# Williams College ECON 523:

# Program Evaluation for International Development

## **Lecture 2: Regression**

Professor: Pamela Jakiela

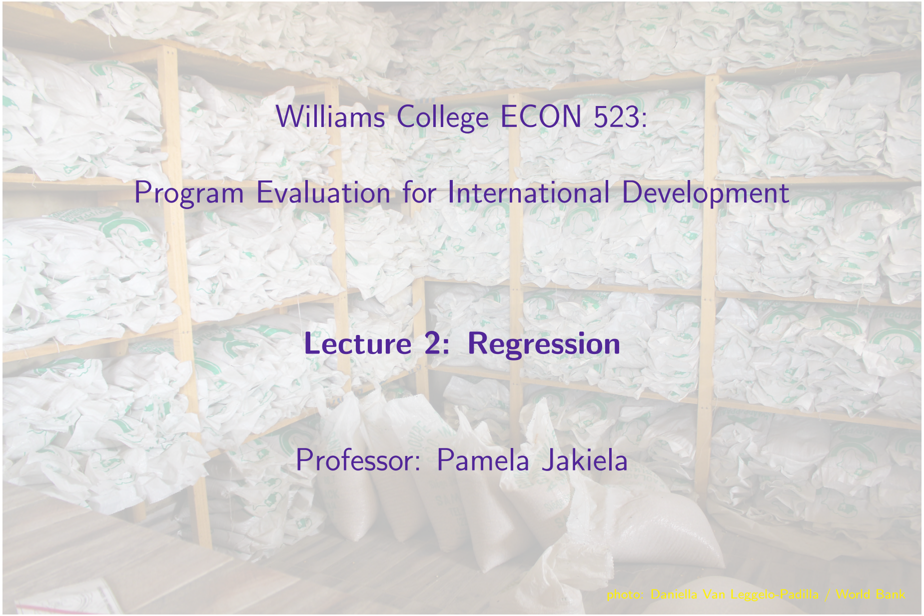
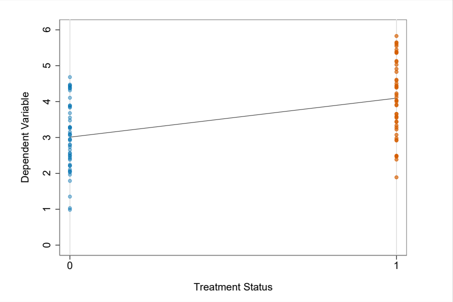


photo: Daniella Van Leggelo-Padilla / World Bank

# Dummy Variables

# OLS Regression on a Binary Independent Variable

$$Y = \alpha + \beta D$$




| control | treatment |
|---|---|
| $D = 0$ | $D = 1$ |
| $\hat{\alpha}$ | $\hat{\alpha} + \hat{\beta}$ |

$\hat{\alpha} = \bar{Y}_C$ (control group mean)

$\hat{\beta} = \bar{Y}_T - \bar{Y}_C$ (difference in means)

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$\begin{aligned}\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} \\ &= \frac{\sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_i \left(X_i - \bar{X}\right)^2}\end{aligned}$$

Notice that the numerator can be re-organized:

$$\sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \;=\; \sum_i X_i Y_i \;-\; \sum_i \bar{X} Y_i$$

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$\hat{\beta}_{OLS} = \frac{COV(X,Y)}{VAR(X)}$$

$$= \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2}$$

Notice that the numerator can be re-organized:

$$\sum_i (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_i X_i Y_i - \sum_i \bar{X} Y_i$$

$$= \sum_i [Y_i (X_i - \bar{X})]$$

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$\hat{\beta}_{OLS} = \frac{COV(X,Y)}{VAR(X)} = \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2}$$

When independent variable is binary:

$\bar{X} = \frac{n_T}{N}$ ($n_T$ is # of treated observations)

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$\hat{\beta}_{OLS} = \frac{COV(X,Y)}{VAR(X)} = \frac{\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2}$$

When independent variable is binary:

$\bar{X} = \frac{n_T}{N}$ ($n_T$ is # of treated observations)

Assume observations are ordered:

$$\{\underbrace{Y_1, Y_2, \ldots, Y_{n_T - 1}, Y_{n_T}}_{\text{treatment group}}, \underbrace{Y_{n_T+1}, Y_{n_{T+2}}, \ldots, Y_N}_{\text{control group}}\}$$

Treatment group: $X_i = 1$ $\Rightarrow X_i - \bar{X} = 1 - \bar{X}$

Control group: $X_i = 0$ $\Rightarrow X_i - \bar{X} = -\bar{X}$

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} \\
&= \frac{\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)}
\end{aligned}
$$

Re-write denominator:

$$
\begin{aligned}
\sum_i \left(X_i - \bar{X}\right)^2 &= \sum_{i=1}^{n_T} \left(1 - \bar{X}\right)^2 + \sum_{i=n_{T+1}}^{N} \left(-\bar{X}\right)^2 \\
&= n_T \left(1 - \bar{X}\right)^2 + \left(N - n_T\right)\left(-\bar{X}\right)^2 \\
&= \ldots = n_T - n_T\bar{X} = N\bar{X}\left(1 - \bar{X}\right)
\end{aligned}
$$

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$\begin{aligned}
\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)}
\end{aligned}$$

$\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right]$

"linear combination of $Y$s"

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} & & \textstyle\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right] \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2} & &= \textstyle\sum_{i=1}^{n_T} \left[Y_i \left(1 - \bar{X}\right)\right] + \sum_{i=n_T+1}^{N} \left[Y_i \left(-\bar{X}\right)\right] \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)} & &= \textstyle\sum_{i=1}^{n_T} Y_i - \sum_{i=1}^{N} \left[Y_i \left(\bar{X}\right)\right] \\
& & &= n_T \bar{Y}_T - \bar{X}\left(N\bar{Y}\right) \\
& & &= N\bar{X}\bar{Y}_T - N\bar{X}\left[\bar{X}\bar{Y}_T + \left(1 - \bar{X}\right)\bar{Y}_C\right] \\
& & &= N\bar{X}\left(1 - \bar{X}\right)\left(\bar{Y}_T - \bar{Y}_C\right)
\end{aligned}
$$

# OLS Regression on a Binary Independent Variable

You may or may not remember that in a bivariate regression:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)} \\
&= \frac{N\bar{X}\left(1 - \bar{X}\right)\left(\bar{Y}_T - \bar{Y}_C\right)}{N\bar{X}\left(1 - \bar{X}\right)}
\end{aligned}
$$

$$
\begin{aligned}
\sum_i \left[Y_i \left(X_i - \bar{X}\right)\right] &= \sum_{i=1}^{n_T} \left[Y_i \left(1 - \bar{X}\right)\right] + \sum_{i=n_T+1}^{N} \left[Y_i \left(-\bar{X}\right)\right] \\
&= \sum_{i=1}^{n_T} Y_i - \sum_{i=1}^{N} \left[Y_i \left(\bar{X}\right)\right] \\
&= n_T \bar{Y}_T - \bar{X}\left(N\bar{Y}\right) \\
&= N\bar{X}\bar{Y}_T - N\bar{X}\left[\bar{X}\bar{Y}_T + \left(1 - \bar{X}\right)\bar{Y}_C\right] \\
&= N\bar{X}\left(1 - \bar{X}\right)\left(\bar{Y}_T - \bar{Y}_C\right)
\end{aligned}
$$

# OLS Regression on a Binary Independent Variable

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \frac{COV(X,Y)}{VAR(X)} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)} \\
&= \frac{N\bar{X}\left(1 - \bar{X}\right)\left(\bar{Y}_T - \bar{Y}_C\right)}{N\bar{X}\left(1 - \bar{X}\right)} \\
&= \bar{Y}_T - \bar{Y}_C
\end{aligned}
$$

# OLS Regression on a Binary Independent Variable

When we regress $Y_i$ on (only) a dummy variable:

$$\hat{\beta}_{OLS} = \bar{Y}_T - \bar{Y}_C$$

- Estimated constant $\hat{\alpha}_{OLS}$ is control group mean, also $\hat{Y}_i$
- Predicted $\hat{Y}_i$ for treated individuals/units is $\hat{\alpha}_{OLS} + \hat{\beta}_{OLS}$

# OLS Regression on Mutually Exclusive Dummy Variables

$$Y = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3$$

| control | treatment 1 | treatment 2 | treatment 3 |
|---|---|---|---|
| $T_1 = T_2 = T_3 = 0$ | | | |
| | | | |

# OLS Regression on Mutually Exclusive Dummy Variables

$$Y = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3$$

| control | treatment 1 | treatment 2 | treatment 3 |
|---|---|---|---|
| $T_1 = T_2 = T_3 = 0$ | $T_1 = 1$ | $T_2 = 1$ | $T_3 = 1$ |
| $\hat{\alpha}$ | $\hat{\alpha} + \hat{\beta}_1$ | $\hat{\alpha} + \hat{\beta}_2$ | $\hat{\alpha} + \hat{\beta}_3$ |

$\hat{\alpha} = \bar{Y}_C$ (control group mean)

$\hat{\beta}_i = \bar{Y}_{T_i} - \bar{Y}_C$ (difference in means between treatment $i$ and control)

# Pooling Treatments

- If we pool treatments to estimate an average effect across treatment arms:
  - Estimated coefficient (and treatment effect) $\beta_{pooled}$ is average of impacts across treatments
  - Average depends on $n_{T_i}$ values: share of treated observations in each treatment arm
  - Also depends on treatment effect of each arm (pooling arms with no impact will matter)
- Estimates of pooled effect more precise because sample size is larger
  - When is pooled effect policy relevant?

# Cross-Cutting Designs

- We often want to estimate the impact of treatments that may work best together
  - Access to credit and vocational training for unemployed youth
  - Nutrition supplements and parenting education for at-risk babies and children
  - Increased enforcement and information campaigns or behavioral nudges for tax compliance
  - Teacher training and additional materials for under-performing schools
  - Management consulting and subsidies for exporting firms
- Cross-cutting designs allow us to estimate impacts of each treatment in isolation as well as the pooled impact, to see whether any program effects are additively separable

# Cross-Cutting Designs

| | control | treatment 2 |
|---|---|---|
| **control** | C | T2 |
| **treatment 1** | T1 | T1×T2 |

# Cross-Cutting Designs and Interaction Terms

$$Y = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 (T_1 \times T_2)$$

| | |
|---|---|
| C | T2 |
| T1 | T1×T2 |

# Challenge Problem: Triple Interactions

$$Y = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \gamma_1 T_1 \times T_2 + \gamma_2 T_2 \times T_3 + \gamma_3 T_1 \times T_3 + \theta T_1 \times T_2 \times T_3$$

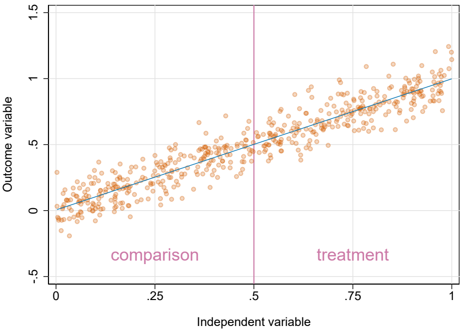
# Continuous Variables

# When Is Treatment a Continuous Variable?

- Sometimes we vary treatment intensity across treatment arms
  - Subsidies for malaria treatment (Cohen, Dupas, and Schaner 2015)
  - Varying the size of grants to entrepreneurs/firms, schools, etc.
  - Proportion treated within clusters (CCTs, job training, etc.)
- Binary treatments might also impact units differently, based on pre-existing conditions
  - Law banning traditional birth attendants in Malawi (Godlonton and Okeke 2016)
  - Impact of eliminating primary school fees on completion (cf. Lucas and Mbiti 2012)
- Should we dichotomize treatment variable or exploit continuous variation in intensity?

# Continuous Variation in Treatment

$$Y_i = X_i + \varepsilon$$

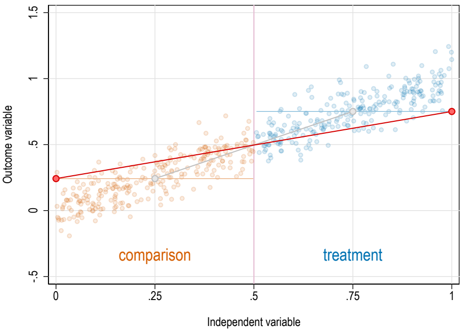



| | (1) OLS |
|---|---|
| Treatment Intensity | 0.994*** |
| | (0.016) |
| Constant | 0.005 |
| | (0.009) |

Standard errors in parentheses.

# Continuous Variation in Treatment

$$Y_i = X_i + \varepsilon$$



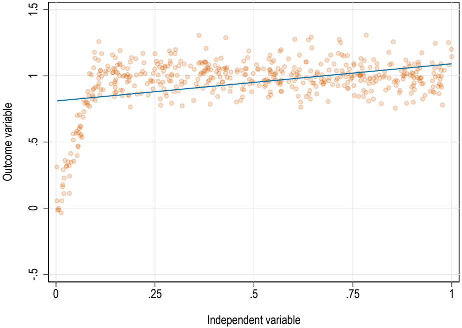

| | (1) |
|---|---|
| | OLS |
| Treatment Dummy | 0.509*** |
| | (0.016) |
| Constant | 0.242*** |
| | (0.011) |

Standard errors in parentheses.

# Dichotomous vs. Continuous Treatment Variables

- When the true dose-response relationship is linear:
  - OLS w/ a continuous treatment variable is correctly specified
  - Uses observed variation increase statistical power (odds of finding an effect)
- The estimand is different when we dichotomize treatment
  - OLS coefficient captures impact of moving from average level of treatment intensity in the control group to average level of treatment intensity in the treatment group
    - Not the same as impact of moving from treatment intensity 0 to treatment intensity 1
- When true dose-response relationship is linear, OLS with continuous $X$ is better

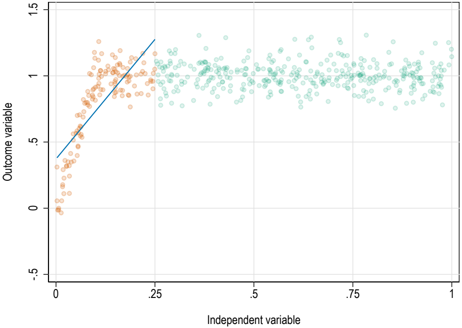
# OLS when the Dose-Response Relationship Is Not Linear




| | (1) OLS |
|---|---|
| Treatment Intensity | 0.281*** |
| | (0.029) |
| Constant | 0.810*** |
| | (0.017) |

Standard errors in parentheses.

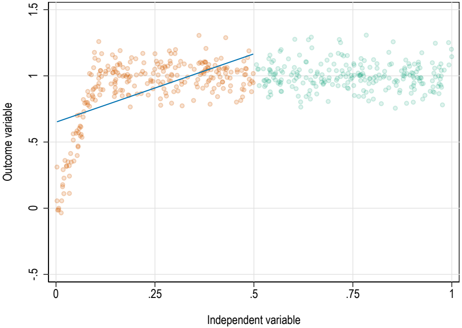
# OLS when the Dose-Response Relationship Is Not Linear




| | (1) OLS |
|---|---|
| Treatment Intensity | 3.614*** |
| | (0.257) |
| Constant | 0.372*** |
| | (0.034) |

Standard errors in parentheses.

# OLS when the Dose-Response Relationship Is Not Linear




| | (1) OLS |
|---|---|
| Treatment Intensity | 1.032*** |
| | (0.092) |
| Constant | 0.650*** |
| | (0.026) |

Standard errors in parentheses.

# Continuous Variation in Treatment Intensity

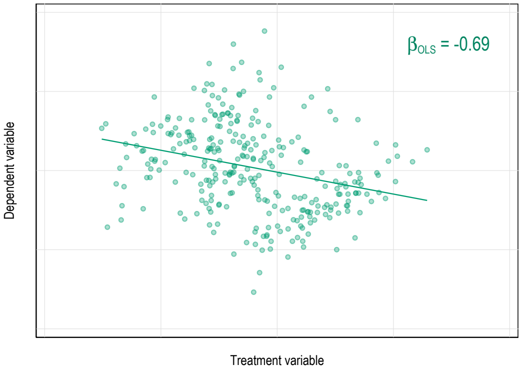
- When the true relationship (i.e. conditional expectation function) is **not** linear:
  - OLS w/ a continuous treatment variable is incorrectly specified (may or may not matter)
  - Estimated treatment effect (i.e. coefficient) depends on choice of sample (values of $X$)
- Graph your data (though often true relationship obscured by noise)
  - Choose to dichotomize (and where to dichotomize)
  - Vary your sample to assess the robustness of your estimates
- Be skeptical of results when treatment assignment process is unclear (observational data!) and you cannot observe the relevant empirical relationships in your data graphically

# Fixed Effects

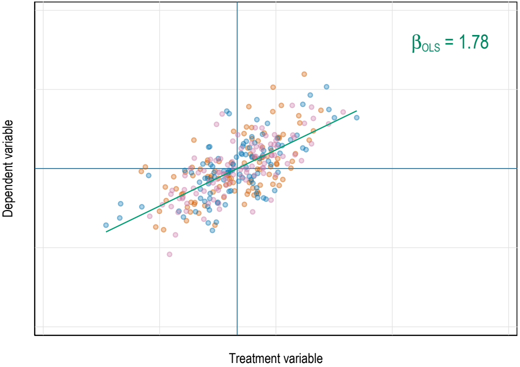
# What Are Fixed Effects?

- Individual dummy variables for mutually exclusive groups in your data
  - Dummy for male or female
  - Age, age group, or year of birth fixed effects
  - Continent/country/state/district/geography fixed effects
  - Year/month/time fixed effects
- Why use fixed effects?
  - Estimation using **within** rather than **between** variation
- We often use multiple sets of fixed effects in empirical work

# Simpson's Paradox



# What Do Fixed Effects Do?




Equivalent regressions:

- $Y = \alpha + \theta X + \beta Orange + \gamma Pink$
- $\tilde{Y} = \alpha + \theta \tilde{X}$

where:

- $\tilde{Y} = Y - \bar{Y}_{within}$
- $\tilde{X} = X - \bar{X}_{within}$

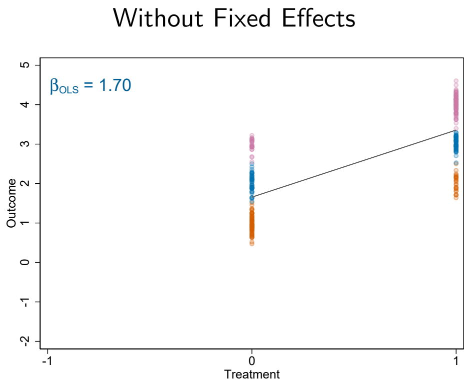
# What Do Fixed Effects Do?

- OLS with fixed effects is equivalent to:
  - Transforming independent and dependent variables by subtracting off the within-group mean
  - Running OLS in your transformed (i.e. re-centered, normalized, de-meaned) data
- Fixed effects changed the coefficient estimate because treatment ($X$) varied across groups
  - When $X$ doesn't vary, FEs can improve precision but won't change slope estimate
  - When randomized/exogenous treatment probability/intensity varies across groups, you must include fixed effects or control directly for probability of treatment (the propensity score)
- If you regress $X$ and $Y$ on only the fixed effects, residuals are the normalized variables
  - Remember: predicted values of $Y$ are the within-group means

# Binary Treatment, Homogeneous Treatment Effects: Example

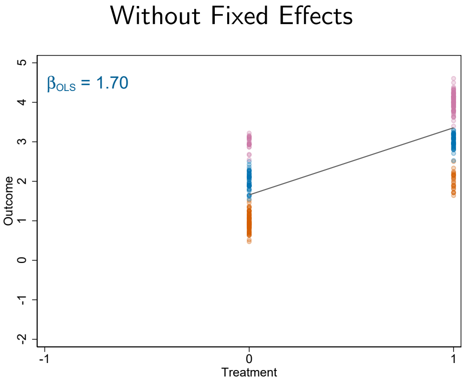
- Consider the following setup:
  - $N = 300$
  - Three groups: $N_1 = N_2 = N_3 = 100$
  - Treatment probability: 25, 50, 75 percent in Groups 1, 2, and 3, respectively
  - Data-generating process: $Y_i = G_i + D_i + \varepsilon_i$
- In this setting?
  - What is the (real) treatment effect?
  - What is the mean in the control group?
  - What is the mean in the treatment group?

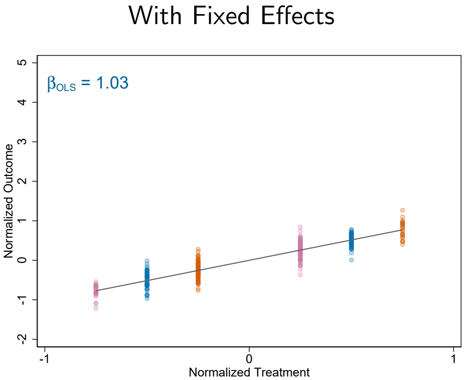
# Binary Treatment, Homogeneous Treatment Effects: Example




- OLS coefficient is $\bar{Y}_T - \bar{Y}_T$
- Biased estimate of ATE because:
  - $\bar{T}$ varies across groups
  - $\bar{Y}$ varies across groups

# Binary Treatment, Homogeneous Treatment Effects: Example



# Binary Treatment, Homogeneous Treatment Effects: Example

Group 1 (orange): $\bar{T}_{orange} = 0.25$

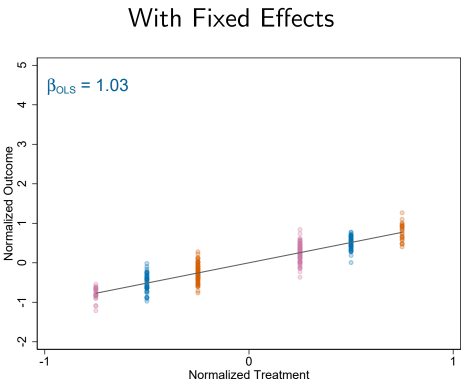
- Treated observations weighted heavily
- Contribute (relatively) more to $\bar{Y}_{T,G=1}$

Group 2 (blue): $\bar{T}_{orange} = 0.5$

- Treated, untreated weighted equally

Group 3 (pink): $\bar{T}_{orange} = 0.75$

- Untreated observations weighted heavily




Coefficient on treatment is a linear combination of the within-group treatment effects, $\bar{Y}_{T,G} - \bar{Y}_{C,G}$

# Fixed Effects with Binary Treatment

- Group-level estimates of treatment effect are not weighted equally
  1. Weights are proportional to $N_G p_G (1 - p_G)$
  2. Weights are normalized by the sum of all country-specific weights
  3. Countries with treatment probability close to one half receive more weight
- Individual observations are also not weighted equally (conditional on $N_G p_G (1 - p_G)$)
  1. Treated units in countries with **few** treated units get **more** weight
  2. Untreated units in countries with **few** treated units get **less** weight
  3. Treated units in countries with **many** treated units get **less** weight
  4. Untreated units in countries with **many** treated units get **more** weight
- OLS coefficient captures ATE across entire sample, not across treated observations

# The Frisch-Waugh-Lovell Theorem

$$Y = \alpha + \beta X + \gamma Z$$

is equivalent to

$$\tilde{Y} = \alpha + \beta \tilde{X}$$

where

$\tilde{Y}$ = residuals from regressing $Y$ on $Z$

$\tilde{X}$ = residuals from regressing $X$ on $Z$

# The Frisch-Waugh-Lovell Theorem

$$Y = \alpha + \beta X + \gamma_1 Z_1 + \ldots \gamma_K Z_K$$

is equivalent to

$$\tilde{Y} = \alpha + \beta \tilde{X}$$

where

$\tilde{Y}$ = residuals from regressing $Y$ on $Z_1, \ldots, Z_K$

$\tilde{X}$ = residuals from regressing $X$ on $Z_1, \ldots, Z_K$

# Frisch-Waugh-Lovell Example: The Impact of Education

- These approaches generate identical estimates of $\beta_{educ}$, the estimated impact of education:

  1. Regress income on education, age dummies, and a dummy for being female
  2. Transform income and education by subtracting off age-specific means, then transform the transformed variables by subtracting off gender-specific means (of transformed variables), then regress transformed income on transformed education
  3. Generate residualized income and education by regressing those variables on the age and gender FEs, then regress residualized income on residualized education (without controls)
  4. Residualize income and education on age dummies first, then gender dummies, then regress

# Frisch-Waugh-Lovell: Why It Matters

- When treatment is binary and plausibly exogenous, the difference in outcome means between the treatment and comparison groups provides an unbiased estimate of impact
  - All treated observations get equal positive weight, all untreated get equal negative weight
- With controls that are correlated with treatment, treatment is (in effect) no longer binary
  - Untreated observations with covariates that predict a high likelihood of treatment get very low negative weights in linear regression; while treated observations with covariates that predict a low likelihood treatment get very high positive weights in multivariate regression
  - Everything is still fine if treatment effects are homogeneous: treatment effect is the same for everyone, so it doesn't matter what weights we use to calculate average treatment effect
  - If effects vary with covariates that predict treatment, mis-specification problems can arise
- What to do: show results with and without covariates, residualize and then plot your data

# Empirical Exercise

# The Graduation Approach

**THE GRADUATION APPROACH**

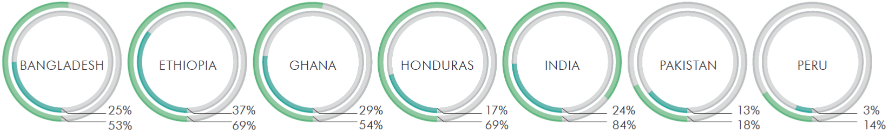
The Graduation approach consists of six complementary components, each designed to address specific constraints facing ultra-poor households.

1. **Productive asset transfer:** One-time transfer of productive assets, such as cows, goats, or supplies for petty trade.
2. **Technical skills training:** Training to manage the productive asset.
3. **Consumption support:** Regular cash or food support for a few months to a year.
4. **Savings:** Access to a savings account, or encouragement to save.
5. **Home visits:** Frequent home visits by implementing partner staff to provide accountability, coaching, and encouragement.
6. **Health:** Health education, health care access, and/or life skills training.

All evaluations in this bulletin include these six components; see Table 1 for country-by-country variation in program design.

# The Graduation Approach: Targeting the Ultra-Poor

FIGURE 1: THE GRADUATION PROGRAM SUCCESSFULLY TARGETED ULTRA-POOR HOUSEHOLDS

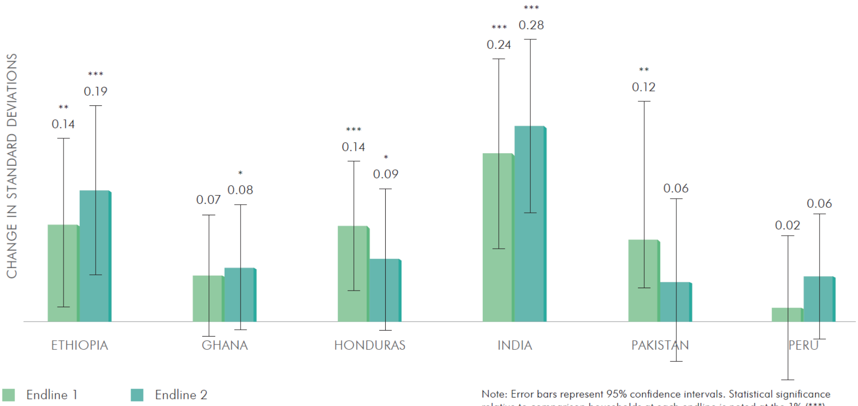



The program successfully targeted ultra-poor households. The proportion of households living below US$1.25 per day that were identified as eligible for the program and included in the study sample exceeded—often by a substantial margin—the proportion of the population living below US$1.25 per day in every country.

- Proportion of households in the program living below PPP US$1.25 per day[5]
- Proportion of population in country below PPP US$1.25 per day[6]

# The Graduation Approach: Impacts on Food Security

FIGURE 3: COUNTRY-BY-COUNTRY IMPACT OF GRADUATION ON INDEX OF FOOD SECURITY[11] [12]




Note: Error bars represent 95% confidence intervals. Statistical significance relative to comparison households at each endline is noted at the 1% (***), 5% (**), or 10% (*) level.

# Empirical Exercise: Takeaways

1. How does the estimate of the treatment effect vary across countries?
2. How are these country-level estimates weighted in fixed effects estimation?
3. When are fixed effects necessary? When are they useful?
4. Which country-level estimates get more/less weight with fixed effects?