Chapter 6. Irregular assignment and principal stratification

Fan Li
Department of Statistical Science, Duke University

November, 2011

Irregular assignment mechanism

- There is no general approach for all irregular assignment mechanisms, there are a number of interesting cases
- We discuss noncompliance in randomized experiments, and by extension, instrumental variables
- In these designs the assignment mechanism is assumed to be *latent regular*, that is, regular given some covariates that are not observed
- To conduct inference in such settings it is often necessary to invoke additional conditions

- Irregular assignment mechanisms in observational studies have an experimental counterpart in randomized experiments with noncompliance
- The analysis of randomized experiments with post-treatment complications plays the role of gold standard also for the analysis of observational studies, by suggesting the assumptions required to identify and estimate causal effects
- The standard analysis for randomized studies with noncompliance is called *Intention to Treat* (ITT), which ignores observed compliance information and compares those assigned to treatment to those assigned to control
- This procedure gives a valid estimate of the effect of treatment assignment on outcome

Compliance Types

- It is useful for our approach to think about the compliance behavior of the different units

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

- The true compliance status is not observed on all units
- Suppose defiers do not exist:

TA	TR	TCS
0	0	? [C,NT]
0	1	AT
1	0	NT
1	1	? [C,AT]

- Compliers (for whom we can identify effects) are not necessarily the subpopulations that are ex ante the most interesting subpopulations, but need extrapolation for others
- The set up here allows the researcher to sharply separate the extrapolation to the (sub-)population of interest from the information contained in the data

- Alternative approaches to estimate the effect of treatment received rather than treatment assignment:
  - *As-treated analysis* compares those who received treatment with those who received control, ignoring treatment assignment
  - *Per protocol analysis* compares people who were assigned to and received treatment with those who were assigned to and received control, i.e., compares those who appeared to comply with the protocol

- The global ITT may be written as the weighted average of the ITT effects across the four subpopulations:

$$ITT = \pi_c ITT_c + \pi_n ITT_n + \pi_a ITT_a + \pi_d ITT_d$$

  where ITT_j is the effect of the treatment assignment on units of type j and π_j is the proportion of units of type j $(j = c, n, a, d)$

- Let Z be the assignment variable

- We postulate the existence of four potential outcomes, $Y_i(z, w)$, corresponding to the outcome that would be observed if the assignment was $Z_i = z$ and the treatment received was $W_i = w$

- **Random Assignment**

$$Z_i \perp (Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), W_i(0), W_i(1))$$

- **Exclusion Restriction**

$$Y_i(z,w) = Y_i(z',w), \quad \text{for all } z, z', w$$

Note that the first of these two assumptions is implied by random assignment of Z_i, but the second is substantive, and randomization has no bearing on it

- **Monotonicity/No-Defiers**

$$W_i(1) \geq W_i(0)$$

This assumption makes sense in a lot of applications. It is implied directly by many (constant coefficient) latent index models of the type: $W_i(z) = 1\{\pi_0 + \pi_1 \cdot z + \epsilon_i > 0\}$ but it is much weaker than that

Distribution of Compliance Types

- Under random assignment and monotonicity we can estimate the distribution of compliance types:

$$\pi_a = Pr(W_i(0) = W_i(1) = 1) = E[W_i|Z_i = 0]$$

$$\pi_c = Pr(W_i(0) = 0, W_i(1) = 1) = E[W_i|Z_i = 1] - E[W_i|Z_i = 0]$$

$$\pi_n = Pr(W_i(0) = W_i(1) = 0) = 1 - E[W_i|Z_i = 1]$$

- The exclusion restriction implies that $ITT_n = ITT_a = 0$; because for never-takers and always takers the assignment does not affect the receipt of the treatment
- The monotonicity of compliance rules out the existence of defiers, $\pi_d = 0$
- These two assumptions allow the identification of the ITT effect for compliers

$$ITT_c = ITT/\pi_c$$

- The global ITT may be viewed as a conservative estimate of the treatment effect: with the implicit assumptions that $\pi_d = 0$ and that both ITT_n and ITT_a are strictly less than ITT_c, it should be expected that $ITT < ITT_c$

- Now consider average outcomes by instrument and treatment:

$$E[Y_i|W_i = 0, Z_i = 0] =$$

$$\frac{\pi_c}{\pi_c + \pi_n} \cdot E[Y_i(0)|\text{complier}] + \frac{\pi_n}{\pi_c + \pi_n} \cdot E[Y_i(0)|\text{never} - \text{taker}]$$

$$E[Y_i|W_i = 0, Z_i = 1] = E[Y_i(0)|\text{never} - \text{taker}]$$

$$E[Y_i|W_i = 1, Z_i = 0] = E[Y_i(1)|\text{always} - \text{taker}]$$

$$E[Y_i|W_i = 1, Z_i = 1] =$$

$$\frac{\pi_c}{\pi_c + \pi_a} \cdot E[Y_i(1)|\text{complier}] + \frac{\pi_a}{\pi_c + \pi_a} \cdot E[Y_i(1)|\text{always} - \text{taker}]$$

- From this we can infer the average outcome for compliers, $E[Y_i(0)|\text{complier}]$, and $E[Y_i(1)|\text{complier}]$

Complier Average Treatment Effect

- Hence the instrumental variables estimand, the ratio of these two reduced form estimands, is equal to the complier average treatment effect (local average treatment effect in Imbens and Angrist, 1994)

$$\beta^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[W_i|Zi = 1] - E[W_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|\text{complier}]$$

Extrapolating to the Full Population

- We can estimate $E[Y_i(0)|\text{never} - \text{taker}]$, and $E[Y_i(1)|\text{always} - \text{taker}]$
- We can learn from these averages whether there is any evidence of heterogeneity in outcomes by compliance status, by comparing the pair of average outcomes of $Y_i(0)$:

$$E[Y_i(0)|\text{never} - \text{taker}] \text{ and } E[Y_i(0)|\text{complier}]$$

  and the pair of average outcomes of $Y_i(1)$:

$$E[Y_i(1)|\text{always} - \text{taker}] \text{ and } E[Y_i(1)|\text{complier}]$$

- If compliers, never-takers and always-takers are found to be substantially different in levels, then it appears much less plausible that the average effect for compliers is indicative of average effects for other compliance types

Principal Stratification

Many scientific problems require that treatment comparisons be adjusted for post-treatment variables

- Treatment noncompliance
- Missing outcomes (dropout)
- Censoring (or truncation) by death
- Surrogate or biomarker endpoints
- Understanding the causal pathways by which a treatment affects an outcome: Direct & Indirect Effects

"Endogenous" selection problems

Throughout we assume that the post-treatment variable is binary:

$$S_i \in \mathcal{S} \equiv \{0, 1\}$$

$$\Downarrow$$

Subjects can be classified into four groups according to the joint potential values of the intermediate potential variable, $(S_i(0), S_i(1))$:

$$\begin{aligned} 00 &= \{i : S_i(0) = 0, S_i(1) = 0\} \\ 10 &= \{i : S_i(0) = 1, S_i(1) = 0\} \\ 01 &= \{i : S_i(0) = 0, S_i(1) = 1\} \\ 11 &= \{i : S_i(0) = 1, S_i(1) = 1\} \end{aligned}$$

This cross-classification of units is the (basic) principal stratification with respect to the (binary) post-treatment variable S. Formally,

Principal Stratification and Principal Strata Causal Effect

(Frangakis and Rubin, 2002)

The ***basic principal stratification*** P_0 with respect to post-treatment variable S is the partition of units $i = 1, \ldots, n$ such that, all units within any set of P_0, all units have the same vector $(S_i(0), S_i(1))$.

A ***principal stratification*** P with respect to post-treatment variable S is a partition of the units whose sets are unions of sets in the basic principal stratification P_0.

Let P be a principal stratification with respect to the post-treatment variable S. Then a ***principal effect*** with respect to that principal stratification is defined as a comparison of potential outcomes under standard versus new treatment within a principal stratum g in P, that is, a comparison between the ordered sets

$$\Big\{Y_i(0) : i \in g\Big\} \qquad \text{and} \qquad \Big\{Y_i(1) : i \in g\Big\}$$

The principal stratum G_i, $G_i \in \{00, 10, 01, 11\}$, to which unit i belongs, is not affected by treatment assignment for any principal stratification P.

- The value of the ordered pair $\big(S_i(0), S_i(1)\big)$ is not affected by treatment assignment
- Principal stratum membership only reflects subject's characteristics: it can be regarded as a pre-treatment variable

By ignorability of treatment assignment, the principal stratum membership G_i is guaranteed to have the same distribution in both treatment arms (within cells defined by pre-treatment variables)

Principal Strata Causal Effects

- Principal effects are properly defined causal effects because they are defined as comparison of potential outcomes $Y_i(0)$ and $Y_i(1)$ on a common set of units: the (union of) principal strata
- Ignorability of treatment assignment implies that

$$\Big(Y_i(0), Y_i(1)\Big) \perp Z_i | S_i(0), S_i(1), X.$$

$$\Downarrow$$

Treated and control units can be compared conditional on a principal stratum

- Average Principal Causal Effect (PCE):

$$PCE = E[Y_i(1) - Y_i(0) | S_i(0) = s_C, S_i(1) = s_T] \quad s_C, s_T \in \{0, 1\}$$

Principal Stratification: Treatment Compliance (Angrist et al., 1996)

$$S_i(z) = \text{ Treatment received given assignment } z, \quad z = 0, 1$$

$$S_i(z) = \begin{cases} 0, \text{ if subject } i \text{ received the control treatment given assignment } z; \\ 1, \text{ if subject } i \text{ received the new treatment given assignment } z. \end{cases}$$

$$\begin{aligned} 00 &= \{i : S_i(0) = 0, S_i(1) = 0\} = \text{ Never Takers} \\ 10 &= \{i : S_i(0) = 1, S_i(1) = 0\} = \text{ Defiers} \\ 01 &= \{i : S_i(0) = 0, S_i(1) = 1\} = \text{ Compliers} \\ 11 &= \{i : S_i(0) = 1, S_i(1) = 1\} = \text{ Always Takers} \end{aligned}$$

Complier Average Causal Effects ($CACE$) =

Causal Effect on the Principal Stratum of Compliers:

$$CACE = E\Big(Y_i(1) - Y_i(0) | S_i(0) = 0, S_i(1) = 1\Big)$$

Principal Stratification: Censoring (or Truncation) by Death (Rubin, 1998, 2000)

Drug Treatment:

$$Z_i = \begin{cases} 0, & \text{if subject } i \text{ is assigned a traditional drug;} \\ 1, & \text{if subject } i \text{ is assigned a new drug.} \end{cases}$$

Primary Outcome:

$$Y_i = \text{ Quality Of Life (QOL) two years post-randomization}$$

Intermediate Outcome:

$$S_i = \text{ Indicator for two-year survival}$$

- Some subjects will not reach the end-point of two years post-randomization survival: they will die before completion of the study.
- QOL for subjects who die would be undefined: it is "censored" or "truncated" by death

$$S_i(z) = \text{ Indicator for two-year survival given assignment } z, \quad z = 0, 1$$

$$S_i(z) = \begin{cases} D, & \text{if subject } i \text{ dies given assignment } z; \\ L, & \text{if subject } i \text{ lives given assignment } z. \end{cases}$$

- Never Survivals: Subjects who will die no matter how treated

$$DD = \{i : S_i(0) = D, S_i(1) = D\}$$

- Defiant Survivals: Subjects who will die if treated but live otherwise

$$LD = \{i : S_i(0) = L, S_i(1) = D\}$$

- Compliant Survivals: Subjects who will live if treated but die otherwise

$$DL = \{i : S_i(0) = D, S_i(1) = L\}$$

- Always Survivals: Subjects who will live no matter how treated

$$LL = \{i : S_i(0) = L, S_i(1) = L\}$$

- A well defined real values for the average causal effect of the active treatment versus the control treatment on QOL exists only for the $LL = \{i : S_i(0) = L, S_i(1) = L\}$:

$$SACE = E\Big[Y_i(1) - Y_i(0) | S_i(0) = L, S_i(1) = L\Big]$$

  where $SACE$ stands for Survival Average Causal Effect

- For the $LD = \{i : S_i(0) = L, S_i(1) = D\}$ and $DL = \{i : S_i(0) = D, S_i(1) = L\}$ groups, the average causal effect on QOL involves to assume we know how to trade off a particular QOL and being dead (and out of misery)

- For the $DD = \{i : S_i(0) = D, S_i(1) = D\}$ group there is no QOL to compare

Censoring (or Truncation) by Death: Additional Examples

- Evaluating the causal effects of job training programs on wages (Zhang et al., 2008, 2009)

$$S(z) = \text{ Indicator of employment given assignment } z$$

- Evaluating the causal effects of a special educational intervention on final test scores (Zhang & Rubin, 2003)

$$S(z) = \text{ Graduation indicator given assignment } z$$

- Evaluating the causal effect of Breast Self-Examination (BSE) teaching courses on quality of execution of BSE (Mattei & Mealli, 2007)

$$S(z) = \text{ Indicator of BSE practice given assignment } z$$

- Evaluating the effectiveness of degree programs on employment status of their graduates (Grilli & Mealli, 2008)

$$S(z) = \text{ Graduation indicator given assignment } z$$

Principal Stratification: Direct and Indirect Causal Effects (Mealli & Rubin, 2003)

Treatment = Socio-economic status (Wealth)

$$Z_i = \begin{cases} 0, & \text{if subject } i\text{'s wealth is low;} \\ 1, & \text{if subject } i\text{'s wealth is high.} \end{cases}$$

Intermediate Outcome:

$$S_i = \text{ General health six months after "assignment" of wealth}$$

$$S_i(z) = \begin{cases} B, & \text{if subject } i\text{'s six-months general health is bad given assignment } z; \\ G, & \text{if subject } i\text{'s six-months general health is good given assignment } z. \end{cases}$$

Primary Outcome:

$$Y_i = \text{ Mortality one year after "assignment" of wealth}$$

The total effect of Wealth will be a combination of the direct effect of Wealth on Mortality and the indirect effect mediated by General Health Status

- Subjects whose general health would be bad and unaffected by their wealth

$$BB = \{i : S_i(0) = B, S_i(1) = B\}$$

- Subjects whose general health would be bad under high wealth and good under low wealth

$$GB = \{i : S_i(0) = G, S_i(1) = B\}$$

- Subjects whose general health would be good under high wealth and bad under low wealth

$$BG = \{i : S_i(0) = B, S_i(1) = G\}$$

- Subjects whose general health would be good and unaffected by their wealth

$$GG = \{i : S_i(0) = G, S_i(1) = G\}$$

A direct causal effect of Wealth, after controlling for current general health, exists if there is a causal effect of Wealth on current mortality for subjects for whom the treatment does not affect general health (i.e., basic principal strata BB and GG)

Principal Stratification: Direct and Indirect Causal Effects (Sjöander et al., 2009; Schwartz et al., 2011)

Treatment = Physical activity (PA)

$$Z_i = \begin{cases} 0, & \text{if subject } i\text{'s PA level is low;} \\ 1, & \text{if subject } i\text{'s PA level is high.} \end{cases}$$

Intermediate Outcome: Body Mass index (BMD) after "assignment" of PA

$$S_i(z) = \begin{cases} H, & \text{if subject } i \text{ is obese (his/her BMI is high) given assignment } z; \\ L, & \text{if subject } i \text{ is not obese (his/her BMI is low) given assignment } z. \end{cases}$$

Primary Outcome: CardioVascular Disease (CVD) after "assignment" of PA

$$Y_i(z) = \begin{cases} 1, & \text{if subject } i \text{ reports at least one CVD event before end of follow-up given assig} \\ 0, & \text{if subject } i \text{ remains undiagnosed through follow-up given assignment } z. \end{cases}$$

The total effect of physical activity will be a combination of the direct effect of PA on CVD and the indirect effect mediated by BMD

- Subjects who would be obese under both PA levels: BMI is unaffected by PA

$$HH = \{i : S_i(0) = H, S_i(1) = H\}$$

- Subjects who would be obese under high PA level and would not be obese under low PA level

$$LH = \{i : S_i(0) = L, S_i(1) = H\}$$

- Subjects who would not be obese under high PA level and would be obese under low PA level

$$HL = \{i : S_i(0) = H, S_i(1) = L\}$$

- Subjects would not be obese under both PA levels: BMI is unaffected by PA

$$LL = \{i : S_i(0) = L, S_i(1) = L\}$$

A direct causal effect of PA, after controlling for BMI, exists if there is a causal effect of PA on CVD for subjects for whom the treatment does not affectBMI (i.e., basic principal strata HH and LL)

Hypothetical Example (1)

Principal Stratum of subject i	Full Data: Post-treatment Variable BMI $S_i(0)$	$S_i(1)$	Potential Outcomes: CVD Events (%) $Y_i(0)$	$Y_i(1)$	Observed Data from a Randomized Study: Average of $(S_i^{\text{obs}}, Y_i^{\text{obs}})$ given assignment $Z_i^{\text{obs}} = 0$	$Z_i^{\text{obs}} = 1$
Not Obese Subjects	L	L	10	10	$(L, 20)$	$(L, 10)$
Normal	H	L	30	50		$(H, 50)$
Obese Subjects	H	H	50	50	$(H, 50)$	

We assume that there are no special subjects: $LH = \emptyset$

Equal proportions for each principal stratum

Hypothetical Example (2)

Principal Stratum of subject i	Full Data: Post-treatment Variable BMI $S_i(0)$	$S_i(1)$	Potential Outcomes: CVD Events (%) $Y_i(0)$	$Y_i(1)$	Observed Data from a Randomized Study: Average of $(S_i^{\text{obs}}, Y_i^{\text{obs}})$ given assignment $Z_i^{\text{obs}} = 0$	$Z_i^{\text{obs}} = 1$
Not Obese Subjects	L	L	10	20	$(L, 20)$	$(L, 20)$
Normal	H	L	30	40		$(H, 50)$
Obese Subjects	H	H	50	60	$(H, 50)$	

We assume that there are no special subjects: $LH = \emptyset$

Equal proportions for each principal stratum

Direct and Indirect Causal Effects: Additional Examples

- Evaluating to what extent the causal effect of birth-control pill on thrombosis in women is mediated by the effect of being on the contraceptive pill on pregnancy (Pearl, 2001)
- Evaluating to what extent the causal effects of a new drug treatment having side-effects is be mediated by the effect of taking additional medication to counter its side-effects (Pearl, 2001)
- The causal effects of a major training program on participants' earnings might be mediated by lock-in effects, that is, the loss of labour market experience (Flores & Flores-Lagunes, 2009)
- Evaluating the extent to which smoking during pregnancy affects the incidence on low birth weight through a shorter gestation time might inform policy makers on the opportunity to promote drugs that lengthen gestation time (Flores & Flores-Lagunes, 2009)
- Evaluating to what extent the effect of military service on veterans' earnings is channelled by subsidized higher education (Angrist & Chen, 2008)

The Faenza Randomized Experiment

- Randomized experiment on Breast Self-Examination (BSE) conducted between January 1988 and December 1990 at the Oncologic Center of the Faenza Health District in Italy.
- In the study, two BSE teaching methods were compared:
  - a *standard treatment* of receiving mailed information only, and
  - a *new treatment* of additional attendance in a self-exam course.
- The question of interest is the effect of an enhanced training class on BSE practices and quality of self-exam execution.

Data Complications

- The Faenza BSE study suffers from complications due to
  - *noncompliance* with the randomly assigned treatment: only 55% of the women assigned to the new treatment complied with their assignment;
  - *"truncation by death"*: quality of self exams is not only unobserved but also undefined on the sample space for women who do not practice BSE.

Faenza BSE study - Summary statistics

		Means				
	Grand mean	$Z_i^{obs}=C$	$Z_i^{obs}=T$	$Z_i^{obs}=T$ $D_i^{obs}=p$	$Z_i^{obs}=T$ $D_i^{obs}=P$	$D_i^{obs}=p$
N	657	327	330	148	182	475
Assignment (Z_i^{obs})	0.502	0	0	1	1	0.312
Course attendance (D_i^{obs})	0.277	0	0.551	0	1	0
Response (S_i^{obs})	0.653	0.688	0.618	0.399	0.797	0.598
BSE practice (S_i^{obs})*	0.785	0.796	0.774	0.475	0.897	0.729
BSE quality (Y_i^{obs})*	0.492	0.402	0.594	0.250	0.669	0.381
Pretreatment variables:						
Prior BSE practice (X_{i1}^{obs})**	0.585	0.591	0.579	0.551	0.601	0.579
Knowledge of breast pathophysiology(X_{i2}^{obs})	0.554	0.560	0.548	0.439	0.637	0.522
Age (X_{i3}^{obs})	41.4	41.5	41.3	41.7	41.0	41.6

(*) Computed on respondents only. (**) Available for 615 women.

Estimands of interest

- *Causal Estimands on BSE practice outcome*
  - Intention-To-Treat (ITT) effect;
  - Complier Average Causal Effect (CACE);
  - Never-taker Average Causal Effect (NACE).
- *Causal Estimands on BSE quality outcome*
  - ITT effect for all women who would practice BSE under both assignments;
  - average causal effect for compliers who would practice BSE who would practice BSE under both treatments.

Potential Outcomes

If woman i in the study ($i = 1, \ldots, N$) is to be assigned to treatment z ($z = 1$ or $z = 0$), we denote the following:

- *Indicator of the treatment received*:

$$D_i(z) = \begin{cases} P, & \text{if the woman attends the training program;} \\ p, & \text{if the woman receives only mailed information on BSE.} \end{cases}$$

- *BSE practice indicator*: $S_i(z) = \begin{cases} B, & \text{if the woman practices BSE;} \\ b, & \text{otherwise.} \end{cases}$

- *Potential quality outcome*:

$$Y_i(z) = \begin{cases} H, & \text{if the woman practices BSE with “High” quality;} \\ L, & \text{if the woman practices BSE with “Low” quality;} \\ *, & \text{if the woman does not practice BSE.} \end{cases}$$

Principal Stratification

- The variable $D_i(1)$ defines the compliance behavior of subject i:
  - If $D_i(1) = P$, then woman i is a "*complier*";
  - if $D_i(1) = p$, then woman i is a "*never-taker*".
- The vector $(S_i(0), S_i(1))$ defines the BSE practice behavior of subject i.

Principal Stratification

PBB $= \{i : D_i(1) = P, S_i(0) = B, S_i(1) = B\}$: compliers who would practice BSE under both treatment arms;

PbB $= \{i : D_i(1) = P, S_i(0) = b, S_i(1) = B\}$: compliers who would not practice BSE under control but would practice BSE under treatment;

PBb $= \{i : D_i(1) = P, S_i(0) = B, S_i(1) = b\}$: compliers who would practice BSE under control but would not practice BSE under treatment;

Pbb $= \{i : D_i(1) = P, S_i(0) = b, S_i(1) = b\}$: compliers who would practice BSE under neither treatment arms;

pBB $= \{i : D_i(1) = p, S_i(0) = B, S_i(1) = B\}$: never-takers who would practice BSE under both treatment arms;

pbB $= \{i : D_i(1) = p, S_i(0) = b, S_i(1) = B\}$: never-takers who would not practice BSE under control but would practice BSE under treatment;

pBb $= \{i : D_i(1) = p, S_i(0) = B, S_i(1) = b\}$: never-takers who would practice BSE under control but would not practice BSE under treatment;

pbb $= \{i : D_i(1) = p, S_i(0) = b, S_i(1) = b\}$: never-takers who would practice BSE under neither treatment arms

Principal stratification and associated pattern for potential outcomes

Principal Stratum	$D_i(1)$	$S_i(0)$	$S_i(1)$	$Y_i(0)$	$Y_i(1)$
PBB	P	B	B	$\in \{L, H\}$	$\in \{L, H\}$
PbB	P	b	B	$*$	$\in \{L, H\}$
PBb	P	B	b	$\in \{L, H\}$	$*$
Pbb	P	b	b	$*$	$*$
pBB	p	B	B	$\in \{L, H\}$	$\in \{L, H\}$
pbB	p	b	B	$*$	$\in \{L, H\}$
pBb	p	B	b	$\in \{L, H\}$	$*$
pbb	p	b	b	$*$	$*$