5. Double Difference

Summary

Double-difference (DD) methods, compared with propensity score matching (PSM), assume that unobserved heterogeneity in participation is present—but that such factors are time invariant. With data on project and control observations before and after the program intervention, therefore, this fixed component can be differenced out.

Some variants of the DD approach have been introduced to account for potential sources of selection bias. Combining PSM with DD methods can help resolve this problem, by matching units in the common support. Controlling for initial area conditions can also resolve nonrandom program placement that might bias the program effect. Where a baseline might not be available, using a triple-difference method with an entirely separate control experiment after program intervention (that is, a separate set of untreated observations) offers an alternate calculation of the program's impact.

Learning Objectives

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

- How to construct the double-difference estimate
- How to address potential violations of the assumption of time-invariant heterogeneity
- How to account for nonrandom program placement
- What to do when a baseline is not available

Addressing Selection Bias from a Different Perspective: Using Differences as Counterfactual

The two methods discussed in the earlier chapters—randomized evaluation and PSM—focus on various single-difference estimators that often require only an appropriate cross-sectional survey. This chapter now discusses the double-difference estimation technique, which typically uses panel data. Note, however, that DD can be used on repeated cross-section data as well, as long as the composition of participant and control groups is fairly stable over time.

In a panel setting, DD estimation resolves the problem of missing data by measuring outcomes and covariates for both participants and nonparticipants in pre- and postintervention periods. DD essentially compares treatment and comparison groups in terms of outcome changes over time relative to the outcomes observed for a preintervention baseline. That is, given a two-period setting where $t = 0$ before the program and $t = 1$ after program implementation, letting Y_t^T and Y_t^C be the respective outcomes for a program beneficiary and nontreated units in time t, the DD method will estimate the average program impact as follows:

$$\mathrm{DD} = E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0). \tag{5.1}$$

In equation 5.1, $T_1 = 1$ denotes treatment or the presence of the program at $t = 1$, whereas $T_1 = 0$ denotes untreated areas. The following section returns to this equation. Unlike PSM alone, the DD estimator allows for *unobserved heterogeneity* (the unobserved difference in mean counterfactual outcomes between treated and untreated units) that may lead to selection bias. For example, one may want to account for factors unobserved by the researcher, such as differences in innate ability or personality across treated and control subjects or the effects of nonrandom program placement at the policy-making level. DD assumes this unobserved heterogeneity is time invariant, so the bias cancels out through differencing. In other words, the outcome changes for nonparticipants reveal the counterfactual outcome changes as shown in equation 5.1.

DD Method: Theory and Application

The DD estimator relies on a comparison of participants and nonparticipants before and after the intervention. For example, after an initial baseline survey of both nonparticipants and (subsequent) participants, a follow-up survey can be conducted of both groups after the intervention. From this information, the difference is calculated between the observed mean outcomes for the treatment and control groups before and after program intervention.

When baseline data are available, one can thus estimate impacts by assuming that unobserved heterogeneity is time invariant and uncorrelated with the treatment over time. This assumption is weaker than conditional exogeneity (described in chapters 2 and 3) and renders the outcome changes for a comparable group of nonparticipants (that is, $E(Y_1^C - Y_0^C \mid T_1 = 0)$) as the appropriate counterfactual, namely, equal to $E(Y_1^C - Y_0^C \mid T_1 = 1)$.[1] Nevertheless, justifiable concerns exist with this assumption that are brought up later in this chapter.

The DD estimate can also be calculated within a regression framework; the regression can be weighted to account for potential biases in DD (discussed in later sections in this chapter). In particular, the estimating equation would be specified as follows:

$$Y_{it} = \alpha + \beta T_{i1} t + \rho T_{i1} + \gamma t + \varepsilon_{it}. \tag{5.2}$$

In equation 5.2, the coefficient β on the interaction between the postprogram treatment variable (T_{i1}) and time ($t = 1\ldots T$) gives the average DD effect of the program. Thus, using the notation from equation 5.1, $\beta = \text{DD}$. In addition to this interaction term, the variables T_{i1} and t are included separately to pick up any separate mean effects of time as well as the effect of being targeted versus not being targeted. Again, as long as data on four different groups are available to compare, panel data are not necessary to implement the DD approach (for example, the t subscript, normally associated with time, can be reinterpreted as a particular geographic area, $k = 1\ldots K$).

To understand the intuition better behind equation 5.2, one can write it out in detail in expectations form (suppressing the subscript i for the moment):

$$E(Y_1^T - Y_0^T \mid T_1 = 1) = (\alpha + \text{DD} + \rho + \gamma) - (\alpha + \rho) \tag{5.3a}$$

$$E(Y_1^C - Y_0^C \mid T_1 = 0) = (\alpha + \gamma) - \alpha. \tag{5.3b}$$

Following equation 5.1, subtracting 5.3b from 5.3a gives DD. Note again that DD is unbiased only if the potential source of selection bias is *additive* and *time invariant.* Using the same approach, if a simple pre- versus postestimation impact on the participant sample is calculated (a reflexive design), the calculated program impact would be DD + γ, and the corresponding bias would be γ.[2] As discussed in chapter 2, without a control group, justifying that other factors were not responsible in affecting participant outcomes is difficult. One might also try comparing just the postprogram difference in outcomes across treatment and control units; however, in this case, the estimated impact of the policy would be DD + ρ, and the bias would be ρ. Systematic, unmeasured differences that could be correlated with treatment cannot be separated easily.

Remember that for the above DD estimator to be interpreted correctly, the following must hold:

1. The model in equation (outcome) is correctly specified. For example, the additive structure imposed is correct.
2. The error term is uncorrelated with the other variables in the equation:

$$Cov(\varepsilon_{it}, T_{i1}) = 0$$

$$Cov(\varepsilon_{it}, t) = 0$$

$$Cov(\varepsilon_{it}, T_{i1}t) = 0.$$

The last of these assumptions, also known as the *parallel-trend* assumption, is the most critical. It means that unobserved characteristics affecting program participation do not vary over time with treatment status.

Panel Fixed-Effects Model

The preceding two-period model can be generalized with multiple time periods, which may be called the *panel fixed-effects model*. This possibility is particularly important for a model that controls not only for the unobserved time-invariant heterogeneity but also for heterogeneity in observed characteristics over a multiple-period setting. More specifically, Y_{it} can be regressed on T_{it}, a range of time-varying covariates X_{it}, and unobserved time-invariant individual heterogeneity η_t that may be correlated with both the treatment and other unobserved characteristics ε_{it}. Consider the following revision of equation 5.2:

$$Y_{it} = \phi T_{it} + \delta X_{it} + \eta_i + \varepsilon_{it}. \tag{5.4}$$

Differencing both the right- and left-hand side of equation 5.4 over time, one would obtain the following differenced equation:

$$(Y_{it} - Y_{it-1}) = \phi(T_{it} - T_{it-1}) + \delta(X_{it} - X_{it-1}) + (\eta_i - \eta_i) + (\varepsilon_{it} - \varepsilon_{it-1}) \tag{5.5a}$$

$$\Rightarrow \Delta Y_{it} = \phi \Delta T_{it} + \delta \Delta X_{it} + \Delta \varepsilon_{it} \tag{5.5b}$$

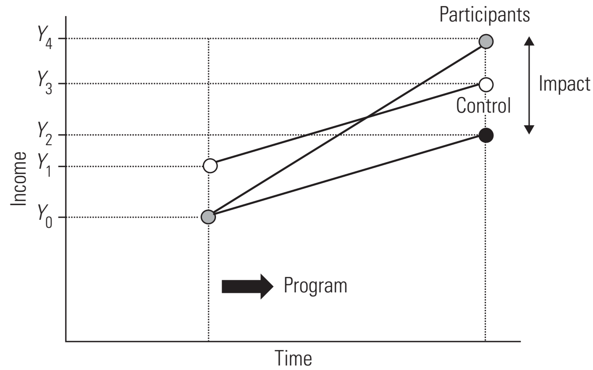
In this case, because the source of endogeneity (that is, the unobserved individual characteristics η_t) is dropped from differencing, ordinary least squares (OLS) can be applied to equation 5.5b to estimate the unbiased effect of the program (ϕ). With two time periods, ϕ is equivalent to the DD estimate in equation 5.2, controlling for the same covariates X_{it}; the standard errors, however, may need to be corrected for serial correlation (Bertrand, Duflo, and Mullainathan 2004). With more than two time periods, the estimate of the program impact will diverge from DD.

Implementing DD

To apply a DD approach using panel data, baseline data need to be collected on program and control areas before program implementation. As described in chapter 2, quantitative as well as qualitative information on these areas will be helpful in determining who is likely to participate. Follow-up surveys after program intervention also should be conducted on the same units.[3] Calculating the average difference in outcomes separately for participants and nonparticipants over the periods and then taking an additional difference between the average changes in outcomes for these two groups will give the DD impact. An example is shown in figure 5.1: DD = $(Y_4 - Y_0) - (Y_3 - Y_1)$.

The lowermost line in figure 5.1 also depicts the true counterfactual outcomes, which are never observed (see chapter 2). Under the DD approach, unobserved characteristics that create a gap between measured control outcomes and true counterfactual outcomes are assumed to be time invariant, such that the gap between the two trends is

Figure 5.1 An Example of DD



Source: Authors' representation.

the same over the period. This assumption implies that $(Y_3 - Y_2) = (Y_1 - Y_0)$. Using this equality in the preceding DD equation, one gets DD = $(Y_4 - Y_2)$.

One application of DD estimation comes from Thomas and others (2004). They examine a program in Indonesia that randomized the distribution of iron supplements to individuals in primarily agricultural households, with half the respondents receiving treatment and controls receiving a placebo. A baseline was also conducted before the intervention. Using DD estimation, the study found that men who were iron deficient before the intervention experienced improved health outcomes, with more muted effects for women. The baseline was also useful in addressing concerns about bias in compliance with the intervention by comparing changes in outcomes among subjects assigned to the treatment group relative to changes among subjects assigned to the control group.

As another example, Khandker, Bakht, and Koolwal (2009) examine the impact of two rural road-paving projects in Bangladesh, using a quasi-experimental household panel data set surveying project and control villages before and after program implementation. Both project and control areas shared similar socioeconomic and community-level characteristics before program implementation; control areas were also targets for future rounds of the road improvement program. Each project had its own survey, covered in two rounds—the first in the mid-1990s before the projects began and the second about five years later after program completion. DD estimation was used to determine the program's impacts across a range of outcomes, including household per capita consumption (a measure of household welfare), prices, employment outcomes for men and women, and children's school enrollment. Using an additional fixed-effects approach that accounted for initial conditions, the study found that households had benefited in a variety of ways from road investment.

Although DD typically exploits a baseline and resulting panel data, repeated cross-section data over time can also be used. Box 5.1 describes the use of different data sources in a study of a conditional cash-transfer program in Pakistan.

Advantages and Disadvantages of Using DD

The advantage of DD is that it relaxes the assumption of conditional exogeneity or selection only on observed characteristics. It also provides a tractable, intuitive way to account for selection on unobserved characteristics. The main drawback, however, rests precisely with this assumption: the notion of time-invariant selection bias is implausible for many targeted programs in developing countries. The case studies discussed here and in earlier chapters, for example, reveal that such programs often have wide-ranging approaches to poverty alleviation that span multiple sectors. Given that such programs are also targeted in areas that are very poor and have low initial growth, one might expect over several years that the behavior and choices of targeted areas would respond dynamically (in both observed and unobserved ways) to the program. Training programs, which are also widely examined in the evaluation literature, provide

BOX 5.1 Case Study: DD with Panel Data and Repeated Cross-Sections

Aside from panel data, repeated cross-section data on a particular area may be pooled to generate a larger sample size and to examine how characteristics of the sample are broadly changing over time. Chaudhury and Parajuli (2006) examined the effects of the Female School Stipend Program in the Punjab province of Pakistan on public school enrollment, using school-level panel data from 2003 (before the program) and 2005 (after the program), as well as repeated cross-section data at the child level between 2001–02 and 2004–05.

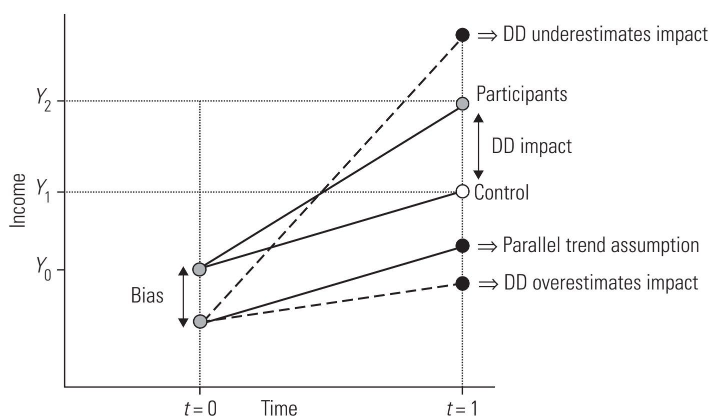
Under the program, girls received a PRs 200 stipend conditional on being enrolled in grades 6 through 8 in a girls' public secondary school within targeted districts and maintaining average class attendance of at least 80 percent. The program was specifically targeted toward low-literacy districts and was not randomly assigned. As part of their analysis, Chaudhury and Parajuli (2006) used both panel and repeated cross-section data to calculate separate difference-in-difference program impacts on girls' enrollment, assuming time-invariant unobserved heterogeneity.

The panel data were drawn from the provincial school censuses across 15 districts receiving the stipend program (covering about 1,780 schools) and 19 control districts (spanning about 3,150 schools) where the program was not available. Using these data, the researchers found the program increased girls' enrollment by about 23 percent. The child-level cross-section data over time were drawn from household surveys and were considered to be potentially more objective relative to the school administrative data. The study found, for 10- to 14-year-old girls, that the average effect of the program ranged from 10 to 13 percentage points. Compared with the panel data regressions, the corresponding regressions with the pooled cross-section data included interaction terms of the stipend program dummy as well as a postprogram time dummy with whether the child was female.

another example. Suppose evaluating the impact of a training program on earnings is of interest. Enrollment may be more likely if a temporary (perhaps shock-induced) slump in earnings occurs just before introduction of the program (this phenomenon is also known as *Ashenfelter's Dip*). Thus, the treated group might have experienced faster growth in earnings even without participation. In this case, a DD method is likely to overestimate the program's effect.[4] Figure 5.2 reflects this potential bias when the difference between nonparticipant and counterfactual outcomes changes over time; time-varying, unobserved heterogeneity could lead to an upward or downward bias.

In practice, ex ante, time-varying unobserved heterogeneity could be accounted for with proper program design, including ensuring that project and control areas share similar preprogram characteristics. If comparison areas are not similar to potential participants in terms of observed and unobserved characteristics, then changes in the outcome over time may be a function of this difference. This factor would also bias the DD. For example, in the context of a school enrollment program, if control areas were selected that were initially much farther away from local schools than targeted areas, DD would overestimate the program's impact on participating localities. Similarly, differences in agroclimatic conditions and initial infrastructural development across treated and control areas may also be correlated with program placement and resulting changes in outcomes over time. Using data from a poverty-alleviation program in China, Jalan and Ravallion (1998) show that a large bias exists in the DD estimate of the project's impact because changes over time are a function of initial conditions that also influence program placement. Controlling for the area characteristics that initially attracted the development projects can correct for this bias; by doing so, Jalan and Ravallion found significant longer-term impacts whereas none had been evident in the standard DD estimator. The next section discusses this issue in more detail.

Figure 5.2 Time-Varying Unobserved Heterogeneity



Source: Authors' representation.

As discussed in chapter 4, applying PSM could help match treatment units with observationally similar control units before estimating the DD impact. Specifically, one would run PSM on the base year and then conduct a DD on the units that remain in the common support. Studies show that weighting the control observations according to their propensity score yields a fully efficient estimator (Hirano, Imbens, and Ridder 2003; also see chapter 4 for a discussion). Because effective PSM depends on a rich baseline, however, during initial data collection careful attention should be given to characteristics that determine participation.

Even if comparability of control and project areas could be ensured before the program, however, the DD approach might falter if macroeconomic changes during the program affected the two groups differently. Suppose some unknown characteristics make treated and nontreated groups react differently to a common macroeconomic shock. In this case, a simple DD might overestimate or underestimate the true effects of a program depending on how the treated and nontreated groups react to the common shock. Bell, Blundell, and van Reenen (1999) suggest a differential time-trend-adjusted DD for such a case. This alternative will be discussed later in terms of the triple-difference method. Another approach might be through instrumental variables, which are discussed in chapter 6. If enough data are available on other exogenous or behavior-independent factors affecting participants and nonparticipants over time, those factors can be exploited to identify impacts when unobserved heterogeneity is not constant. An instrumental variables panel fixed-effects approach could be implemented, for example; chapter 6 provides more detail.

Alternative DD Models

The double-difference approach described in the previous section yields consistent estimates of project impacts if unobserved community and individual heterogeneity are time invariant. However, one can conceive of several cases where unobserved characteristics of a population may indeed change over time—stemming, for example, from changes in preferences or norms over a longer time series. A few variants of the DD method have therefore been proposed to control for factors affecting these changes in unobservables.

Do Initial Conditions Matter?

One case in which unobserved heterogeneity may not remain constant over time is where public investments depend on initial (preprogram) local area conditions. Not controlling for initial area conditions when assessing the impact of an antipoverty program may lead to significant omitted variable bias—if local conditions were also responsible for the improvement of household outcomes or program targeting was correlated with such area characteristics. Approaches to controlling for initial area conditions in a DD approach, using data over multiple years as well as data covering only two time periods, are discussed in box 5.2.

BOX 5.2 **Case Study: Accounting for Initial Conditions with a DD Estimator—Applications for Survey Data of Varying Lengths**

Long-Term Data with Multiple Rounds

Jalan and Ravallion (1998) examined the impact of a development program in a poor area on growth in household consumption by using panel data from targeted and nontargeted areas across four contiguous provinces in southwest China. Using data on about 6,650 households between 1985 and 1990 (supplemented by additional field visits in 1994–95), they employed a generalized method-of-moments time-series estimation model for household consumption growth, including initial area conditions on the right-hand side and using second and higher lags of consumption as instruments for lagged consumption to obtain consistent estimates of a dynamic growth model with panel data.

Their results show that program effects are indeed influenced by initial household and community wealth; dropping initial area conditions (such as initial wealth and fertilizer use) caused the national program effect to lose significance completely, with provincial program effects changing sign and becoming slightly negative. In particular, after correcting for the area characteristics that initially attracted the development projects, Jalan and Ravallion (1998) found significant longer-term impacts than those obtained using simple fixed-effects methods. Thus, failing to control for factors underlying potential differences in local and regional growth trajectories can lead to a substantial underestimation of the welfare gains from the program.

Data with Two Time Periods

With fewer time periods (for example, with two years) a simpler OLS-first difference model can be applied on the data, incorporating a range of initial area characteristics across project and control areas prior to program implementation. In their study on rural roads (discussed later in this chapter), Khandker, Bakht, and Koolwal (2009) used two rounds of data—namely, baseline and postprogram data on treated and control areas—to compare DD results based on a household fixed-effects approach with OLS-first difference estimations on the same outcomes and covariates. These OLS-first difference estimates control for a number of preproject characteristics of villages where households were located. These initial area characteristics included local agroclimatic factors; the number of banks, schools, and hospitals serving the village; the distance from the village to the nearest paved road; the average short-term interest rate in the village; and the number of active microfinance institutions in the village.

Although the project estimates are similar across both specifications for a number of outcomes, the study found that the beneficial household impact of the program was also strengthened for many outcomes when initial area conditions were controlled for. Because the project's effect did not disappear for most outcomes after initial area conditions were controlled for, the study provides one indication that program targeting was not entirely predisposed toward certain areas with particular initial development characteristics.

PSM with DD

As mentioned earlier, provided that rich data on control and treatment areas exist, PSM can be combined with DD methods to better match control and project units on preprogram characteristics. Specifically, recalling the discussion in chapter 4, one notes that the propensity score can be used to match participant and control units in the base (preprogram) year, and the treatment impact is calculated across participant and matched control units within the common support. For two time periods $t = \{1,2\}$, the DD estimate for each treatment area i will be calculated as $\mathrm{D}D_i = (Y_{i2}^T - Y_{i1}^T) - \sum_{j \in C} \omega(i,j)(Y_{j2}^C - Y_{j1}^C)$, where $\omega(i,j)$ is the weight (using a PSM approach) given to the jth control area matched to treatment area i. Different types of matching approaches discussed in chapter 4 can be applied.

In terms of a regression framework (also discussed in chapter 4), Hirano, Imbens, and Ridder (2003) show that a weighted least squares regression, by weighting the control observations according to their propensity score, yields a fully efficient estimator:

$$\Delta Y_{it} = \alpha + \beta T_i + \gamma \Delta X_{it} + \varepsilon_{it}, \beta = DD. \quad (5.6)$$

The weights in the regression in equation 5.6 are equal to 1 for treated units and to $\hat{P}(X)/(1 - \hat{P}(X))$ for comparison units. See box 5.3 for a case study.

Triple-Difference Method

What if baseline data are not available? Such might be the case during an economic crisis, for example, where a program or safety net has to be set up quickly. In this context, a triple-difference method can be used. In addition to a "first experiment" comparing certain project and control groups, this method exploits the use of an entirely separate control experiment after program intervention. That is, this separate control group reflects a set of nonparticipants in treated and nontreated areas that are not part of the

BOX 5.3 Case Study: PSM with DD

In a study on rural road rehabilitation in Vietnam, van de Walle and Mu (2008) controlled for time-invariant unobserved heterogeneity and potential time-varying selection bias attributable to differences in initial observable characteristics by combining DD and PSM using data from 94 project and 95 control communes over three periods: a baseline survey in 1997, followed by surveys in 2001 and 2003.

Highlighting the importance of comparing short-term versus long-term impacts, the study found that most outcomes were realized at different stages over the period. Primary school completion, for example, reflected sustained growth between 1997 and 2003, increasing by 15 to 25 percent. Other outcomes, such as the expansion of markets and availability of non-food-related goods, took a longer time to emerge (markets, for example, developed in about 10 percent more project than control communes after seven years) than did short-run effects such as the number of secondary schools and availability of food-related goods. Moreover, van de Walle and Mu found that market impacts were greater if the commune initially was poorly developed.

first control group. These new control units may be different from the first control group in socioeconomic characteristics if evaluators want to examine the project's impact on participants relative to another socioeconomic group. Another difference from the first experiment would then be taken from the change in the additional control sample to examine the impact of the project, accounting for other factors changing over time (see, for example, Gruber 1994). This method would therefore require data on multiple years after program intervention, even though baseline data were missing.

Box 5.4 discusses an example of a triple-difference approach from Argentina, where Ravallion and others (2005) examine program impacts on income for "stayers" versus "leavers" in the Trabajar workfare program in Argentina (see chapter 4 for a discussion of the program). Given that the program was set up shortly after the 1997 financial crisis, baseline data were not available. Ravallion and others therefore

BOX 5.4 **Case Study: Triple-Difference Method—Trabajar Program in Argentina**

Lacking baseline data for the Trabajar program, and to avoid making the assumption that stayers and leavers had similar opportunities before joining the program, Ravallion and others (2005) proposed a triple-difference estimator, using an entirely separate control group that never participated in the program (referred to as *nonparticipants* here). The triple-difference estimator is first calculated by taking the DD between matched stayers and nonparticipants and then the DD for matched leavers and nonparticipants. Finally, the DD of these two sets of groups is calculated across matched stayers and leavers.

Specifically, letting $D_{it} = 1$ and $D_{it} = 0$ correspond to participants and matched nonparticipants, respectively, in round t, $t = \{1,2\}$, the study first calculated the DD estimates $A = \left[(\bar{Y}_2^T - \bar{Y}_1^T) - (\bar{Y}_2^C - \bar{Y}_1^C) \middle| D_{i2} = 1\right]$ (corresponding to the stayers in period 2, matched with nonparticipants from the separate urban survey) and $B = \left[(\bar{Y}_2^T - \bar{Y}_1^T) - (\bar{Y}_2^C - \bar{Y}_1^C) \middle| D_{i2} = 0\right]$ (corresponding to the leavers in period 2, matched with nonparticipants). The triple-difference estimator was then calculated as $A - B$.

Ravallion and others (2005) used a sample of 419 stayers matched with 400 leavers (originally taken from a random sample of 1,500 Trabajar workers), surveyed in May and June 1999, October and November 1999, and May and June 2000. Nonparticipants were drawn from a separate urban household survey conducted around the same time, covering a range of socioeconomic characteristics; this survey was conducted twice a year and covered about 27,000 households.

Interpreting the triple-difference estimate as a measure of the average gains to participants, however, requires that (a) there was no selection bias in dropping out from the program and (b) there were no current income gains to nonparticipants. Ravallion and others (2005) used a third round of the survey to test these conditions jointly, comparing the triple-difference estimate for those who dropped out and for those who stayed in the program. Using this test, they were not able to reject the conditions required for using the triple-difference measure as an estimate of the gains to current participants. They also found evidence of an Ashenfelter's Dip, where people were able to recover an increasing share of the Trabajar wage after dropping out of the program as time went on.

examine the difference in incomes for participants leaving the program and those still participating, after differencing out aggregate economywide changes by using an entirely separate, matched group of nonparticipants. Without the matched group of nonparticipants, a simple DD between stayers and leavers will be unbiased only if counterfactual earnings opportunities outside of the program were the same for each group. However, as Ravallion and others (2005) point out, individuals who choose to remain in the program may intuitively be less likely to have better earnings opportunities outside the program than those who dropped out early. As a result, a DD estimate comparing just these two groups will underestimate the program's impact. Only in circumstances such as an exogenous program contraction, for example, can a simple DD between stayers and leavers work well.

Adjusting for Differential Time Trends

As mentioned earlier, suppose one wants to evaluate a program such as employment training introduced during a macroeconomic crisis. With data available for treated and nontreated groups before and after the program, one could use a DD approach to estimate the program's effect on earnings, for example. However, such events are likely to create conditions where the treated and nontreated groups would respond differently to the shock. Bell, Blundell, and van Reenen (1999) have constructed a DD method that accounts for these differential time-trend effects. Apart from the data on treated and nontreated groups before and after treatment, another time interval is needed (say, $t - 1$ to t) for the same treated and nontreated groups. The recent past cycle is likely the most appropriate time interval for such comparison. More formally, the time-trend-adjusted DD is defined as follows:

$$\begin{aligned} DD = {} & [E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0)] \\ & - [E(Y_t^T - Y_{t-1}^T \mid T_1 = 1) - E(Y_t^C - Y_{t-1}^C \mid T_1 = 0)] \end{aligned} \tag{5.7}$$

Notes

1. Refer to chapter 2 for an introductory discussion of the role of the counterfactual in specifying the treatment effect of a program.
2. Note that when the counterfactual means are time invariant ($E[Y_1^C - Y_0^C \mid T_1 = 1] = 0$), the DD estimate in equation 5.1 becomes a reflexive comparison where only outcomes for the treatment units are monitored. Chapter 2 also discusses reflexive comparisons in more detail. This approach, however, is limited in practice because it is unlikely that the mean outcomes for the counterfactual do not change.
3. Although some large-scale studies are not able to revisit the same households or individuals after program intervention, they can survey the same villages or communities and thus are able to calculate DD program impacts at the local or community level. Concurrent surveys at the beneficiary and community levels are important in maintaining this flexibility, particularly because surveys before and after program intervention can span several years, making panel data collection more difficult.
4. A similar argument against the DD method applies in the case of evaluating a program using repeated cross-sectional survey data. That is, if individuals self-select into a program according to some unknown rule and repeated cross-section data are used, the assumption of time-invariant heterogeneity may fail if the composition of the group changes and the intervention affects the composition of treated versus nontreated groups.

References

Bell, Brian, Richard Blundell, and John van Reenen. 1999. "Getting the Unemployed Back to Work: An Evaluation of the New Deal Proposals." *International Tax and Public Finance* 6 (3): 339–60.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.

Chaudhury, Nazmul, and Dilip Parajuli. 2006. "Conditional Cash Transfers and Female Schooling: The Impact of the Female School Stipend Program on Public School Enrollments in Punjab, Pakistan." Policy Research Working Paper 4102, World Bank, Washington, DC.

Gruber, Jonathan, 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84 (3): 622–41.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Jalan, Jyotsna, and Martin Ravallion. 1998. "Are There Dynamic Gains from a Poor-Area Development Program?" *Journal of Public Economics* 67 (1):65–85.

Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal. 2009. "The Poverty Impacts of Rural Roads: Evidence from Bangladesh." *Economic Development and Cultural Change* 57 (4): 685–722.

Ravallion, Martin, Emanuela Galasso, Teodoro Lazo, and Ernesto Philipp. 2005. "What Can Ex-Participants Reveal about a Program's Impact?" *Journal of Human Resources* 40 (1): 208–30.

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohammed Hakimi, Jaswadi, Nathan Jones, Christopher McKelvey, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, and Siswanto Wilopo. 2004. "Iron Deficiency and the Well-Being of Older Adults: Preliminary Results from a Randomized Nutrition Intervention." University of California–Los Angeles, Los Angeles, California.

van de Walle, Dominique, and Ren Mu. 2008. "Rural Roads and Poor Area Development in Vietnam." Policy Research Working Paper 4340, World Bank, Washington, DC.