

Randomized Experiments with noncompliance: Instrumental Variable Approach

instrumental
variable
approach

Complications in Randomized Experiments

- Randomized experiments are the gold standard in many fields.
- But they are often not feasible in practice due to practical or/and ethical reasons.
- Even whether they are feasible, complications can arise, break the initial randomization.
- Complications: 并发症
 - (1) Noncompliance to assigned treatment
 - (2) Loss to follow-up

Noncompliance

- Noncompliance: subjects take treatment different from the assigned one.
- Noncompliance can arise because, e.g., side effects, perception of the effect of the treatment (believe the effect is more/less effective)
- Noncompliance behavior is self-selected – breaks the initial randomization.

Notations

- Random *assignment*: Z_i (0 assigned to control, 1 assigned to treatment)
- Actual treatment *received*: W_i (0 received control, 1 received treatment)
- Outcome: Y_i
- Covariates (not essential here): X_i
- Noncompliance occurs when $Z \neq W$.
- One-sided compliance: the control group is restricted access to treatment, so that noncompliance is only on the treatment group

Example: Sommer and Zeger (1991)

- Goal: Study the effect of vitamin A supplements on infant mortality in Indonesia.
- The vitamin supplements were randomly assigned to villages, but some of the individuals in villages assigned to the treatment group failed to receive them.
- None of the individuals assigned to the control group received the supplements.
- So noncompliance is one-sided.
- Outcome Y : binary, survival of infant.
- Z , W are also binary.

Two Naïve Approaches

1. Per-protocol: discarding non-complying units ($Z \neq W$).
 2. As-treated: ignoring the initial random assignment, comparing units per their actual treatment status (W)
- Both approaches are invalid. Why?
 - Per-protocol: compliance is self-selected, the remaining subsample is not representative of the whole study population.
 - As-treated: randomization is broken.

Intention-to-treat (ITT) Approach

- Intention-to-treat (ITT): ignore the compliance status W completely. Get the difference in the outcome between the assigned trt and con groups. ITT
- Rationale:
 - (1) Preserve the randomization
 - (2) Estimate the “effectiveness”
- Drawback: not estimate efficacy
- ITT gives a valid estimate of the effect of the assignment, but not the actual treatment.

Effectiveness and Efficacy

功效 效力

- Effectiveness: the effect of a treatment work in practice
- Efficacy: the effect of a treatment in ideal situations
- Example: In the clinical development of a vaccine, an efficacy study asks the question, "Does the vaccine work?" In contrast, an effectiveness study asks the question "Does vaccination help people?".
- Effectiveness is more of policy interest (population level); efficacy is more of clinical or scientific interest (individual level).
- Randomized experiments are usually designed to study efficacy, but noncompliance and other complications render this difficult.

Instrumental Variable Approach

- Angrist, Imbens, and Rubin (1996) proposed an instrumental variable approach to non-compliance.
- Potential outcomes: $Y(z)$, $z=0,1$
- W is post-assignment, also has two potential outcomes: $W(z)$, $z=0,1$.
- Observed outcomes: $Y_i=Y_i(Z_i)$, $W_i=W_i(Z_i)$.

ITT estimands

- The ITT effect of assignment on outcome (Y):

$$ITT_Y = E(Y(1) - Y(0))$$

This ITT effect is not the effect of the treatment, but only assignment.

- The ITT effect of assignment on treatment received (W):

$$ITT_W = E(W(1) - W(0))$$

Compliance Type

- The central idea is to divide units into latent subgroups based on their compliance behavior.
- Defining compliance type: $S=(W(0), W(1))$
- Four compliance type:

		$W_i(0)$	
		0	1
$W_i(1)$	0	never-taker	defier
	1	<u>complier</u>	always-taker

control group

treatment group

- Compliance type is not fully observed in all units.
- Under one-sided compliance ($W_i(0) = 1$), only compliers and never-takers.



Compliers and Never-takers

- Under one-sided compliance, three possible combinations of observed (Z , W):
 - $(1,0)$: assigned trt, receive con – never-takers
 - $(1,1)$: assigned trt, receive trt – compliers
 - $(0,0)$: assigned con, receive con – can be either never-takers or compliers!
- The control group ($Z=1$) is a mixture of never-takers and compliers.

Causal Estimands

- Never-takers: units who would not receive the treatment regardless of the assignment.
- Compliers: units who would receive the treatment that they are assigned to.
- Compliers are the only units that we can obtain information about the effect of the treatment.
- **Estimand: Complier Average Causal Effect (CACE).**
$$\text{CACE} = E(Y(1) - Y(0) | \text{compliers})$$
- Similarly, we can define Never-taker Average Causal Effect: $\text{NACE} = E(Y(1) - Y(0) | \text{never-takers})$
- CACE is the estimand of interest.

Estimating CACE (and NACE)

- CACE and NACE are causal effects defined on a subgroup – “local” effects.
- The ITT effect on the outcome is a weighted average of the CACE and NACE.

$$ITT_Y = CACE * Pr(\text{compliers}) + NACE * Pr(\text{never-takers})$$

- But we cannot observe directly who are the compliers or never-takers. Need to disentangle the effect.
- Need some assumptions

Assumptions

- Monotonicity: $W_i(0) = 0$ - this is true by design.
- Randomized assignment:
 $Y(0), Y(1), W(0), W(1) \perp Z$
- Exclusion Restriction (ER) for never-takers:
 $NACE=0$
- ER assumes that there is no effect of assignment for never-takers. Is this always plausible?

Estimate CACE

- Under ER: $ITT_Y = CACE * Pr(\text{compliers})$
- That is: $CACE = ITT_Y / Pr(\text{compliers})$
- ITT_Y can be estimated from comparing the difference in the average outcome between $Z=0$ and $Z=1$ group.

$$ITT_Y = \sum_i (Y_i Z_i) / \sum_i (Z_i) - \sum_i (Y_i (1 - Z_i)) / \sum_i (1 - Z_i)$$

- How to estimate the proportion of compliers?
- In the $Z=1$ group, compliers are known – those who receive the treatment.
- $Pr(\text{compliers}) = \sum_i (W_i Z_i) / \sum_i (Z_i)$

Estimate CACE

- Remember: $ITT_w = E(W(1) - W(0)) = E(W(1))$
- The quantity $\text{sum}(W_i Z_i) / \text{sum}(Z_i)$ is in fact an estimate of ITT_w
- $CACE = ITT_y / ITT_w$
- That is, CACE estimate is a ratio of the estimate of the ITT effect on Y and the estimate of the ITT effect on W.

Example: Sommer and Zeger (1991)

- Total sample size: $N = 23,682$.
- Observed data

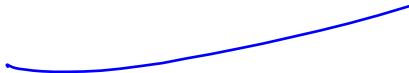
Table 23.1: SOMMER–ZEGGER VITAMIN SUPPLEMENT DATA

Compliance Type	Assignment $Z_{\text{obs},i}$	Vitamin Supplements $W_{\text{obs},i}$	Survival $Y_{\text{obs},i}$	Number of Units (Total 23,682)
co or nc	0	0	0	74
co or nc	0	0	1	11,514
nc	1	0	0	34
nc	1	0	1	2385
co	1	1	0	12
co	1	1	1	9663

Example: Sommer and Zeger (1991)

- Estimate the CACE using the IV approach.
- $ITT_Y = (34+12)/(34+12+2385+9663) - 74/(74+11514) = -0.00258$
- $ITT_w = (12+9663)/(34+2385+12+9663) = 0.79998$
- $CACE = ITT_Y / ITT_w = -0.00258 / 0.79998 = -0.00323$
- CACE: reduction in infant death of 0.323%
- $CACE > ITT_Y$
- How to calculate the confidence intervals?

Two-sided noncompliance: Randomized Encouragement Design

- Another example of randomized experiment: randomized encouragement design.
 - Z : randomized encouragement to take a treatment/intervention.
 - W : actual receipt of the treatment.
 - Y : outcome
 - Units who got encouraged are more likely to take the treatment, but still some of them might not take it.
 - The encouragement itself does not directly affect the outcome, it affects the outcome only through its effect on the taking the treatment.
- 

Randomized Encouragement Design

- An example of randomized encouragement design: flu vaccine encouragement.
- Doctor received mails that remind them to encourage patients to take vaccine. Which doctors the mails (Z) are sent to are randomized.
- But not all patients who got encouraged take flu vaccine (W), and some patients who did get encouraged still get the vaccine. The noncompliance is actually **two-sided** (will discuss next).
- Outcome (Y) is the hospitalization of flu-related disease.

Two-sided Noncompliance

- In many randomized experiments, noncompliance is two-sided. That is, the control group has access to treatment, and vice versa.
- In the flu vaccine encouragement example, noncompliance is two sided.
- Need to extend the previous discussion to incorporate the two-sided noncompliance.

Example: Hirano, Imbens, Rubin, Zhou (2001, Biostatistics)

Table 1: SUMMARY STATISTICS, FLU DATA (SAMPLE SIZE 2893)

	Grand Mean	Means			Means		
		no letter $Z_i^{\text{obs}} = 0$	letter $Z_i^{\text{obs}} = 1$	t-stat.	no flu shot $D_i^{\text{obs}} = 0$	flu shot $D_i^{\text{obs}} = 1$	t-stat.
Letter (Z_i^{obs})	0.514	0	1	—	0.475	0.631	-7.5
Flu Shot (D_i^{obs})	0.250	0.190	0.307	-7.3	0	1	—
Hospitalization (Y_i^{obs})	0.085	0.092	0.078	1.4	0.085	0.084	0.1
Age (X_{i1}^{obs})	65.2	65.0	65.4	-0.8	64.7	66.8	-4.1
COPD (X_{i2}^{obs})	0.283	0.290	0.277	0.8	0.264	0.343	-4.0

Compliance Type

- Defining compliance type: $S=(W(0), W(1))$
- Four compliance type:

		$W_i(0)$	
		0	1
$W_i(1)$	0	never-taker	defier
	1	complier	always-taker

- With two-sided noncompliance, all four types are possible.

Compliance Type

- Never-takers (0,0): units who would not receive the treatment regardless of the assignment.
- Compliers (0,1): units who would receive the treatment that they are assigned to.
- Always-takers (1,1): units who would receive the treatment regardless of the assignment.
- Defiers (1,0): units who receive the opposite treatment he/she is assigned.

Causal Estimand

- As in one-sided noncompliance, we can define the “local” average treatment effect for each compliance type:
- C(complier)ACE, N(never taker)CAE, D(defier)ACE, A(always)ACE
- Similarly, the ITT effect on outcome (Y):

$$ITT_Y = E(Y(1) - Y(0))$$

- Weighted average of ITT

$$ITT_Y = CACE * Pr(C) + NACE * Pr(N) + AACE * Pr(A) + DACE * Pr(D)$$

Assumptions for Estimation

- Monotonicity: $W_i(1) \geq W_i(0)$. Essentially assuming no defiers.

- Randomized assignment:

$$Y(0), Y(1), W(0), W(1) \perp Z$$

- Exclusion Restriction (ER) for never-takers and always-takers: $NACE=0$, $AACE=0$
- Under ER: $ITT_Y = CACE * Pr(\text{compliers})$
- That is: $CACE = ITT_Y / Pr(\text{compliers})$

Estimate the proportion of compliers

Due to monotonicity:

- In the observed $Z=0$ group, the units who received treatment ($W=1$) must be always-takers.

$$Pr(a) = \sum_i W_i (1 - Z_i) / \sum_i (1 - Z_i)$$

- In the observed $Z=1$ group, the units who did not receive treatment ($W=0$) must be compliers.

$$Pr(n) = \sum_i (1 - W_i) Z_i / \sum_i Z_i$$

Due to randomization:

- The proportions of compliers, always-takers, never-takers are the same between $Z=0$ and $Z=1$ group.

Therefore, the proportion of compliers can be estimated

$$Pr(c) = 1 - Pr(a) - Pr(n)$$

Estimate the CACE

- Remember ITT_Y can be estimated from the difference in the outcome Y between $Z=0$ and 1 group.

$$ITT_Y = \sum_i (Y_i Z_i) / \sum_i (Z_i) - \sum_i (Y_i (1 - Z_i)) / \sum_i (1 - Z_i)$$

- Then under all the previous assumptions, the CACE can be estimated from $ITT_Y / Pr(c)$, where $Pr(c)$ is estimated using the method in the previous page.

How to calculate variance?

- We have obtained *point estimate* of the CACE.
- In practice you usually also want to have an interval estimate (e.g., 95%CI or standard error) of the CACE.
- Same for the estimate of ATE.
- Direct large sample calculation (see the formulas in the book) or bootstrap.

Instrumental Variables

- What is an instrumental variable?
- A variable that has a causal effect on the treatment, W , but (is assumed to) have no “direct” causal effect on the outcome of interest Y , with any effect on Y “channeling through” an effect of the instrument on the treatment.
- In the case of randomized experiment with noncompliance, the instrument is the assignment Z : Z does not directly affect Y , but strongly affect the treatment W , which in turn affects Y .

IV: other examples

- Most IV used in economics are not in the context of noncompliance to randomized experiment.
- Instead, IV is often viewed as a natural experiment.
- Example 1: Study the effect of education on income. Clearly the relationship between the year of education and income can be highly confounded by factors like family, social-economics background, etc.
- In econ terminology, education is endogenous, not exogenous – it is correlated with the error term of a regression model of income.

IV example 1: quarter of birth

- An instrumental variable commonly used here is: half/quarter of the year of birth.
- When one was born in the year (Jan or Dec) is largely randomized, decided by nature. It clearly does not affect your later income directly.
- However, it directly affects when you goes to school first – it can create one year of difference in the year of school entrance.
- Further, due to the compulsory education requirement, it can create one year difference in education, which in turn affects income.

IV example 2: tobacco tax

- Goal: Study the effect of smoking (W) on health (Y).
- Impossible to do randomized experiment.
- Instrumental variable: tobacco tax.
- Reasoning: tobacco tax rate (Z) is controlled by government, it does not directly affect health. But it affects the price of tobacco, thus in turn affects how much one smokes, which affects one's health.

IV: Two-stage Least Square

- In standard econometrics literature (no causal inference), inference of IV is usually based on a two-stage least square (2SLS) regression.

$$\text{Model 1: } Y = a_0 + a_1 W + u$$

$$\text{Model 2: } W = b_0 + b_1 Z + v$$

- Here u , v are error terms. In Model 1, u is correlated with W , that is, W is endogenous. OLS estimate of a_1 is a biased estimate of the effect of W on Y .
- Instrument Z is uncorrelated with u , but correlated with W . The IV 2SLS estimate of a_1 is a ratio:

$$a_{1,iv} = \text{cov}(Y, W) / \text{cov}(W, Z)$$

- Remember the IV estimator for the CACE in the noncompliance case, it has this form of ratio.