Microeconometrics MECT2
Lecture 9: Evaluation Methods I

Richard Blundell
http://www.ucl.ac.uk/~uctp39a/

University College London

February-March 2016

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,
5. discontinuity design methods

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,
5. discontinuity design methods
6. control function methods.

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,
5. discontinuity design methods
6. control function methods.

- All are an attempt to deal with endogenous selection (assignment).

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,
5. discontinuity design methods
6. control function methods.

- All are an attempt to deal with endogenous selection (assignment).
- Not directly dealing with 'fully' structural simultaneous models, which are also used to address the evaluation problem in empirical microeconometrics - see Blundell and MaCurdy (1999), for example.

Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.

Six distinct, but related, approaches:

1. social experiments methods (RCTs),
2. natural experiments,
3. matching methods,
4. instrumental methods,
5. discontinuity design methods
6. control function methods.

- All are an attempt to deal with endogenous selection (assignment).
- Not directly dealing with 'fully' structural simultaneous models, which are also used to address the evaluation problem in empirical microeconometrics - see Blundell and MaCurdy (1999), for example.
- Q: Under what conditions will the models (and methods) we consider here recover parameters of interest that are consistent with structural simultaneous models?

- The **random experiment (R)** is closest to the 'theory' free method of a clinical trial, relying on the availability of a randomized assignment rule.
- **Natural experiments (DiD)** mimic the randomized assignment of the experimental setting but do so with non-experimental data and some 'natural' randomisation.
- **Matching (M)** attempts to reproduce the treatment group among the non-treated, re-establishing the experimental conditions in a non-experimental setting, but relies on observable variables to account for selection bias.
- **Instrumental variables (IV)** is a closer to the structural method, relying on exclusion restrictions to achieve identification.
- **Discontinuity design (RD)** methods are closest in spirit to the natural experiment as they exploit discreteness in the rules used to assign individuals to receive a treatment.
- The **control function (CF)** approach is closest to the structural econometric approach, directly modelling the assignment rule in order to control for selection in observational data.

Which Treatment Parameter?

- In the *homogeneous linear model*, common in elementary econometrics, there is only one impact of a programme and it is one that would be common to participants and nonparticipants a like.

- In the *heterogeneous response model*, the treated and non-treated may benefit differently from programme participation.
  - In this case, the treatment on the treated parameter will differ from the treatment on the untreated parameter or the average treatment effect.
  - We can now define a whole distribution of the treatment effects.

- A central issue in understanding evaluation methods relates to the aspects of this distribution that can be recovered by the different approaches.

- Suppose we wish to measure the impact of treatment on an outcome, y. For the moment, we abstract from other covariates that may impact on y.
- Denote by d the treatment indicator: a dummy variable assuming the value 1 if the individual has been treated and 0 otherwise.
- The potential outcomes for individual i at any time t are denoted by y_{it}^1 and y_{it}^0.
- These outcomes are specified as

$$\begin{aligned} y_{it}^1 &= \beta + \alpha_i + u_{it} \quad \text{if } d_{it} = 1 \\ y_{it}^0 &= \beta + u_{it} \quad \text{if } d_{it} = 0 \end{aligned} \tag{1}$$

where β is the intercept parameter, α_i is the effect of treatment on individual i and u is the unobservable component of y.

The observable outcome is then

$$y_{it} = d_{it} y_{it}^1 + (1 - d_{it})\, y_{it}^0. \tag{2}$$

so that

$$y_{it} = \beta + \alpha_i d_{it} + u_{it}. \tag{3}$$

Selection into treatment (assignment) determines the treatment status, d.

- We assume this assignment occurs at a fixed moment in time, say k, and depends on the information available at that time summarised by the set of variables, Z_k, and unobservables, v_k.

Assignment to treatment is then assumed to be made on the basis of an index function, d^*

$$\begin{aligned} d_{ik}^* &= Z_{ik}\gamma + v_{ik} && (4) \\ &= g(Z_{ik}, v_{ik}) && (5) \end{aligned}$$

where γ is the vector of coefficients and v_{ik} is the unobservable term.

- The treatment status is then defined as

$$d_{it} = \begin{cases} 1 & \text{if } d_{ik}^* > 0 \text{ and } t > k, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

- As before the structural function for the outcome variable y and the assignment equation for d are assumed to have a triangular structure.
- General question: when is the triangular structure a reasonable formulation of the endogeneity in microeconometrics?

Estimation methods typically identify some average impact of treatment over some sub-population.
The three most commonly used parameters are:

1. the population average treatment effect (ATE), which would be the outcome if individuals were assigned at random to treatment,
2. the average effect on individuals that were assigned to treatment (ATT), and
3. the average effect on non-participants (ATNT).

Using the model specification above, we can express these three average parameters at time $t > k$ as follows

$$\alpha^{ATE} = E(\alpha_i) \tag{7}$$
$$\alpha^{ATT} = E(\alpha_i | d_{it} = 1) = E(\alpha_i | g(Z_{ik}, v_{ik}) \geqslant 0) \tag{8}$$
$$\alpha^{ATNT} = E(\alpha_i | d_{it} = 0) = E(\alpha_i | g(Z_{ik}, v_{ik}) < 0). \tag{9}$$

Historically an increasing interest on the distribution of treatment effects led to the study of additional treatment effects in the literature (Bjorklund and Moffitt, 1987, Imbens and Angrist, 1994, Heckman and Vytlacil, 1999).

- Two particularly important parameters are the local average treatment effect (LATE) and the marginal treatment effect (MTE).
- To introduce them we need to assume that d^* is a non-trivial function of Z, meaning that it changes with Z.
- Now suppose there exist two distinct values of Z, say Z' and Z'', for which only a subgroup of participants under Z'' will also participate if having experienced Z'.

The average impact of treatment on individuals that move for non-participants to participants when Z changes from Z' to Z'' is the LATE parameter

$$\alpha^{LATE}\left(Z', Z''\right) = E\left(\alpha_i \left| d_i(Z'') = 1, d_i(Z') = 0\right.\right)$$

where $d_i(Z)$ is a dichotomous random variable representing the treatment status for an individual i drawing observables Z.
The MTE measures the change in aggregate outcome due to an infinitesimal change in the participation rate,

$$\alpha^{MTE}\left(P\right) = \frac{\partial E\left(y|P\right)}{\partial P}.$$

- Under certain conditions, to be explored in the next lecture, the MTE is a limit version of LATE.

- All these parameters will be identical under homogeneous treatment effects.
- Under heterogeneous treatment effects, however, a non-random process of selection into treatment may lead to differences between them.
- However, whether the impact of treatment is homogeneous or heterogeneous, *selection* bias may be present.

Selection and Assignment.

Collecting all the unobserved heterogeneity terms together we can rewrite the outcome equation (2) as

$$\begin{aligned} y_{it} &= \beta + \alpha^{ATE} d_{it} + \left(u_{it} + d_{it} \left(\alpha_i - \alpha^{ATE} \right) \right) \\ &= \beta + \alpha^{ATE} d_{it} + e_{it}. \end{aligned} \tag{10}$$

- Non-random selection occurs if the unobservable term e in (10) is correlated with d.
- This implies that e is either correlated with the regressors determining assignment, Z, or correlated with the unobservable component in the selection or assignment equation, v.
- Consequently there are two forms of non-random selection: *selection on the observables* and *selection on the unobservables*.
- Different estimators use different assumptions about assignment to identify the impact of treatment.

As a result of selection, the relationship between y and d is not directly observable from the data since participants and non-participants are not comparable.

- Under homogeneous treatment effects, selection bias occurs only if d is correlated with u since the outcome equation is reduced to

$$y_{it} = \beta + \alpha d_{it} + u_{it}$$

where α is the impact of treatment on any individual since this is constant across the population in this case.

- The OLS estimator will then identify

$$E\left[\widehat{\alpha}^{OLS}\right] = \alpha + E\left[u_{it}|d_{it} = 1\right] - E\left[u_{it}|d_{it} = 0\right]$$

which is in general different from α if d and u are dependent (prove this as an exercise).

The selection process is expected to be more severe in the presence of heterogeneous treatment effects.

- The correlation between e and d may now arise through u or through the idiosyncratic gains from treatment, $\alpha_i - \alpha^{ATE}$ (selection on gains).
- The parameter identified by the OLS estimator will now be

$$E\left[\widehat{\alpha}^{OLS}\right] = \overline{\alpha} + E\left[\alpha_i - \overline{\alpha} | d_{it} = 1\right] + E\left[u_{it} | d_{it} = 1\right] - E\left[u_{it} | d_{it} = 0\right]$$

  Note that the first term, $\alpha^{ATE} + E\left[\alpha_{it} - \alpha^{ATE} | d_{it} = 1\right]$, is the ATT.
- Thus, even if d and u are independent, as long as $E\left[d_{it}\left(\alpha_i - \alpha^{ATE}\right)\right] \neq 0$, OLS will not recover the ATE. $E\left[d_{it}\left(\alpha_i - \alpha^{ATE}\right)\right] \neq 0$ implies that the idiosyncratic gains to treatment, α_i, are correlated with the participation decision itself.
- What does OLS identify when d and u are independent?

An example: returns to education

- Individuals differ with respect to educational attainment, which is partly determined by a scholarship (tuition) subsidy policy and partly determined by other factors.
- The model, described in full detail in the Blundell and Costa-Dias (2009), http://www.ucl.ac.uk/~uctp39a, is used to generate a simulated dataset.
- This model is then used as a common framework to analyse the properties of each of the estimators.

(i) The Social Experiment Approach (R)

- Suppose it would be possible to run a social experiment - that is a randomised control trial.
- In this case, assignment to treatment would be random, and thus independent from the outcome or the treatment effect. This ensures that the treated and the non-treated groups are equal in all aspects apart from the treatment status.
- The following are the randomization assumptions:

$$\text{R1: } E[u_i|d_i = 1] = E[u_i|d_i = 0] = E[u_i]$$
$$\text{R2: } E[\alpha_i|d_i = 1] = E[\alpha_i|d_i = 0] = E[\alpha_i]$$

- These conditions are enough to identify the average returns in the experimental population using OLS which is the ATE.

(ii) Natural Experiment: difference-in-differences (DID)

- The natural experiment method makes use of naturally occurring phenomena that can be argued to induce some form of randomization across individuals in the eligibility or the assignment to treatment.
- Typically this method is implemented using a before and after comparison across groups. It is formally equivalent to a difference-in-differences approach which uses some naturally occurring event to create a 'policy' shift for one group and not another.
- The policy shift may refer to a change of law in one jurisdiction but not another, to some natural disaster which changes a policy of interest in one area but not another, or to a change in policy that makes a certain group eligible to some treatment but keeps a similar group ineligible.
- The difference between the two groups before and after the policy change is contrasted - thereby creating a difference-in-differences (DID) estimator of the policy impact.

(iii) The matching estimator (M)

- The main purpose of matching is to reproduce the treatment group among the non-treated, this way re-establishing the experimental conditions in a non-experimental setting.
- Under certain assumptions, the matching method constructs *the* correct sample counterpart for the missing information on the treated outcomes had they not been treated by pairing each participant with members of non-treated group.
- The matching assumptions ensure that the only remaining difference between the two groups is programme participation.
- Matching can be used with cross-sectional or longitudinal data. In its standard formulation, however, the longitudinal dimension is not explored. We therefore initially exclude the time subscript and will focus first on the appropriate choice of the matching variables in what follows.

To start we need to include some observable regressors in the outcome equation in a very general way.

- The covariates X explain part of the 'residual' term u and part of the idiosyncratic gains from treatment:

$$\begin{aligned} y_i^1 &= \beta + u(X_i) + \alpha(X_i) + [(u_i - u(X_i)) + (\alpha_i - \alpha(X_i))] \\ y_i^0 &= \beta + u(X_i) + (u_i - u(X_i)) \end{aligned} \tag{11}$$

- where $u(X)$ is the predictable part of y^0, $(u_i - u(X_i))$ is what is left over of the error u after conditioning for X,
- $\alpha(X)$ is some average treatment effect over individuals with observable characteristics X and
- α_i is an individual i specific 'gain', which differs from $\alpha(X_i)$ by an unobservable heterogeneity term $(\alpha_i - \alpha(X_i))$.

The choice of the appropriate matching variables, X, is a delicate issue.

- To the extent that the goal of evaluation methods is to control for selection, the correct information is that available to the individual at the time of deciding about participation. What information was used when assignment took place?

- What remains unexplained is random with respect to treatment status.

The solution advanced by matching is based on the following assumption:

M1: (*conditional independence assumption - CIA)* Conditional on the set of observables X, the non-treated outcomes are independent of the participation status,

$$y_i^0 \perp d_i \mid X_i$$

which is equivalent to the unobservable in the non-treated outcome equation being independent of the participation status conditional on X,

$$(u_i - u(X_i)) \perp d_i \mid X_i.$$

- This means that, conditional on X, treated and non-treated individuals are comparable with respect to the outcome y in the non-treatment case.
- Thus, there is no remaining selection on the unobservable term u_i.

Assumption M1 implies a conditional version of the randomization hypothesis (R1)

$$E\left[u_i|d_i, X_i\right] = E\left[u_i|X_i\right]$$

which, under the usual hypothesis of exogeneity of X yields $E\left[u_i\right]$.

- Note, nothing like the randomization hypothesis (R2) is required to identify the ATT. This implies that selection on the unobservable gains can be accommodated by matching when identifying the ATT.

- The implication of (M1) is that for each treated observation (y^1), if we can find a non-treated (set of) observation(s) (y^0) with the same X-realization, we can be certain that such y^0 constitutes the correct counterfactual.

Matching is explicitly a process of re-building an experimental data set.
The ability to do so, however, depends on the availability of the counterfactual.
That is, we need to ensure that each treated observation can be reproduced among the non-treated.

- This is captured in the second matching assumption.

M2 All treated individuals have a counterpart on the non-treated population and anyone constitutes a possible participant:

$$0 < P(d_i = 1 \mid X_i) < 1$$

Let S represent the common support of X, that is, the subspace of the distribution of X that is both represented among the treated and the control groups.

- Under assumption (M2), S is the whole domain of X.
- The matching estimator for the ATT is the empirical counterpart of

$$\begin{aligned} \alpha^{ATT}(S) &= E\left[y^1 - y^0 | d = 1, X \in S\right] \\ &= \frac{\int_S E(y^1 - y^0 \mid X, d = 1)\, dF_{X|d}(X \mid d = 1)}{\int_S dF_{X|d}(X \mid d = 1)} \end{aligned}$$

  - where $F_{X|d}$ is the cumulative distribution function of X conditional on d
  - and $\alpha^{ATT}(S)$ is the mean of impact on participants with observable characteristics X in S.

- In general, the form of the matching estimator is given by

$$\widehat{\alpha}^{M} = \sum_{i \in T} \left\{ y_i - \sum_{j \in C} \varpi_{ij} y_j \right\} \omega_i \tag{12}$$

  - where T and C represent the treatment and comparison groups respectively,
  - ϖ_{ij} is the weight placed on comparison observation j for individual i
  - and ω_i accounts for the re-weighting that reconstructs the outcome distribution for the treated sample.

- Note, the parameter identified by matching, α^{M}, may differ from the actual ATT if the common support is not the whole domain of X represented among the treated.

Identification of ATE requires a strengthened version of (M1),

M1' (*conditional independence assumption - CIA)* Conditional on the set of observables X, the two potential outcomes are independent of the participation status,

$$\left(y_i^0, y_i^1\right) \perp d_i \mid X_i$$

- That is, in addition to (M1), identification of ATE using matching requires no selection on the unobservable idiosyncratic gain.
- Under (M1), the matching estimator of ATE is the sample counterpart of,

$$\begin{aligned} \alpha^M &= E\left[y^1 - y^0 | X \in S\right] \\ &= \frac{\int_S E(y^1 - y^0 \mid X)\, dF(X)}{\int_S dF(X)} \end{aligned}$$

where, as before, S is the common support and the average is now weighted with the distribution of X over the whole population.

Propensity score matching

- A serious limitation to the implementation of matching is the dimension of the matching space as defined by X.
- A more feasible alternative is to match on a function of X. Usually, this is carried out on the propensity to participate given the set of characteristics X:

$$P(X_i) = P(d_i = 1 \mid X_i)$$

  the *propensity score*.
- Its use is usually motivated by Rosenbaum and Rubin's result (1983, 1984), which shows that the CIA remains valid if controlling for $P(X_i)$ instead of X_i:

$$y_i^0 \perp d_i \mid P(X_i)$$

- However, it is also shown that knowledge of $P(X)$ may improve the efficiency of the estimates of ATT, its value lying on the "dimension reduction" feature.

- When using $P(X)$, the comparison group for each treated individual is chosen with a pre-defined criteria (established in terms of a pre-defined metric) of proximity between the propensity scores for the each treated and the controls.
- Having defined the neighborhood for each treated observation, the next issue is that of choosing the appropriate weights to associate the selected set of non-treated observations for each participant one.
- Several possibilities are commonly used, e.g.:

1. **Nearest Neighbor matching** assigns a weight 1 to the closest non-treated observation and 0 to all others. (But can we use the bootstrap?).
2. **Kernel matching** defines a neighborhood for each treated observation and constructs the counterfactual using all control observations within the neighborhood, not only the closest one. It assigns a positive weight to all observations within the neighbour while the weight is zero otherwise.

Weaknesses of matching

- The main weakness of matching is the ability to select the correct matching information.
- The common support assumption (M2) ensures that the missing counterfactual can be constructed from the population of non-treated.
- What (M2) does not ensure is that the same counterfactual exists in the sample.
- If some of the treated observations cannot be matched, the definition of the estimated parameter becomes unclear.
    - It is the average impact over some subgroup of the treated, but such subgroup may be difficult to define.
    - For ATT or ATE we require that, in the limit,we can always find a matching observation.

Differences-in-Differences

The DID estimator can make use of longitudinal data, where the same individuals are followed over time, or repeated cross section data, where samples are drawn from the same population before and after the intervention being studied.

- We start by considering the evaluation problem when longitudinal data is available.
- Assume a change in policy occurs at time $t = k$ and each individual is observed before and after the policy change, at times $t = t_0 < k$ and $t = t_1 > k$, respectively. For simplicity of notation, we denote by d_i (without the time subscript) the treatment group to which individual i belongs to.
- This is identified by the treatment status at $t = t_1$:

$$d_i = \begin{cases} 1 & \text{if } d_{it} = 1 \text{ for } t > k \text{ (in particular, } d_{it_1} = 1) \\ 0 & \text{otherwise} \end{cases}$$

The DID estimator uses a common trend assumption

$$\begin{aligned} y_{it} &= \beta + \alpha_i d_{it} + u_{it} && (13) \\ \text{where } E\left(u_{it} \mid d_i, t\right) &= E\left(n_i \mid d_i\right) + m_t. \\ \text{or } u_{it} &= \eta_i + m_t + \varepsilon_{it} \text{ with } E(\varepsilon_{it}|d_{it}) = 0. && (14) \end{aligned}$$

- In the above equation, η is an unobservable individual fixed effect, m is an aggregate(common) macro shock and ε is a transitory shock. Thus, DID is based on the assumption that the randomization hypothesis (R1) holds in first differences

$$E\left[u_{it_1} - u_{it_0}|d_i = 1\right] = E\left[u_{it_1} - u_{it_0}|d_i = 0\right] = E\left[u_{it_1} - u_{it_0}\right].$$

- This assumption does not rule out selection on the unobservables but restricts its source by ruling out the possibility of selection based on transitory individual-specific effects ε_{it}.

Note, DID does not impose any conditions about selection on idiosyncratic gains from treatment that would mimic the randomization hypothesis (R2). As a consequence, and as will be seen, it will only identify ATT in general.
Under the DID assumption we can write,

$$E\left[y_{it}|d_i,t\right]=\begin{cases}\beta+E\left[\alpha_i|d_i=1\right]+E\left[n_i|d_i=1\right]+m_{t_1} & \text{if } d_i=1 \text{ and } t=\\ \beta+E\left[n_i|d_i\right]+m_t & \text{otherwise}\end{cases}$$

- It is now clear that we can eliminate both β and the error components by sequential differences

$$\begin{aligned}\alpha^{ATT} &= E\left(\alpha_i|d_i=1\right) \\ &= \left[E\left(y_{it}|d_i=1,t=t_1\right)-E\left(y_{it}|d_i=1,t=t_0\right)\right] \\ &\quad -\left[E\left(y_{it}|d_i=0,t=t_1\right)-E\left(y_{it}|d_i=0,t=t_0\right)\right]\end{aligned} \tag{15}$$

This is precisely the DID identification strategy.

The sample analog of equation (15) is the DID estimator:

$$\widehat{\alpha}^{DID} = \left[\overline{y}^{1}_{t_1} - \overline{y}^{1}_{t_0}\right] - \left[\overline{y}^{0}_{t_1} - \overline{y}^{0}_{t_0}\right] \tag{16}$$

where $\overline{y}^{d}_{t}$ is the average outcome over group d at time t.

- DID measures the excess outcome change for the treated as compared to the non-treated, this way identifying the ATT.
- The large sample properties of each of the elements in (16) implies

$$\text{p lim}\left[\widehat{\alpha}^{DID}\right] = \alpha^{ATT}.$$

- Prove this last statement.

- Notice that, the DID estimator is just the first differences estimator commonly applied to panel data when the presence of fixed effects is suspected.
- This means that an alternative way of obtaining $\widehat{\alpha}^{DID}$ is to take the first differences of (13) to obtain

$$y_{it_1} - y_{it_0} = \alpha_i d_{it_1} + (m_{t_1} - m_{t_0}) + (\varepsilon_{it_1} - \varepsilon_{it_0})$$

  where the ε terms represent transitory idiosyncratic shocks.
- Under the DID assumptions, the above regression equation can be consistently estimated using OLS. Notice also that the DID assumption implies that the transitory shocks, ε_{it}, are uncorrelated with the treatment variable. This is an exogeneity assumption.
- Therefore, the standard within groups panel data estimator is analytically identical to the DID estimator of the ATT under these assumptions (see, for example, Blundell and MaCurdy (1999)).

- It follows that repeated cross-sectional data would be enough to identify ATT, as long as treatment and control groups can be separated before the policy change, in period $t = t_0$.

- Such information is sufficient for the average fixed effect per group to cancel out in the before after differences.

Weaknesses of DID

(a) Selection on idiosyncratic temporary shocks: Ashenfelter's dip

The DID procedure does not control for unobserved temporary individual-specific shocks that influence the participation decision.
If ε is not unrelated to d, DID is inconsistent for the estimation of ATT

$$E\left(\hat{\alpha}^{DID}\right) = \alpha^{ATT} + E(\varepsilon_{it_1} - \varepsilon_{it_0} \mid d_{it_1} = 1) - E(\varepsilon_{it_1} - \varepsilon_{it_0} \mid d_{it_1} = 0)$$

- To illustrate the conditions such inconsistency might arise, suppose a training programme is being evaluated in which enrolment is more likely if a temporary dip in earnings occurs just before the programme takes place - the so-called Ashenfelter's dip (see Heckman and Smith (1994)).
- A faster earnings growth is expected among the treated, even without programme participation.
- Thus, the DID estimator is likely to over-estimate the impact of treatment.

(b) Differential macro trends

The identification of ATT using DID relies on the assumption the treatment and controls experience the same macro shocks.

If this is not the case, the DID approach will yield a biased and inconsistent estimate of ATT. For example, differential trends might arise in the evaluation of training programs if treated and controls operate in different labour markets.

(c) Compositional changes over time

Although DID does not require longitudinal data to identify the true ATT parameter, it does require that the same group treatment and control to be followed over time.

In particular, the composition of the groups with respect to the fixed effects term must remain unchanged to ensure before-after comparability.

Combining matching and DID (MDID)

- Decompose the unobservable term u into a fixed effect (η), macro shock (m) and an idiosyncratic transitory shock (ε)

$$\begin{aligned} y_{it}^{1} &= \beta + u(X_i) + \overline{\alpha}(X_i) + (\eta_i + m_t + \varepsilon_{it} - u(X_i)) + (\alpha_i(X_i) - \overline{\alpha}(X_i \\ y_{it}^{0} &= \beta + u(X_i) + (\eta_i + m_t + \varepsilon_{it} - u(X_i)) \end{aligned} \tag{17}$$

Under this specification, the following transformation of the CIA can be used achieve identification of ATT:

MDID1: Conditional on the set of observables X, the before-after difference in the unobservable u is independent of the participation status,

$$\left(u_{it_1}^{0} - u_{it_0}^{1}\right) \perp d_{it_1} \mid X_i$$

which, under specification (17) is the same as assuming

$$\varepsilon_{it} \perp d_{it_1} \mid X_i$$

where $t_0 < k < t_1$.

Assumption (MDID1) is not enough to ensure identifiability of ATT. Just as in the matching case, we also need to impose a common support hypothesis. This will be the same as (M2) when longitudinal data is available.
If we only have repeated cross-section data, however, we will need to strengthen it to ensure that the treated group can be reproduced in all three control groups characterized by treatment status before and after the program:

MDID2: All treated individuals have a counterpart on the non-treated population before and after the treatment and anyone constitutes a possible participant,

$$0 < P(d_{it_1} = 1 \mid X_i, t) < 1$$

where $P(d_{it_1} = 1 \mid X_i, t)$ is the probability that an individual observed at time t with characteristics X_i had been treated would the same observation correspond to time t_1.

- The effect of the treatment on the treated can now be estimated over the common support of X.

- The following estimator is adequate to the use of propensity score matching with longitudinal data

$$\widehat{\alpha}^{MDID,L} = \sum_{i \in T} \left\{ \left[y_{it_1} - y_{it_0} \right] - \sum_{j \in C} \varpi_{ij} \left[y_{jt_1} - y_{jt_0} \right] \right\} \omega_i$$

where the notation is similar to what has been used before.

With repeated cross-section data, however, matching must be performed over the three control groups: treated and non-treated at t_0 and non-treated at t_1.

- In this case, the matching-DID estimator would be

$$\widehat{\alpha}^{MDID,RCS} = \sum_{i \in T_1} \left\{ \left[y_{it_1} - \sum_{j \in T_0} \omega^{T}_{ijt_0} y_{jt_0} \right] - \left[\sum_{j \in C_1} \omega^{C}_{ijt_1} y_{jt_1} - \sum_{j \in C_1} \omega^{C}_{ijt_0} y_{jt_0} \right] \right.$$

  where T_0, T_1, C_0 and C_1 stand for the treatment and comparison groups before and after the programme, respectively, and ω^{G}_{ijt} represent the weights attributed to individual j in group G (where $G = C$ or T) and time t when comparing with treated individual i.
- What are the likely issues with matching-DiD in the cross-section case?
- Randomisation, Matching, DID and MDiD are all different ways of dealing with the endogenous selection (assignment) problem. How do these compare to IV, RD, and control function methods?