Variance Weights and Heterogeneous Effects

Gov 2001: Quantitative Social Science Methods I

Week 13, Lecture 25

Spring 2026

For Today

Required Reading

- Angrist (1998): “Estimating the Labor Market Impact...”
- Słoczyński (2022): “Interpreting OLS Estimands...”

Today: What does OLS estimate when treatment effects vary?

Roadmap

1. Heterogeneous treatment effects
2. OLS with a binary treatment
3. Variance weighting
4. When OLS $\neq$ ATE
5. Implications for applied work

Part I: Heterogeneous Treatment Effects

The Constant Effects Assumption

So far, we often assumed:

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

where β is the **same for everyone**.

Interpretation: Moving from $D = 0$ to $D = 1$ increases Y by β for all individuals.

This is often called the "constant treatment effect" assumption.

Treatment Effect Heterogeneity

Reality: Effects often vary across individuals.

Examples:

- GOTV campaigns work better in some districts than others
- Campaign ads have larger effects in competitive races
- Foreign aid impacts vary by regime type

Let β_i = treatment effect for individual i.
Different individuals have different β_i's.

What We Might Want to Estimate

Average Treatment Effect (ATE):

$$\tau_{ATE} = \mathbb{E}[\beta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$$

Average effect across the whole population.

Average Treatment Effect on the Treated (ATT):

$$\tau_{ATT} = \mathbb{E}[\beta_i | D_i = 1]$$

Average effect among those who actually receive treatment.

With heterogeneous effects, $ATE \neq ATT$ in general.

The Key Question

When treatment effects are heterogeneous:

What does OLS estimate?

Is it ATE? ATT? Something else?

Spoiler: OLS estimates a *variance-weighted* average of effects—which may be none of the above.

Part II: OLS with a Binary Treatment

The Saturated Model

Consider the simplest case:

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

where $D_i \in \{0, 1\}$ (binary treatment).

OLS gives:

$$\hat{\beta} = \bar{Y}_1 - \bar{Y}_0$$

Difference in means between treated and control.

With random assignment (and constant effects), this is unbiased for β.

With Control Variables

Now add controls:

$$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i$$

Question: What does $\hat{\beta}$ estimate now?

FWL tells us:

$$\hat{\beta} = \frac{\sum_i \tilde{D}_i Y_i}{\sum_i \tilde{D}_i^2}$$

where $\tilde{D}_i$ = residual from regressing D on X.

$\hat{\beta}$ uses variation in D that is **orthogonal to** X.

Part III: Variance Weighting

Angrist (1998): Variance Weights

Key insight: OLS puts more weight on cells with more "residual variance" in treatment.

The formula:

$$\hat{\beta} = \sum_x w_x \cdot \hat{\beta}_x$$

where:

- $\hat{\beta}_x$ = effect estimate for subgroup with $X = x$
- $w_x \propto p_x(1 - p_x) \cdot n_x$
- p_x = proportion treated in cell x

The Variance Weights

Weight for cell x:

$$w_x = \frac{n_x \cdot p_x(1-p_x)}{\sum_{x'} n_{x'} \cdot p_{x'}(1-p_{x'})}$$

$p_x(1-p_x)$ is maximized when $p_x = 0.5$!

Cells with 50-50 treatment split get the most weight.
Cells with p_x near 0 or 1 get almost no weight.

Why Variance Weights?

Intuition:
OLS identifies β from variation in D.

- If everyone in cell x is treated ($p_x = 1$), no variation $\Rightarrow$ no information
- If no one in cell x is treated ($p_x = 0$), no variation $\Rightarrow$ no information
- Maximum variation when $p_x = 0.5$

OLS naturally uses information-rich cells more.

Example: Two Cells

Suppose two cells with different treatment probabilities:

Cell	n_x	p_x	$p_x(1-p_x)$
A	100	0.50	0.25
B	100	0.90	0.09

Cell A gets weight: $\frac{100\times0.25}{100\times0.25+100\times0.09} = 0.74$
Cell B gets weight: $\frac{100\times0.09}{100\times0.25+100\times0.09} = 0.26$

Cell A gets nearly 3x the weight, even with same n.

Part IV: When OLS ≠ ATE

When Does This Matter?

If treatment effects are **constant** ($\beta_i = \beta$ for all i):
$\Rightarrow$ OLS = ATE (regardless of weights)

If treatment effects are **heterogeneous**:
$\Rightarrow$ OLS = weighted average of cell-specific effects
$\Rightarrow$ Weights depend on treatment propensities
$\Rightarrow$ May differ from ATE

Słoczyński (2022): A Decomposition

With controls and heterogeneous effects, OLS estimates:

$$\hat{\beta} = \omega \cdot \tau_{ATT} + (1 - \omega) \cdot \tau_{ATC}$$

where:

- τ_{ATT} = Average Treatment Effect on Treated
- τ_{ATC} = Average Treatment Effect on Controls
- ω = proportion of treated in the comparison "overlap" region

OLS is a **convex combination** of ATT and ATC.
But the weights may not be what you want!

An Extreme Case

Suppose:

- Half the population has $p_x = 0.9$ (effect = +10)
- Half the population has $p_x = 0.1$ (effect = −10)

True ATE: $0.5 \times 10 + 0.5 \times (-10) = 0$

OLS estimate:

- Group with $p = 0.9$ gets weight 0.09
- Group with $p = 0.1$ gets weight 0.09
- Weights are equal! $\Rightarrow$ OLS ≈ 0

In this case, OLS happens to get ATE right.

Another Example

Suppose:

- Half the population has $p_x = 0.5$ (effect = +10)
- Half the population has $p_x = 0.9$ (effect = +2)

True ATE: $0.5 \times 10 + 0.5 \times 2 = 6$

OLS weights:

- $p = 0.5$ group: weight $\propto 0.25$
- $p = 0.9$ group: weight $\propto 0.09$

OLS weight: $\frac{0.25}{0.34} \approx 0.74$ on first group.
OLS estimate: $0.74 \times 10 + 0.26 \times 2 \approx 7.9$
OLS **overstates** the ATE!

Part V: Implications

When Should You Worry?

Worry more when:

1. Treatment propensities vary a lot across cells
2. Treatment effects vary a lot across cells
3. Propensity and effect variation are correlated

Worry less when:

1. Propensities are similar across cells (all near 0.5)
2. Effects are similar across cells (homogeneity)
3. You're doing a randomized experiment with balanced assignment

What Can You Do?

1. Understand the weights
Calculate and report variance weights for your regression.

2. Re-weight if needed
Inverse propensity weighting can recover ATE.

3. Consider the estimand
Maybe variance-weighted effect *is* what you want.
(Focuses on where there's most "action")

4. Report heterogeneity
Show effects by subgroup; don't just report one number.

Summary

Key points:

1. Treatment effects often vary across individuals
2. OLS uses **variance weights**—cells with $p \approx 0.5$ count more
3. With heterogeneous effects, OLS may differ from ATE
4. OLS estimates a convex combination of ATT and ATC
5. Understanding weights helps interpret what OLS gives you

Looking Ahead

Next lecture: Regression Adjustment and Causality

- Conditional independence assumption
- When regression gives causal estimates
- Preview of causal inference methods

OLS uses variance weights:

Cells with 50-50 treatment split get most weight.

With heterogeneous effects,
OLS may not equal the ATE.

Understand what your regression is estimating.