IV in observational studies and "imperfect" experiments: what, why and how

Di Porto Edoardo[°]

[°] CSEF-University of Naples Federico II and UCFS Uppsala

Pisa, November , 2019

- This lecture focus on straightforward causal questions that are ideally addressed with randomized experiments.

  1. In practice, however, traditional randomized trials are difficult to implement in the untidy/complex world of social science.
  2. This lecture argues that the instrumental variables methods (IV) used to solve omitted variables bias problems in observational studies also solve the major statistical problems that arise in imperfect experiments.
  3. In general, IV methods estimate causal effects on subjects who comply with a randomly assigned treatment

- The use of IV is illustrated through a re-analysis of the Minneapolis domestic violence experiment (1984)

Design failures

- The use of randomized trials (RCT) in social sciences has continued to grow (i.e. criminology, tax evasion, education, epidemiology, development)
- Two sorts of design failure seem especially likely:
  1. *treatment dilution*, is when subjects or units assigned to the treatment group do not get treated.
  2. *treatment migration*, is when subjects or units in the control group nevertheless obtain the experimental treatment
- These scenarios are indeed potential threats to the validity of a randomized trial

- With non-random crossovers, the group that ends up receiving treatment may no longer be comparable to the remaining pool of untreated controls.
- In addition, if intended treatment is only an imperfect proxy for treatment received, it seems clear that an analysis based on the original intention-to-treat probably understates the causal effect of treatment per se.
- These problems arise when neither subjects nor those delivering treatment can be blinded and, must, in any case, be given some discretion as to program participation for both practical and ethical reasons.

IV might solve these failures

- The purpose of this lecture is to show how the instrumental variables (IV) methods solve both the treatment dilution and treatment migration problems.
- the IV framework also opens up the possibility of a wide range of flexible experimental research designs.
- These designs are unlikely to raise the sort of ethical questions that are seen as limiting the applicability of traditional experimental designs.
- Finally, the logic of IV suggests a number of promising quasi-experimental research designs that may provide a reasonably credible (and inexpensive) substitute for RCT.

Minneapolis domestic violence experiment (MDVE), Sherman and Berk (1984)

- one of the most influential experiment in criminological research
- The MDVE was motivated by debate over the importance of deterrence effects in the police response to domestic violence.
- Police are often reluctant to make arrests for domestic violence unless the victim demands an arrest or the suspect does something that warrants arrest

Motivation

- this attitude has many sources: a general reluctance to intervene in family disputes,
- the fact that domestic violence cases may not be prosecuted,
- genuine uncertainty as to what the best course of action is, and
- an incorrect perception that domestic assault cases are especially dangerous for arresting officers

Experimental Design

- The research design incorporated three treatments:
  1. arrest,
  2. ordering the offender off the premises for 8 hours,
  3. and some form of advice that might include mediation.
- The research design called for one of these three treatments to be randomly selected each time participating Minneapolis police officers encountered a situation meeting the experimental criteria:
- some kind of apparent misdemeanour domestic assault where there was probable cause to believe that a cohabitant or spouse had committed an assault against the other party in the past 4 hours.
- Cases of life-threatening or severe injury, i.e., felony assault, were excluded. Both suspect and victim had to be present upon the intervening officers‘ arrival.

Devices and Adherence to the protocol

- The randomization device was a pad of report forms that were randomly color coded for each of the three possible responses.
- Officers who encountered a situation that met the experimental criteria were to act according to the color of the form on top of the pad.
- The police officers who participated in the experiment had volunteered to take part, and were therefore expected to comply with the research design.
- On the other hand, strict adherence to the protocol was understood by the experimenters to be both unrealistic and inappropriate

RCT in practice

- In practice, officers often deviated from the responses called for by the color of the report form drawn at the time of an incident.
  1. In some cases, suspects were arrested when random assignment called for separation or advice.
  2. Most arrests in these cases came about when a suspect attempted to assault an officer, a victim persistently demanded an arrest, or if both parties were injured.
  3. In one case where random assignment called for arrest, officers separated instead. In a few cases, advice was swapped for separation and vice versa.
  4. However most deviations from the intended treatment reflected purposeful action on the part of the officers involved, sometimes deviations arose when officers simply forgot to bring their report forms.

Non compliance

- As noted above, non-compliance with random assignment is not unique to the MDVE or criminological research.
- Any experimental intervention where ethical or practical considerations lead to a deviation from the original research protocol is likely to have this feature.
- It seems fair to say that non-compliance is usually unavoidable in research using human subjects.
- In the MDVE, the most common deviation from random assignment was the failure to separate or advise when random assignment called for this.

Non compliance

Table 1. Assigned and delivered treatments in spousal assault cases.

	Delivered treatment			
		Coddled		
Assigned treatment	*Arrest*	*Advise*	*Separate*	*Total*
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0 (314)

The table shows statistics from Sherman and Berk (1984), Table 1.

Endogenous Non compliance

- The random assignment of intended treatments in the MDVE does not appear to have been subverted.
- At the same time, it is clear that delivered treatments had a substantial behavioral component.

  The variable *treatment delivered* is, **endogenous**.
- *delivered treatment* were determined by unobserved features that were very likely correlated with outcome variables: i.e. re-offense.
- ... some of the suspects who were arrested in spite of having been randomly assigned to receive advice (or be separated) were especially violent.

The Intention To Treat: ITT

- A simple approach to the analysis of randomized clinical trials with imperfect compliance is:
- to compare subjects according to original random assignment, ignoring compliance entirely.
- This is known as an **intention-to-treat (ITT) analysis**.

The Intention To Treat: ITT

- ITT comparisons use only the original random assignment, and ignore information on *treatments delivered*,
- they indeed provide unbiased estimates of the causal effect of intention to treat.
- ITT estimates are (almost) always too small relative to the effect of treatment.
- **ITT is not ATE**

ITT almost always lower than ATE

- ITT effect is, except under very unusual circumstances, diluted by *non-compliance*. This dilution diminishes as compliance rates go up.
- ITT effect provides a poor predictor of the ATE of similar interventions in the future
- if compliance rates substantially growth because of intervention, compliance growth and ITT of future experiments might growth, this means that the first ITT is very misleading .

- The simplest and most robust solution to the treatment-dilution and treatment migration problems is instrumental variables (Angrist, 2006).

- This can be seen most easily using a conceptual framework that postulates a set of potential outcomes that could be observed in alternative states of the world.

Let's simplify our experiment

- Because the policy discussion in the domestic assault context focuses primarily on the decision to arrest and possible alternatives,
- We define a binary (dummy) treatment variable for not arresting, which We call **coddling**.
- A suspect was randomly assigned to be coddled if the officer on the scene was instructed by the random assignment protocol to **advise or separate**.

The outcome variable

- The most important outcome variable in the MDVE was recidivism, i.e., the occurrence of post-treatment domestic assault by the same suspect.

- Let Y_i denote the observed re-offense status of suspect i.

- The potential outcomes in the binary treatment version of MDVE are the re-offense status of suspect i if he were coddled, denoted Y_{1i},
- and the re-offense status of suspect i if he were not coddled, denoted Y_{0i}.

- Both of these potential outcomes are assumed to be well-defined for each suspect even though only one is ever observed.

The outcome variable in potential outcome

- Let Di denote the treatment delivered to subject i. Then we can write the observed recidivism outcome as:

$$Y_i = Y_{0i}(1 - D_i) + Y_{1i}(D_i) \tag{1}$$

- In words, this means we get to see the Y_{1i} for any subject who was coddled, but we do not know whether he would have re-offended if he had been arrested.

- Likewise, we get to see Y_{0i} for any subject who was arrested, but we do not know whether he would have re-offended had he been coddled.

- It is usual to start an empirical analysis is by comparing outcomes on the basis of treatment delivered.

- however in case of non-random nature treatment assignment, such naive comparisons are likely to be misleading. This can be seen formally by writing:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] = \qquad (2)$$

$$= E[Y_{1i} - Y_{0i}|D_i = 1] + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\} \qquad (3)$$

- the last eq. shows ATET + {selection bias} **in case of non random assignment**

The analysis: perfect compliance

- Selection bias disappears when delivered treatment is determined in a manner independent of potential outcomes, as in a randomized trial with perfect compliance. We then have:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}] \quad (4)$$

- With perfect compliance, the simple treatment-control comparison recovers ATET.

- Moreover, because $\{Y_{1i}, Y_{0i}\}$ is assumed to be independent of D_i in this case, ATET is also the population average treatment effect, $E[Y_{1i} - Y_{0i}]$.

- The most important consequence of non-compliance is *the likelihood of a relation* between **potential outcomes** and **delivered treatments**.

- This relation confounds analyses based on delivered treatments because of selection bias.

- But we have an **ace in the hole**: the compliance problem does not compromise the independence of potential outcomes and randomly assigned **intended treatments**.

IV provides a solution

- The easiest way to see how IV solves the compliance problem is in the context of a model with constant treatment effects, i.e., Y_{1i} - $Y_{0i} = \alpha$, for some constant, α .

- Also, let $Y_{0i} = \beta + \epsilon_i$, where $\beta = E[Y_{0i}]$.

- The potential outcomes model can now be written

$$Y_i = \beta + \alpha D_i + \epsilon_i \tag{5}$$

  where α is the treatment effect of interest. Note that because D_i is a dummy variable, the regression of Y_i on D_i is just the difference in mean outcomes by delivered treatment status.

- As noted above, this difference does not consistently estimate α because Y_{0i} and D_i are not independent (equivalently, ϵ_i and D_i are correlated).

- The random assignment of intended treatment status, **which we will call** Z_i, provides the key to untangling causal effects in the face of non compliance.

- By virtue of random assignment, and the assumption that assigned treatments have no direct effect on potential outcomes other than through delivered treatments, Y_{0i} and Z_i are independent. It therefore follows that

$$E[\epsilon_i, Z_i] = 0 \tag{6}$$

- by the way ϵ_i and D_i are not indipendent

IV provides a solution

- Taking conditional expectations of

$$Y_i = \beta + \alpha D_i + \epsilon_{\mathrm{i}} \tag{7}$$

with Z_i switched off and on, we obtain a simple formula for an interesting treatment effect :

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]/E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = \alpha \tag{8}$$

- Thus, the causal effect of delivered treatments is given by the causal effect of assigned treatments (the ITT effect) divided by

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \tag{9}$$

IV provides a solution

- Note that in experiments where there is complete compliance in the comparison group (i.e., no controls get treated), eq. (8) is just the ITT effect divided by the compliance rate in the originally assigned treatment group.
- More generally, the denominator in eq. (8) is the difference in compliance rates by assignment status. In the MDVE: $E[D_i|Z_i = 1] = P[D_i = 1|Z_i = 1] = .77$, that is, a little over three-fourths of those assigned to be coddled were coddled.
- Note that, almost no one assigned to be arrested was coddled compliance rate $= .01$: Hence, the denominator of eq. (8) is estimated to be about .76.

Wald estimator

- The sample analog of eq. (8) is called a Wald estimator (Wald, 1940).
- The law of large numbers, which says that sample means converge in probability to population means, ensures that the Wald estimator of α is consistent (i.e., converges in probability to α).

- The constant-effects assumption is clearly unrealistic.
- There is also important heterogeneity in treatment delivery. Some suspects would have been coddled with or without the experimental manipulation, while others were coddled only because the police were instructed to treat them this way. **The MDVE is informative about causal effects only on this latter group**.

Not for everybody, just for compliers

- Imbens and Angrist (1994) showed that in a world of heterogeneous treatment effects, IV methods capture the average causal effect of delivered treatments on the subset of treated men whose delivered treatment status can be changed by the random assignment of intended treatment status.
- The men in this group are called **compliers**,
- In a randomized drug trial compliers are those who take their medicine when randomly assigned to do so, but not otherwise
- In the MDVE, compliers were coddled when randomly assigned to be coddled but would not have been coddled otherwise

- The average causal effect for compliers is called a local average treatment effect (LATE).

- A formal description of LATE requires more notations.

- Define potential treatment assignments, D_{0i} and D_{1i}, to be individual i treatment status when Z_i equals 0 or 1. Note that one of D_{0i} or D_{1i} is necessarily counterfactual since observed treatment status is:

$$D_i = D_{0i} + Z_i(D_{1i} - D_{0i}) \tag{10}$$

- Assumptions for LATE:
  1. conditional indipendence the joint distribution of $\{D_{1i}, D_{0i}, Y_{1i}, Y_{0i}\}$ is independent of Z_i
  2. monotonicity, which requires that either $D_{1i} \geq D_{0i}$ for all i or vice versa.
- Monotonicity requires that, while the instrument might have no effect on some individuals, all of those affected are affected in the same way.
- Monotonicity in the MDVE amounts to assuming that random assignment to be coddled can only make coddling more likely. Given these two identifying assumptions, the Wald estimator consistently estimates LATE, which is written formally as:

$$E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}] \tag{11}$$

LATE: the compliers

- Compliers are those with $D_{1i} > D_{0i}$, i.e., they have $D_{1i} = 1$ and $D_{0i} = 0$.

- The monotonicity assumption partitions the world of experimental subjects into three groups:
  1. compliers who are affected by random assignment
  2. always-takers, i.e., subjects with $D_{1i} = D_{0i} = 1$.
  3. never-takers, i.e., subjects with $D_{1i} = D_{0i} = 0$.

- Because the treatment status of always-takers and never-takers is invariant to random assignment, IV estimates are uninformative about treatment effects for subjects in these groups.

- LATE is not the same as ATET,

$$D_i = D_{0i} + Z_i(D_{1i} - D_{0i}) \tag{12}$$

- shows that the treated can be divided into two groups:
  1. the set of subjects with $D_{0i} = 1$,
  2. and the set of subjects with $D_{0i} = 0, D_{1i} = 1$, and $Z_i = 1$.
- Subjects with $D_{0i} = 1$, are always-takers since $D_{0i} = 1$ implies $D_{1i} = 1$ by monotonicity.
- The other are compliers with $Z_i = 1$.
- By virtue of the random assignment of Z_i, the average causal effect on compliers with $Z_i = 1$ is the same as the average causal effects for all compliers.
- In general, therefore, ATET differs from LATE because it is a weighted average of two effects: one on always-takers and one on compliers.

LATE and ATET

- An special case when LATE is ATET is when D_{0i} equals zero for everybody, i.e., there are no always-takers.
- If no one in the control group receives treatment, then by definition there can be no always-takers. Hence, all treated subjects must be compliers.
- The MDVE is (approximately) this sort of experiment. Since we have defined treatment as coddling, and (almost) no one in the group assigned to be arrested was coddled, there are (almost) no always-takers. LATE in this case is ATET, the effect of coddling on the population coddled

The language of 2SLS

- Applied economists typically discuss IV using the language of two-stage least squares (2SLS), a generalized IV estimator introduced by Theil (1953) in the context of simultaneous equation models.
- In models without covariates, the 2SLS estimator using a dummy instrument is the same as the Wald estimator.
- In models with exogenous covariates, 2SLS provides a simple and easily implemented generalization that allows for multiple instruments and multiple treatments.

- The setup is the same as before, we only add some covariates X_i
- if we suppose that is randomly assigned as intended D_i then, we can regress via OLS

$$Y_i = X_i'\beta + \alpha D_i + \epsilon_i \tag{13}$$

- in 2SLS language this is called structural equation
- Note that the causal effect here is the effect of being coddled on recidivism, relative to the baseline recidivism rate when arrested.

2SLS: first stage

- We can construct 2SLS estimates in two steps, each by OLS.
- In the first stage, the endogenous right-hand side variable (treatment delivered) is regressed on the exogenous covariates plus the instrument (or instruments).

$$D_i = X_i'\phi_0 + \phi_1 Z_i + \eta_i \tag{14}$$

- ϕ_1, is called the first-stage effect of the instrument.
- Note that the first-stage equation must include exactly the same exogenous covariates as appear in the structural equation.
- The size ϕ_1 is the major determinant of the statistical precision of IV estimates.
- if D_i is a dummy, ϕ_1 measures the proportion of the population that are compliers

2SLS: second stage

- In the second stage, fitted values from the first-stage are plugged directly into the structural equation in place of the endogenous regressor.
- Note, that although the term 2SLS arises from the fact that estimates can be constructed from two OLS regressions, we do not usually compute them this way. This is because the resulting standard errors are incorrect.
- Best practice therefore is to use a packaged 2SLS routine such as may be found in software like SAS or Stata

- In addition to the first-stage, an important auxiliary equation that is often discussed in the context of 2SLS is the *reduced form*.
- The reduced form for Y_i is the regression obtained by substituting the first-stage into the causal model,

$$Y_i = X_i'\beta + \alpha[X_i'\phi_0 + \phi_1 Z_i + \eta_i] + \epsilon_i = X_i'\gamma_0 + \gamma_1 Z_i + \psi_i \tag{15}$$

- The coefficient γ_1 is called reduced-form effect of the instrument and can be estimated by OLS

2SLS: second stage estimates

- Note that with a single endogenous variable and a single instrument, the effect of D_i in the causal model is the ratio of reduced-form to first-stage effects:

$$\alpha = \gamma_1 \; / \; \phi_1 \tag{16}$$

- 2SLS second-stage estimates can therefore be understood as a re-scaling of the reduced form.
- It can also be demonstrated that the significance levels for the reduced-form and the second-stage are asymptotically the same under the null hypothesis of no treatment effect.
- Hence, the workingmans' IV motto: **If you can not see your causal effect in the reduced form, it is not there**

First stage and reduced form

Table 2. First stage and reduced forms for Model 1.

Endogenous variable is coddled				
	First stage		*Reduced form (ITT)*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled-assigned	0.786 (0.043)	0.773 (0.043)	0.114 (0.047)	0.108 (0.041)
Weapon		−0.064 (0.045)		−0.004 (0.042)
Chem. influence		−0.088 (0.040)		0.052 (0.038)
Dep. var. mean	0.567 (Coddled–delivered)		0.178 (Re-arrested)	

The table reports OLS estimates of the first-stage and reduced form for Model 1 in the text.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

OLS and 2SLS

Table 3. OLS and 2SLS estimates for Model 1.

Endogenous variable is coddled				
	OLS		*IV/2SLS*	
	(1)	*(2)**	*(3)*	*(4)**
Coddled–delivered	0.087 (0.044)	0.070 (0.038)	0.145 (0.060)	0.140 (0.053)
Weapon		0.010 (0.043)		0.005 (0.043)
Chem. influence		0.057 (0.039)		0.064 (0.039)

The Table reports OLS and 2SLS estimates of the structural equation in Model 1.
*Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

2 endogenous and 2 instruments

Table 4. First stage and reduced forms for Model 2.

Two endogenous variables: Advise, separate						
	First stages					
	Advised		*Separated*		*Reduced form (ITT)*	
	(1)	*(2)**	*(3)*	*(4)**	*(5)*	*(6)**
Advise-assigned	0.778 (0.039)	0.766 (0.039)	0.035 (0.043)	0.035 (0.043)	0.097 (0.054)	0.088 (0.046)
Separate-assigned	0.044 (0.038)	0.031 (0.039)	0.717 (0.042)	0.715 (0.043)	0.130 (0.053)	0.127 (0.046)
Weapon		−0.038 (0.036)		−0.031 (0.039)		−0.001 (0.042)
Chem. influence		−0.068 (0.032)		−0.018 (0.035)		0.051 (0.038)
Dep. var. mean	0.283 (Adv.-deliver)		0.283 (Sep.-deliver)		0.178 (Re-arrested)	

The table reports OLS estimates of the first-stage and reduced form for Model 2 in the text.
*In addition to the covariates reported in the table, these models include year and quarter dummies, and dummies for non-white and mixed race.

Some final remarks

- IV methods are not limited to the estimation of the effects of binary on-or-off treatments like coddling, separation, or advice in the MDVE. Many experimental evaluations are concerned with the effects of interventions with variable treatment intensity

- The IV framework goes beyond randomized trials and can be used to exploit quasi-experimental variation in observational studies as in Angrist (1990), which uses the draft lottery numbers that were randomly assigned in the early 1970s as instrumental variables for the effect of Vietnam era veteran status on post-service earnings.

- **Instrumental variables need not be randomly assigned to be useful.**
- Angrist and Lavy (1999) constructed instrumental variables estimates of the effects of class size on test scores. The instrument in this case is the class size predicted using Maimonides rule, a mathematical formula derived from the practice in Israeli elementary schools of dividing grade cohorts by integer multiples of 40, the maximum class size

- A pioneering illustration of this point from criminology is Levitt (1997) study of the effects of extra policing using municipal election cycles to create instruments for numbers of police.

final remark

- there is much more to know about IV and LATE
- i.e. an instrument must be relevant, but how much ?
- and there is a lot that have to be added on the error term
- ... to be continued ...

References

- Angrist, J. D., Imbens, G. W. Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association 91(434)
- Angrist, Instrumental variables methods in experimental criminological research: what, why and how, Journal of Experimental Criminology (2006) 2: 23?44, Springer 2006
- Angrist, J. D. Krueger, A. B. (2001). Instrumental variables and the search for identification. Journal of Economic Perspectives 15(4).
- Angrist, J. D. Lavy, V. C. (1999). Using Maimonides rule to estimate the effect of class size on student achievement. Quarterly Journal of Economics 114(2).
- Berk, R. A. Sherman, L. W. (1988). Police response to family violence incidents: An analysis of an experimental design with incomplete randomization. Journal of the American Statistical Association 83(401).
- Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effects of police on crime. American Economic Review 87(3).

- `http://masteringmetrics.com`