Econometrics II

Lecture 7: Instrumental Variables with Heterogeneous Treatment Effects

Konrad Burchardi

Stockholm University

23rd of April 2024

Literature

1. **”Mostly Harmless Econometrics”**, Angrist and Pischke
   Chapters 4.4 and 4.5

These notes draw on those books. All mistakes are mine.

Plan for Today

1. Local Average Treatment Effect
2. Characterizing Compliers
3. Generalisations of LATE

The LATE

Traditional IV framework is useful:

- Think about estimating constant causal effects.
- Think clearly about source of variation in the Z used to identify causal effects.

Today: $Y_i(1) - Y_i(0)$ does not have to be same across individuals.

$\rightarrow$ Return to the potential outcomes framework.

Focus on simple case where:

- we estimate effect of a *binary* treatment, D_i, on an outcome Y_i;
- the treatment is endogenous, but have *binary* instrument, Z_i.

(Later we will discuss generalizations.)

The LATE

Setup:

1. Let $Y_i(d, z)$ denote the potential outcome[1] when $D_i = d$ and $Z_i = z$, and $D_i(z)$ potential treatment status.
2. We think of the instrument as causally affecting treatment, but this effect too is allowed to be heterogeneous!

Potential treatment status: $D_i(0)$ when $Z_i = 0$; $D_i(1)$ when $Z_i = 1$

Observed treatment status: $D_i = D_i(0) + (D_i(1) - D_i(0))\, Z_i$

[1] Double indexing: candidate instruments might have a direct effect on outcomes.

LATE: Compliance Types

Therefore there are four 'compliance types':

	$D_i(1)$	$D_i(0)$
Compliers:	1	0
Never-takers:	0	0
Always-takers:	1	1
Defiers:	0	1

Just like with potential outcomes, the compliance type is unobserved.

What *is* observed? Z_i and D_i

	$D_i = 0$	$D_i = 1$
$Z_i = 0$	Compliers, Never-takers	Defier, Always-takers
$Z_i = 1$	Defier, Never-takers	Compliers, Always-takers

LATE: Assumptions

1. **Random Assignment** of instrument: $[\{Y_i(d,z)\forall d,z\}, D_i(1), D_i(0)] \perp Z_i$.
2. **Exclusion Restriction:** $Y_i(d,1) = Y_i(d,0), d = 0,1$.
3. **Relevance:** $\mathbb{E}[D_i(1) - D_i(0)] \neq 0$.
4. **Monotonicity:** $D_i(1) \geq D_i(0)$ [no defiers], or $D_i(1) \leq D_i(0)$ [no compliers].

LATE Assumptions: Random Assignment

What does random assignment assumption buy us?

$$[\{Y_i(d,z), \forall d, z\}, D_i(1), D_i(0)] \perp Z_i$$

- The treatment being as good as randomly assigned means that Z_i is independent of potential outcomes and potential treatments.
- Independence implies that **first stage** is average causal effect of Z_i on D_i:

$$\begin{aligned}\mathbb{E}[D_i|Z_i=1] - \mathbb{E}[D_i|Z_i=0] &= \mathbb{E}[D_i(1)|Z_i=1] - \mathbb{E}[D_i(0)|Z_i=0] \\ &= \mathbb{E}[D_i(1) - D_i(0)]\end{aligned}$$

- Independence is sufficient for a causal interpretation of the **reduced form**:

$$\mathbb{E}[Y_i|Z_i=1] - \mathbb{E}[Y_i|Z_i=0] = \mathbb{E}[Y_i(D_i(1),1)] - \mathbb{E}[Y_i(D_i(0),0)]$$

but this does not link the effect to treatment.

LATE Assumptions: Exclusion Restriction

What does exclusion restriction assumption buy us?

$$Y_i(d,1) = Y_i(d,0) \equiv Y_i(d)$$

The exclusion restriction means that Z_i affects Y_i only through D_i.

Technically, we can write Y_i as:

$$\begin{aligned} Y_i &= Y_i(0,z) + (Y_i(1,z) - Y_i(0,z))D_i(z) \\ &= Y_i(0) + (Y_i(1) - Y_i(0))D_i(z) \end{aligned}$$

'Random assignment' and the 'exclusion restriction' should look familiar!

LATE Assumptions: Monotonicity

What does monotonicity assumption buy us?

Observed outcome:

$$\begin{aligned} Y_i &= Y_i(0) + (Y_i(1) - Y_i(0))D_i \\ &= Y_i(0) + (Y_i(1) - Y_i(0))[D(0) + (D_i(1) - D_i(0))Z_i] \\ &= Y_i(0) + (Y_i(1) - Y_i(0))D_i(0) + Z_i(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))] \end{aligned}$$

Then the difference in population means is, using randomization:

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] = \mathbb{E}[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))]$$

Notice that $D_i(1) - D_i(0) \in \{-1, 0, 1\}$.

- For always-takers and never-takers $D_i(1) - D_i(0) = 0$!
  Do not impact mean difference.
- One of the other two, -1 and 1, is ruled out my monotonicity.

LATE Assumptions: Monotonicity

What does monotonicity assumption buy us? Focus on 'no defiers' case.

- The **reduced form** coefficient estimate is:

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$$
$$= \mathbb{E}[Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1] * Pr(D_i(1) - D_i(0) = 1)$$

- The **first-stage** regression estimates, using randomisation:

$$\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = \mathbb{E}[D_i(1) - D_i(0)] = Pr(D_i(1) - D_i(0) = 1)$$

The ratio of these two is the IV/Wald-estimator:

$$\beta^{LATE} = \mathbb{E}[Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1]$$

LATE: Interpretation

$$\beta^{LATE} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i(1) - D_i(0) = 1]$$

IV thus identifiers a ’**Local Average Treatment Effect**’:

- The average effect of treatment in the sub-population whose behaviour was changed because of the value of the instrument, the *compliers*.
- With treatment heterogeneity, this average depends on the instrument.
    1. With > 1 instrument, use to test whether treatment effects are homogenous.
    2. Highlights difference between internal and external validity.

Illustration: Angrist and Evans (AER, 1998)

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

Illustration: Angrist and Evans (AER, 1998)

An aside: **What do you think of the same-sex instrument?**

1. As good as randomly assigned?
   Most likely.
2. Excludable?
   What do you think?
3. Relevant?
   Check the first stage.
4. Monotonicity?
   What do you think?

LATE Intuition: Attempt 1

What is the intuition for the LATE result?

Observed outcome:

$$Y_i = \overbrace{\underbrace{Y_i(0)}_{\text{For never-taker: only this.}} + (Y_i(1) - Y_i(0))D_i(0)}^{\text{For always-taker: only this.}} + Z_i(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))$$

By randomization:

1. Always- and never-takers are just as frequent and have same outcomes in $Z_i = 1$ and $Z_i = 0$ group.
2. When calculating mean-difference $\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$, they drop out.
3. Just need to correct for mean-difference being over all observations, not just compliers: first stage.

LATE Intuition: Attempt 2

What is the intuition for the LATE result?

Without monotonicity:

	$D_i = 0$	$D_i = 1$
$Z_i = 0$	Compliers, Never-takers	Defier, Always-takers
$Z_i = 1$	Defier, Never-takers	Compliers, Always-takers

With monotonicity:

	$D_i = 0$	$D_i = 1$
$Z_i = 0$	Compliers, Never-takers	Always-takers
$Z_i = 1$	Never-takers	Compliers, Always-takers

By Randomization:

1. Know fraction of never-takers and their outcome $\mathbb{E}[Y_i(0)|\text{never taker}]$.
2. Know fraction of compliers. This allows us to back out $\mathbb{E}[Y_i(0)|\text{complier}]$.
3. Same with always-takers and $\mathbb{E}[Y_i(1)|\text{complier}]$.

LATE Intuition: Attempt 3

What is the intuition for the LATE result?

- What is the propensity score of always-takers?
- What is the propensity score of never-takers?
- What is the propensity score of compliers?

For whom can you estimate causal effects?

If you are interested in this interpretation, see
Heckman and Vytlacil (2005, Econometrica) and Zhou and Xie (2019, JPE).

LATE: Monotonicity

Is Monotonicity a sensible assumption? Sometimes.

- A Latent Index model would directly impose monotonicity:

$$D_i(Z_i) = \mathbb{I}(\gamma_0 + \gamma_1 Z_i > v_i). \tag{1}$$

- In many situations it is also a natural assumption:
  Think of Z as assignment to a treatment (e.g. credit access). Few individuals would do the exact opposite **because** they were assigned to a treatment.

Special Case: Bloom (1984) Estimator

Many experiments randomize the *offer* of a treatment.

- As-treated analysis (OLS) is contaminated by selection bias.
- Intention-to-treat (ITT) analyses diluted by non-compliance.

IV solves problem: IV estimand is average treatment effect of the treated (ATT).[2]

$$\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1]} = \frac{\text{ITT Estimate}}{\text{Compliance Rate}} = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$$

[2] Intuitively: *always-takers* are no longer present in $D_i = 1$ group; the treated group can only consist of compliers.)

Plan for Today

Counting Compliers

What fraction of the sample are compliers?

As you saw in Intuition 2, we can count compliers.

- Let $C_i = \mathbb{I}(D_i(1) > D_i(0))$ indicate whether an individual is a complier.
- We can say how many there are: simply check the first-stage coefficient.

$$Pr(C_i = 1) = \mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]$$

Note:
We can pin-point some of the always- and never-takers. How?
This is not true for compliers. Why?

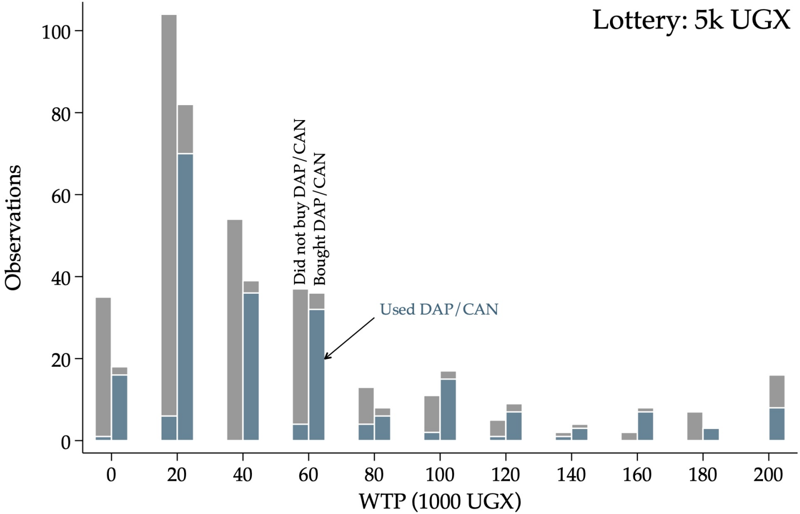
Counting Compliers: Examples

TABLE 4.4.2
Probabilities of compliance in instrumental variables studies

Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	Compliance Probabilities $P[D_1 > D_0 \mid D=1]$ (8)	$P[D_1 > D_0 \mid D=0]$ (9)
Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
		Non-white men born in 1950	.163	.060	.534	.197	.033
More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
	First two children are same sex		.381	.060	.506	.080	.048
High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

From Mostly Harmless Econometrics.

Illustration: own work



Characterizing Compliers

What are characteristics of compliers relative to sample?

Take a characteristic $X_i = x$.

$$
\begin{aligned}
\frac{Pr(X_i = x | C_i = 1)}{Pr(X_i = x)} &= \frac{Pr(C_i = 1 | X_i = x)}{Pr(C_i = 1)} \\
&= \frac{\mathbb{E}[D_i | Z_i = 1, X_i = x] - \mathbb{E}[D_i | Z_i = 0, X_i = x]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]}
\end{aligned}
$$

This is useful to generalise from sample to other population.

Illustration: Angrist and Evans (AER, 1998)

Table 21: Complier analysis

	A. # compliers			
	$\Pr(D_i = 1)$	$\Pr(Z_i = 1)$	$\Pr(C_i = 1)$	$\Pr(C_i = 1 \mid D_i = 1)$
	0.378	0.506	0.068	0.091
	(0.001)	(0.001)	(0.002)	(0.003)

	B. Complier characteristics		
	$\Pr(X_i = 1)$	$\Pr(X_i = 1 \mid C_i = 1)$	$\Pr(X_i = 1 \mid C_i = 1) / \Pr(X_i = 1)$
Age at 1st birth≥ 25	0.124	0.096	0.772
	(0.001)	(0.009)	(0.073)
Black	0.0502	0.038	0.761
	(0.0004)	(0.007)	(0.135)
Yrs. of schooling≥ 12	0.487	0.513	1.054
	(0.001)	(0.025)	(0.051)
Yrs. of schooling≥ 16	0.135	0.0940	0.698
	(0.001)	(0.010)	(0.077)

Notes: Standard errors in parentheses are robust. Data are from the 1980 US Census (PUMS) and correspond (almost) to the sample used in Angrist and Evans (1998). SE:s in Panel A, col. (4), and Panel B, cols. (2) and (3), are calculated using the delta method.

From Peter Fredrikssom's old slides.

Mean Potential Outcomes

Mean complier potential outcomes are also identified. (See Intuition 2.)

	$D_i = 0$	$D_i = 1$
$Z_i = 0$	Compliers, Never-takers	Always-takers
$Z_i = 1$	Never-takers	Compliers, Always-takers

Let , π^j, $j = a, n, c$, denote shares of always- takers, never-takers, and compliers.

With randomisation of Z_i they are identified:

- $\pi^a = \mathbb{E}(D_i = 1 | Z_i = 0)$
- $\pi^n = \mathbb{E}(D_i = 0 | Z_i = 1)$
- $\pi^c = \mathbb{E}(D_i = 1 | Z_i = 1) - \mathbb{E}(D_i = 1 | Z_i = 0)$

Mean Potential Outcomes

- Define the identified quantity

$$\mu_{wz} = \mathbb{E}[Y_i | D_i = d, Z_i = z].$$

- Then

$$\mu_{10} = \mathbb{E}[Y_i(1) | \text{Always-taker}], \quad \mu_{01} = \mathbb{E}[Y_i(0) | \text{Never-taker}]$$

$$\mu_{11} = \frac{\pi^c}{\pi^c + \pi^a} \mathbb{E}[Y_i(1) | \text{Complier}] + \frac{\pi^a}{\pi^c + \pi^a} \mathbb{E}[Y_i(1) | \text{Always-taker}]$$

$$\mu_{00} = \frac{\pi^c}{\pi^c + \pi^n} \mathbb{E}[Y_i(0) | \text{Complier}] + \frac{\pi^n}{\pi^c + \pi^n} \mathbb{E}[Y_i(1) | \text{Never-taker}].$$

- Solve for the mean potential outcomes of compliers:

$$\mathbb{E}[Y_i(0) | \text{Complier}] = \frac{\pi^c + \pi^n}{\pi^c} \mu_{00} - \frac{\pi^n}{\pi^c} \mu_{01}$$

$$\mathbb{E}[Y_i(1) | \text{Complier}] = \frac{\pi^c + \pi^a}{\pi^c} \mu_{11} - \frac{\pi^a}{\pi^c} \mu_{10}$$

Potential Outcomes Distribution

More generally true for distribution of potential outcomes of compliers!

- Let $g_{c0}(y)$ and $g_{c1}(y)$ be distributions of $Y_i(0)$ and $Y_i(1)$ amongst compliers.
- Let $f_{dz}(y)$ denote the directly observed distribution when $D_i = d$ and $Z_i = z$.

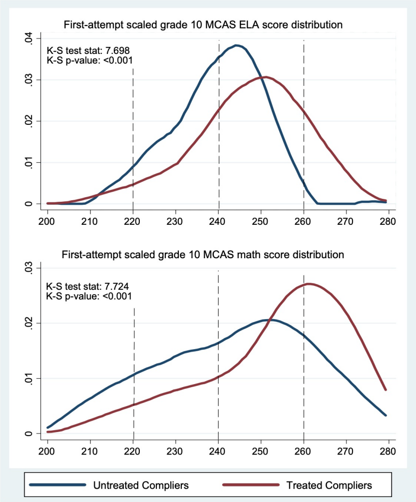
Then

$$g_{c0}(y) = \frac{\pi^c + \pi^n}{\pi^c} f_{00}(y) - \frac{\pi^n}{\pi^c} f_{01}(y)$$
$$g_{c1}(y) = \frac{\pi^c + \pi^a}{\pi^c} f_{11}(y) - \frac{\pi^a}{\pi^c} f_{10}(y)$$

(Imbens and Rubin, 1997; see Abadie, 2002 for convenient alternative.)

Potential Outcomes Distribution

Figure 1: Complier Distributions for MCAS Scaled Scores



From Angrist's lecture notes (MIT 14.387)

Plan for Today

LATE with Multiple Instruments

Scenario: two (mutually exclusive) binary instruments Z_{1i} and Z_{2i}.

- Each instrument can be used to construct separate Wald estimates.
  Both Wald estimates have a LATE interpretation, although the complier population generally differs across the two instruments.
- Alternatively, two instruments can be combined into a single 2SLS estimate.

  $\rightarrow$ This is a weighted average of the underlying Wald estimates.

LATE with Multiple Instruments

- Note that LATE defined by each separate instrument is given by

$$\beta_j = \frac{Cov(Y_i, Z_{ji})}{Cov(D_i, Z_{ji})}, j = 1, 2$$

  when instruments are mutually exclusive.[3]

- Denote the first stage as:

$$\hat{D}_i = \gamma_1 Z_{1i} + \gamma_2 Z_{2i}.$$

- The 2SLS estimand is given by:

$$\beta_{2SLS} = \frac{Cov(Y_i, \hat{D}_i)}{Var(\hat{D}_i)} = \gamma_1 \frac{Cov(Y_i, Z_{1i})}{Cov(D_i, \hat{D}_i)} + \gamma_2 \frac{Cov(Y_i, Z_{2i})}{Cov(D_i, \hat{D}_i)}.$$

[3] See Mogstad, Torgovitsky, Walters, AER, 2021.

LATE with Multiple Instruments

$$\beta_{2SLS} = \frac{Cov(Y_i, \hat{D}_i)}{Var(\hat{D}_i)} = \gamma_1 \frac{Cov(Y_i, Z_{1i})}{Cov(D_i, \hat{D}_i)} + \gamma_2 \frac{Cov(Y_i, Z_{2i})}{Cov(D_i, \hat{D}_i)}$$

Use the definition of β_j:

$$\beta_{2SLS} = \gamma_1 \frac{Cov(D_i, Z_{1i})}{Cov(D_i, \hat{D}_i)} \beta_1 + \gamma_2 \frac{Cov(D_i, Z_{2i})}{Cov(D_i, \hat{D}_i)} \beta_2$$

or

$$\beta_{2SLS} = \varphi \beta_1 + (1 - \varphi) \beta_2,$$

where

$$\varphi = \gamma_1 \frac{Cov(D_i, Z_{1i})}{Cov(D_i, \hat{D}_i)} = \gamma_1 \frac{Cov(D_i, Z_{1i})}{\gamma_1 Cov(D_i, Z_{1i}) + \gamma_2 Cov(D_i, Z_{2i})}$$

Thus 2SLS produces a weighted average of LATE(Z_1) and LATE(Z_2).

LATE with Covariates

A fully saturated first stage, and saturated in covariates second stage produce a weighted average of covariate-specific LATEs:

$$\beta_{2SLS} = \mathbb{E}[\omega(X_i)\beta^X_{LATE}]$$

where

$$\beta^X_{LATE} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i, C_i = 1]$$

and

$$\omega(X_i) = \frac{Var(\mathbb{E}[D_i|Z_i, X_i]|X_i)]}{\mathbb{E}[Var(\mathbb{E}[D_i|Z_i, X_i]|X_i)]}$$

Remember the regression/matching comparison in Lecture 5?

- Regression: covariate-specific weights were conditional variance of D given X.
- Here, variability in variance term comes from Z. More weight to covariate terms where the instrument creates more variation in fitted values.

Average Causal Response

Consider a multi-valued treatment.

- Define the average causal response function as

$$Y_i(S) \equiv f_i(S),$$

  defines potential outcomes for each individual at any treatment value.
- Take case with binary instrument:
  $\rightarrow$ 2SLS provides weighted average of unit causal effects.

Average Causal Response: Assumptions

1. Randomization and exclusion: $\{Y_i(0), Y_i(1), ..., Y_i(\bar{S}), S_i(1), S_i(0)\} \perp Z_i$
2. Relevance: $\mathbb{E}[S_i(1) - S_i(0)] \neq 0]$
3. Monotonicity: $S_i(1) \geq S_i(0) \quad$ (or $S_i(1) \leq S_i(0)$)

Then

$$\beta_{2SLS} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[S_i|Z_i = 1] - \mathbb{E}[S_i|Z_i = 0]}$$
$$= \sum_{s=1}^{S} \omega_s \mathbb{E}[Y_i(s) - Y_i(s-1)|S_i(1) \geq s \geq S_i(0)]$$

where

$$\omega_s = \frac{Pr(S_i < s|Z_i = 0) - Pr(S_i < s|Z_i = 1)}{\mathbb{E}[S_i|Z_i = 1] - \mathbb{E}[S_i|Z_i = 0]}, \text{ and } \sum_s \omega_s = 1.$$

and $\mathbb{E}[Y_i(s) - Y_i(s-1)|S_i(1) \geq s \geq S_i(0)]$ is the unit causal effect: average difference in potential outcomes for compliers at point s.

Questions?