Matching and selection models: a comparison

Pierre Mohnen

Univ. of Maastricht and UNU-MERIT

Matching methods

- Matching estimators
- Difference in differences
- Regression discontinuity design

Example in Ravallion (1991)

- Evaluation of an anti-poverty program that provides cash transfers to poor families with school-age children. To receive the transfer families have to show that they are poor and must keep their children at school until they are 18.
- Notations:
  - D_i=participation dummy (0/1)
  - S_{0i}= schooling if D_i=0
  - S_{1i}= schooling if D_i=1

Counterfactuals

- ATE= $E(S_{i1}|D_i=1) - E(S_{i0}|D_i=0)$ is not quite right because individuals who choose $D_i=1$ may differ from those who choose $D_i=0$ and these individual characteristics may also affect choice of schooling
- Better way: $E(S_{i1}|D_i=1) - E(S_{i0}|D_i=1)$= ATET
  - ATET (average treatment effect on the treated)
  - ATET = ATE if random assignment, i.e. $E(S_{i0}|D_i=0) = E(S_{i0}|D_i=1)$
  - The program will probably pick poor families who would not have put their children to school without the program
  - Hence $E(S_{i0}|D_i=1) < E(S_{i0}|D_i=0)$ and ATE is underestimated
- $S_{i0}|D_i=1$ is not observed, hence counterfactual needs to be estimated

Another way to look at it: regression analysis

- $S_i = a + bD_i + \varepsilon_i$

  – where ε_i captures unobservable influences on choice of schooling with $E(\varepsilon_i|D_i)=0$

- $E(S_{i1}|D_i=1) = a + bD_i$
- $E(S_{i0}|D_i=0) = a$
- Hence ATE = b

Biases in ATE

- Heterogeneity between the two groups:
  - $S_{i0} = a_0 + \varepsilon_{i0}$ if $D_i=0$
  - $S_{i1} = a_1 + b_{1Di} + \varepsilon_{i1}$ if $D_i=1$
  - $ATE = (a_1-a_0) + b_{1Di} + E(\varepsilon_{i1}- \varepsilon_{i0}|D_i)$

- Omitted variables
  - $S_i = a + bD_i + cX_i + \varepsilon_i$ with $E(\varepsilon_i|X_i,D_i)=0$

- D is not exogenous, i.e. $E(\varepsilon_i|X_i,D_i)\neq 0$
  - $D_i = d + eZ_i + v_i$ with $E(\varepsilon_i|Z_i)\neq 0$ or $E(\varepsilon_i v_i |Z_i,X_i)\neq 0$

IV and omitted variables

- example: regress earnings on schooling knowing that ability plays also a role, but ability cannot be measured
- The true model: $Y_i = \alpha + \rho S_i + \beta A_i + \varepsilon_i$
- The regression of Y on S (omitting A) would yield $\rho + \beta\gamma$ as the coefficient of S where γ is the regression coefficient from a regression of A on S.
- Absence of a bias if β=0 or if γ=0.

IV and omitted variables (2)

- Solution: instrumental variables
- Instead of exploiting $X'\varepsilon=0=X'(Y-X\beta)=0$
  - Hence $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$
- We exploit the non-correlation between Z and ε: $Z'\varepsilon=0=Z'(Y-X\beta)=0$
  - Hence $\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$

Good instruments

- “Good instrument is correlated with the endogenous regressor for reasons the researcher can verify and explain, but uncorrelated with the outcome variable for reasons beyond its effect on the endogenous regressor.” (Angrist and Krueger, p. 73)

- If Z are weak instruments, i.e. not strongly correlated with X, or if the instruments are correlated with the outcome variable (i.e. with ε), IV can lead to a bias maybe even larger than OLS

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1}Z'\varepsilon \approx \beta + \mathrm{cov}(Z,\varepsilon)/\mathrm{cov}(Z,X)$$

Choice of instruments

- “In our view, good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest.” (Angrist and Krueger , p.73)
- Natural experiments (environment similar to randomized experiment)

Example: Angrist and Krueger, QJE, 1991

- differences in length of schooling because of different quarter of birth: those born in the quarter before Dec 31 enter at 5 ¾ years at school, those born in the quarter after Dec 31 enter with 6 ¼ years, and they all have to stay in school until they reach 16.
- Hence those born before Dec 31, benefit from one more year of schooling.

Intuition behind instrumental variables

- "Instrumental variables solve the omitted variable problem by using only part of the variability in schooling – specifically, a part that is uncorrelated with the omitted variables – to estimate the relationship between schooling and earnings." (p. 39)
- Of course, this instrument only works for students leaving school just after reaching 16. After that, the length of schooling becomes again endogenous.

Endogeneity of treatment or selection bias

- D is not exogenous, i.e. $E(\varepsilon_i|X_i,D_i)\neq 0$
  - $D_i = d + eZ_i + v_i$ with $E(\varepsilon_i|Z_i)\neq 0$ or $E(\varepsilon_i v_i|Z_i,X_i)\neq 0$
- There is no selection bias if
  - either all observable variables Z are completely included in X
  - or there is no correlation between the unobservables v_i and ε_i

Matching estimators: assumptions

- Conditional independence assumption
  - $S_o,S_1 \perp D|X$
  - once we control for covariates X (some of which may relate to D) treatment and outcome are independent
  - Also known as unconfoundedness assumption (no omitted variables) or ignorability assumption (treatment assignment ignores outcome)
- Common support assumption, also known as overlap or matching assumption
  - $0 < P[D=1|X] < 1$
  - For a given X, you should always be able to find an observation in both the treated and the untreated
- Stable unit treatment value assumption
  - Treatment does not affect untreated, i.e. no externalities, general equilibrium effects

Idea of matching

- For every treated observation find an observation from the control group (non-treated) that is similar in all respects (i.e. observables that matter, X) and then compare the means of the outcome variable among the two groups

- Selection on observables

Matching methods (2)

- Exact matching (impossible if X has high dimension or if there are continuous variations)
- Propensity score matching
  - Estimate $P[D_i=1|X_i]$ by probit or logit
  - Match on the basis of the propensity score
  - Useful if high dimension of X
  - With or without replacement? (depends on size of comparison group)
  - Nearest neighbor matching, kernel matching, stratification or interval matching, radius or caliper matching (question of bias versus variance)

Differences in differences

- Allow for an unobserved individual effect that is time independent
- Requires data before and after treatment
- Compare treatment and comparison group in difference in outcome before and after the treatment
- ATET= $[E(S_{i1}-S_{i0})|Di=1]-[E(S_{i1}-S_{i0})|Di=0]$
- $S_{iB} = a + cX_{iB} + \eta_i + \varepsilon_{iB}$
- $S_{iA} = a + bD_i + cX_{iA} + \eta_i + \varepsilon_{iA}$
- $S_{iA}- S_{iB} = bD_i + c(X_{iA}-X_{iB}) + (\varepsilon_{iA}- \varepsilon_{iB})$
- again you could control for different a and c before and after
- Problem of endogenous attrition: attrition bias

Regression discontinuity design

- Quasi-experimental design where probability of receiving treatment is a discontinuous function of a variable at an exogenous cut-off point
- Example threshold for receiving treatment
- Matching below and above the threshold

Instrumental variable methods

- What if selection depends on unobservables possibly correlated with unobservables in outcome equation?
- Selection on unobservables
- e.g. intelligence, motivation
- Maximum likelihood estimation of selection and outcome equation with joint distribution of the two error terms

Selection on unobservables

- D=1 if $Z\gamma+v > 0$
- D=0 if $Z\gamma+v <= 0$

- $Y_j = a_j + b_j X_j + \varepsilon_j \quad j = 0,1$
- $[v\ \varepsilon_0\ \varepsilon_1]'$ distributed as $N(0,\Sigma)$

Comparison

- <u>Matching</u>
- Propensity score
- Differences in differences

- <u>Parametric approach</u>
- Selection equation
- Control for time-invariant heterogeneity
- Control for time-variant correlation between errors in selection and outcome equations

Criteria for choosing a method

- Identify factors that can affect outcome and the selection
- Common factors behind selection and outcome observable or unobservable
- Quantity and kind of data available conditions what is feasible: the more you need to control for, the more data you need
- The less restrictive the assumptions of independence, the more you need to make assumptions about functional form of outcome and selection equations and error distribution

References

- Angrist, J. and A. Krueger (2001), “Instrumental variables and the search for identification: from supply and demand to natural experiments”, Journal of Economic Perspectives, 15(4), 69-85.
- Ravallion, M. (2001), “The mystery of the vanishing benefits: An introduction to impact evaluation”, The World Bank Economic Review, 15(1), 115-140.
- Cameron, A. Colin and Pravin K. Trivedi