 2005):

$$Y_{it} = \delta Q_{it} + \eta_i + v_{it}, t = 1, \ldots, T, \tag{6.6}$$

In equation 6.6, η_i is the unobserved fixed effect (discussed in chapter 5) that may be correlated with participation in the program, v_{it} represents a time-varying idiosyncratic error, and Q_{it} is a vector of covariates that includes exogenous variables X as well as the program T. In this specification, therefore, correlation between η_i and the treatment variable in Q_{it} is accounted for through the fixed-effects or differencing approach, and instruments Z_{it} are introduced to allow for correlation between some of the regressors in Q_{it} (such as T) and v_{it}. The idea here would be to find instruments correlated with program uptake (but not outcomes) over time. The remaining assumptions and interpretation of the estimate are similar.

Concerns with IVs

Concerns with IVs include weak instruments and correlation with unobserved characteristics.

Implications of Weak Instruments on Estimates

A drawback of the IV approach is the potential difficulty in finding a good instrument. When the instrument is correlated with the unobserved characteristics affecting the outcome (that is, $\text{cov}(Z, \varepsilon) \neq 0$), the estimates of the program effect will be biased. Furthermore, if the instrument only weakly correlates with the treatment variable T, the standard error of the IV estimate is likely to increase because the predicted impact on the outcome will be measured less precisely. Consistency of the IV estimate (that is, asymptotic bias) is also likely to be large when Z and T are weakly correlated, even if the correlation between Z and ε is low. This problem can violate the assumption underlying IV estimation as seen here. As mentioned in the previous section, asymptotically, $\beta_{\hat{IV}} = \beta + \text{cov}(Z, \varepsilon)/\text{cov}(Z, T)$; thus, the lower $\text{cov}(Z, T)$, the greater the asymptotic bias of $\hat{\beta}$ away from the true β.

Testing for Weak Instruments

One cannot test for whether a specific instrument satisfies the exclusion restriction; as mentioned earlier, justifications can be made only through direct evidence of how the program and participation evolved. With multiple instruments, however, quantitative tests (also known as *tests of overidentifying restrictions*) exist. They involve the following steps:

1. Estimate the structural equation by 2SLS, and obtain the residuals $\hat{\varepsilon}_i$.
2. Regress $\hat{\varepsilon}_i$ (which embody all heterogeneity not explained by the instruments Z and other exogenous variables X) on X and Z. Obtain the R^2.

3. Use the null hypothesis that all the instrumental variables are uncorrelated with the residuals, $nR^2 \sim \chi_q^2$, where q is the number of instrumental variables from outside the model minus the total number of endogenous explanatory variables. If nR^2 is statistically greater than the critical value at a certain significance level (say, 5 percent) in the χ_q^2 distribution, then the null hypothesis is rejected, and one can conclude that at least one of the instrumental variables is not exogenous.

Local Average Treatment Effects

As mentioned earlier, the IV estimate of the program effect is ultimately an intent-to-treat impact, where the measured effect of the program will apply to only a subset of participants. Imperfect targeting is one case where only intent-to-treat impacts can be measured; the researcher then has to search for an exogenous indicator of participation that can account for unobserved heterogeneity. A good instrument in this case would satisfy the exclusion restriction and be well correlated with participation. However, the instrument would very unlikely be perfectly correlated with participation, so only a subset of participants would be picked up by the instrument and resulting IV effect. The same holds where an instrument is needed to correct for errors in measuring participation; similar ITT impacts relating to a subset of participants would result. The resulting IV program effect would therefore apply only to the subset of participants whose behavior would be affected by the instrument.

One difficulty arises with the standard IV estimate if individuals know more about their expected gains from the program than the evaluator or researcher does. That is, individuals are anticipating gains from the program that the evaluator or researcher cannot observe. Consequently, unobserved selection occurs in participation, because those individuals that end up benefiting more from the program, given their characteristics X, may also be more likely to participate. Because the instrument Z affects participation, unobserved characteristics driving participation will also correlate with Z, and the IV estimate will be biased.

Heckman (1997), for example, brings up a study by Angrist (1990) that examines the effect of military service on earnings. As an instrument for joining the military, Angrist uses the 1969 U.S. military draft lottery, which randomly assigned priority numbers to individuals with different dates of birth. A higher number meant the person was less likely to be drafted. However, even if a person received a high number, if he nevertheless enrolled in military service, one could assume that his unobserved anticipated gains from military service were also likely to be higher. Thus, the instrument causes systematic changes in participation rates that relate to unobserved anticipated gains from the program. This change creates bias in comparing participants and nonparticipants.

Imbens and Angrist (1994) address this problem by introducing the local average treatment effect. In the special case where heterogeneity exists in individuals' response to the program, IV methods consistently estimate the average effect of the program only for those whose participation changes because of changes in instrument Z. Specifically, the LATE estimates the treatment effect only for those who decide to participate because of a change in Z (see, for example, Imbens and Angrist 1994). In the context of schooling, for example, if outcome Y is a test score, T is an indicator for whether a student is in a Catholic high school, and instrument Z is an indicator for whether the student is Catholic, then the LATE is the mean effect on test scores for students who choose to go to a Catholic high school because they are Catholic (see Wooldridge 2001). The LATE avoids the problem of unobserved forecasting of program gains by limiting the analysis to individuals whose behavior is changed by local changes in Z in a way unrelated to potential outcomes. In the previous military service example, for instance, those with high anticipated gains from participating are unlikely to be among the shifters. Note that, as a result, the LATE does not measure the treatment effect for individuals whose behavior is not changed by the instrument.

One of the underlying assumptions for the LATE is monotonicity, or that an increase in Z from $Z = z$ to $Z = z'$ leads some to participate but no one to drop out of the program. Participation T in this case depends on certain values of the instruments Z (say, $Z = z$ versus $Z = z'$), such that $P(T = 1|Z = z)$ is the probability of participating when $Z = z$, and $P(T = 1|Z = z')$ is the probability of participating when $Z = z'$.[2] Note that, recalling chapter 4, $P(T = 1|Z = z)$ and $P(T = 1|Z = z')$ can also be interpreted as the propensity scores for participation based on instruments Z—that is, $P(z)$ and $P(z')$, respectively.

The LATE, $\beta_{\text{IV}},_{\text{LATE}}$, can then be written as

$$\beta_{\text{IV, LATE}} = \frac{E(Y|P(Z)=P(z)) - E(Y|P(Z)=P(z'))}{P(z)-P(z')}. \tag{6.7}$$

The denominator in equation 6.7 is the difference in the probability of participating in the program (probability of $T = 1$) under the different values of the instrument, $Z = z$ and $Z = z'$.

Using equation 6.7, one can estimate the LATE using linear IV methods. In the first stage, program participation T is estimated as a function of the instruments Z to obtain the propensity score, $\hat{P}(Z) = \hat{P}(T = 1|Z)$. Second, a linear regression can be estimated of the outcome $Y_i = [T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)]$ on $\hat{P}(Z)$. The interpretation of the estimated program effect $\hat{\beta}_{\text{IV}}$ is the average change in outcomes Y from a change in the estimated propensity score of participating $\hat{P}(Z)$, holding other observed covariates X fixed.

Recent Approaches: Marginal Treatment Effect

The marginal treatment effect (MTE), introduced in chapter 3, is the limit form of the LATE and has been discussed recently (see Heckman and Vytlacil 2005; Todd 2007) as a method for estimating treatment effects when conditional exogeneity does not hold. As mentioned earlier, the MTE is the average gain in outcomes for participants near the threshold or at the margin of participating, given a set of observed characteristics and conditioning on a set of unobserved characteristics in the participation equation. Following Heckman and Vytlacil (2000), the MTE can be written as

$$\text{MTE} = E(Y_i(1) - Y_i(0)|X_i = x,\ U_i = u). \tag{6.8}$$

In equation 6.8, $Y_i(1)$ is the outcome for those under treatment, $Y_i(0)$ is the outcome for those not receiving treatment, $X_i = x$ are observed characteristics for individual i, and $U_i = u$, $U_i \in (0,1)$ are unobserved characteristics for individual i that also determine participation. Looking at the effect of U_i on participation T_i (recall from earlier chapters that $T_i = 1$ for participants and $T_i = 0$ for nonparticipants), Heckman and Vytlacil (2000) assume that T_i is generated by a latent variable T_i^*:[3]

$$\begin{gathered} T_i^* = \mu_T(Z_i) - U_i \\ T_i = 1 \text{ if } T_i^* > 0,\ T_i = 0 \text{ if } T_i^* \leq 0, \end{gathered} \tag{6.9}$$

where Z_i are observed instruments affecting participation and $\mu_T(Z_i)$ is a function determining potential outcomes Y from Z that are conditional on participation. Individuals with unobserved characteristics u close to zero, therefore, are the most likely to participate in the program (T_i closer to 1), and individuals with u close to one are the least likely to participate. The MTE for individuals with $U_i = u$ close to zero therefore reflects the average treatment effect (ATE) for individuals most inclined to participate, and the MTE for individuals with $U_i = u$ close to one represents the ATE for individuals least likely to participate.

Why is the MTE helpful in understanding treatment effects? Also, if both the MTE and the LATE examine the varying impact of unobserved characteristics on participation, what is the difference between them? Both the MTE and the LATE allow for individuals to anticipate gains in Y on the basis of unobserved characteristics. However, just as the LATE is a finer version of the treatment effect on the treated (TOT) (Heckman 1997), the MTE is the limit form of the LATE and defines the treatment effect much more precisely as the LATE for an infinitesimal change in Z (Blundell and Dias 2008; Heckman and Vytlacil 2000).

A useful property of the MTE (see Heckman and Vytlacil 2000, 2005) is that the ATE, TOT, and LATE can all be obtained by integrating under different regions of the MTE. The ATE, which, as discussed in chapter 3, is the average effect for the entire population (that is, the effect of the program for a person randomly drawn from the population), can be obtained by integrating the MTE over the entire support ($u = 0$ to $u = 1$).

The TOT, which is the average treatment effect for those who choose to participate, can be obtained by integrating MTE from $u = 0$ to $u = P(z)$. As described earlier, $P(z)$

is the propensity score, or probability, of participating when the instrument $Z = z$. Thus, the TOT is the treatment effect for individuals whose unobserved characteristics make them most likely to participate in the program.

Finally, if one assumes (as previously) that the instrument Z can take values $Z = z'$ and $Z = z$, and one also assumes that $P(z') < P(z)$, then LATE integrates MTE from $u = P(z')$ to $u = P(z)$. This outcome occurs because, when $P(z') < P(z)$, some individuals who would not have participated when $Z = z'$ will participate when $Z = z$, but no individual who was participating at $Z = z'$ will drop out of the program when $Z = z$.

How, then, to estimate the MTE? Heckman and Vytlacil (2000) propose a two-stage local instrumental variable estimator:

$$\beta_{\text{LIV, MTE}} = \lim_{P(z')\to P(z)} \frac{E(Y|P(Z)=P(z)) - E(Y|P(Z)=P(z'))}{P(z)-P(z')}. \tag{6.10}$$

The approach is similar to the estimation of the LATE previously discussed. In the first stage, program participation is still estimated as a function of the instruments Z to obtain the propensity score $\hat{P}(Z)$. In the second stage, however, a nonparametric local linear regression can be estimated of the outcome $Y_i = [T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)]$ on $\hat{P}(Z)$. Evaluating this function at different values of the propensity score yields the MTE function. Local IV is differevnt from the IV approach used to estimate the LATE, in the sense that local IV estimates the average change in Y around a local neighborhood of $P(Z)$, whereas the LATE is estimated globally over the support (this difference can be seen as well by comparing equations 6.7 and 6.10).

Approaches to estimating the MTE are new and evolving. Moffitt (2008) also proposes a nonparametric method for estimating the MTE. Instead of a two-step procedure where participation is first instrumented and then the average change Y is calculated on the basis of predicted participation, Moffitt estimates the outcome and participation equations jointly through nonlinear least squares. This method relaxes some of the assumptions embedded in the IV and latent linear index models. Very few applications of MTE exist thus far, however, particularly in developing countries.

Sources of IVs

Understanding the factors underlying program targeting and take-up can help in finding appropriate instruments. For example, collecting detailed information on how the program was targeted and implemented can reveal sources of exogenous variation in the program's evolution. This information can be collected for both the baseline and the follow-up quantitative surveys together with qualitative information (stemming from interviews with program officials, for example).

Randomization as a Source of IVs

As discussed in chapter 3, randomization may not perfectly identify participants. Even when randomization takes place at an aggregate (say, regional) level, selection bias may persist in individual take-up. Randomization also does not ensure that targeted subjects will all participate. Nevertheless, if program targeting under this scheme is highly correlated with participation, randomized assignment (which by definition satisfies the exclusion restriction) can still act as an IV. Box 3.2 in chapter 3 describes the use of randomization, even when intention to treat is different from actual take-up of the program.

Nonexperimental Instruments Used in Prior Evaluations: Case Studies

Within a nonrandomized setting, common sources of instruments have included geographic variation, correlation of the program with other policies, and exogenous shocks affecting program placement. Box 6.1 describes how, in the context of the Food for Education program in Bangladesh, geography can be used as a source of instruments. Box 6.2 presents a study from Ghana of improved child health on schooling outcomes. It uses different approaches to address the endogeneity of estimates, including an IV reflecting geographic distance to medical facilities.

Instruments might also be determined from program design, such as eligibility rules or the nature of treatment. Boxes 6.3 and 6.4 discuss examples from Bangladesh and Pakistan, and chapter 7 on discontinuity designs discusses this concept further.

BOX 6.1 Case Study: Using Geography of Program Placement as an Instrument in Bangladesh

In a study on the Food for Education program in Bangladesh, Ravallion and Wodon (2000) examined the claim that child labor displaces schooling and so perpetuates poverty in the longer term. The Food for Education program, in which 2.2 million children were participating in 1995 and 1996, involved targeted subsidies to households to enroll their children in school and was used in the study as the source of a change in the price of schooling in the study's model of schooling and child labor. To address the endogeneity of program placement at the individual level, Ravallion and Wodon used prior program placement at the village level as the IV.

To counter the concern that village placement correlated with geographic factors that might also affect outcomes, Ravallion and Wodon (2000) used administrative assignment rules to construct exogeneity tests that supported their identification strategy. Using a sample of about 2,400 boys and 2,300 girls from the rural sample of the 1995–96 Bangladesh Household Expenditure Survey, the study indicated that the subsidy increased schooling (at the mean of the sample, an extra 100 kilograms of rice increased the probability of going to school by 0.17 for a boy and by 0.16 for a girl) by far more than it reduced child labor. Substitution effects appear to have helped protect current incomes from the higher school attendance induced by the subsidy.

BOX 6.2 **Case Study: Different Approaches and IVs in Examining the Effects of Child Health on Schooling in Ghana**

Glewwe and Jacoby (1995) examined the effects of child health and nutrition on education outcomes in Ghana, including age of enrollment and years of completed schooling. They used cross-sectional data on about 1,760 children 6 to 15 years of age, from 1988 to 1989. In the process, they showed what the options and challenges are for using cross-sections to identify effects.

Given the cross-section data, unobserved characteristics of parents (such as preferences) may correlate across both child health and education. One of the approaches in the study by Glewwe and Jacoby (1995) was to seek instruments that affect child health characteristics (such as height-for-age anthropometric outcomes) but are not correlated with unobserved family characteristics affecting child education. They proposed as instruments for child health (a) distance to the closest medical facility and (b) maternal height. Both justifiably correlate with child health, but Glewwe and Jacoby also point out that mother's height could affect her labor productivity and, hence, household income and the resulting time she has to spend on her children's education. Distance to nearby medical facilities could also correlate with other community characteristics, such as presence of schools. Both of these caveats weaken the assumption that cov(Z, ε) = 0. From the IV estimates, as well as alternate estimates specifying fixed effects for families, Glewwe and Jacoby found strong negative effects of child health on delayed enrollment but no statistically significant effect on completed years of schooling.

BOX 6.3 **Case Study: A Cross-Section and Panel Data Analysis Using Eligibility Rules for Microfinance Participation in Bangladesh**

Pitt and Khandker (1998) studied the impact of microfinance programs in Bangladesh to assess the impact of participation by men versus women on per capita expenditure, schooling enrollment of children, and other household outcomes. They used a quasi-experimental data set from 1991 to 1992 of about 1,800 households across a random sample of 29 thanas (about 1,540 households from 24 thanas targeted by credit initiatives, and the remainder from 5 nontargeted thanas). Of the targeted households, about 60 percent were participating in microcredit programs.

As the source of identification, Pitt and Khandker (1998) relied on exogenous eligibility conditions based on household landholding (specifically, an eligibility cutoff of one-half acre of land owned) as a way of identifying program effects. The fact that men could participate only in men's groups and women only in women's groups added another constraint on which impacts could be identified. Village fixed effects (for example, to account for why some villages have just men-only groups and other villages have just female-only groups) were also included in the estimations. Pitt and Khandker found that when women are the program participants, program credit has a larger impact on household outcomes, including an increase in annual household expenditure of Tk 18, compared with Tk 11 for men.

Some of the conditions, however, are restrictive and might not be reliable (for example, the nonenforceability of the landholding criterion for program participation). An impact assessment can be carried out using a follow-up survey to test the sensitivity of the findings. As discussed at the beginning of this chapter, Khandker (2006) used the 1998–99 follow-up survey to the 1991–92

(*Box continues on the following page.*)

BOX 6.3 Case Study: A Cross-Section and Panel Data Analysis Using Eligibility Rules for Microfinance Participation in Bangladesh (continued)

survey to assess the sensitivity of the earlier findings on the poverty effects of microfinance in rural Bangladesh. The panel data analysis helps to estimate the effects on poverty using an alternative estimation technique and also helps to estimate the impacts of past and current borrowing, assuming that gains from borrowing, such as consumption gains, vary over time. The instrument is whether the household qualifies to participate in the program on the basis of the landholding criteria. The instrumented decision to participate is then interacted with household-level exogenous variables and village fixed effects.

Khandker's (2006) follow-up study found that the average returns to cumulative borrowing for female members of microfinance programs are as much as 21 percent in 1998–99, up from 18 percent in 1991–92. However, the impact on poverty reduction among program participants was lower in 1998–99 (2 percentage points) than in 1991–92 (5 percentage points). This result is due to diminishing returns to additional borrowing, so that despite the increase in the stock of borrowing by female members, the resulting increases in consumption were not large enough to reduce poverty as expected.

BOX 6.4 Case Study: Using Policy Design as Instruments to Study Private Schooling in Pakistan

As another example, Andrabi, Das, and Khwaja (2006) examined the effect of private schooling expansion in Pakistan during the 1990s on primary school enrollment. The growth in private schools exhibited variation that the study exploited to determine causal impacts. Specifically, using data from a sample of about 18,000 villages in rural Punjab province (spanning data from national censuses of private schools, village-level socioeconomic characteristics from 1981 and 2001, and administrative data on the location and date of public schools), Andrabi, Das, and Khwaja found that private schools were much more likely to set up in villages where public girls' secondary schools (GSS) had already been established.

To obtain an identifying instrument for private school expansion, Andrabi, Das, and Khwaja (2006) therefore exploited official eligibility rules for placement of GSS across villages. Specifically, villages with larger population were given preference for construction of GSS, as long as no other GSS were located within a 10-kilometer radius. The study also exploited an administrative unit called a *Patwar Circle* (PC), which was four or five contiguous villages roughly spanning a 10-kilometer radius. From historical records, Andrabi, Das, and Khwaja determined that PCs were primarily defined for revenue purposes. The IV estimate would be unbiased if (a) private school placement did not follow the same discontinuous relationship with local population and (b) unobserved characteristics of PCs with the highest population rank were also not correlated with private school expansion as well as educational market outcomes. If the latter were not true, for example, then $\text{cov}(Z, \varepsilon) \neq 0$.

Andrabi, Das, and Khwaja (2006) found that a public girls' secondary school increased the likelihood of a private school in the village by 35 percent. However, they found little or no relationship between the placement of these private schools and preexisting coeducational primary

BOX 6.4 Case Study: Using Policy Design as Instruments to Study Private Schooling in Pakistan (continued)

schools or secondary schools for boys. Robustness checks using propensity score matching on the baseline data compared the change in private schools and GSS for matching villages; the existence of GSS raised the probability that private schools would be introduced by 11 to 14 percent. Regarding the program effect on outcomes, using data from about 7,000 villages, they found that preexisting public GSS roughly doubled the supply of local skilled women. However, with few earning opportunities for women, overall wages for women fell by about 18 percent, as did teaching costs for private schools.

Notes

1. These data sets are also used in the Stata exercises in part 2 of the handbook.
2. As discussed earlier, T is the treatment variable equal to 1 for participants and equal to 0 for nonparticipants. Outcomes Y and participation T are also functions of other observed covariates X, which have been suppressed for simplicity in equation 6.7.
3. This equation is also known as a linear latent index model (see Heckman and Hotz 1989; Heckman and Robb 1985; Imbens and Angrist 1994).

References

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2006. "Students Today, Teachers Tomorrow? Identifying Constraints on the Provision of Education." Harvard University, Cambridge, MA.

Angrist, Joshua. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records." *American Economic Review* 80 (3): 313–35.

Blundell, Richard, and Monica Costa Dias. 2008. "Alternative Approaches to Evaluation in Empirical Microeconomics." CeMMAP Working Paper 26/08, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, London.

Glewwe, Paul, and Hanan G. Jacoby. 1995. "An Economic Analysis of Delayed Primary School Enrollment in a Low Income Country: The Role of Early Childhood Nutrition." *Review of Economic Statistics* 77 (1): 156–69.

Heckman, James J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–62.

Heckman, James J., and V. Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84 (408): 862–74.

Heckman, James J., and Richard Robb. 1985. "Alternative Methods for Estimating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, ed. James Heckman and Burton Singer, 156–245. New York: Cambridge University Press.

Heckman, James J., and Edward J. Vytlacil. 2000. "Causal Parameters, Structural Equations, Treatment Effects, and Randomized Evaluations of Social Programs." University of Chicago, Chicago, IL.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3): 669–738.

Imbens, Guido, and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–76.

Khandker, Shahidur R. 2006. "Microfinance and Poverty: Evidence Using Panel Data from Bangladesh." *World Bank Economic Review* 19 (2): 263–86.

Moffitt, Robert. 2008. "Estimating Marginal Treatment Effects in Heterogeneous Populations." Economic Working Paper Archive 539, Johns Hopkins University, Baltimore, MD. http://www.econ.jhu.edu/people/moffitt/welfls0_v4b.pdf.

Morduch, Jonathan. 1998. "Does Microfinance Really Help the Poor? New Evidence on Flagship Programs in Bangladesh." Princeton University, Princeton, NJ.

Pitt, Mark, and Shahidur Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy* 106 (5): 958–98.

Ravallion, Martin, and Quentin Wodon. 2000. "Does Child Labour Displace Schooling? Evidence on Behavioural Responses to an Enrollment Subsidy." *Economic Journal* 110 (462): 158–75.

Semykina, Anastasia, and Jeffrey M. Wooldridge. 2005. "Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application." Working Paper, Michigan State University, East Lansing, MI.

Todd, Petra. 2007. "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated." In *Handbook of Development Economics*, vol. 4, ed. T. Paul Schultz and John Strauss, 3847–94. Amsterdam: North-Holland.

Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.