“Causal Inference” by Hernan & Robins
Ch.15-16: PS, IV

Svetlana Belitser

17 dec 2018

Outcome regression and various versions of ***propensity score*** analyses are the most commonly used parametric methods for causal inference.
They work fine in simpler settings.
But they do not work in general: causal inference for time-varying treatment.

Outcome regression (Ch.15.1)

Goal: to estimate the average causal effect (ACE) as

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$$

(the average causal effect of smoking cessation (the treatment) A on weight gain (the outcome) Y).

Assumptions:
* conditional exchangeability (CE) ($Y^a \perp\!\!\!\perp A \,|\, L \quad \forall a$)
(exchangeability of the treated and the untreated conditional on the L variables:
sex, age, race, education, intensity and duration of smoking, physical activity, recreational exercises, weight).
* Positivity (P): $\mathrm{P}[A = a|L = l] > 0 \quad \forall l : \ \mathrm{P}[L = l] \neq 0$ in the population of interest.
* Well-defined interventions(C).

In the three previous chapters we described how to estimate the ACE

- in population (*IP weighting*, *standardization* and *g-estimation*) and
- within subsets of the population (p.43 Ch.15)
  - – by *restricting* the analysis to the subset of interest, or
  - – by adding product terms in *marginal structural models* (MSM) (Ch.12), or
  - – by adding product terms in *structural nested models* (SNM) (Ch.14).

Structural nested models (Ch.14):

the causal effect of A on Y within levels of L under (C), (CE), (P):

$$\mathrm{E}[Y^a - Y^{a=0}|L] = \mathrm{E}[Y^a - Y^{a=0}|A = a, L] = \beta_1 a + \beta_2 aL$$

Outcome regression

The outcome regression model: to specify the $L-Y$ association within levels of A:

$$\mathrm{E}[Y^a|L] = \beta_0 + \beta_1 a + \beta_2\, a\, L + \beta_3 L, \quad \beta_2, \beta_3 \; are \; vectors.$$

The outcome regression adjusts for confounding by estimating the causal effect of treatment in each stratum of L (like stratification in Ch.3).

We refer to this model as “faux marginal structural model” (in a slightly humorous vein) because it has form of *marginal structural model* but *without weights* because $SW^A(L) = 1$:
Definition of the stabilized weights $SW^A(V)$ with effect modifier[s] V (Ch.12.5):
$SW^A(V) = \dfrac{f(A|V)}{f(A|L)}$, $V \in L \;\Rightarrow\; SW^A(L) = 1.$ □

CE, P, C $\Rightarrow$ $\mathrm{E}[Y^a|L] = \mathrm{E}[Y|A, L] = \alpha_0 + \alpha_1 a + \alpha_2 aL + \alpha_3 L$
For dichotomous Y: a logistic regression.

Nuisance parameters (Fine Point 15.1.)

Consider model: $\mathrm{E}[Y^a|L] = \beta_0 + \beta_1 a + \beta_2\, a\, L + \beta_3 L.$

$\Longrightarrow \mathrm{E}[Y^{a=0}|L] = \beta_0 + \beta_3 L$ and $\mathrm{E}[Y^a - Y^{a=0} \mid L] = \beta_1 a + \beta_2 aL.$

Goal: to estimate the causal parameters β_1 and β_2.
β_1 and β_2 are in general ***consistent*** estimated only if

- in **the outcome regression**:
  the model $\mathrm{E}[Y^{a=0}|L] = \beta_0 + \beta_3 L$ is correctly specified.
  β_0 and β_3 - *nuisance parameters* (do not have a causal interpretation, not of primary interest).
- in g**-estimation of the structural nested model (SNM)**:
  the model $logit\, \mathrm{P}[A = 1|L] = \alpha_0 + \alpha_1 L$ is correctly specified.
  α_0 and α_1 are *nuisance parameters*.

noncorrectly specified model: e.g., should be $\beta_3 L + \beta_4 L^2$ instead of $\beta_3 L$.
Deciding what method (the *outcome model* or *structural model*) to use:
$=$ deciding which nuisance parameters we believe can be more accurately estimated;
or to use *doubly-robust methods* (Tech.Point 14.2).

in 13.2: IP weighting, standardization $\Rightarrow$ est. ACE = 3,5 kg
in 15.1: outcome regeression: est. ACE = 2.6 + 0.05 * (smoking intensity)

A common approach to outcome regression:

- no effect modification $\Rightarrow$ no product terms
- β_1 is an estimate of both: **conditional** and **marginal** average causal effects.

When outcome regression is an intermediate step to estimate the counterfactual outcomes
correct specification of model is required. $\Rightarrow$ β_3 become necessary too.

(unweighted) outcome regression model if correctly specified, fully adjusts for all confounding by L (with all effect modifiers) (p.20, Ch.12.5)

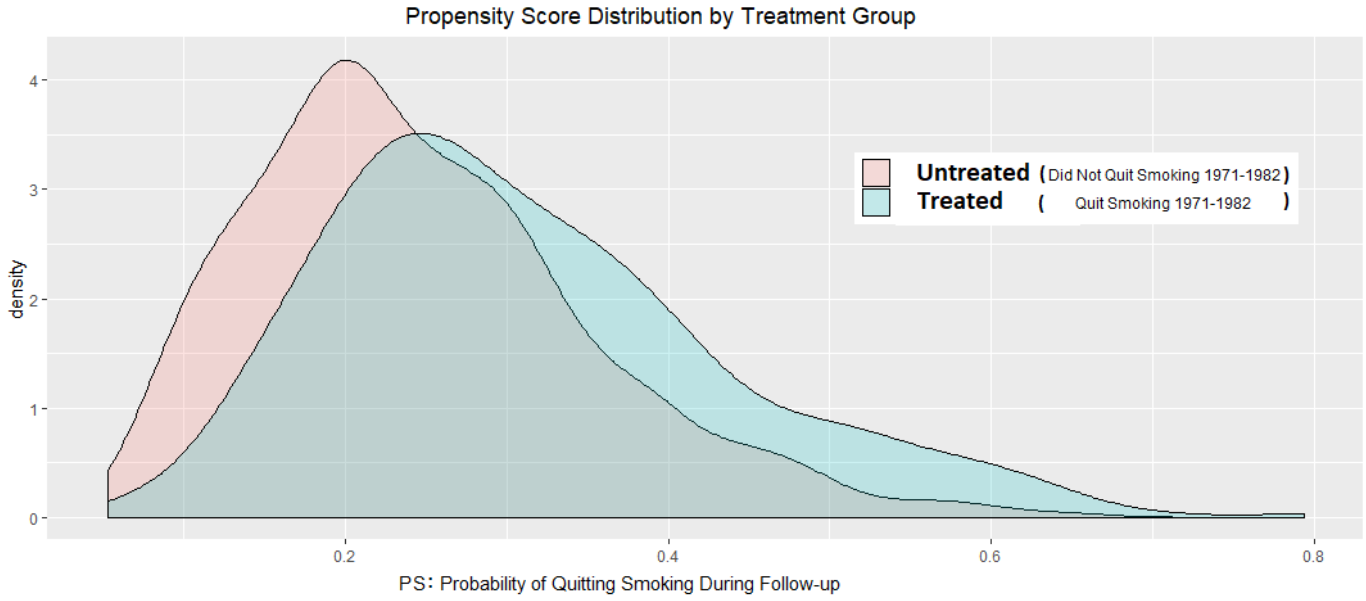
Propensity scores (PS) (Ch.15)

Consider PS only for dichotomous treatment $A \in \{0, 1\}$.
PS methods (other than IP weighting and g-estimation and doubly-robust estimators) are difficult to generalize to non-dichotomous A.

Propensity score $p(L)$:

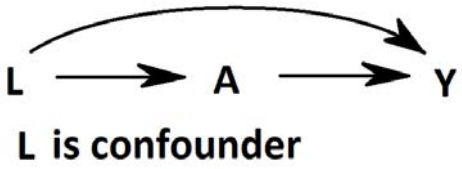
$$p(L) = \mathrm{P}[A = 1|L]$$

measures the propensity of individuals to receive treatment given the information available in the covariates L.



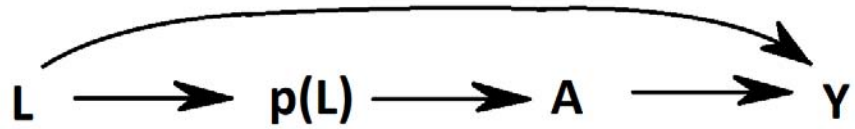
An ideal 50/50 randomized trial: $p(L) = 0.5$ for all individuals and any choise of L.

If the distribution of $p(L)$ is the same for $A = 1$ and $A = 0$
$\Rightarrow$ no confounding due to L (no open path $L \rightarrow A$):



L is confounder

A⟶Y - causal effect
A⟵L⟶Y - backdoor path (noncausal)



L - confounder

p(L) - an intermediate between L and A

A⟶Y - causal effect

A⟵ p(L) ⟵ L⟶Y - backdoor path (noncausal)

Balancing scores $b(L)$ is any function of L (Tech.Point 15.1):

$$A \perp\!\!\!\perp L \mid b(L) .$$

- $p(L)$ is the simplest example of a balancing score.
- For each value of $b(L)$, the distribution of L is the same in the treated and the untreated.
- Rosenbaum & Rubin (1983) proved:
  – (CE) and (P) based on $L \implies$ (CE) and (P) based on $b(L)$;
  – if it is sufficient to adjust for $L \implies$ sufficient to adjust for $b(L)$ (or $p(L)$).

Prognostic scores $s(L)$ is any function of L (Tech.Point 15.1):

$$Y^{a=0} \perp\!\!\!\perp L \mid s(L) .$$

- stronger assumptions than for $b(L)$;
- cannot be easily extended to time-varying treatments.

Assumptions:

- exchangeability within levels of PS (CE):
$$Y^a \perp\!\!\!\perp A|L \implies Y^a \perp\!\!\!\perp A \,|\, p(L)\,.$$
- Positivity (P) within levels of PS:
$$\mathrm{P}[A = a \,|\, L = l] > 0 \iff 0 < \mathrm{P}[A = a \,|\, p(L)] < 1\,.$$
- Well-defined interventions (C).

Propensity scores approach

The average causal effect (ACE) under (CE), (P), (C): for a particular s,

$$\mathrm{E}[Y^{a=1}|\, p(L)=s] - \mathrm{E}[Y^{a=0}|\, p(L)=s] \;=\; \mathrm{E}[Y|A=1, p(L)=s] - \mathrm{E}[Y|A=0, p(L)=s].$$

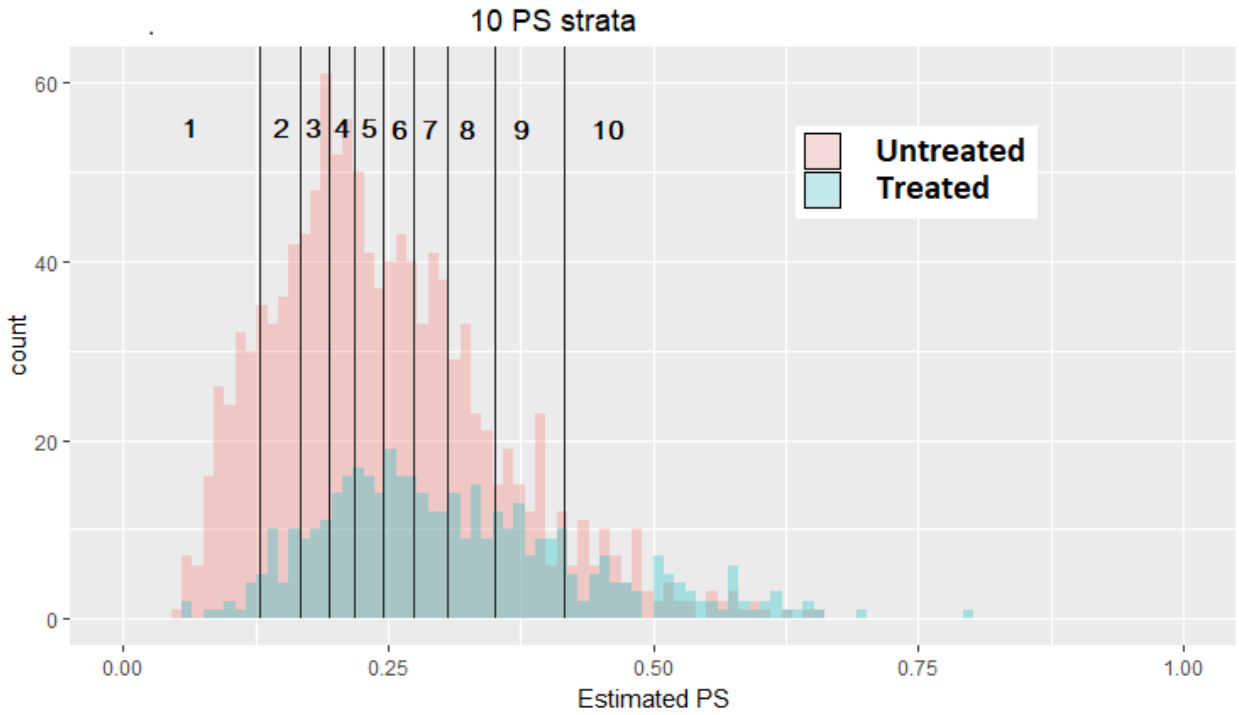
1. the ***true*** $p(L)$ is unknown $\Rightarrow \hat{p}(L)$:
   – For example, by logistic regression, bagging or boosting, recursive partitioning or tree-based methods, random forests, and neural networks.
   – The same models (for $p(L)$) as for *IP weighting* and *g-estimation*.
2. $p(L)$ is continuous $\Rightarrow$ unlikely find individuals with the same $p(L) = s$.
   $\Rightarrow$ Compare individuals with similar values of $p(L)$ (not exact s).

PS methods:

- stratification;
- as covariate in the outcome regression model;
- standardization;
- matching;
- weighting (the same as *IP weighting*).

Propensity stratification and standardization (Ch.15.3)

Stratification on PS *(Program 15.3):*
based on estimated propensity scores individuals are classified in M strata of approximately equal size (often: $M = 10$), then the causal effect is estimated in each of the strata.



Potential problem:

- exchangebility assumption within all strata
  – the average $p(L)$ were greater in the treated than in the untreated in some strata.
  – Balance diagnostics should be used for comparing the distribution of covariates between treatment groups within PS-strata: $A \perp\!\!\!\perp L \,|\, p_M(L)$ and assessing whether the propensity score model has been adequately specified
- the correct specification of the relationship between $p(L)$ and Y.
  – we assume: linear, but it can be, e.g., cubic splines.
  – no such problem in *IP weighting* or *g-estimation* (no resgression for $Y - p(L)$ relation).

Stratification on $p_M(L)$ (M strata of PS):

- $p_M(L)$ as effect modifier:

$$\mathrm{E}[Y \mid A, p_M(L)] = \alpha_0 + \alpha_A A + \alpha_1\, p_M(L) + \alpha_2\, A\, p_M(L),$$

  M - number of strata $\Rightarrow$ $\alpha_1 = (\alpha_{1,1}, \ldots, \alpha_{1,M-1})$ and $\alpha_2 = (\alpha_{2,1}, \ldots, \alpha_{2,M-1})$.
- no effect modification by $p_M(L)$:

$$\mathrm{E}[Y \mid A, p_M(L)] = \alpha_0 + \alpha_A A + \alpha_1\, p_M(L).$$

PS as a covariate in the outcome regression model:
$p(L)$ in the previous models instead of $p_M(L)$.

Standardization on $p(L)$ (not on L)
to estimate ACE in the entire population $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$ (Program 15.4):

Standardized mean of Y under (CE),(P),(C):

$\mathrm{E}[Y^a] = \mathrm{E}[Y \mid A = a] = \int \mathrm{E}[Y \mid A = a, p(L) = s] dF_{p(L)}(s)$ for continuous $p(L)$, $\mathrm{E}[Y^a] = \mathrm{E}[Y \mid A = a] = \sum_s \mathrm{E}[Y \mid A = a, p(L) = s]\, \mathrm{P}[p(L) = s]$ for discrete $p(L)$.

$\Rightarrow$ The standardized mean difference in the entire population:

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=1}] = \mathrm{E}[Y \mid A = 1] - \mathrm{E}[Y \mid A = 0]$$
$$= \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathrm{E}}[Y \mid A = 1,\, p(L_i)] - \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathrm{E}}[Y \mid A = 0,\, p(L_i)],$$

where $\widehat{\mathrm{E}}[Y \mid A = a,\, p(L_i)]$ - predicted outcome for individual i and treatment a

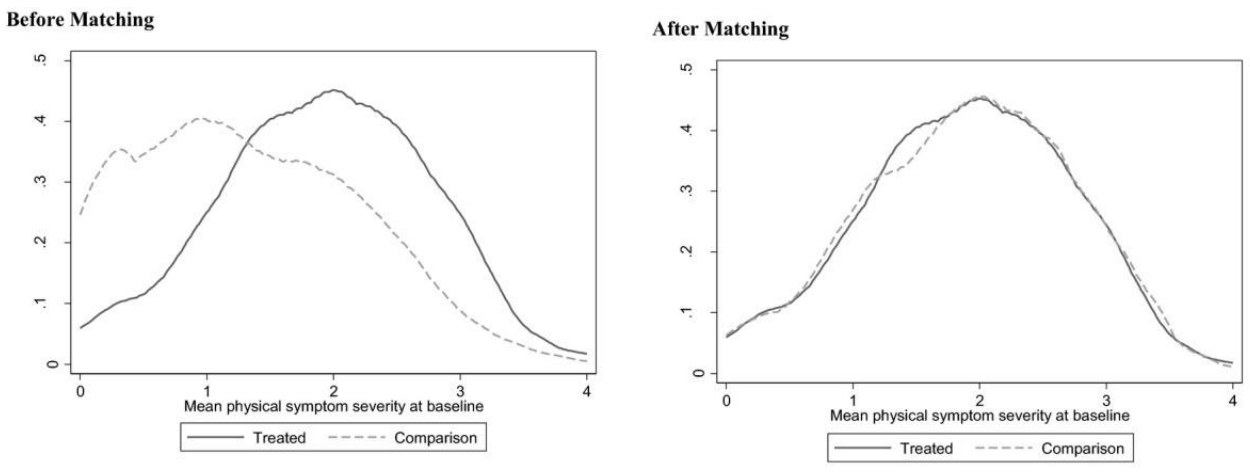
(estimated using regression: *stratification* on the propensity score, or *covariate adjustment* using the propensity score).

Propensity matching

Matching treated and untreated individuals with similar propensity score $p(L)$:

- many forms of propensity matching:
  nearest neighbor matching, caliper matching, mahalanobis metric matching, stratification matching, difference-in-differences matching (kernel and local linear weights), exact matching, …
- Propensity matching:
  often almost all treated individuals included and many untreated individuals excluded.
- *The matched population*:
  – subset of the original population with the treated-untreated pairs (or sets),
  – may be very different from the original study population $\Rightarrow$ hard-to-describe.
- (C), (CE), (P) given $p(L)$ $\Rightarrow$ consistent estimates of effect measures.

Nobody knew how to compute the variance of the effect estimate till 2006 (Abadie and Imbens,2006).



from site: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213057/ (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213057/)

Instrumental variables (IV) (Ch.16)

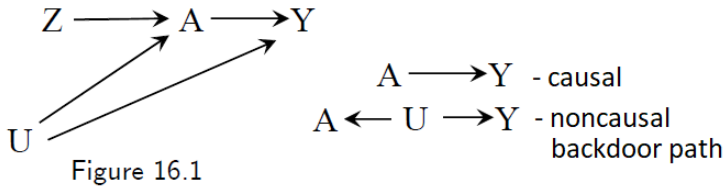


Figure 16.1

E.g., Z - randomization, A - treatment use, $Z = A$ (compliance), $Z \neq A$ (noncompliance), U - unmeasured.

Z is an instrument because IV conditions hold:

(i): most patients compliant; (ii): double-blind design, (iii): randomization.

(CE) fails, because we can not block path $A \leftarrow U \rightarrow Y$.

$\Rightarrow$ ***no*** *IP weighting, standardization, g-estimation, stratification*, or *matching*.

Three ***instrumental conditions*** for IV:

i. $Z \not\perp\!\!\!\perp A$ (***relevance condition***).

ii. Z does not affect Y except through its potential effect on A (***exclusion restriction***: no direct effect of Z on Y).

iii. Z and Y do not share causes (***marginal exchangeability***: $Y^{a,z} \perp\!\!\!\perp Z \ \ \forall a, z$).

.(ii) & (iii) $\Rightarrow Y^a \perp\!\!\!\perp Z$.

Two-Stage least squares regression model:

$$Y = \alpha + \beta \widehat{A} + E$$

$$A = \gamma + \delta Z + F$$

Three conditions for IV Z:

i. $\rho_{Z,A} \neq 0$,

ii. $\rho_{Z,E} = 0$ (no correlation between IV and other factors explaining Y, i.e. E),

iii. $\rho_{Z,F} = 0$ (no correlation between IV and other factors explaining A, i.e. F).

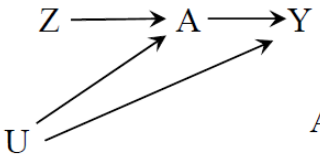


Figure 16.1

A → Y - causal

A ← U → Y - noncausal backdoor path

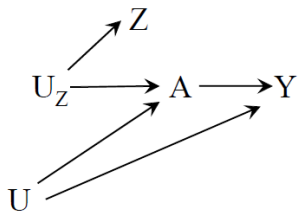


Figure 16.2

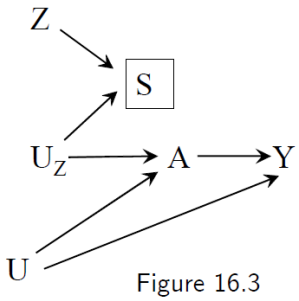


Figure 16.3

Candidate instruments in observational studies: genetic factors, preference, access.

Not to prove that Z is truly IV because:
(i) only (i) can be verified empirically ($\Rightarrow$ *weak* or *strong instrument*),
(ii) and (iii) cannot be verified.

$\Rightarrow$ assume that (ii) and (iii) hold.

Z - instrument if (i),(ii),(iii) hold.

An instrument Z (i.e.(i)-(iii) hold):

- not enough to estimate ACE (need one more [unverifiable] assumption);
- enough to estimate *upper* and *lower* bounds of ACE.

The **usual IV estimand** (16.2):

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] = \frac{cov(Y,Z)}{cov(A,Z)}, \quad Z \text{ is continuos},$$

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] = \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]}, \quad Z \in \{0,1\}.$$

Additive structural mean models (SMM) and **IV estimation** (Tech.point 16.3)
Saturated additive **structural mean models**
(*semiparametric str.mean model*, or *structural nested mean model*):

$$\mathrm{E}[Y^{a} - Y^{a=0} \mid Z\,] = \mathrm{E}[Y^{a} - Y^{a=0} | A = a, Z] = \beta_0 a + \beta_1 a Z$$

$$\mathrm{E}[Y - Y^{a=0} | A, Z] = A(\beta_0 + \beta_1 Z)$$

If $\beta_1 = 0$ (no effect modification by Z) $\Rightarrow \beta_0$ is the **usual IV estimand** (Robins, 1994).

Proof:
Z is IV and (ii) holds. $\Rightarrow \mathrm{E}[Y^{a=0} | Z = 1] = \mathrm{E}[Y^{a=0} | Z = 0]$.
via SMM:

$$\mathrm{E}[Y - Y^{a=0} | A, Z] = \mathrm{E}[Y | A, Z] - \mathrm{E}[Y^{a=0} | A, Z] = A(\beta_0 + \beta_1 Z)$$

$$\mathrm{E}[Y^{a=0} | A, Z] = \mathrm{E}[Y | A, Z] - A(\beta_0 + \beta_1 Z) = \mathrm{E}[Y - A(\beta_0 + \beta_1 Z) | A, Z]$$

$$\mathrm{E}[Y^{a=0} | Z] = \mathrm{E}_A\, \mathrm{E}[Y^{a=0} | A, Z] = \mathrm{E}_A\, \mathrm{E}[Y - A(\beta_0 + \beta_1 Z) | A, Z] = \mathrm{E}[Y - A(\beta_0 + \beta_1 Z) | Z]$$

$$\text{(ii) holds } \Rightarrow \mathrm{E}[Y - A(\beta_0 + \beta_1) | Z = 1] = \mathrm{E}[Y - A\,\beta_0 | Z = 0]$$

$$\text{no effect modifier} \Rightarrow \beta_1 = 0: \quad \mathrm{E}[Y - A\,\beta_0 | Z = 1] = \mathrm{E}[Y - A\,\beta_0 | Z = 0]$$

$$\mathrm{E}[Y | Z = 1] - \beta_0\ \mathrm{E}[A | Z = 1] = \mathrm{E}[Y | Z = 0] - \beta_0\ \mathrm{E}[A | Z = 0]$$

$$\beta_0 = \frac{\mathrm{E}[Y | Z = 1] - \mathrm{E}[Y | Z = 0]}{\mathrm{E}[A | Z = 1] - \mathrm{E}[A | Z = 0]} \quad \square$$

A fourth identifying condition: effect homogeneity (Ch.16.3)

One version of ***effect homogeneity*** condition (extreme unrealistic version):

iv. constant ACE across individuals (= additive rank preservation in Section 14.4).

Yet additive rank preservation was implicitly assumed in many early IV analyses using the two-stage-least squares estimator.

Another less extreme version of ***effect homogeneity*** condition (additive or multiplicative):

iv. constant ACE within levels of Z and A.
$\mathrm{E}[Y^{a=1} - Y^{a=0} \mid Z = 1, A = a] = \mathrm{E}[Y^{a=1} - Y^{a=0} \mid Z = 0, A = a]$,
$a = \{0, 1\}, Y \in \{0, 1\}$.

- in practice: no unmeasured effect modifiers.

More alternatives:

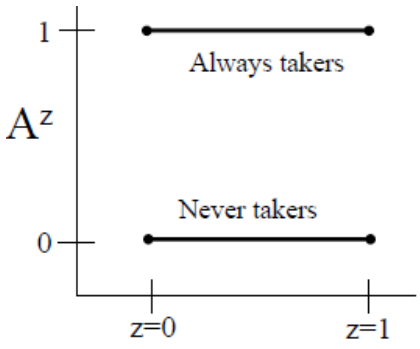
- baseline covariates in the models for IV estimation;
  (Also with multiple proposed instruments simultaneously),
- monotonicity condtition.

An alternative fourth condition: monotonicity (Ch.16.4)

Consider counterfactuals for A: $A^{z=1}$ and $A^{z=0}$.

Classify all individuals into 4 ***compliance types*** or ***principal strata***:

1. ***Always-takes***: $A^{z=1} = 1$ and $A^{z=0} = 1$.
2. ***Never-takers***: $A^{z=1} = 0$ and $A^{z=0} = 0$.
3. ***Compliers*** : $A^{z=1} = 1$ and $A^{z=0} = 0$.
4. ***Defiers*** : $A^{z=1} = 0$ and $A^{z=0} = 1$.



Figures 16.4, 16.5

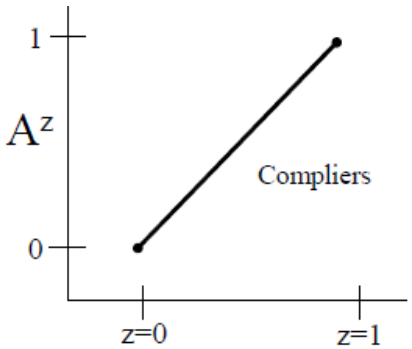


Figure 16.6

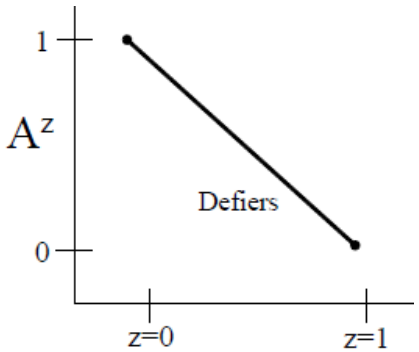


Figure 16.7

iv. ***Monotonicity condition***:
$A^{z=1} \geq A^{z=0}$ for all individuals, i.e., no defiers exist.

Under (i)-(iv), ACE of A on Y in the compliers is the usual IV estimand:

$$\mathrm{E}[Y^{a=1} - Y^{a=0} \mid A^{z=1} = 1, A^{z=0} = 0] = \mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$$

$$= \frac{cov(Y, Z)}{cov(A, Z)} = \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]}, \quad Z \in \{0, 1\}$$

Proof (Tech.point 16.6) for $Z \in \{0; 1\}$:

$$\mathrm{E}[Y^{z=1} - Y^{z=0}] =$$

$$= \mathrm{E}[Y^{z=1} - Y^{z=0} \mid A^{z=1} = 1, A^{z=0} = 1]\ \mathrm{P}[A^{z=1} = 1, A^{z=0} = 1] \text{ (=0 (ii), always-takes)}$$

$$+ \mathrm{E}[Y^{z=1} - Y^{z=0} \mid A^{z=1} = 0, A^{z=0} = 0]\ \mathrm{P}[A^{z=1} = 0, A^{z=0} = 0] \text{ (=0 (ii), never-takers)}$$

$$+ \mathrm{E}[Y^{z=1} - Y^{z=0} \mid A^{z=1} = 1, A^{z=0} = 0]\ \mathrm{P}[A^{z=1} = 1, A^{z=0} = 0] \quad \text{(compliers)}$$

$$+ \mathrm{E}[Y^{z=1} - Y^{z=0} \mid A^{z=1} = 0, A^{z=0} = 1]\ \mathrm{P}[A^{z=1} = 0, A^{z=0} = 1] \quad \text{(defiers)}$$

Assumption of no defiers, and because in the compliers $Z = A \Rightarrow$:

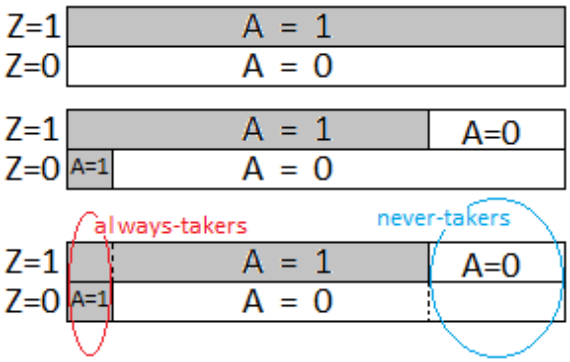
$$\mathrm{E}[Y^{z=1} - Y^{z=0}] = \mathrm{E}[Y^{a=1} - Y^{a=0} \mid A^{z=1} = 1, A^{z=0} = 0]\ \mathrm{P}[A^{z=1} = 1, A^{z=0} = 0]$$

$$\mathrm{E}[Y^{a=1} - Y^{a=0} \mid A^{z=1} = 1, A^{z=0} = 0] = \frac{\mathrm{E}[Y^{z=1} - Y^{z=0}]}{\mathrm{P}[A^{z=1} = 1, A^{z=0} = 0]}$$

Ass: Z is randomization $\Rightarrow Z \perp\!\!\!\perp \{Y^{a,z},\ A^z;\ z = 0, 1;\ a = 0, 1\}$ (joint independence and consistency) $\Rightarrow$.

$$\mathrm{E}[Y^{z=1} - Y^{z=0}] = \mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]$$

$\mathrm{P}[A^{z=1} = 1, A^{z=0} = 0] = \mathrm{P}[A = 1|Z = 1] - \mathrm{P}[A = 1|Z = 0]$, because:



$$\mathrm{E}[Y^{a=1} - Y^{a=0} \mid A^{z=1} = 1, A^{z=0} = 0] = \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{P}[A=1|Z=1] - \mathrm{P}[A=1|Z=0]}\ \square$$