

# MECT Microeconometrics

## Blundell Lecture 3

### Evaluation Methods II

**Richard Blundell**

<http://www.ucl.ac.uk/~uctp39a/>

University College London

February-March 2016

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,

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,
- 2 natural experiments,

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,
- 2 natural experiments,
- 3 matching 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,
- 2 natural experiments,
- 3 matching methods,
- 4 **instrumental 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,
- 2 natural experiments,
- 3 matching methods,
- 4 **instrumental methods,**
- 5 **discontinuity design 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,
- 2 natural experiments,
- 3 matching methods,
- 4 **instrumental methods,**
- 5 **discontinuity design methods**
- 6 **control function methods.**

## (iv) The instrumental variables (IV) estimator

Unlike the matching method, Instrumental Variables deals directly with selection on the *unobservables*.

Consider the outcome model

$$y_i = X_i' \beta + \alpha_i d_i + u_i \quad (1)$$

$$d_i^* = Z_i \gamma + v_i \quad (2)$$

$$d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

IV requires at least one regressor exclusive to the decision rule - a variable  $z$  in  $Z$  but not in  $X$ .



To begin with we make the following assumptions:

IV0: The impact of treatment is homogeneous

$$\alpha_i = \alpha \text{ for all } i$$

IV1: Conditional on  $d$ ,  $y$  is mean-independent of  $z$

$$E[y|d, z] = E[y|d]$$

which under (IV0) is to say that  $u$  is mean independent of  $z$

$$E[u|d, z] = E[u|d]$$

IV2: Conditional on the remaining regressors in  $Z$  (which we denote by  $Z_{-z}$ ), the decision rule is a non-trivial (non-constant) function of  $z$

$$P[d = 1|Z_{-z}, z] \neq P[d = 1|Z_{-z}]$$

Assumption (IV1) is the exclusion restriction, meaning that  $z$  has no impact on outcomes apart from through the treatment status,  $d$ . Under homogeneous treatment effects - assumption (IV0) - this implies that  $z$  affects the level only, not the difference between the outcomes in the treated and non-treated scenarios.

The variable  $z$  is the (excluded) instrument:

- the source of exogenous variation used to approximate randomized trials.
- It provides variation that is correlated with the participation decision but does not affect the potential outcomes directly.

Under assumptions (IV0) and (IV1) we can write

$$\begin{aligned} E(y_i | z_i) &= \alpha P(d_i = 1 | z_i) + E(u_i | z_i) \\ &= \alpha P(d_i = 1 | z_i) + E(u_i) \end{aligned}$$

which when used with two different values for  $z$ , say  $z^*$  and  $z^{**}$ , yields

$$\begin{aligned} &E(y_i | z_i = z^*) - E(y_i | z_i = z^{**}) \\ &= \alpha [P(d_i = 1 | z_i = z^*) - P(d_i = 1 | z_i = z^{**})] \end{aligned}$$

thus identifying the treatment effect from the ratio

$$\alpha^{IV} = \frac{E(y_i | z_i = z^*) - E(y_i | z_i = z^{**})}{P(d_i = 1 | z_i = z^*) - P(d_i = 1 | z_i = z^{**})} \quad (4)$$

as long as  $P(d_i = 1 | z_i = z^*) \neq P(d_i = 1 | z_i = z^{**})$  (IV2).

The IV estimator replaces these expectations and probabilities with their empirical counterparts:

$$\hat{\alpha}^{IV} = \frac{\hat{E}(y_i | z_i = z^*) - \hat{E}(y_i | z_i = z^{**})}{\hat{P}(d_i = 1 | z_i = z^*) - \hat{P}(d_i = 1 | z_i = z^{**})}. \quad (5)$$

Think of this as an analog estimator.

- This ratio is the standard form for the IV estimator for the model

$$y_i = \beta + \alpha d_i + u_i \quad (6)$$

using instrumental variable  $z$ .

- Typically of the form

$$\hat{\alpha}^{IV} = \frac{\widehat{\text{cov}}(y, z)}{\widehat{\text{cov}}(d, z)}.$$

## Weaknesses of IV

- A key issue in implementation of IV is the choice of the instrument. Very frequently, it is impossible to find a variable that satisfies (IV1) and (IV2), in which case IV is of no practical use.
- In many cases, the instrument  $z$  has insufficient variation, which means that the estimation must rely on two very close values of  $z$ . In such case, the denominator in (??) can be very small, leading to very imprecise estimates of the treatment effect.

# Heterogeneous Effects

If there is no selection on the idiosyncratic gains,  $\alpha_i - \bar{\alpha}$ , then  $e$  is mean independent from  $z$

$$E[e|d, z] = E[u|d, z] + P[d = 1|z] E(\alpha_i - \bar{\alpha}|d = 1, z) = E[u].$$

In this case, IV still identifies ATE which is not different from ATT given that individuals do not use information on their idiosyncratic gains to decide about participation.

- However, in the more general case of heterogeneous effects with selection on idiosyncratic gains, IV will not identify ATE or ATT. If individuals are aware of their own idiosyncratic gains from treatment, they are expected to make a more informed participation decision.
- The resulting selection process generates some correlation between  $\alpha_i$  and  $z$ . This is easily understood given that  $z$  determines  $d$ , facilitating or inhibiting participation.

For example, it may be that participants with values of  $z$  that make participation more unlikely gain on average more from treatment than participants with values of  $z$  that make participation more likely to occur.

- Consider the education example:
- Suppose we use family background to instrument the level of education under the model assumption that family background is uncorrelated with ability. In such case, family background will be uncorrelated with potential earnings under the two treatment scenarios,  $y^0$  and  $y^1$ .
- However, in the data family background will be related with the idiosyncratic component of the returns to education, determined by ability, since individuals with a “good family background” (facing relatively low educational costs) are more likely to invest in education than individuals with “low family background” (facing high educational costs), and do so even if expecting relatively low returns.

# The LATE parameter

The solution advanced by Imbens and Angrist (1994) is to identify the impact of treatment from local changes in the instrument  $z$ .

- The rationale is that some local changes in the instrument  $z$  reproduce random assignment by inducing individuals to decide differently as they face different conditions unrelated to potential outcomes.
- To discuss this parameter we define  $y_i(z)$  as the outcome of individual  $i$  at a given point of the instrument  $z$ . Thus, we can rewrite the outcome equation by taking the instrument explicitly into account

$$y_i(z) = d_i(z)y_i^1 + (1 - d_i(z))y_i^0$$

where  $d_i(z)$  is the random variable representing the treatment status of individual  $i$  at a *given point*  $z$ .



This use of IV requires a strengthened version of (IV1) and (IV2).

- Start by considering the following transformation of (IV1)

IV1':  $(y_i^1, y_i^0, d_i(z))$  is jointly independent of  $z_i$ .

which means that  $z$  is uncorrelated with the unobservable in the selection equation  $v$ , the unobservable in the outcome equation  $u_i$ , and the idiosyncratic gain  $\alpha_i$ .

- Under (IV1') we can write

$$\begin{aligned} E[y_i(z)|z] &= PE[y_i^1|d_i(z) = 1, z] + (1 - P)E[y_i^0|d_i(z) = 0, z] \\ &= PE[y_i^1|d_i(z) = 1] + (1 - P)E[y_i^0|d_i(z) = 0] \\ &= E[y_i(z)] \end{aligned} \tag{7}$$

since, conditional on  $d$ ,  $z$  contains no extra information about the potential outcomes,  $y^0$  and  $y^1$ .

We can now use this result together with two possible values of  $z$ , say  $z^*$  and  $z^{**}$ , to write

$$\begin{aligned} & \bullet E [y_i(z)|z = z^{**}] - E [y_i(z)|z = z^*] \\ &= E [y_i(z^{**}) - y_i(z^*)] \\ &= E [(d_i(z^{**}) - d_i(z^*)) (y_i^1 - y_i^0)] \\ &= P [d_i(z^{**}) > d_i(z^*)] E [y_i^1 - y_i^0 | d_i(z^{**}) > d_i(z^*)] \\ &\quad - P [d_i(z^{**}) < d_i(z^*)] E [y_i^1 - y_i^0 | d_i(z^{**}) < d_i(z^*)] \end{aligned}$$

where the second equality is obtained by substituting in the expression for  $y(z)$  and the third equality uses the fact that whenever  $d_i(z^*) = d_i(z^{**})$  the expression in the expectations operator is nil.

- This expression means that, under (IV1'), any change in the average outcome  $y$  when  $z$  changes is solely due to changes in the treatment status of a subset of the population.

Identification of the impact of treatment on individuals that change their participation decision still requires another assumption

IV2': The decision rule is a non-trivial monotonic function of  $z$ .

- Under (IV2'), one of the terms in the last equality vanishes.

- wlog, suppose  $d$  is increasing in  $z$  and  $z^{**} > z^*$ . Then  $P[d_i(z^{**}) < d_i(z^*)] = 0$  and

$$\begin{aligned} & E[y_i(z)|z = z^{**}] - E[y_i(z)|z = z^*] \\ &= P[d_i(z^{**}) > d_i(z^*)] E[y_i^1 - y_i^0 | d_i(z^{**}) > d_i(z^*)]. \end{aligned} \quad (8)$$

- This equation can be rearranged to yield the LATE parameter

$$\begin{aligned} \alpha^{LATE}(z^*, z^{**}) &= E[y_i^1 - y_i^0 | d_i(z^{**}) > d_i(z^*)] \\ &= \frac{E[y_i(z)|z = z^{**}] - E[y_i(z)|z = z^*]}{P[d_i(z^{**}) > d_i(z^*)]} \\ &= \frac{E[y_i(z)|z = z^{**}] - E[y_i(z)|z = z^*]}{P[d_i = 1|z^{**}] - P[d_i = 1|z^*]} \end{aligned} \quad (9)$$

which equals the IV estimator as in (4), it measures the impact of treatment on individuals that move from non-treated to treated when  $z$  changes from  $z^*$  to  $z^{**}$ .

To illustrate the LATE approach, take the education example and suppose  $z$  is family background.

- Participation is assumed to become more likely as  $z$  increases.
- To estimate the effect of education, consider a group of individuals that differ only in the family background dimension.
- Among those that enroll in education when the family background  $z$  equals  $z^{**}$  some would not do so if  $z = z^*$ .
- LATE measures the impact of education on the “movers” by attributing any difference in the average outcomes between the two groups defined by different family backgrounds to the different enrollment rates.

## The LATE assumptions

Assumption (IV1') is required to establish the result in equation (??).

- Even if  $(y^0, y^1)$  are not directly related with  $z_i$ , some relation may arise if  $d(z)$  is not independent of  $z_i$ .  $z_i$  is related to  $d(z)$  if it is related with the unobservable in the selection rule,  $v$ .
- If, furthermore,  $v$  is related with the unobservable in the outcome's equation,  $u$ , then the potential outcomes will in general be correlated with  $z$ .
- In the education example, take  $z$  to be family background and assume it has an impact on the taste for education, included in  $v$ . Thus, a change in  $z$  affects  $v$ . If the taste for education is related with the taste for working, which is included in  $u$ , a change in  $z$  may affect  $u$ , thus altering the potential outcomes even among those that do not change treatment status.

In such case, the population average outcome responds to a change in  $z$  not only through individuals altering their treatment status but also through changes in potential outcomes for the whole population, irrespective of their treatment status in the two  $z$ -scenarios.

- The monotonicity assumption in (IV2') is required for interpretation purposes. Under monotonicity of  $d$  with respect to  $z$ , the LATE parameter measures the impact of treatment on individuals that move from non-treated to treated as  $z$  changes.
- If monotonicity does not hold, LATE measures the change in average outcome caused by a change in the instrument, which, under (IV1'), is due to individuals moving in and out of participation.
- However, it is not possible to separate the effect of treatment on individuals that move in from that on individuals that move out as a consequence of a change in  $z$  (see Heckman, 1997).

## What does LATE measure?

- Although very similar to the IV estimator presented in (??), LATE is intrinsically different since it does not represent ATT or ATE. LATE depends on the particular values of  $z$  used to evaluate the treatment and on the particular instrument chosen.
- The group of “movers” is not in general representative of the whole treated or, even less, the whole population. For instance, individuals benefiting the most from participation are more unlikely to be observed among the movers.
- The LATE parameter answers a different question, of how much individuals at the margin of participating benefit from participation given a change in policy.
- That is, it measures the effect of treatment on the sub-group of treated at the margin of participating for a given value of  $z$ .



## (v) The (regression) discontinuity design estimator (RD)

- Certain non-experimental policy designs provide sources of randomization that can be explored to estimate treatment effects under less restrictive assumptions.
- This is really the motivation for the natural experiment approach discussed earlier. However, a special case that has attracted considerable attention occurs when the probability of enrollment into treatment changes discontinuously with some continuous variable  $z$ .
- The variable  $z$  is an observable instrument, typically used to determine eligibility. The discontinuity design approach uses the discontinuous dependence of  $d$  on  $z$  to identify the local average treatment effect even when the instrument does not satisfy the IV assumptions discussed before.
- As will be discussed, the parameter identified by discontinuity design is a local average treatment effect like the LATE parameter discussed under IV but is not necessarily the same parameter.

# The (regression) discontinuity design estimator (RD)

The discontinuity design approach applies when  $d$ , or more generally  $E(d|z) = P(d = 1|z)$ , is a discontinuous function of  $z$  at a certain point  $z = z^*$ .

- Suppose, therefore, that the following condition holds

$$\lim_{z \rightarrow z^{*-}} P(d = 1|z) \neq \lim_{z \rightarrow z^{*+}} P(d = 1|z) \quad (10)$$

where both limits exist.

- The superscripts  $-$  and  $+$  mean that the limit is taken as  $z$  approaches  $z^*$  from below and above, respectively.
- Equation (??) represents the requirement for  $P(d = 1|z)$  to have a discontinuity at  $z^*$ .

# The (regression) discontinuity design estimator (RD)

Instead of requiring the independence assumptions used in IV, the discontinuity design is based on continuity assumptions.

- Consider the simplified version of equation

$$\begin{aligned}y_i &= \beta + \alpha_i d_i + u_i \\ &= \beta_i + \alpha_i d_i \\ &= \beta + \alpha d_i + e_i\end{aligned}$$

where

$$e_i = \beta_i - \beta + d_i (\alpha_i - \alpha)$$

- RD1:  $E(\beta_i|z)$  as a function of  $z$  is continuous at  $z = z^*$ .
- RD2:  $E(\alpha_i|z)$  as a function of  $z$  is continuous at  $z = z^*$ .
- RD3: The participation decision,  $d$ , is independent from the participation gain,  $\alpha_i$ , in the neighborhood of  $z^*$ .

# The (regression) discontinuity design estimator (RD)

- In the special homogeneous treatment effects case, assumptions (RD2) and (RD3) are always true, thus only (RD1) being required to identify  $\alpha$ .
- In the general heterogeneous case, conditions (RD1) and (RD2) ensure that the expected values of the potential outcomes

$$E(y_i^1|z) = E(\beta_i|z) + E(\alpha_i|z)$$
$$E(y_i^0|z) = E(\beta_i|z)$$

are both continuous functions of  $z$  at  $z = z^*$ .

# The (regression) discontinuity design estimator (RD)

Condition (RD3) is an independence assumption but only applies locally:

- Under (RD3) we can write in the neighbourhood of  $z^*$

$$\begin{aligned} E(y_i|z) &= E(\beta_i|z) + P(d_i = 1|z) E(\alpha_i|d = 1, z) \\ &= E(\beta_i|z) + P(d_i = 1|z) E(\alpha_i|z) \end{aligned}$$

since the gain  $\alpha$  is independent of  $z$  for  $z$  sufficiently close to  $z^*$ .

- This ensures that  $E(y|z)$  is discontinuous at  $z^*$  as a consequence of the discontinuity in the odds of participation at that point only.

# The (regression) discontinuity design estimator (RD)

- Using a small  $\delta > 0$ , we can now write

$$\begin{aligned} & E(y_i|z + \delta) - E(y_i|z - \delta) \\ &= [E(\beta_i|z + \delta) - E(\beta_i|z - \delta)] \\ &\quad + [P(d_i = 1|z + \delta) E(\alpha_i|z + \delta) - P(d_i = 1|z - \delta) E(\alpha_i|z - \delta)] \end{aligned}$$

- which, taking the limits as  $\delta \rightarrow 0$  at  $z = z^*$ , yields

$$\begin{aligned} & \lim_{z \rightarrow z^{*+}} E(y_i|z) - \lim_{z \rightarrow z^{*-}} E(y_i|z) \\ &= E(\alpha_i|z^*) \left[ \lim_{z \rightarrow z^{*+}} P(d_i = 1|z) - \lim_{z \rightarrow z^{*-}} P(d_i = 1|z) \right]. \end{aligned}$$

# The (regression) discontinuity design estimator (RD)

The RD estimator of the impact of treatment at  $z = z^*$  is the sample analog of

$$\alpha^{RD}(z^*) = \frac{\lim_{z \rightarrow z^{*+}} E(y_i|z) - \lim_{z \rightarrow z^{*-}} E(y_i|z)}{\lim_{z \rightarrow z^{*+}} P(d_i = 1|z) - \lim_{z \rightarrow z^{*-}} P(d_i = 1|z)}$$

which identifies the local average treatment effect,  $E(\alpha_i|z^*)$ .

- It measures the impact of treatment on individuals with characteristics  $z$  close to  $z^*$ .
- This is an average treatment effect at the local level since selection on idiosyncratic gains is excluded at the local level by assumption (RD3).
- The importance of assumptions (RD1) and (RD2) is made clear from the derivation of the RD estimator above.
- Would  $\beta$  or  $\alpha$  had a discontinuity at  $z^*$  and we would not be able to separate the average impact.

# The (regression) discontinuity design estimator (RD)

- In **the sharp design** case, the denominator in the expression

$$\alpha^{RD}(z^*) = \frac{\lim_{z \rightarrow z^{*+}} E(y_i|z) - \lim_{z \rightarrow z^{*-}} E(y_i|z)}{\lim_{z \rightarrow z^{*+}} P(d_i = 1|z) - \lim_{z \rightarrow z^{*-}} P(d_i = 1|z)}$$

reduces to 1.

- In this case the parameter identified by the discontinuity design approach is simply:

$$\begin{aligned}\alpha^{RD}(z^*) &= E(\alpha_i|z^*) \\ &= \lim_{z \rightarrow z^{*+}} E(y_i|z) - \lim_{z \rightarrow z^{*-}} E(y_i|z)\end{aligned}\tag{11}$$



# The (regression) discontinuity design estimator (RD)

- Notice how (RD1)-(RD3) recover randomisation under discontinuity in the odds of participation at the discontinuity point.
- Assumption (RD3) is precisely a local version of (RD2), meaning that ATE is identifiable locally by discontinuity design.
- Under (RD3), ATE and ATT are locally equal.
- Assumption (RD1) is not guaranteed to hold but instead the error term for the non-treated,  $u$ , is required to be a continuous function of  $z$  at  $z^*$ .
- Continuity ensures that it vanishes by differencing and taking the limits, thus ceasing to be a problem.

## Drawbacks to discontinuity design

A major drawback of discontinuity design is its dependence on discontinuous changes in the odds of participation.

- This means that only the average parameter at a given point in the distribution of  $z$  is identifiable.
- As in the binary instrument case of LATE, the discontinuity design is restricted to the discontinuity point which is dictated by the design of the policy.
- This can be a problem whenever the treatment effect,  $\alpha$ , changes with  $z$ .
- To see why, consider the context of our educational example.

# The (regression) discontinuity design estimator (RD)

- Suppose a subsidy is available for individuals willing to enroll in high education for as long as they score above a certain threshold  $z^*$  in a given test.
- The introduction of such subsidy together with the eligibility rule creates a discontinuity in the odds of participation.
- On the other hand, the test score,  $z$ , and the returns to education,  $\alpha$ , are expected to be (positively) correlated if both depend on, say, ability.
- Thus, by restricting the analysis to the neighborhood of  $z^*$ , we only consider a specific subpopulation with a particular distribution of ability which is not that of the whole population or of the treated population.
- That is, the returns to education are estimated at a certain margin from where other more general parameters cannot be inferred.

## (vi) The Control Function Estimator (CF)

- If selection occurs (partly) on the unobservables, an alternative solution to the IV estimator is to take the selection model explicitly into consideration in the estimation process.
- The control function method does exactly this, treating the endogeneity of  $d$  as an omitted variable problem.
- There is also an interesting link between CF and IV methods in the binary treatment evaluation model considered here.

# The Control Function Estimator (CF)

- Let's return to the Triangular Structure of the heterogeneous treatments effect model

$$y_i = d_i y_i^1 + (1 - d_i) y_i^0. \quad (12)$$

so that

$$y_i = \beta + \alpha_i d_i + u_i. \quad (13)$$

- Assignment to treatment is given by the reduced form binary response

$$\Pr[d_i = 1] = \Pr[g(Z_i, v_i) > 0] \quad (14)$$

$$= \Pr[Z_i \gamma + v_i > 0] \quad (15)$$

# The Control Function Estimator (CF)

- The control function approach is based on the following assumptions:

$$\text{CF1: } (u, \alpha) \perp (d, Z) | v$$

that is conditional on  $v$ ,  $u$  and  $\alpha$  are independent of  $d$  and  $Z$ , and

$$\text{CF2: } P[d = 1 | Z_{-z}, z] \neq P[d = 1 | Z_{-z}]$$

that is conditional on the remaining regressors in  $Z$  (which we denote by  $Z_{-z}$ ), the decision rule is a non-trivial (non-constant) function of  $z$ .

- Assumption (CF1) allows for the variation in  $d$  to be separated from that in  $u$  and  $\alpha$  by conditioning on  $v$  (see, for example, Blundell and Powell, 2003).

# The Control Function Estimator (CF)

- Under the CF1 condition

$$E[u|d, Z, v] = h_u(v)$$

$$E[\alpha|d, Z, v] = h_\alpha(v)$$

- If we knew these functions or could estimate them we could fully correct for selection on observables and recover the distribution of treatment effects.

# The Control Function Estimator (CF)

- The CF1 condition is often weakened to:

$$\text{CF1a: } (u) \perp (d, Z) | v$$

- Which will be sufficient to recover the ATT.
- Indeed many applications of the control function approach typically make a parametric assumption on the joint distribution of the error terms,  $u$  and  $v$ .



# The Control Function Estimator (CF)

- The most commonly encountered assumption imposes joint normality so that

$$\begin{aligned} E[u|d=1, Z] &= \rho\lambda_1(Z\gamma) \\ E[u|d=0, Z] &= \rho\lambda_0(Z\gamma) \end{aligned} \quad (16)$$

- where  $\rho = \sigma_u \text{corr}(u, v)$ ,  $\sigma_u$  is the standard error of  $u$ , and the control function are (adopting the standardization  $\sigma_v = 1$ )

$$\lambda_1(Z\gamma) = \frac{\phi(Z\gamma)}{\Phi(Z\gamma)} \quad \text{and} \quad \lambda_0(Z\gamma) = \frac{-\phi(Z\gamma)}{1 - \Phi(Z\gamma)}$$

- implying that the conditional expectation of  $u$  on  $d$  and  $Z$  is a known function of the threshold,  $Z\gamma$ , that determines assignment:  
 $P(d_i = 1|Z_i) = P(v_i > -Z_i\gamma|Z_i)$ .

- Estimates of the control functions specified above can be obtained from a first stage binary response regression of  $d$  on  $Z$ .
- Then including these estimates in the outcome equation

$$y_i = \beta + d_i (\bar{\alpha} + E[\alpha_i - \bar{\alpha} | d_i = 1]) \quad (17)$$

$$+ \left[ \rho d_i \frac{\phi(Z_i \hat{\gamma})}{\Phi(Z_i \hat{\gamma})} + \rho (1 - d_i) \frac{-\phi(Z_i \hat{\gamma})}{1 - \Phi(Z_i \hat{\gamma})} \right] + \delta_i \quad (18)$$

where  $\delta$  is what remains of the error term in the outcome equation

$$\delta_i = u_i + d_i \left( [\alpha_i - \bar{\alpha}] - E[\alpha_i - \bar{\alpha} | d_i = 1] - \hat{E}[u_i | Z, d_i = 1] \right) - (1 - d_i) \hat{E}[u_i | Z, d_i = 0] \quad (19)$$

which is mean independent from  $d$ .

- It is clear from the regression equation (??) that all that is identified when the impact of treatment is heterogeneous is  $\alpha^{ATT} = \bar{\alpha} + E[\alpha_i - \bar{\alpha} | d_i = 1]$ .

# The Control Function Estimator (CF)

- In general we might also write

$$E[\alpha | d = 1, Z] = \rho_a \lambda_1(Z\gamma)$$

- Then including these estimates in the outcome equation

$$y_i = \beta + d_i \left[ \rho_a \frac{\phi(Z_i \hat{\gamma})}{\Phi(Z_i \hat{\gamma})} (d_i = 1) \right] \quad (20)$$

$$+ \left[ \rho d_i \frac{\phi(Z_i \hat{\gamma})}{\Phi(Z_i \hat{\gamma})} + \rho (1 - d_i) \frac{-\phi(Z_i \hat{\gamma})}{1 - \Phi(Z_i \hat{\gamma})} \right] + \tilde{\delta}_i \quad (21)$$

- Given  $\rho_a$  we can recover the ATE.

# The Control Function Estimator (CF)

There are two key assumptions in this parametric CF approach

- 1 the assumption for the joint distribution of unobservables and
  - 2 the linear index assumption  $Z\gamma$ .
- These assumptions can be relaxed and we can apply the control function approach in a nonparametric setting. But we do require the triangularity assumption.
  - The general CF estimator can be shown to have the local IV interpretation.

# The Control Function Estimator (CF)

- The control function method is close to a fully structural approach in the sense that it *explicitly* incorporates the decision process for the assignment rule in the estimation of the impact of the treatment.
- The problem is how to identify the unobservable term,  $v$ , in order to include it in the outcome equation.
- If  $d$  is a *continuous* variable and the decision rule is invertible, then  $d$  and  $Z$  are sufficient to identify  $v$ . In such case,  $v$  is a deterministic function of  $(d, z)$ .
- However, if  $d$  is *discrete*, and  $z$  is *continuous* then we can still recover the complete distribution of treatment effects.
- In this case the probability of  $d = 1$  is a continuous function of  $z$ , say  $P(z)$ . Recall that the MTE is given by

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

## Six distinct, but related, approaches:

1. social experiments methods,
  2. natural experiments,
  3. matching methods,
  4. instrumental methods,
  5. discontinuity design methods
  6. control function methods.
- Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method.
  - The idea has been to relate the different approaches and to set them within the standard semiparametric microeconomic framework.