Lecture 2: Beyond Matching

Costas Meghir

2008-09

What is they key problem with Matching

- Individuals may be selected into treatment on the basis of unobserved characteristics.
- To see this consider the simple choice problem based on the Roy model

$$
\begin{aligned}
Y_i^0 &= \alpha' X_i + u_1^0 \\
Y_i^1 &= \beta' X_i + u_1^1 \\
T_i = 1(Y_i^1 - Y_i^0 > 0) &= 1(u_i^1 - u_i^0 > (\alpha - \beta)' X_i)
\end{aligned}
$$

- Here the selection into treatment is based on the gain and there is a trade-ff between observed and unobserved characteristics in this determination
- Individuals with high values of u^0 will tend not to be allocated to treatment, everything else being equal.
- Matching will not lead to consistent estimates.
- We now investigate alternatives

- Drop the Xs for simplicity and use the regression representation:

$$Y_i = \alpha + \beta T_i + [u_i + (\beta_i - \beta) T_i]$$

- Drop the Xs for simplicity and use the regression representation:

$$Y_i = \alpha + \beta T_i + [u_i + (\beta_i - \beta) T_i]$$

- Define an instrument Z_i such that $E(Z_i T_i) \neq 0$ and $E(Z_i u_i) = E(Z_i(\beta_i - \beta)) = 0$

- Drop the Xs for simplicity and use the regression representation:

$$Y_i = \alpha + \beta T_i + [u_i + (\beta_i - \beta) T_i]$$

- Define an instrument Z_i such that $E(Z_i T_i) \neq 0$ and $E(Z_i u_i) = E(Z_i(\beta_i - \beta)) = 0$
- The IV estimator is defined as $\beta^{IV} = cov(Y, Z)/cov(T, Z)$

- Drop the Xs for simplicity and use the regression representation:

$$Y_i = \alpha + \beta T_i + [u_i + (\beta_i - \beta) T_i]$$

- Define an instrument Z_i such that $E(Z_i T_i) \neq 0$ and $E(Z_i u_i) = E(Z_i(\beta_i - \beta)) = 0$
- The IV estimator is defined as $\beta^{IV} = cov(Y, Z)/cov(T, Z)$
- For a discrete Z this becomes $\beta^{IV} = [E(Y|Z=1) - E(Y|Z=0)] \, / \, [E(T|Z=1) - E(T|Z=0)]$

- Drop the Xs for simplicity and use the regression representation:

$$Y_i = \alpha + \beta T_i + [u_i + (\beta_i - \beta) T_i]$$

- Define an instrument Z_i such that $E(Z_i T_i) \neq 0$ and $E(Z_i u_i) = E(Z_i(\beta_i - \beta)) = 0$
- The IV estimator is defined as $\beta^{IV} = cov(Y, Z)/cov(T, Z)$
- For a discrete Z this becomes $\beta^{IV} = [E(Y|Z=1) - E(Y|Z=0)] \,/\, [E(T|Z=1) - E(T|Z=0)]$
- What does this give?

What does the IV estimator identify?

We can write

$$\beta^{IV} = \beta + \frac{E(Z_i v_i | T_i = 1)\Pr(T_i = 1|Z_i)}{cov(T, Z)}$$

Thus for IV to be consistent for the average treatment effect we need the covariance between the return to treatment and the instrument to be zero *among the treated.* This will genrally not be true.
Under an additional assumption Imbens and Angrist offer an alternative interpretation to Instrumental Variables: this is known as the Local Average Treatment Effect (LATE)

Local Average Treatment Effect (LATE: Imbens and Angrist, Econometrica 1994)

- Define the treatment allocation as a function of an instrument Z : $T_i(Z)$
- The key additional assumption is that of *monotonicity*: Take two values of the instrument $Z = a$ and $Z = b$ such that $a > b$ (or just think of the instrument as discrete). Then for each and every individual $T(Z = a) \geq T(Z = b)$ (or for all i $T_i(Z = a) \leq T_i(Z = b)$).
- The exclusion and rank conditions are defined as follows: For any value of the instrument $Z = a$ $\left\{Y_i^1, Y_i^0, T(Z = a)\right\}$ is independent of Z.Rank condition: The probability of being allocated to treatement $P(T = 1|Z = a) = E(T|Z = a)$ is a *nontrivial* function of a.

Local Average Treatment Effect

- Using this notation and assuming the intrument Z is discrete we can now rewrite what IV identifies.

Local Average Treatment Effect

- Using this notation and assuming the intrument Z is discrete we can now rewrite what IV identifies.
- Take first the difference in mean outcomes for the two values of the instrument, which we take to be $1, 0$:

$$E(Y|Z=1) - E(Y|Z=0) =$$

$$E\left[T(1)Y^1 + (1-T(1))Y^0 | Z=1\right]$$

$$-E\left[T(0)Y^1 + (1-T(0))Y^0 | Z=0\right]$$

Local Average Treatment Effect

- Because of the independence assumption the conditioning on $Z = 1$ or 0 can be ignored. So we can rewrite the above as

Local Average Treatment Effect

- Because of the independence assumption the conditioning on $Z = 1$ or 0 can be ignored. So we can rewrite the above as

$$E(Y|Z=1) - E(Y|Z=0) =$$

$$E\left[(T(1) - T(0))(Y^1 - Y^0)\right] =$$

-

$$E\left[Y^1 - Y^0 | T(1) - T(0) = 1\right] \Pr\left[T(1) - T(0) = 1\right]$$

$$-E\left[Y^1 - Y^0 | T(1) - T(0) = -1\right] \Pr\left[T(1) - T(0) = -1\right]$$

Local Average Treatment Effect

- Because of the independence assumption the conditioning on $Z = 1$ or 0 can be ignored. So we can rewrite the above as
$$E(Y|Z = 1) - E(Y|Z = 0) =$$
$$E\left[(T(1) - T(0))(Y^1 - Y^0)\right] =$$
- $$E\left[Y^1 - Y^0 | T(1) - T(0) = 1\right] \Pr\left[T(1) - T(0) = 1\right]$$
$$-E\left[Y^1 - Y^0 | T(1) - T(0) = -1\right] \Pr\left[T(1) - T(0) = -1\right]$$
- Thus instrumental variables is a weighted difference of the effect for those who move into treatment as a result of Z and those "moving out".

Local Average Treatment Effect

- Because of the independence assumption the conditioning on $Z = 1$ or 0 can be ignored. So we can rewrite the above as

$$E(Y|Z=1) - E(Y|Z=0) =$$

$$E\left[(T(1) - T(0))(Y^1 - Y^0)\right] =$$

-

$$E\left[Y^1 - Y^0 | T(1) - T(0) = 1\right] \Pr\left[T(1) - T(0) = 1\right]$$

$$-E\left[Y^1 - Y^0 | T(1) - T(0) = -1\right] \Pr\left[T(1) - T(0) = -1\right]$$

- Thus instrumental variables is a weighted difference of the effect for those who move into treatment as a result of Z and those "moving out".
- This is meaningless as a parameter

Local Average Treatment Effect

- Because of the independence assumption the conditioning on $Z = 1$ or 0 can be ignored. So we can rewrite the above as
$$E(Y|Z=1) - E(Y|Z=0) =$$
$$E\left[(T(1) - T(0))(Y^1 - Y^0)\right] =$$
- $$E\left[Y^1 - Y^0 | T(1) - T(0) = 1\right] \Pr\left[T(1) - T(0) = 1\right]$$
$$-E\left[Y^1 - Y^0 | T(1) - T(0) = -1\right] \Pr\left[T(1) - T(0) = -1\right]$$
- Thus instrumental variables is a weighted difference of the effect for those who move into treatment as a result of Z and those "moving out".
- This is meaningless as a parameter
- With Homogeneous treatment effects this does indeed become equal to the average treatment effect (=constant treatment effect)

Local Average Treatment Effect

- The *monotonicity assumption* however makes IV interpretable.

Local Average Treatment Effect

- The *monotonicity assumption* however makes IV interpretable.
- In this case $\Pr\left[T(1) - T(0) = -1\right] = 0$

Local Average Treatment Effect

- The *monotonicity assumption* however makes IV interpretable.
- In this case $\Pr\left[T(1) - T(0) = -1\right] = 0$
- Now consider what happens to the probability$\Pr\left[T(1) - T(0) = 1\right]$ which under monotonicity is the probability of the event $T(1) = 1$ and $T(0) = 0$

	T(0)		
T(1)	1,0: Pr(T(1)=1)-Pr(T(0)=1)	1,1: Pr(T(0)=1)	Pr(T(1)=1)
	0,0	0,1: probability=0	Pr(T(1)=0)

Local Average Treatment Effect

- The *monotonicity assumption* however makes IV interpretable.
- In this case $\Pr\left[T(1) - T(0) = -1\right] = 0$
- Now consider what happens to the probability$\Pr\left[T(1) - T(0) = 1\right]$ which under monotonicity is the probability of the event $T(1) = 1$ and $T(0) = 0$

	T(0)		
T(1)	1,0: Pr(T(1)=1)-Pr(T(0)=1)	1,1: Pr(T(0)=1)	Pr(T(1)=1)
	0,0	0,1: probability=0	Pr(T(1)=0)

- Hence *IV* which is

$$\left[E(Y|Z=1) - E(Y|Z=0)\right] / \left[Pr(T(1)=1) - Pr(T(0)=1)\right]$$

is now equal to the average effect of the treatment for those induced into training by the instrument Z.

Local Average Treatment Effect

- So how is this result useful?

Local Average Treatment Effect

- So how is this result useful?
  1. It shows why IV can be meaningless when effects are heterogeneous

Local Average Treatment Effect

- So how is this result useful?
    1. It shows why IV can be meaningless when effects are heterogeneous
    2. It shows that if the monotonicity assumption can be justified IV estimates the effect for a particular subset of the population.

Local Average Treatment Effect

- So how is this result useful?
  1. It shows why IV can be meaningless when effects are heterogeneous
  2. It shows that if the monotonicity assumption can be justified IV estimates the effect for a particular subset of the population.
  3. In general the estimates are specific to that instrument and are not generalisable to other contexts.

Local Average Treatment Effect

- So how is this result useful?
  1. It shows why IV can be meaningless when effects are heterogeneous
  2. It shows that if the monotonicity assumption can be justified IV estimates the effect for a particular subset of the population.
  3. In general the estimates are specific to that instrument and are not generalisable to other contexts.
  4. As an example consider two alternative policies that can increase participation in Higher education

Local Average Treatment Effect

- So how is this result useful?
  1. It shows why IV can be meaningless when effects are heterogeneous
  2. It shows that if the monotonicity assumption can be justified IV estimates the effect for a particular subset of the population.
  3. In general the estimates are specific to that instrument and are not generalisable to other contexts.
  4. As an example consider two alternative policies that can increase participation in Higher education
     1. Free tuition is randomly allocated to young people to attend College ($Z_1 = 1$ means that the subsidy is available).

Local Average Treatment Effect

- So how is this result useful?
  1. It shows why IV can be meaningless when effects are heterogeneous
  2. It shows that if the monotonicity assumption can be justified IV estimates the effect for a particular subset of the population.
  3. In general the estimates are specific to that instrument and are not generalisable to other contexts.
  4. As an example consider two alternative policies that can increase participation in Higher education
     1. Free tuition is randomly allocated to young people to attend College ($Z_1 = 1$ means that the subsidy is available).
     2. The possibility of a competitive scholarship is available for free tuition ($Z_1 = 1$ means that the individual is allowed to compete for the scholarship)

Local Average Treatment Effect

- Suppose the aim is to use these two policies to estimate the returns to College education. In this case the pair $\{Y^1, Y^0\}$ are log earnings, the treatment is going to College and the instrument is one of the two randomly allocated programmes

Local Average Treatment Effect

- Suppose the aim is to use these two policies to estimate the returns to College education. In this case the pair $\{Y^1, Y^0\}$ are log earnings, the treatment is going to College and the instrument is one of the two randomly allocated programmes
- First we need to assume that no one who intended to go to College will be discouraged from doing so as a result of the policy (*monotonicity*)

Local Average Treatment Effect

- Suppose the aim is to use these two policies to estimate the returns to College education. In this case the pair $\{Y^1, Y^0\}$ are log earnings, the treatment is going to College and the instrument is one of the two randomly allocated programmes
- First we need to assume that no one who intended to go to College will be discouraged from doing so as a result of the policy (*monotonicity*)
- This could fail as a result of a General Equilibrium response of the policy; for example if it is perceived that the returns to College decline as a result of the increased supply, those with better outside opportunities may drop out.

- Now compare the two instruments.

- Now compare the two instruments.
- The subsidy is likely to draw poorer liquidity constrained students into College but not necessarily those with the highest returns.

- Now compare the two instruments.
- The subsidy is likely to draw poorer liquidity constrained students into College but not necessarily those with the highest returns.
- The scholarship is likely to draw in the best students, who may also have higher returns.

- Now compare the two instruments.
- The subsidy is likely to draw poorer liquidity constrained students into College but not necessarily those with the highest returns.
- The scholarship is likely to draw in the best students, who may also have higher returns.
- It is not a priori possible to believe that the two policies will identify the same parameter, or that one experiment will allow us to learn about the returns for a broader/different group of individuals

Finally, we need to understand what monotonicity means in terms of restrictions on economic theory

- To quote from Vytlacil (2002) Econometrica:
  "*The LATE assumptions are not weaker than the assumptions of a latent index model, but instead impose the same restrictions on the counterfactual data as the classical selection model if one does not impose parametric functional form or distributional assumptions on the latter.*"

Finally, we need to understand what monotonicity means in terms of restrictions on economic theory

- To quote from Vytlacil (2002) Econometrica:
  "*The LATE assumptions are not weaker than the assumptions of a latent index model, but instead impose the same restrictions on the counterfactual data as the classical selection model if one does not impose parametric functional form or distributional assumptions on the latter.*"
- This is important because it shows that the LATE assumptions are equivalent to whatever economic modelling assumptions are required to justify the standard Heckman selection model and has no claim to greater generality.

Finally, we need to understand what monotonicity means in terms of restrictions on economic theory

- To quote from Vytlacil (2002) Econometrica:
  "*The LATE assumptions are not weaker than the assumptions of a latent index model, but instead impose the same restrictions on the counterfactual data as the classical selection model if one does not impose parametric functional form or distributional assumptions on the latter.*"
- This is important because it shows that the LATE assumptions are equivalent to whatever economic modelling assumptions are required to justify the standard Heckman selection model and has no claim to greater generality.
- On the other hand there are no magical solutions to identifying effects when endogeneity/selection is present; this problem is exacerbated when the effects are heterogeneous and individuals select into treatment on the basis of the returns

- Sometimes we may feel we can impose more structure on the problem
- Suppose in particular that we can write the outcome equation as

$$Y_{it} = \alpha_i + d_t + \beta_i T_{it} + u_{it}$$

- In the above we have now introduced a time dimension $t = \{1, 2\}$.
- Now suppose that $T_{i1} = 0$ for all i and $T_{i2} = 1$ for a well defined group of individuals in our population. Denote the groups by G.

This framework allows us to identify the ATT effect under the assumption that the growth of the outcome in the non-treatment state is independent of treatment allocation:

$$E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$$

Difference in Differences

We have that

$$E\left[Y_{i2} - Y_{i1} \middle| T_{i2} = 1\right] = E\left[Y^1_{i2} - Y^1_{i1} \middle| T_{i2} = 1\right]$$

$$= d_2 - d_1 + E\left[\beta_i \middle| T_{i2} = 1\right]$$

- The above is the basis for the *before and after* estimator. However, this is not consistent because in general the time effects d_t differ with t.Thus if there is overall growth in the outcome variable this will be attributed to the treatment erroneously.

- Now consider:

$$E\left[Y_{i2}^{0}-Y_{i1}^{0}\middle|T_{i2}=1\right]=E\left[Y_{i2}^{0}-Y_{i1}^{0}\middle|T_{i2}=0\right]$$

$$E\left[Y_{i2}-Y_{i1}\middle|T_{i2}=0\right]=d_2-d_1$$

- Now consider:

$$E\left[Y_{i2}^0 - Y_{i1}^0 \middle| T_{i2} = 1\right] = E\left[Y_{i2}^0 - Y_{i1}^0 \middle| T_{i2} = 0\right]$$

$$E\left[Y_{i2} - Y_{i1} \middle| T_{i2} = 0\right] = d_2 - d_1$$

- We need an extra assumption: We can then obtain immediately an estimator for ATT as

$$E\left[\beta_i \middle| T_{i2} = 1\right] = E\left[Y_{i2} - Y_{i1} \middle| T_{i2} = 1\right] - E\left[Y_{i2} - Y_{i1} \middle| T_{i2} = 0\right]$$

which can be estimated by the difference in the growth between the treatment and the control group.

Difference in Differences

- Now consider the following problem:

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group
- Those with higher earnings form the control group.

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons
  1. They have low permanent earnings (a_i is low) - this is accounted for by diff in Diffs

Difference in Differences

- Now consider the following problem:
- Suppose we wish to evaluate a training programme for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons
  1. They have low permanent earnings (a_i is low) - this is accounted for by diff in Diffs
  2. They have a negative transitory shock (u_{i1} is low) - this is *not* accounted for by diff in Diffs

- Number 2 above violates the assumption that $E\left[Y_{i2}^0 - Y_{i1}^0 \middle| T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.

- Number 2 above violates the assumption that $E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.
- To see why note that those participating into the programme are such that $Y_{i1}^0 < B$. Assume for simplicity that the shocks u are *iid*. Hence $u_{i1} < B - a_i - d_1$. This implies that $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 1\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} < B - a_i - d_1\right]$. For the control group $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 0\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} > B - a_i - d_1\right]$.

- Number 2 above violates the assumption that $E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.
- To see why note that those participating into the programme are such that $Y_{i1}^0 < B$. Assume for simplicity that the shocks u are *iid*. Hence $u_{i1} < B - a_i - d_1$. This implies that $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 1\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} < B - a_i - d_1\right]$. For the control group $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 0\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} > B - a_i - d_1\right]$.
- Hence

$$\begin{aligned} E\left[Y_{i2}^0 - Y_{i1}^0 | T = 1\right] - E\left[Y_{i2}^0 - Y_{i1}^0 | T = 0\right] &= \\ E\left[u_{i1} | u_{i1} > B - a_i - d_1\right] - E\left[u_{i1} | u_{i1} < B - a_i - d_1\right] &> 0 \end{aligned}$$

- Number 2 above violates the assumption that $E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.
- To see why note that those participating into the programme are such that $Y_{i1}^0 < B$. Assume for simplicity that the shocks u are *iid*. Hence $u_{i1} < B - a_i - d_1$. This implies that $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 1\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} < B - a_i - d_1\right]$. For the control group $E\left[Y_{i2}^0 - Y_{i1}^0 | T = 0\right] = d_2 - d_1 - E\left[u_{i1} | u_{i1} > B - a_i - d_1\right]$.
- Hence

$$\begin{aligned} E\left[Y_{i2}^0 - Y_{i1}^0 | T = 1\right] - E\left[Y_{i2}^0 - Y_{i1}^0 | T = 0\right] &= \\ E\left[u_{i1} | u_{i1} > B - a_i - d_1\right] - E\left[u_{i1} | u_{i1} < B - a_i - d_1\right] &> 0 \end{aligned}$$

- This is effectively regression to the mean: those unlucky enough to have a bad shock recover and show greater growth relative to those with a good shock. The nature of the bias depends on the stochastic properties of the shocks and how individuals select into training.

- Ashefelter (1978) was one of the first to consider difference in differences to evaluate training programmes

TABLE 1.—MEAN EARNINGS PRIOR, DURING, AND SUBSEQUENT TO TRAINING FOR 1964 MDTA CLASSROOM TRAINEES AND A COMPARISON GROUP

	White Males		Black Males		White Females		Black Females	
	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group
1959	$1,443	$2,588	$ 904	$1,438	$ 635	$ 987	$ 384	$ 616
1960	1,533	2,699	976	1,521	687	1,076	440	693
1961	1,572	2,782	1,017	1,573	719	1,163	471	737
1962	1,843	2,963	1,211	1,742	813	1,308	566	843
1963	1,810	3,108	1,182	1,896	748	1,433	531	937
1964	1,551	3,275	1,273	2,121	838	1,580	688	1,060
1965	2,923	3,458	2,327	2,338	1,747	1,698	1,441	1,198
1966	3,750	4,351	2,983	2,919	2,024	1,990	1,794	1,461
1967	3,964	4,430	3,048	3,097	2,244	2,144	1,977	1,678
1968	4,401	4,955	3,409	3,487	2,398	2,339	2,160	1,920
1969	$4,717	$5,033	$3,714	$3,681	$2,646	$2,444	$2,457	$2,133
Number of Observations	7,326	40,921	2,133	6,472	2,730	28,142	1,356	5,192

Ashenfelter (1978) reports the following results

TABLE 2.—CRUDE ESTIMATES (AND ESTIMATED STANDARD ERRORS), ASSUMING $B=0$ AND $\beta_j'=0$ FOR $j>1$, OF THE EFFECT OF TRAINING ON EARNINGS DURING AND AFTER TRAINING, WHITE MALE MDTA 1964 CLASSROOM TRAINEES

Effect in (value of t)	Value of Effects for		
	$t-s=1963$	$t-s=1962$	$t-s=1961$
1962	—	—	91 (13)
1963	—	−179 (14)	−88 (17)
1964	−426 (16)	−605 (18)	−514 (20)
1965	763 (20)	584 (22)	675 (23)
1966	697 (25)	518 (27)	609 (28)
1967	833 (28)	655 (30)	746 (31)
1968	745 (34)	566 (35)	657 (36)
1969	984 (37)	805 (39)	896 (40)

- The assumption on growth of the non-tratment outcome being independent of assignment to treatment may be violated, but it may still be tru conditional on X

- The assumption on growth of the non-tratment outcome being independent of assignment to treatment may be violated, but it may still be tru conditional on X
- Consider the assumption

$$E\left[Y_{i2}^0 - Y_{i1}^0 | X, T\right] = E\left[Y_{i2}^0 - Y_{i1}^0 | X\right]$$

- The assumption on growth of the non-tratment outcome being independent of assignment to treatment may be violated, but it may still be tru conditional on X
- Consider the assumption

$$E\left[Y_{i2}^{0}-Y_{i1}^{0}|X,T\right]=E\left[Y_{i2}^{0}-Y_{i1}^{0}|X\right]$$

- This is just matching assumption on a redefined outcome variable, namely the growth in the outcomes. In its simplest form the approach is implemented by running the regression

$$Y_{it}=\alpha_i+d_t+\beta_i T_{it}+\gamma_t' X_i+u_{it}$$

which allows for differential trends in the non-treatment growth depending on X_i.More generally one can implement propensity score matching on the growth of outcome variable when panle data is available.

Difference in Differences with repeated cross sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment a post-treatment period. We can still use difference in differences.

Difference in Differences with repeated cross sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment a post-treatment period. We can still use difference in differences.
- First consider the simpler case where $E\left[Y_{i2}^{0}-Y_{i1}^{0}|T\right]=E\left[Y_{i2}^{0}-Y_{i1}^{0}\right]$.

Difference in Differences with repeated cross sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment a post-treatment period. We can still use difference in differences.
- First consider the simpler case where $E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.
- We need to modify slightly the assumption to

$$E\left[Y_{i2}^0 | \text{Group receiving training}\right] - E\left[Y_{i1}^0 | \text{Group to receive training in next period}\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$$

  which essentially requires, in addition to the original independence assumption that conditioned on particular individuals that population we will be sampling from does not change composition

Difference in Differences with repeated cross sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment a post-treatment period. We can still use difference in differences.
- First consider the simpler case where $E\left[Y_{i2}^0 - Y_{i1}^0 | T\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$.
- We need to modify slightly the assumption to

$$E\left[Y_{i2}^0|\text{Group receiving training}\right] - E\left[Y_{i1}^0|\text{Group to receive training in next period}\right] = E\left[Y_{i2}^0 - Y_{i1}^0\right]$$

which essentially requires, in addition to the original independence assumption that conditioned on particular individuals that population we will be sampling from does not change composition

- We can then obtain immediately an estimator for ATT as

$$E\left[\beta_i | T_{i2} = 1\right] =$$
$$E\left[Y_{i2}|\text{Group receiving training}\right] - E\left[Y_{i1}|\text{Group to receive training in next period}\right]$$
$$-\left\{E\left[Y_{i2}|\text{Non-trainees}\right] - E\left[Y_{i1}|\text{Group not to receive training in next period}\right]\right\}$$

Difference in Differences with repeated cross sections

- More generally we need an assumption of conditional independence of the form

$$E\left[Y_{i2}^{0}|X, \text{Group receiving training}\right] - E\left[Y_{i1}^{0}|X, \text{Group to receive training in next per}\right.$$
$$= E\left[Y_{i2}^{0}|X\right] - E\left[Y_{i1}^{0}|X\right]$$

Difference in Differences with repeated cross sections

- More generally we need an assumption of conditional independence of the form

$$E\left[Y_{i2}^0|X, \text{Group receiving training}\right] - E\left[Y_{i1}^0|X, \text{Group to receive training in next per}\right.$$
$$= E\left[Y_{i2}^0|X\right] - E\left[Y_{i1}^0|X\right]$$

- Under this assumption (and some auxiliary parametric assumptions) we can obtain an estimate of the effect of treatment on the treated by the regression

$$Y_{it} = \alpha_g + d_t + \beta T_{it} + \gamma' X_{it} + u_{it}$$

Difference in Differences with repeated cross sections

- More generally we can first run the regression

$$Y_{it} = \alpha_g + d_t + \beta(X_{it})T_{it} + \gamma_t' X_{it} + u_{it}$$

where α_g is a dummy for the treatment of comparison group, and where $\beta_i(X_{it})$ can be parameterised as $\beta(X_{it}) = \beta' X_{it}$. The ATT can then be estimated as the average of $\beta' X_{it}$over the (empirical) distribution of X.

Difference in Differences with repeated cross sections

- More generally we can first run the regression

$$Y_{it} = \alpha_g + d_t + \beta(X_{it})T_{it} + \gamma_t' X_{it} + u_{it}$$

  where α_g is a dummy for the treatment of comparison group, and where $\beta_i(X_{it})$ can be parameterised as $\beta(X_{it}) = \beta' X_{it}$. The ATT can then be estimated as the average of $\beta' X_{it}$over the (empirical) distribution of X.

- A non parametric alernative is offered by Blundell, Dias, Meghir and van Reenen (2004)

Difference in Differences and selection on unobservables

- Suppose we relax the assumption of *no selection* on unobservables

Difference in Differences and selection on unobservables

- Suppose we relax the assumption of *no selection* on unobservables
- Instead we can start by assume that

$$E\left[Y_{i2}^{0}|X,Z\right]-E\left[Y_{i1}^{0}|X,Z\right]$$
$$=E\left[Y_{i2}^{0}|X\right]-E\left[Y_{i1}^{0}|X\right]$$

where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0)

Difference in Differences and selection on unobservables

- Suppose we relax the assumption of *no selection* on unobservables
- Instead we can start by assume that

$$E\left[Y_{i2}^{0}|X,Z\right]-E\left[Y_{i1}^{0}|X,Z\right]$$
$$=E\left[Y_{i2}^{0}|X\right]-E\left[Y_{i1}^{0}|X\right]$$

  where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0)

- Non-Compliance: not all members of the eligible group ($Z=1$) will take up training and some of those ineligible ($Z=0$) may obtain training by other means.

Difference in Differences and selection on unobservables

- Suppose we relax the assumption of *no selection* on unobservables
- Instead we can start by assume that

$$E\left[Y_{i2}^{0}|X,Z\right]-E\left[Y_{i1}^{0}|X,Z\right]$$
$$=E\left[Y_{i2}^{0}|X\right]-E\left[Y_{i1}^{0}|X\right]$$

  where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0)
- Non-Compliance: not all members of the eligible group $(Z = 1)$ will take up training and some of those ineligible $(Z = 0)$ may obtain training by other means.
- A difference in differences approach based on grouping by Z will estimate the impact of being allocated to the eligible group, but not the impact of training itself.

Difference in Differences and selection on unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)

Difference in Differences and selection on unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions

Difference in Differences and selection on unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
  - Independence: for $Z = a$, $\left\{ Y^0_{i2} - Y^0_{i1}, Y^1_{i2} - Y^1_1, T(Z = a) \right\}$ is independent of Z

Difference in Differences and selection on unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
  - Independence: for $Z = a$, $\left\{ Y_{i2}^{0} - Y_{i1}^{0}, Y_{i2}^{1} - Y_{1}^{1}, T(Z = a) \right\}$ is independent of Z
  - Monotonicity $T_i(1) \geq T_i(0) \; \forall \; i$

Difference in Differences and selection on unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
  - Independence: for $Z = a$, $\left\{Y^0_{i2} - Y^0_{i1}, Y^1_{i2} - Y^1_1, T(Z = a)\right\}$ is independent of Z
  - Monotonicity $T_i(1) \geq T_i(0) \ \forall \ i$
- In this case LATE is defined by

$$[E(\Delta Y|Z = 1) - E(\Delta Y|Z = 0)] \,/\, [Pr(T(1) = 1) - Pr(T(0) = 1)]$$

assuming that the probability of training in the first period is zero