Bristol, 28th June 2011

Causal inference with non compliance

Leonardo Grilli grilli@ds.unifi.it
www.ds.unifi.it/grilli

Department of Statistics
University of Florence

Experiments with non-compliance

2

- Randomisation implies that the treatment assignment Z is independent of all the pre-assignment variables (including the potential outcomes) thus unconfoundedness holds
- However, in case of non-compliance the actual treatment D is not always the same as the assigned treatment Z and it is a post-assignment variables: conditioning on D is harmful since it violates the unconfoundedness assumption

Types of non-compliance

3

- Compliance can be
  - **Binary (all-or-none)**: e.g. the patient assigned to a new therapy only has two options: to participate or to not participate
  - **Partial**: e.g. the patient assigned to a treatment entailing the take of one pill per day for 30 days may take 0 pills (she does not participate), or 30 pills (full compliance), or 0<k<30 pills (partial compliance)
- We focus on the following case:
  - Binary compliance (all-or-none)
  - Randomised binary treatment Z (randomised → ignorable)

Types of non-compliance /cont

4

- In the case of binary treatment and binary compliance (all-or-none) there are two instances of non-compliance:
  1. The patient **is** assigned to treatment but she **does not take** the treatment
  2. The patient **is not** assigned to treatment but she **does take** the treatment
- Case 2 is usually ruled out in clinical trials (the treatment is available outside the trial), while it is common in encouragement experiments (e.g. some individuals receive a letter recommending to take a flu vaccine)

Variables

5

- In time order $Z \rightarrow D \rightarrow Y$
- Z (treatment assignment)
- D (actual treatment)
  - Potential versions D(0) and D(1)
  - Observed treatment $D^{obs}=D(Z)$
- Y (outcome)
  - Potential outcomes Y(z,d), i.e. Y(0,0), Y(0,1), Y(1,0), Y(1,1)
  - Observed outcome $Y^{obs}=Y(Z, D(Z))$

We wish to estimate the causal effect of D on Y

Causal effects

6

- For a given individual in a given experiment, all the causal effects of Z on post-assignment variables are defined a priori in terms of potential outcomes
  - Z on D: D(1) – D(0)
  - Z on Y: Y(1, D(1)) – Y(0, D(0))
- Since Z is randomised both effects are estimable
  - The mean causal effect of Z on Y is known as **Intention-To-Treat** (***ITT***)

$$ITT = E\left[Y\left(1, D(1)\right) - Y\left(0, D(0)\right)\right]$$

Compliance behaviour

7

Z and D binary → 4 compliance states (types of reaction to treatment assignment)

D(1)	D(0)	D(1)- D(0)	Type	Prob.
1	0	1	Complier (C)	π_C
0	0	0	Never taker (NT)	π_{NT}
1	1	0	Always taker (AT)	π_{AT}
0	1	-1	Defier (D)	π_D

Compliance behaviour /cont

8

- The labelling of individuals as C, NT, AT and D on the basis of the reaction to treatment assignment is
  - Unobservable, i.e. the groups C, NT, AT and D are **latent groups**
  - Relative to the behaviour of individuals in the **actual experiment**, not in general (e.g. a Never Taker in an experiment may be a Complier in another experiment)
  - It is not affected by the actual assignment so it is a **pre-assignment variable** (this is a key property: since Z is randomised, the composition of the population in terms of compliance behaviour is the same at any level of Z)
- For AT and NT there is no information to estimate the causal effect of D on Y since Z is unable to change D

Latent and observable groups

9

General

Z \ D^{obs}	0	1
0	NT/C	AT/D
1	NT/D	AT/C

Monotonicity (→ no Defiers)

Z \ D^{obs}	0	1
0	NT/C	AT
1	NT	AT/C

Treatment available only to individuals assigned to treatment (→ no D, no AT)

Z \ D^{obs}	0	1
0	NT/C	.
1	NT	C

Membership to latent groups cannot be established (except for certain combinations of Z e D^{obs} depending on the design)

Nevertheless, if Z is randomised and the latent groups are no more than 3, we can estimate the probabilities of group membership

Latent and observable groups /cont

10

Hypothetical experiment with random assignment Pr(Z=1)=0.5 and monotonicity

		Percentages of groups			
	Assigned	*Control arm*		*Treatment arm*	
Population	*Taken*	*Control*	*Treatment*	*Control*	*Treatment*
50	*Compliers*	25			25
30	*Never Takers*	15		15	
20	*Always Takers*		10		10
100	Observed	40	10	15	35
		50		50	

The observed proportions allow us to recover the proportions of the 3 latent groups:

Prop(AT) = 10/50 = 0.2 Prop (NT) = 15/50 = 0.3

Prop(C) = 1-0.2-0.3 = 0.5

11

Complier Average Causal Effect

- Definition and assumptions
- Estimation
- Examples
- Related estimands
- Identifiability and sensitivity analysis
- Structural equations and Instrumental Variables

CACE (or LATE)

12

- Data contain information to estimate the causal effect on the compliers, known as
  - CACE = Complier Average Causal Effect
  - LATE = Local Average Treatment Effect

- Angrist J., Imbens G.W, Rubin D.B. (1996) Identification of Causal Effects using Instrumental Variables. *JASA* 91, 444-472.
- Imbens, G.W. and Rubin, D.B. (1997) Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* 25, 305–327, 1997.
- Mealli F., and D.B. Rubin (2002) Assumptions when Analyzing Randomized Experiments with Noncompliance and Missing Outcomes. *Health Services and Outcomes Research Methodology* 3, 225-232.

CACE and ITT /1

13

Assume **SUTVA**

Decomposition of ITT (Intention-To-Treat effect)

$$ITT = ITT_C \cdot \pi_C + ITT_{NT} \cdot \pi_{NT} + ITT_{AT} \cdot \pi_{AT} + ITT_D \cdot \pi_D$$

$$ITT_{group} = E\left[Y_i\left(1, D_i(1)\right) - Y_i\left(0, D_i(0)\right) \mid i \in group\right]$$

$$\pi_{group} = P\left[i \in group\right]$$

CACE and ITT /2

14

Assume **Exclusion restriction for NT and AT**:

$ITT_{NT} = ITT_{AT} = 0$

(it concerns the relationship between Z and Y)

$$ITT = ITT_C \cdot \pi_C + ITT_D \cdot \pi_D$$

CACE and ITT /3

15

Assume **Strong monotonicity**:

1) Non-null average effect of Z on D

$$E\left[D_i(1) - D_i(0)\right] = \pi_C - \pi_D \neq 0$$

2) Monotonicity (No defiers): $\pi_D = 0$

Remark: 1 & 2 imply $\pi_C > 0$

(it concerns the relationship between Z and D)

$$ITT = ITT_C \cdot \pi_C \quad \text{with } \pi_C > 0$$

CACE and ITT /4

16

$$ITT_C = \frac{ITT}{\pi_C} = \frac{E\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right]}{E\left[D_i(1) - D_i(0)\right]}$$

Remark: the ITT for compliers is the ratio between to causal effects: the effect of Z on Y (ITT) and the effect of Z on D

If treatment is random (or at least ignorable) each of the two causal effects is estimable as the difference between the means of the individuals assigned to treatment and those assigned to control

Interpretation of CACE

17

Assume **Exclusion restriction for compliers**:

$$Y_i\left(0, D_i(0)\right) = Y_i\left(1, D_i(0)\right)$$
$$Y_i\left(1, D_i(1)\right) = Y_i\left(0, D_i(1)\right)$$

With this assumption Y is no longer a function of two variables: Y only depends on the actual treatment D

$$ITT_C = E\left[Y_i\left(1, D_i(1)\right) - Y_i\left(0, D_i(0)\right) \mid i \in complier\right]$$
$$= E\left[Y_i(1) - Y_i(0) \mid i \in complier\right]$$

ITT for compliers = CACE = causal effect of D on Y for the unobservable group of compliers

Estimation of CACE

18

We showed that under the assumptions SUTVA, strong monotonicity and exclusion restrictions

$$CACE = \frac{ITT}{\pi_C} = \frac{E\left[Y_i(1) - Y_i(0)\right]}{E\left[D_i(1) - D_i(0)\right]} = \frac{E\left[Y_i(1)\right] - E\left[Y_i(0)\right]}{E\left[D_i(1)\right] - E\left[D_i(0)\right]}$$

Under random assignment, a consistent estimator of CACE is obtained by estimating the expected values by means of the corresponding sampling mean:

$$\widehat{CACE} = \frac{\bar{Y}^{obs}_{[Z=1]} - \bar{Y}^{obs}_{[Z=0]}}{\bar{D}^{obs}_{[Z=1]} - \bar{D}^{obs}_{[Z=0]}}$$

CACE and instrumental variables (IV)

19

- The CACE estimator is equal to the IV estimator
- Indeed, when the IV variable is binary the IV estimator of the effect of D on Y has the following expression (Wald)

$$\widehat{IV} = \frac{\widehat{\text{cov}}(Y_i^{obs}, Z_i^{obs})}{\widehat{\text{cov}}(D_i^{obs}, Z_i^{obs})} = \frac{\bar{Y}_{[Z=1]}^{obs} - \bar{Y}_{[Z=0]}^{obs}}{\bar{D}_{[Z=1]}^{obs} - \bar{D}_{[Z=0]}^{obs}} = \widehat{CACE}$$

- The IV estimator is defined with no reference to the paradigm of causal inference via potential outcomes
- However, under the assumptions of SUTVA, exclusion restriction and strong monotonicity, the IV estimator is estimating the CACE, a quantity with a valid casual interpretation

ITT vs CACE

20

- If Z is randomised we can consistently estimate
  - ITT (effect of Z on Y)
    - Whole population
    - Assumptions: SUTVA
  - CACE (effect of D on Y)
    - Complier (non-observable sub-population)
    - Assumptions: SUTVA, Exclusion restriction and Strong monotonicity

If we are interested in the effect of D on Y, then CACE is more relevant than ITT, even if it requires stronger assumptions and it refers to a non-observable sub-population

Remark: under the CACE assumptions it follows that

CACE = ITT/pr(complier) → |CACE| > |ITT|

thus the ITT estimate is a conservative estimate of CACE!

Remark: CACE is more apt to be generalised than ITT (since it does not depend on the proportion of compliers)

Example

Hypothetical experiment with random assignment Pr(Z=1)=0.5 and monotonicity

		Percentages of groups				Mean outcome			
	Assigned	Control arm		Treatment arm		Control arm		Treatment arm	
Population	Taken	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
50	Compliers	25		25		500		600	
30	Never Takers	15		15		480		480	
20	Always Takers		10		10		550		550
100	Observed	40	10	15	35	492.5	550.0	480.0	585.7
		50		50		504.0		554.0	

As-treated: $\overline{Y}^{obs}_{[D=1]} - \overline{Y}^{obs}_{[D=0]}$

Per-protocol: $\overline{Y}^{obs}_{[Z=1,D=1]} - \overline{Y}^{obs}_{[Z=0,D=0]}$

Intention-to-treat: $\overline{Y}^{obs}_{[Z=1]} - \overline{Y}^{obs}_{[Z=0]}$

Mean of treated	577.8
Mean of control	489.1
As-treated effect	88.7
Per-protocol effect	93.2
Intention-to-treat effect	50.0
Proportion of compliers	0.5
CACE	100.0

ITT is an underestimate of CACE (higher proportion of compliers → higher bias)

The *As-treated* and *Per-protocol* effects are over- or under-estimates of CACE depending on the mean outcome of Never Takers and Always Takers

Example (mean outcome AT: from 550 to 650)

Hypothetical experiment with random assignment Pr(Z=1)=0.5 and monotonicity

		Percentages of groups				Mean outcome			
	Assigned	Control arm		Treatment arm		Control arm		Treatment arm	
Population	Taken	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
50	Compliers	25		25		500		600	
30	Never Takers	15		15		480		480	
20	Always Takers		10		10		650		650
100	Observed	40	10	15	35	492.5	650.0	480.0	614.3
		50		50		524.0		574.0	

As-treated: $\overline{Y}^{obs}_{[D=1]} - \overline{Y}^{obs}_{[D=0]}$

Per-protocol: $\overline{Y}^{obs}_{[Z=1,D=1]} - \overline{Y}^{obs}_{[Z=0,D=0]}$

Intention-to-treat: $\overline{Y}^{obs}_{[Z=1]} - \overline{Y}^{obs}_{[Z=0]}$

Mean of treated	622.2
Mean of control	489.1
As-treated effect	133.1
Per-protocol effect	121.8
Intention-to-treat effect	50.0
Proportion of compliers	0.5
CACE	100.0

Example: lottery for draft – U.S. army /1

23

- AIR 1996: Angrist J., Imbens G.W, Rubin D.B., 1996. Identification of Causal Effects using Instrumental Variables. *JASA* 91, 444-472.
- In the years1970-1973, draft to U.S. army for Vietnam war was based on a lottery:
  - For any cohort, each date of birth was randomly assigned a number from 1 to 365
  - Males with a number lower than a certain threshold were drafted, e.g. males born in 1950 were called to be drafted until the number195
- Variables:
  - Z = 1 if the individual had a number lower than the threshold (i.e. he was called to be drafted)
  - D = 1 if the individual actually drafted
  - Y = 1 if the individual died in the period 1974-1983 (civilian mortality, i.e. for causes not directly related to war)

Example: lottery for draft – U.S. army /2

24

- The assumptions for a valid causal interpretation of the IV estimand are (AIR 1996 p 452):
  - **SUTVA**: The veteran status of any man at risk of being drafted in the lottery was not affected by the draft status of others at risk of being drafted, and, similarly, that the civilian mortality of any such man was not affected by the draft status of others
  - **Ignorable Assignment**: Assignment of draft status was random
  - **Exclusion restriction**: Civilian mortality risk was not affected by draft status once veteran status is taken into account
  - **Nonzero Average Causal Effect of Z on D**: Having a low lottery number increases the average probability of service
  - **Monotonicity assumption**: There is no one who would have served if given a high lottery, but not if given a low lottery number

Some men with low lottery numbers changed their educational plans so as to retain draft deferments and avoid the conscription. If so, then ***the exclusion restriction could be violated***, because draft status may have affected civilian outcomes through channels other than veteran status.

Estimates and comments from AIR 1996 p 453

Year	Draft eligibility[a]	Number of deaths[b]	Probability of death[d]	Probability of military service[e]
1950 *(men born in 1950)*	Yes	2,601	.0204 (.0004)	.3527 (.0325)
	No	2,169	.0195 (.0004)	.1934 (.0233)
Difference (Yes minus No)			.0009 (.0006)	.1593 (.0401)
IV estimates[f]			.0056 (.0040)	

Of the men with low lottery numbers (Z = 1), 35.3% actually served in the military. Of those who had high lottery numbers (Z = 1), only 19.3% served in the military. Random assignment of draft status suggests that draft status had a causal effect that increased the probability of serving by an estimated **15.9%** on average.

Similarly, of those with low lottery numbers, 2.04% died between 1974 and 1983, compared to 1.95% of those who had high lottery numbers. The difference of **.09%** can be interpreted as an estimate of the average causal effect of draft status on civilian mortality.

Assuming that these estimated causal effects are population averages, the ratio of these two causal effects of draft status is, under the Assumptions 1-5, the causal effect of military service on civilian mortality for the 15.9% who were induced by the draft to serve in the military. For this group, the average causal effect is **.56%**.

Example: lottery for draft – U.S. army /4

26

- These estimates highlight the fact that **the IV estimator does not require observations on individuals;** sample averages of outcomes and treatment indicators by values of the instruments are sufficient
- In applications like the one discussed here, **these moments are drawn from different data sets**

Example: Influenza Vaccine /1

27

- Hirano K., Imbens G.W., Rubin D.B., Zhou X. (2000) Assessing the effect of an Influenza Vaccine in an Encouragement Design, *Biostatistics* 1, 69-88.
- **Encouragement design**: a randomly selected group of physicians receives a letter encouraging them to inoculate patients at risk for flu.
- A standard ITT analysis shows that ***encouragement decreases hospitalization rates***: hospitalization rate is
  - 9.2% among patients whose doctors did not receive the encouraging letter
  - 7.8% among patients whose doctors received the encouragement
- Hirano et. al. performed a full Bayesian analysis maintaining SUVTA and monotonicity but **relaxing exclusion restrictions**

Example: Influenza Vaccine /2

Group	Estimated proportion (%)	Estimand	Estimated effect	Standard error
C	11.7	ITT_C	−0.037	0.078
NT	69.3	ITT_{NT}	0	0
AT	19.0	ITT_{AT}	−0.053	0.032
All	100	ITT	−0.014	0.008

Comments from Mealli & Rubin (2002)

- The analysis with $ITT_{NT} = 0$ (but $ITT_{AT} \neq 0$) suggests that encouragement has a similar beneficial effect on people who would have received the flu shot regardless of the assignment, the AT, as on the compliers, $ITT_C \approx ITT_{AT}$.
- There is thus little evidence in this experiment that the flu shot itself had beneficial effects, perhaps because *encouragement caused AT to get their flu shot earlier*, and the earlier flu shot for them was as beneficial as the flu shot itself for compliers.
- Incidentally, ITT_{AT} is more precisely estimated than ITT_C for two reasons: the AT group is larger than the C group, and there are some individuals in the AT group who are directly identifiable (i.e., those who were assigned to control), which is not true for any individuals in the C group.

Other estimands

29

- CACE, namely the causal effect on compliers, is not always the more interesting estimand, but it is **the only one estimable without further untestable assumptions**
- In may applications interesting estimands are
  - ATT $E\left[Y_i(1)-Y_i(0) \mid D_i^{obs}=1\right]$ Observable subset
  - ATE $E\left[Y_i(1)-Y_i(0)\right]$ Whole population

To estimate ATT or ATE we need further untestable assumptions

Identifiability of ATT (Heckman) /1

30

- In the discussion of AIR 1996, James Heckman criticizes the potential outcomes approach and the implied focus on CACE
- Heckman relies on an approach based on structural equation models ("switching regime models", widely used in econometrics) to show that the ATT (which is more interesting than CACE) can be easily estimated under general conditions, including

$$E\left[Y^1-Y^0 \mid \mathbf{Z}, D=1\right]=E\left[Y^1-Y^0 \mid D=1\right] \qquad \text{(A-2')}$$

- However, assumption (A-2') in terms of potential outcomes is equivalent to

$$E\left[Y_i(1)-Y_i(0) \mid Z_i, D_i(Z_i)=1\right]=E\left[Y_i(1)-Y_i(0) \mid D_i(Z_i)=1\right]$$

- AIR 1996, p 469: "Simple manipulation shows that Heckman's assumption (A-2') […] amounts to assuming that *the effect for always-takers is the same as that for compliers*. Given this assumption, Heckman claims that he can identify a more interesting parameter: the average effect for those who receive the treatment. But because those who receive the treatment are a mixture of always-takers and compliers, ***Heckman's assumptions simply assume the answer***"

Identifiability of ATT (Heckman) /2

31

AIR 1996, p 469, conclude as follows:

"We also view Heckman's assumption (A-2') as lacking in scientific (economic) content. Our assumptions restrict outcomes at the unit level given different assignments, so that-like Fisher (1918), Neyman (1923), Tinbergen (1930) and Haavelmo (1944) - we compare ***for a specific unit*** the outcomes that would be observed given different environments.

Thus our assumptions can be immediately interpreted as comparisons of outcomes in behavioral models of utility maximizing behavior given different sets of constraints. In contrast, Heckman's key assumption (A-2') compares ***average*** outcomes for ***different*** groups of individuals. He provides no examples where this assumption is plausible or can be related to the economic behavior of agents."

Identifiability of ATT & ATE (Robins)

32

Ass. I: $E\left[Y_i(1)-Y_i(0)\,|\,Z_i=1, D_i^{obs}=1\right]$ — **C** or **AT**

$= E\left[Y_i(1)-Y_i(0)\,|\,Z_i=0, D_i^{obs}=1\right]$ — **D** or **AT**

Ass. II: $E\left[Y_i(1)-Y_i(0)\,|\,Z_i=z, D_i^{obs}=1\right]$

$= E\left[Y_i(1)-Y_i(0)\,|\,Z_i=z, D_i^{obs}=0\right] \quad \forall z$

Robins (1989): under the assumptions of AIR (1996), except for monotonicity,

Ass. I $\Rightarrow$ IV estimand = ATT

Ass. I & Ass. II $\Rightarrow$ IV estimand = ATE

Bounds for ATE

33

Under the assumptions allowing the estimation of CACE

$$ATE = \pi_C \cdot E\left[Y_i(1) - Y_i(0) \mid i \in C\right]$$
$$+\pi_{AT} \cdot E\left[Y_i(1) - Y_i(0) \mid i \in AT\right]$$
$$+\pi_{NT} \cdot E\left[Y_i(1) - Y_i(0) \mid i \in NT\right]$$

$E[Y_i(0) \mid i \in AT]$ non-estimable

$E[Y_i(1) \mid i \in NT]$ non-estimable

- The probabilities π_C , π_{AT} , π_{NT} are estimable
- The causal effect is estimable for C, but not for AT and NT
- If we let the two non-estimable means to vary in their admissible sets we get non-parametric bounds for ATE

Sensitivity analysis /1

34

- Let us keep SUTVA and the assumption of a non-null mean effect of Z on D
- The ignorability of Z is not relevant for the definition of the estimands (it is relevant for estimation)
- Let us evaluate the consequences of relaxing (one at a time)
  - Exclusion restriction
    - For non-compliers
    - For compliers
  - Monotonicity

Details in AIR 1996 §6

Sensitivity analysis /2

Monotonicity yes

Exclusion restriction for non-compliers no

$$ITT = ITT_C \cdot \pi_C + ITT_{NT} \cdot \pi_{NT} + ITT_{AT} \cdot \pi_{AT}$$

$$\frac{ITT}{\pi_C} = ITT_C + ITT_{NT} \cdot \frac{\pi_{NT}}{\pi_C} + ITT_{AT} \cdot \frac{\pi_{AT}}{\pi_C}$$

$$= ITT_C + ITT_{AT\&NT} \cdot \frac{\pi_{AT} + \pi_{NT}}{\pi_C}$$

IV estimand

CACE

$$ITT_{AT\&NT} = \frac{ITT_{AT} \cdot \pi_{AT} + ITT_{NT} \cdot \pi_{NT}}{\pi_{AT} + \pi_{NT}}$$

Sensitivity analysis /3

36

Monotonicity yes

Exclusion restriction for compliers no

Assuming additive effects, for any $i \in C$

$$Y_i(1, D_i(1)) - Y_i(0, D_i(0)) = Y_i(1,1) - Y_i(0,0)$$
$$= \underbrace{\left[Y_i(1,1) - Y_i(0,1)\right]}_{H_i} + \underbrace{\left[Y_i(0,1) - Y_i(0,0)\right]}_{G_i}$$

H_i causal effect of Z on Y controlling for D

G_i causal effect of D on Y controlling for Z

$$ITT_C = E[Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \mid i \in C]$$
$$= E\left[H_i \mid i \in C\right] + E\left[G_i \mid i \in C\right]$$

Sensitivity analysis /4

37

Monotonicity no

Exclusion restriction (per everyone, even defiers) yes

$$ITT = ITT_C \cdot \pi_C + ITT_D \cdot \pi_D$$

$$\frac{ITT}{\pi_C - \pi_D} = ITT_C + \frac{\pi_D}{\pi_C - \pi_D}\left[ITT_C - (-ITT_D)\right]$$

IV estimand

CACE

Causal effect of D on Y for defiers

$$-ITT_D = -E[Y_i(1,0) - Y_i(0,1) \mid i \in D]$$
$$= E[Y_i(1) - Y_i(0) \mid i \in D]$$

Testability of the assumptions

38

- **Exclusion restriction**
  - For AT and NT: if the groups were identifiable, the exclusion restriction would be testable by comparing the mean of Y across levels of Z
  - For compliers: it concerns a comparison between unobservable results (a priori counterfactuals)
    $Y_i(0, D_i(0)) = Y_i(1, D_i(0))$ $\quad Y_i(1, D_i(1)) = Y_i(0, D_i(1))$
  - It is made plausible through tricks such as *placebo*, *blinding* and *double-blinding*
- **Monotonicity**: untestable, often plausible, certainly true in experiments where the individuals assigned to control do not have access to the treatment

Structural equations & IV estimation /1

39

Dummy endogenous variable model

β_1 represents the causal effect of D on Y

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$
$$D_i^* = \alpha_0 + \alpha_1 Z_i + \nu_i$$
$$D_i = I(D_i^* > 0)$$

In this version the effect is the same for any individual (but it can be generalized by letting β_1 to be random)

$$corr(\varepsilon_i, \nu_i) \neq 0 \quad \Rightarrow \quad corr(\varepsilon_i, D_i) \neq 0 \quad \Rightarrow \quad \hat{\beta}_1^{OLS} \quad \text{biased}$$

In econometric jargon, D is an **endogenous regressor**

In causal inference jargon, D is a **non-ignorable treatment**

Structural equations & IV estimation /2

40

The assumptions for the IV estimation of β_1 are

(1) $\text{cov}(Z_i, D_i) \neq 0$ (relevance)

(2) $\text{cov}(Z_i, \varepsilon_i) = 0$ (exclusion)

In words: Z has effect on D

Z has effect on Y only through D

If Z is randomly assigned then it also holds (even if it is not a necessary assumption) $\text{cov}(Z_i, \nu_i) = 0$

$$\hat{\beta}_1^{IV} = \frac{\widehat{\text{cov}}(Y_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)}$$

Under assumptions (1) and (2) the IV estimator is consistent for β_1

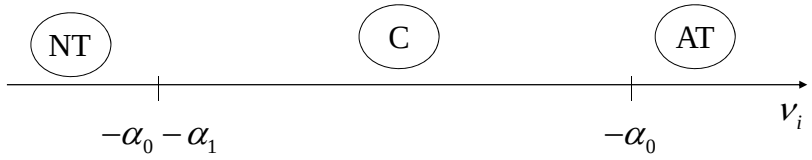
Structural equations & IV estimation /3

41

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$
$$D_i^* = \alpha_0 + \alpha_1 Z_i + \nu_i$$
$$D_i = I(D_i^* > 0)$$

$\alpha_1 > 0 \Rightarrow$ monotonicity (no defiers)

$$\{i : i \in AT\} = \{i : \nu_i > -\alpha_0\}$$
$$\{i : i \in NT\} = \{i : \nu_i \leq -\alpha_0 - \alpha_1\}$$



Structural equations & IV estimation /4

42

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$
$$D_i^* = \alpha_0 + \alpha_1 Z_i + \nu_i$$
$$D_i = I(D_i^* > 0)$$

The model assumes the same causal effect β_1 for all individuals

However, only for individuals i such that

$$\{i : -\alpha_0 - \alpha_1 < \nu_i \leq -\alpha_0\}$$

we have information to estimate β_1

For the other individuals the use of the model to estimate β_1 entails an **extrapolation**

Mealli F., and Pacini B. (2008) Comparing principal stratification and selection models in parametric causal inference with nonignorable missingness. *Computational Statistics and Data Analysis*, 53, 507-516.

Stating the assumptions /1

43

- For the aims of causal inference the potential outcomes approach is preferable over the structural equation approach (and implied IV estimator) for the way to formulate the assumptions
  - Separation between assumptions on the **phenomenon of interest** (SCIENCE) and assumptions on the **assignment mechanism** (OBSERVATION DESIGN)
  - Formulation at the level of **individual behaviour** (e.g. interpretation in terms of behavioural models for economic agents based on utility maximization)

Stating the assumptions /2

44

Angrist, Imbens & Rubin (1996) p 446:

… [in the structural equation approach] critical assumptions are cast in terms of disturbances from incompletely specified regression functions (i.e., ε_i and ν_i), rather than in terms of intrinsically meaningful and potentially observable variables.

Typically the researcher does not have a firm idea what these disturbances really represent, and therefore it is difficult to draw realistic conclusions or communicate results based on their properties.

Stating the assumptions /3

45

Mealli & Rubin (2002) p 232:

… there are no universally good assumptions when faced with noncompliance or missing outcomes.

The most plausible assumptions are specific to each context. But *explicit restrictions* can be made to be *scientifically more plausible* than the implicit assumptions underlying standard ITT approaches to noncompliance, or ad hoc methods for imputing missing data, or stylized IV structural models with scientifically remote assumptions.

In particular, we feel that the practice of specifying assumptions through functional and distributional restrictions on error terms should be eschewed.

Further references

46

- Little, R., Long, Q., & Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in **randomized clinical trials** subject to noncompliance. Biometrics, 65, 640–649.
- Schochet P.Z. & Chiang H.S. (2011) Estimation and Identification of the Complier Average Causal Effect Parameter in **Education RCTs**. JEBS 36, pp. 307–345.

47

Latent class models for CACE

- Principles
- JOBS II trial

Latent class models for CACE /1

48

- In case of covariates the estimation of CACE can be carried out through a latent class model (as far as the compliance states are discrete):
  - Skrondal A. & Rabe-Hesketh S. (2004) Generalized Latent Variable Modeling. CRC.
  - Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. JEBS, 27, 385–409.
  - Jo B., Asparouhov T. & Muthen B. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. Statistics in Medicine, 27, 5565–5577.
  - Grilli L. (2011) Causal inference through principal stratification: a special type of latent class modelling. In Fichet B, Piccolo D, Verde R, Vichi M (Eds) Classification and Multivariate Analysis for Complex Data Structures. 265-270.
- Software:
  - Mplus (www.statmodel.com)
  - Latent Gold (www.statisticalinnovations.com/products/choice.html)
  - gllamm (www.gllamm.org)

Latent class models for CACE /2

49

- The model has two components
  1. **Model for the compliance status**: a multinomial response model (binary in case of two states)
     - Subjects with **unobserved** compliance status → latent class
     - Subjects with **observed** compliance status (e.g. when drug is available only to subjects assigned to treatment) → observed variable
  2. **Model for the outcome conditional on compliance status** (if observed → covariate, otherwise latent class)

JOBS II trial: data

50

- Little, R. J. and Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model, Psychol. Methods 3, pp. 147–159.
- We follow the re-analysis of Skrondal & Rabe-Hesketh (2004)
- JOBS II is an intervention trial: unemployed individuals who had lost their jobs within the last 13 weeks and were looking for a job were randomized to receive
  - either five half-day sessions of job training plus a booklet briefly describing search methods and tips (**treatment group**) or
  - just the booklet (**control group**)
- One of the aims is to prevent poor mental health → the outcome is the **change in depression score**
- Dataset: 'high risk' group of 502 subjects
  - 167 randomized to control
  - 335 randomized to job training, but only 183 (55%) participated

JOBS II trial: structure

51

- Individuals in the control group did not have access to the treatment
  - There are no Defiers and no Always Takers
  - Individuals assigned to treatment ***and*** taking treatment are certainly Compliers
  - Individuals assigned to treatment ***and not*** taking treatment are certainly Never Takers
  - Individuals assigned to control are a mixture of Compliers and Never Takers
- CACE = mean outcome of compliers in the treatment group - mean outcome of compliers in the control group

$$\delta_C = \mu_C(1) - \mu_C(0)$$

- Exclusion restriction:

$$\mu_{NT}(1) = \mu_{NT}(0)$$

D^{obs} / Z	0	1
0	NT/C	.
1	NT	C

JOBS II trial: compliance model

52

- Treatment variable: z_i (= 1 if assigned to job training)
- The compliance status is observed in the treatment group (z_i = 1)

$$c_i = 1 \quad \Leftrightarrow \quad \text{individual } i \text{ is a complier}$$

- The compliance status is not observed in the control group (z_i = 0) → two latent classes

$$\eta_i = 1 \quad \Leftrightarrow \quad \text{individual } i \text{ is a complier}$$

- Compliance model (a logit model for the probability of being a complier):

$$\text{logit}\left[\Pr(\eta_i = 1 \mid z_i = 0)\right] = \mathbf{w}_i^T \boldsymbol{\rho} = \text{logit}\left[\Pr(c_i = 1 \mid z_i = 1)\right]$$

JOBS II trial: outcome model

53

- Outcome model (a normal linear model for the change in depression score)
  - Version without covariates

$$y_i = \beta_0 + \beta_1 \eta_i (1 - z_i) + \beta_2 c_i z_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \theta)$$

Mean for Never Takers $\mu_{NT}(0) = \mu_{NT}(1) = \beta_0$

Mean for Compliers under control $\mu_C(0) = \beta_0 + \beta_1$

Mean for Compliers under treatment $\mu_C(1) = \beta_0 + \beta_2$

CACE $\delta_C = \mu_C(1) - \mu_C(0) = \beta_2 - \beta_1$

  - Version **with covariates** (with constant effects across groups):

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\alpha} + \beta_1 \eta_i (1 - z_i) + \beta_2 c_i z_i + \varepsilon_i$$

JOBS II trial: covariates

54

- **Covariates for the compliance model**
  - [Age] age in years
  - [Motivate] motivation to attend
  - [Educ] school grade completed
  - [Assert] assertiveness
  - [Single] dummy for being single
  - [Econ] economic hardship
  - [Nonwhite] dummy variable for not being white versus white
- **Covariates for the outcome model**
  - [Basedep] baseline depression score
  - [Risk] baseline risk score (index based on depression, financial strain and assertiveness)

JOBS II trial: ML estimates

Parameter	No covariates Est	SE	Covariates Est	SE
Compliance model				
ϱ_0 [Cons]	−0.19	(0.11)	−8.74	(1.58)
ϱ_1 [Age]			0.08	(0.01)
ϱ_2 [Motivate]			0.67	(0.16)
ϱ_3 [Educ]			0.30	(0.07)
ϱ_4 [Assert]			−0.38	(0.15)
ϱ_5 [Single]			0.54	(0.28)
ϱ_6 [Econ]			−0.16	(0.16)
ϱ_6 [Nonwhite]			−0.50	(0.31)
Depression model				
β_0 [Cons]	−0.39	(0.07)	1.63	(0.28)
β_1	0.02	(0.17)	0.18	(0.13)
β_2	−0.12	(0.09)	−0.13	(0.08)
$\delta_c = \beta_2 - \beta_1$	−0.14	(0.14)	−0.31	(0.12)
α_1 [Basedep]			−1.46	(0.18)
α_1 [Risk]			0.91	(0.26)
θ	0.60	(0.04)	0.51	(0.03)
Log likelihood	−815.15		−729.41	

Skrondal & Rabe-Hesketh (2004) p 433

Stata gllamm

WITHOUT COVARIATES: depression goes down for subjects in the control arm (–0.39 for Never Takers and –0.37 for Compliers)

CACE = –0.14 (0.14) → job training seems to further reduce depression but it is not significant

WITH COVARIATES:

CACE = –0.31 (0.12) is now stronger and significant

Compliance status is affected by age (+), motivation (+), education (+), assertiveness (-)