

PhD Program: Development Economics (Micro)

Lecture 1: Methods

Måns Söderbom*

4 November 2009

1. Introduction & Organization

- This is a course in development economics with a focus on issues that lend themselves to micro-economic analysis - outcomes and decisions observed at the level of the household, the farm, or the enterprise.
- For example:
 - Household decisions with regards to schooling and health
 - Consequences of such decisions for income (earnings) and fertility
 - Causes and consequences of differences in the economic performance across firms and farms
- We pay special attention to policy in this context - what can policy makers do; how can we assess the outcomes; what are the outcomes?
- Macro development is discussed as part of a separate course, in the spring.
- Most of the research that we will discuss is empirically oriented. This reflects the situation in the field today. Grand theoretical models of development (e.g. the Lewis model of structural change) appears to be a thing of the past.
- In recent years, the 'treatment' framework has become very popular in development micro economics, especially when it comes to analyzing household decisions and outcomes. This lecture discusses this approach in some detail. Those of you who took my course in Applied Econometrics will recognize a lot of the material here. However, it is not assumed that you have background knowledge in this area.

1.1. Mechanics & Examination

- To get the course credits you need to:
 - Complete two computer exercises (one on Progresia; one on firm performance)

- Participate in two workshops (one on randomization; one on agricultural development)
- Pass a take home exam (January)
- Grades (Fail, Pass or High Pass) will be based on the performance in the above components.
- Some changes to the schedule:
 - Workshop: Student presentations 1 moves from 14-16 to 10-12, November 18th.
 - Workshop: Student presentations 2 moves from 14-16 to 10-12, November 25th.
 - Peter Martinsson will give a 2-hour lecture on field experiments in development. Details to come.
- Check the course web page for updates every now and then.

2. Program Evaluation in Development

References:

Ravallion, Martin (2008). "Evaluating Anti-Poverty Programs," in T. Paul Schultz and John Strauss (eds.), *Handbook of Development Economics*, Amsterdam: North-Holland, pp. 3788-3846.

Duflo, Esther, Rachel Glennerster and Michael Kremer (2008), 'Using Randomization in Development Economics Research: a Toolkit', in T. Paul Schultz and John Strauss (eds.), *Handbook of Development Economics*, Amsterdam: North-Holland, pp. 3895-3962.

Optional:

Söderbom, Måns (2009a). "Applied Econometrics. Lecture 11: Treatment Effects Part I," downloadable at http://www.soderbom.net/lecture11_notes.pdf.

Söderbom, Måns (2009b). "Applied Econometrics. Lecture 12: Treatment Effects Part II," downloadable at <http://www.soderbom.net/lec12notesfinal.pdf>

- Important policy instrument for reducing poverty: public programs, or interventions. For example, Progresa (Mexico): Poor households are given grants provided they enroll their children in school.

- Do we know if such interventions really work? How much impact do they have? Answering questions such as these is central for researchers in development.
- Note: these are questions to do with **causality**: we want to learn whether the outcome of interest (e.g. poverty rate) would have been different in the absence of the program, everything else equal. That is, we want to compare what actually happened (e.g. with respect to poverty) to the counterfactual; the difference is, by definition, the causal effect
- In practice, the approach boils down to comparing outcomes across two groups of units (units = individuals, households, villages etc.):
 - the group made up of units that got the program
 - and that consisting of units that did not.
- The latter group of individuals - the comparison group - is used to compute the **counterfactual** for those that were treated. But for this to be a valid approach, it must be that the comparison group has outcomes similar to those that the treated would have, had in the absence of treatment.
- However, those individuals who are exposed to a treatment may well be different from those who are not.
 - Programs are placed in specific areas (poorer areas)
 - Individuals are screened for participation (e.g. on the basis of poverty)
 - The decision to participate is typically voluntary.
- Therefore, simply comparing unconditional average outcomes across the two groups is not, in general, a convincing approach for establishing the causal effect.

2.1. Formulation of the evaluation problem

Note: This section draws on my applied econometrics notes which in turn draw on Wooldridge. In effect, this replaces Section 2 in Ravallion which, in my view, is a rather poor exposition that contains several

ambiguities and some mistakes. The other sections in Ravallion's chapter are fine.

- **The potential-outcomes framework:** For each individual there is a potential outcome with treatment, denoted y_1 (Y^T in Ravallion's exposition), and another potential outcome without treatment, denoted y_0 (Y^C in Ravallion). Think of these as outcomes in alternative states of the world; the treatment (causal) effect is the difference between these two quantities: $y_1 - y_0$.
- We can't measure treatment effects at the **individual** level, as we can never observe the full set of potential outcomes in alternative states of the world. Researchers therefore focus on various forms of **average treatment effects**.
- Define w as a binary treatment indicator (a dummy variable), where

$$w = 1 \text{ if treatment,}$$

$$w = 0 \text{ if no treatment.}$$

The outcome variables, y_1 and y_0 , as well as the difference $y_1 - y_0$, are random variables that potentially vary across individuals. In seeking to estimate the effect of treatment on outcomes, it is therefore natural to focus on estimating the **average treatment effect**. It is useful to distinguish between:

1. The average treatment effect (ATE):

$$ATE = E(y_1 - y_0)$$

2. The average treatment effect on the treated (ATE_1 , or TT):

$$ATE = E(y_1 - y_0 | w = 1)$$

3. The average treatment effect on the untreated (TU):

$$TU = E(y_1 - y_0 | w = 0)$$

- ATE is the expected effect of treatment for a randomly drawn individual from the population
- ATE_1 is the expected effect of treatment for a randomly drawn individual from those individuals in the population that have undergone treatment.
- TU is the expected effect of treatment for a randomly drawn individual from those individuals in the population that have not undergone treatment.
- In the data, the observed outcome y can be written

$$y = (1 - w) y_0 + w y_1 = y_0 + w (y_1 - y_0).$$

This complicates the estimation of treatment effects. How can we estimate ATE or ATE_1 , if this is all the data available?

3. Randomization: Experimental data

- **Randomization:** can be thought of as a process in which the outcome of a toss of a coin determines whether an individual get treatment ($w_i = 1$) or not ($w_i = 0$). If treatment is randomized across individuals, then estimation of the average treatment effect is simple, despite the unobservability problem just discussed.
- Suppose your sample consists of N observations, and your goal is to calculate $E(y_1)$ and $E(y_0)$. Your problem is that for each individual, either y_{1i} or y_{0i} is unobserved. Might it still be valid to calculate $E(y_1)$ by taking the average of the **observed** values of y_1 , and vice versa for $E(y_0)$?

- Yes it would, since randomization ensures the potential outcomes (y_1, y_0) are statistically independent of treatment status.
- The reason is that independence implies $E(y_1|w = 1) = E(y_1|w = 0) = E(y_1)$, and so

$$\begin{aligned}ATE &= E(y_1 - y_0) \\ATE &= E(y_1) - E(y_0) \\ATE &= E(y|w = 1) - E(y|w = 0),\end{aligned}$$

where independence allows us to go from the second to the third line. It also follows that

$$\begin{aligned}ATE_1 &= E(y_1 - y_0|w = 1), \\&= E(y_1|w = 1) - E(y_0|w = 1), \\&= E(y_1) - E(y_0), \\&= E(y|w = 1) - E(y|w = 0),\end{aligned}$$

where independence allows us to go from the second to the third line, and from the third to the fourth line. Notice that, in this case,

$$ATE = ATE_1.$$

- Thus, a randomized experiment guarantees that the **difference-in-means** estimator is fine (unbiased and consistent). Notice that this estimator can be obtained by running the following simple OLS regression:

$$y_i = \beta_0 + \beta_1 w_i + u_i,$$

where the estimate of β_1 is the estimated ATE (and, by implication, ATE_1).

- You see how powerful the method of randomization is. Provided you get the design of your experiment right, all you need to do is to compare mean values across the two groups ($w = 0, w = 1$). This

provides an unbiased estimate of the impact of the program in the sample under study (internal validity).

4. Non-experimental data

- For a number of reasons, most economic research still uses **non-experimental** (observational) data. In this case, we must assume that individuals at least partly determine whether they receive treatment. This may lead to problems with the simple difference-in-means estimator if the individual's decision to get treatment depends on the benefits of treatment.
- To see this, consider again the difference-in-means, now without the independence assumption:

$$\begin{aligned} D &= E(y|w = 1) - E(y|w = 0) \\ D &= E(y_1|w = 1) - E(y_0|w = 0) - E(y_0|w = 1) + E(y_0|w = 1) \\ D &= E(y_1 - y_0|w = 1) + \{E(y_0|w = 1) - E(y_0|w = 0)\}. \end{aligned} \tag{4.1}$$

The first term here is the ATE_1 , by definition. The second term (that within $\{ \}$) is the **selection bias**, capturing the difference in potential untreated outcomes between the treatment group and the control group.

- For example, if the treatment variable is free textbooks to schools and the outcome variable is pupil test scores (see Section 2.1 in Duflo et al.), then it may well be that, if the intervention is not randomized, treatment schools may have had different test scores on average even if they had not been treated.
- Sticking to this example: Why might the selection bias be positive? Why might it be negative?
- Addressing the problem of **self-selection** of treatment is largely what the literature on treatment effect estimation based on non-experimental data is about. Notice that this is precisely the problem solved - in principle - by randomization. Look at (4.1) again: had neither group received treatment,

randomization implies their expected outcomes would have been the same, i.e.

$$E(y_0|w = 1) = E(y_0|w = 0),$$

and so the selection bias vanishes.

- We now focus on the case where individuals potentially self-select into treatment. This breaks independence between (y_1, y_0) and w , and so the simple difference-in-means estimator (4.1) does not estimate the average treatment effects consistently.

4.1. Selection on observables

We argued above that the difference-in-means,

$$D = E(y_1 - y_0|w = 1) + \{E(y_0|w = 1) - E(y_0|w = 0)\}$$

may be a biased estimate of the average treatment effect because of selection. Think about the schools and textbooks example introduced above, and suppose it's primarily schools in poor areas that receive the treatment. Suppose test scores are worse in poor areas (perhaps because of child labour and poor attendance rates), so that the selection bias is negative.

Now consider the difference-in-means conditional on some vector of observable variables X (e.g. local poverty):

$$D(X) = E(y_1 - y_0|X, w = 1) + \{E(y_0|X, w = 1) - E(y_0|X, w = 0)\}.$$

The term in $\{ \}$ is still defined as selection bias, but now that we have conditioned on X - local poverty, for example - it is no longer obvious that the selection bias will be as serious as before.

Indeed, if we assume that

$$E(y_0|X, w = 1) = E(y_0|X, w = 0) = E(y_0|X),$$

it is clear that the difference in means conditional on X , i.e. $D(X)$, identifies $ATE_1(X)$. Notice that the $ATE_1(X)$ coincides with $ATE_1(X)$ (by the same logic as in the section on randomization; but conditional on X). This assumption is known as **selection on observables**, or conditional mean independence.

Given a credible estimate of $ATE_1(X)$, finding the unconditional average treatment effects is straightforward:

$$ATE = \frac{1}{N} \sum_{i=1}^N ATE(X)_i,$$

$$ATE_1 = \frac{1}{N_1} \sum_{j=1}^{N_1} ATE(X)_j,$$

where the notation implies individuals $j = 1, 2, \dots, N_1$ are treated individuals.

So far I haven't said anything about **how** we can condition on X . Let's consider the most straightforward (and often used) approach.

Example: Regression Suppose potential outcomes are modelled as follows:

$$y_0 = \mu_0 + v_0,$$

$$y_1 = \mu_1 + v_1,$$

where $E(v_0) = E(v_1) = 0$, so that $E(y_1 - y_0) = \mu_1 - \mu_0$ is the average treatment effect. Recall y_0, y_1 are unobserved; denote the observed outcome variable by y and write this in this form of a **switching regression**:

$$y = w(\mu_1 + v_1) + (1 - w)(\mu_0 + v_0),$$

$$y = \mu_0 + (\mu_1 - \mu_0)w + v_0 + w(v_1 - v_0).$$

Now suppose there is selection on observables, so that we need to condition on observables. More specifically, suppose the following equations hold:

$$E(v_1|x) = \eta_1 + x\beta_1,$$

$$E(v_0|x) = \eta_0 + x\beta_0.$$

We can then show that

$$E(y|w, x) = \mu_0 + \alpha w + x\beta_0 + w(x - \bar{x})\delta,$$

where $\alpha = (\mu_1 - \mu_0)$ and $\delta = \beta_1 - \beta_0$. This equation can be estimated by means of OLS:

$$y = \mu_0 + \alpha w + x\beta_0 + w(x - \bar{x})\delta + \varepsilon,$$

in which case

$$\begin{aligned} \hat{ATE} &= \hat{\alpha}, \\ \hat{ATE}_1 &= \hat{\alpha} + \left(\sum_{i=1}^N w_i \right)^{-1} \left(\sum_{i=1}^N w_i (x_i - \bar{x}) \hat{\delta} \right). \end{aligned}$$

4.2. Doing an impact evaluation: Generic issues in practice

- Get the key stakeholders to agree to doing an impact evaluation. This can sometimes be done with the government, but collaborating with an NGO is a more common approach. Either way, you are likely to encounter problems here.
- The most commonly heard objection to an impact evaluation is that it's **unethical**. The argument is basically like this: "If you can find a valid comparison group, then this must include equally needy people to the participants, in which case the only ethically acceptable option is to help them, rather than just observe them passively for the purposes of an evaluation."
- The counter argument is that a) resources are limited, and inadequate to cover everyone in need

anyway; b) evaluations benefit the poor in the medium and long run.

- **Design problems.** Think about selection bias.
- The assignment of an anti-poverty program typically involves **purposive placement**, and it's likely that the factors that influence placement (e.g. poverty prevalence) also influence counterfactual outcomes (e.g. pupil test scores in the absence of treatment). So the basic assumption must be: if placement is not random, then selection bias needs to be tackled.
- Suppose your way of tackling this problem is to appeal to the ignorability of treatment assumption, and thus control for observables X . Whether this will be convincing depends on whether you've managed to condition on all relevant elements of X . Without deep knowledge about the specific program and the general context, it's hard to know if you're missing out on one or several relevant variables in the X vector.
- And even if you know a lot about the program and the context, it would seem unlikely that you can rule out the possibility that you are missing some relevant variables in X . That is, if you observe X^{obs} which is a subset of X , then it may well be that

$$E(y_0|X^{obs}, w = 1) - E(y_0|X^{obs}, w = 0) \neq 0$$

In that case, conditioning on observables does not, in general, give you unbiased estimates of the average treatment effect. In such a case, estimation by means of instrumental variables may be a solution.

- We saw above that one statistical approach for dealing with selection bias is regression. Many researchers working on impact evaluation feel the functional form assumptions that you accept when running regressions are too arbitrary. An alternative approach is **propensity score matching** (PSM). I'll come back to this below; at this point I just want to discuss how it links to practical issues.

- The basic idea behind PSM is to model the **likelihood** of being treated as dependent on the variables in the X vector (typically by means of probit or logit); and then compare outcomes for treated and non-treated individuals with the same (or similar) treatment likelihoods. Provided you've done a good job in predicting what types of individual gets, and doesn't get, treatment, then if you are comparing two individuals with the same likelihood of getting treatment, it's basically a question of 'luck' whether one individual got treatment and not the other person.
- Now think about the practical issues involved here. An important feature of PSM is that a valid comparison group can only be found in the region of common support - see Figure 1 in Ravallion's paper.
- A potentially serious problem when evaluating anti-poverty programs is that the placement may be determined by a proxy-means test - basically, villages or households will only be treated if they are sufficiently poor. That may sound fine from the point of view of fairness, but it causes a headache for the researcher hoping to use PSM because the region of common support can get very small.
- Clearly, the smaller the region of common support, the more data you have to throw away (so you collected some data unnecessarily) and, potentially, the less representative your estimation sample becomes.
- Bottom line: In practice, you want the treatment group and comparison group to be as similar as possible with regards to the average characteristics that you think determine selection into treatment.
- **Hidden impacts** for 'non-participants'. Another practical problem that may arise is dubbed 'hidden impacts', or 'contamination' Everything we've said above is based on the implicit assumption that the non-treated are not affected by the program. This may not always be an innocuous assumption, however.
- If you are analyzing a large public program, for example, it may be that the control group is affected by the intervention due to, say, market response or treatment externalities.

- A classic example is the analysis of treatment of intestinal worms in children (Kenya) by Miguel and Kremer: even if you are not treated, you are benefiting if pupils in your class are treated since that reduces the risk of becoming infected.
- More examples of hidden impacts:
 - Fungible external aid - donor implements program in village A but not B; local authorities 'compensate' village B by increasing public spending targeted at that village.
 - Households cut back on education expenditures when grants are provided to schools (Das, Krishnan, Habyarimana, and Dercon, 2004; Zambia).
- In many cases, hidden impacts **should** be taken into account when assessing a program. If the measured (total) impact of the program is low as a result of hidden impacts, it doesn't necessarily follow that the program has had no effect. If, for example, school grants enable parents to cut back on education expenditures, then that (presumably) is a welfare gain for them.

4.3. Propensity score matching

- Recall: ignorability of treatment implies that, conditional on X , we can construct counterfactuals from the comparison group. In principle, it would be best to compare outcomes for treated and nontreated individuals with exactly the same values of all the variables in the X vector. In practice this is not feasible: there are too many x -variables and at least some of them are likely continuous variables (in which case it's unlikely we would observe two observations with the same values).
- Fortunately, there is a way around that, by matching on the **propensity score** instead. The reason matching on the propensity score is more attractive than matching on k different x -variables, is that the propensity score is a single (estimated) "variable" for each individual.
- Here's how PSM works.
- **First**, model the **likelihood of being treated** by means of a binary choice model (e.g. logit or

probit):

$$\Pr(w_i = 1|x) = G(x\beta) \equiv p(x).$$

The function $p(x)$ is known as the **propensity score** - it's simply the predicted probability implied by the logit/probit results.

- **Second**, for each treated observation, compute the **counterfactual** based on the outcomes for nontreated observations with similar propensity scores:

$$\hat{y}_{0,i} = \sum_{j \in \{w=0\}} \phi(i,j) y_{0,j},$$

where $\{w = 0\}$ is the set of nontreated individuals (the control group), and $\phi(i, j)$ is a **weight** based on the propensity scores of the treated individual i and the comparison observation j . The simplest weighting scheme is nearest-neighbour: $\phi(i, j) = 1$ for the nontreated observation that has the propensity score nearest to that of individual i . Other schemes use more information, e.g. kernel matching (see Section 4.1.2 in Söderbom, 2009 for details and an explicit example).

- **Third**, compute the difference between the actual outcome and the counterfactual for each treated observation, and calculate the average of these differences; that's your average treatment effect on the treated:

$$ATE_1^M = \frac{1}{N_T} \sum_{i \in \{w=1\}} (y_{1,i} - \hat{y}_{0,i})$$

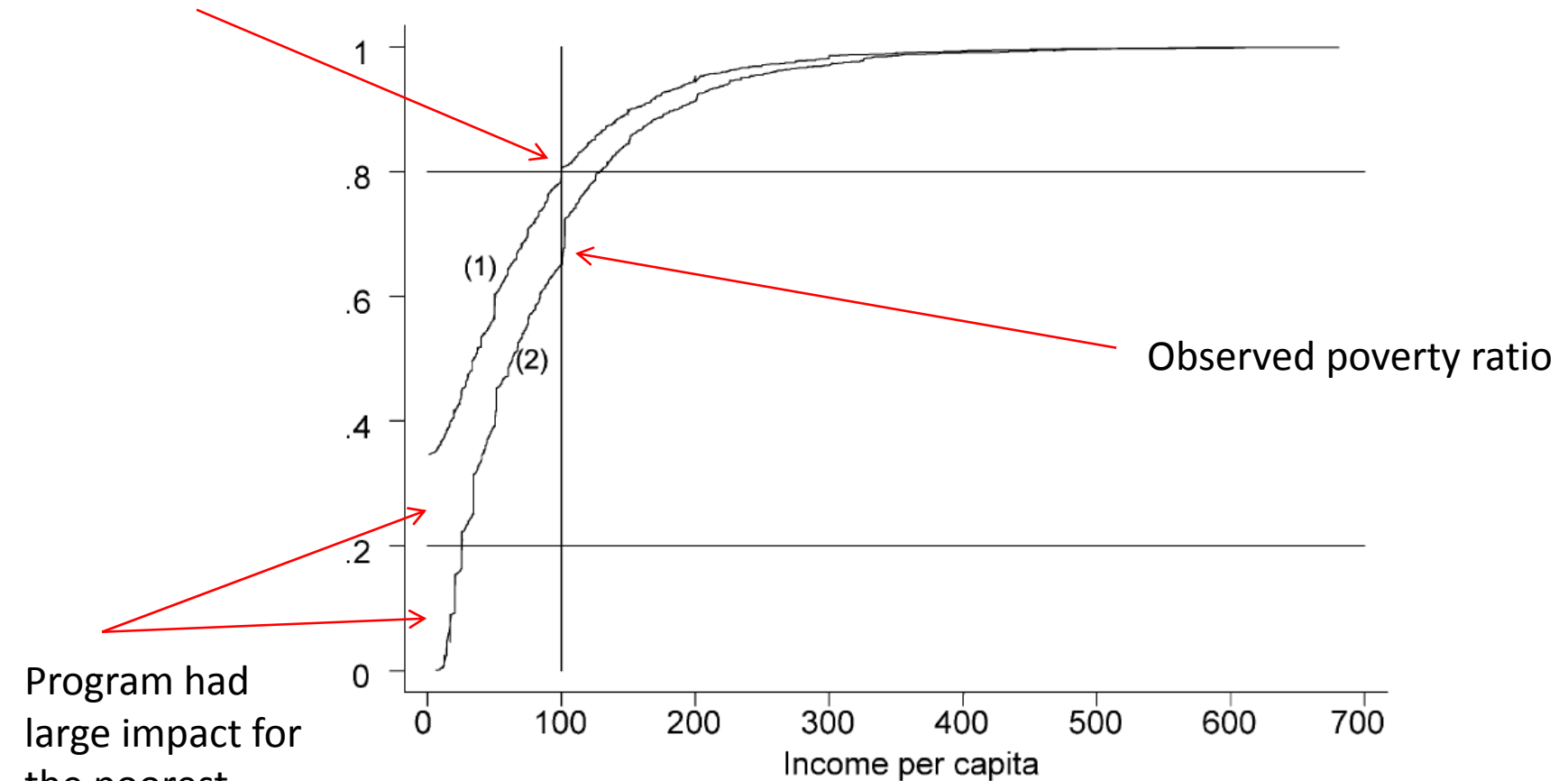
where $\{w = 1\}$ is the set of treated individuals and N_T is the number of treated observations.

- Of course, we don't have to stop at this stage. Indeed, a lot of potentially useful information may be hidden in the average treatment effect. For example, we don't know from looking at the ATE how the program impacts on **different types** of individuals or households.
- For example, how does the program impact vary across households with differing incomes? To find out, simply relate the different $(y_{1,i} - \hat{y}_{0,i})$ computed above to different income groups. Or, show,

for treated observations, the actual CDF of income and compare that to the counterfactual CDF of income (by making use of the $\hat{y}_{0,i}$).

[Figure 2 in Ravallion (p.3808). Draws on Argentina's Trabajar program, which provides low-wage work, targeted to the unemployed, on various small-scale community-level projects. The overall aim is to reduce poverty.]

Poverty ratio in the absence of the program



Program had large impact for the poorest people.

(1) Participant sample pre-intervention (estimated)

(2) Participant sample post-intervention (observed)

Source: Jalan and Ravallion (2003b).

Figure 2. Poverty impacts of disbursements under Argentina's Trabajar program.

- The basic idea behind propensity score matching is quite appealing: To estimate the counterfactual y_{0i} (i.e. the outcome that individual i , who was treated, would have recorded had s/he not been treated), use one or several observations in the (nontreated) control group that are similar to individual i , in terms of the propensity score (i.e. the propensity to select into treatment).

4.3.1. Practical issues

- **Predicting treatment.** For PSM to work, you need to use **all** relevant variables for selection when modeling treatment - otherwise there will be selection on unobservables, which violates the ignorability of treatment assumption. On the other hand you must not include variables that are **caused** by treatment status.
- Of course, it's hard to know a priori what the right set of explanatory variables in the first stage are. Should draw on economic theory. The more you know about the process determining treatment, the more convincing is this particular identification strategy.
- **Inference.** Once we have calculated the average treatment effects of interest, we want estimates of the associated standard errors in order to do inference (e.g. we want to know if the estimated average treatment effect is significantly different from zero or not). The computation of the standard errors needs to take into account the fact that the propensity score is affected by sampling error. One popular and reasonably simple way of doing this is by means of bootstrapping, however it's not clear that this is always an appropriate method - and it's also not clear what to do instead.

4.4. Selection on unobservables

- We have concentrated on calculating average treatment effects when there is selection on observables. When there is selection on **unobservables**, however, the methods that we have reviewed will not yield consistent estimates of average treatment effects.
- In the case where the relevant unobserved variable is time invariant, we may be able to use longitudinal (panel) data and remove the unobserved term by differencing. The most common estimator

in this situation is known as the **Difference-in-Differences** estimator (Ravallion refers to it as the **double-difference estimator**).

4.4.1. Difference-in-Differences

- Consider, again, the selection bias term

$$\{E(y_0|w = 1) - E(y_0|w = 0)\},$$

which is non-zero if actual treatment is correlated with potential outcomes. Now suppose you observe individuals twice: before and after the intervention. Furthermore, suppose that the selection bias is driven by time invariant unobserved heterogeneity across individuals. The latter assumption implies the **growth rate** in actual outcomes for the nontreated group is a valid counterfactual for the growth rate of the treated group. This is because the selection mechanism is constant over time.

- The difference-in-differences estimator, in its simplest form, is defined as the difference in the change in average earnings for the treatment group and the control group:

$$ATE_1 = (\bar{y}_a^T - \bar{y}_b^T) - (\bar{y}_a^U - \bar{y}_b^U),$$

where

$$\bar{y}_a^U = \text{average outcome for nontreated after treatment}$$

$$\bar{y}_b^U = \text{average outcome for nontreated before treatment.}$$

- What about selection into treatment? Consider the following equations:

$$y_{it}^U = \phi_i + \delta_t + \varepsilon_{it} \quad (\text{no treatment})$$

$$y_{it}^T = y_{it}^U + \alpha, \quad (\text{treatment})$$

for $t = a, b$ (after and before), where ϕ_i is an individual-specific, possibly **unobserved**, time invariant fixed effect (thus a source of heterogeneity across individuals in observed outcomes - you could think of this as x_i), δ_t is a dummy variable equal to 1 in the time period after treatment and zero in the period before treatment ($\delta_a = 1, \delta_b = 0$), ε_{it} is a zero-mean residual, and α is the treatment effect (for everyone, as well as for the treated).

- Provided ε_{it} is **uncorrelated** with treatment status, it follows that

$$(\bar{y}_a^T - \bar{y}_b^T) = \alpha + \delta_a,$$

$$(\bar{y}_a^U - \bar{y}_b^U) = \delta_a,$$

thus α is the treatment effect:

$$ATE_1 = \alpha.$$

- Notice that even if ϕ_i is correlated with treatment, this would not lead to bias. In effect, we are exploiting the time dimension of the data to define the counterfactual. But also notice that we require the source of 'selection bias' to be constant over time - otherwise the assumption of ignorability of treatment does not hold. If ε_{it} - which is time varying - is correlated with treatment, then the above DiD estimate of the treatment effect will be biased. In such a case we need to do more work - either by trying to control for the time varying factors driving selection directly (i.e. condition on observable variables - regression or PSM) or by means of IV techniques.
- In practice, implementing the DiD estimator is often done by means of regression. To see the connection, note that

$$y_{it}^U = \phi_i + \delta_t + \varepsilon_{it} \quad (\text{no treatment}),$$

$$y_{it}^T = y_{it}^U + \alpha, \quad (\text{treatment}),$$

implies

$$y_{it} = w_{it} (y_{it}^U + \alpha) + (1 - w_{it}) y_{it}^U,$$

$$y_{it} = \phi_i + \alpha w_{it} + \delta_t + \varepsilon_{it},$$

and so, in differences,

$$\Delta y_{it} = \alpha \Delta w_{it} + \Delta \delta_t + \Delta \varepsilon_{it},$$

which becomes

$$\Delta y_{it} = \alpha w_{it} + \delta_t + \Delta \varepsilon_{it},$$

if there are only two time periods (before and after), so that $\delta_a = 1$ and $\delta_b = 0$, and treatment happens after time b but before time a . Clearly heterogeneity in the form of individual fixed effects **will not** bias our estimate of α - but non-zero correlation between $\Delta \varepsilon_{it}$ and treatment **will**.

- Sometimes you see an alternative way of implementing the DiD estimator, as follows:

$$y_{it} = c_0 + c_1 \mathit{after} + c_2 w_i + c_3 (w_i \times \mathit{after}) + \varepsilon_{it},$$

where the treatment dummy is now defined as time invariant (either you belong to this group; or not), and after is a dummy equal to one in the period after the intervention. You would run this regression pooling both waves of the data (no differencing involved thus). Note that

$$E(y_{it} | \mathit{after} = 0, w_i = 0) = c_0$$

$$E(y_{it} | \mathit{after} = 1, w_i = 0) = c_0 + c_1$$

$$E(y_{it} | \mathit{after} = 0, w_i = 1) = c_0 + c_2$$

$$E(y_{it} | \mathit{after} = 1, w_i = 1) = c_0 + c_1 + c_2 + c_3$$

Hence, the DiD effect is:

$$\begin{aligned} & E(y_{it}|after = 1, w_i = 1) - E(y_{it}|after = 0, w_i = 1) \\ & \text{MINUS} \\ & E(y_{it}|after = 1, w_i = 0) - E(y_{it}|after = 0, w_i = 0), \end{aligned}$$

thus

$$\begin{aligned} & c_0 + c_1 + c_2 + c_3 - (c_0 + c_2) \\ & \text{MINUS} \\ & c_0 + c_1 - c_0 \\ & = c_3 \end{aligned}$$

It follows that the DiD effect is equal to the coefficient on the **interaction** term ($w_i \times after$).

- Adding observable control variables to the DiD regression is straightforward.

Practical issues: DiD

- You need a **baseline survey** - i.e. data from a period before the intervention was implemented. If you are lucky, such data already exist (e.g. because of regular household surveys). If not, you need to organize one; then observe the intervention; then do a follow-up survey once sufficient time has passed after the intervention, to enable you to measure the impact. This is a time consuming and costly process
- Time-invariant selection bias may be a restrictive assumption. For instance, if the attributes on which program placement are based influence subsequent growth rates the DiD estimate will be biased.
- Example: poor villages are targeted; but suppose one reason these are poor is that they've been

exposed to a temporary negative shock from which they would have recovered in the absence of the program (mean-reversion). Hence they record relatively high growth rates during the evaluation period. This will show up as indicating a high DiD estimate, but part of it is spurious, due to mean reversion.

- Why might the bias go in the opposite direction as well?

4.4.2. Instrumental variables

Consider the following regression model:

$$y_i = ATE \times w_i + X_i\beta + e_i,$$

where e_i is unobserved (the residual) and ATE is common across all individuals (i.e. no heterogeneity in treatment effects) Suppose we are concerned that v_i is correlated with treatment, i.e. there is selection on **unobservables**.

- Suppose we have an instrumental variable Z_i , which influences program placement independently of X :

$$w_i = \gamma Z_i + X_i\delta + v_i$$

where $\gamma \neq 0$ (instrument relevance); and $cov(Z_i e_i) = 0$ (instrument validity).

- Under these assumptions, ATE can be estimated consistently using instrumental variables techniques.
- But the assumptions, as you will appreciate, are not totally innocent ones.
- **Instrument validity** is not testable and so has to be taken on faith - basically, how much you should trust IV estimates depends on how much you believe the economic arguments underlying the IV approach.

- Moreover, the assumption of a **common impact** of the treatment across everyone is restrictive. If you don't believe in this, instead thinking the impact is heterogeneous then that changes the way you should **interpret** your IV estimates.
- In particular, with heterogeneous impacts, IV identifies the impact of the program for a specific population sub-group, namely those induced to take up the program by the exogenous variation attributable to the instrument Z . This is known as the local average treatment effect (LATE). Unfortunately, the sub-group for which the LATE applies is rarely identified in practice. So how useful is the LATE? As Ravallion says, "it is presumably the impact for someone, but who"?
- In the first student workshop we will discuss the usefulness of LATE in development. If you have never encountered this concept before, you may find it helpful to have a look at section 2.2 in Söderbom (2009b).