Williams College ECON 523:

Program Evaluation for International Development

Lecture 2: Regression

Professor: Pamela Jakiela

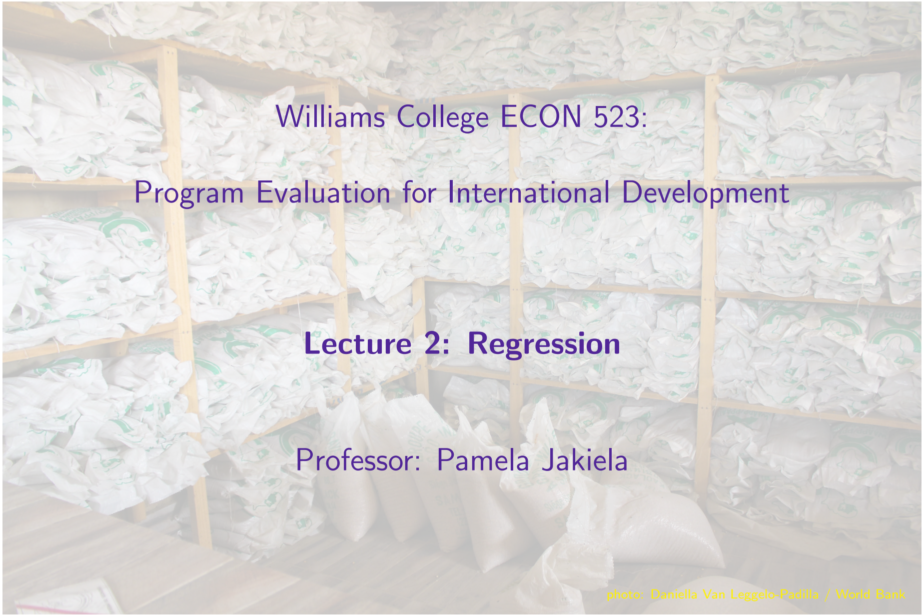
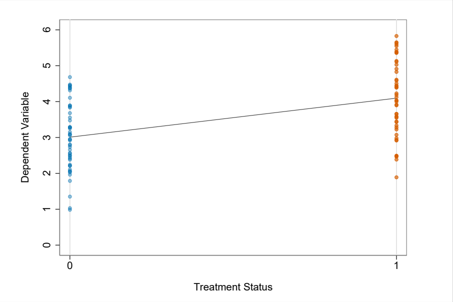


photo: Daniella Van Leggelo-Padilla / World Bank

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:

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

Notice that the numerator can be re-organized:

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

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:

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

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

OLS Regression on a Binary Independent Variable

Always remember that in a bivariate regression:

$$\hat{\beta}_{OLS} = \frac{\sum_i [Y_i (X_i - \bar{X})]}{\sum_i (X_i - \bar{X})^2} = \sum_i Y_i w_i$$

OLS Regression on a Binary Independent Variable

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

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

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:

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

$$= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{\sum_i \left(X_i - \bar{X}\right)^2}$$

$$= \frac{\sum_i \left[Y_i\left(X_i - \bar{X}\right)\right]}{N\bar{X}\left(1 - \bar{X}\right)}$$

Re-write denominator:

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

$$= n_T \left(1 - \bar{X}\right)^2 + \left(N - n_T\right)\left(-\bar{X}\right)^2$$

$$= \ldots = n_T - n_T\bar{X} = N\bar{X}\left(1 - \bar{X}\right)$$

OLS Regression on a Binary Independent Variable

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

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

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

OLS Regression on a Binary Independent Variable

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

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

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

OLS Regression on a Binary Independent Variable

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

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

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

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

- We can re-write the bivariate regression specification as

$$Y = \alpha + \beta_{pooled}\,(T_1 + T_2 + T3)$$

- If we pool treatments to estimate an average effect across treatment arms:
  - Estimated coefficient (and treatment effect) $\hat{\beta}_{pooled}$ 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)
    - (Show this by expressing $\hat{\beta}_{pooled}$ in terms of N_{T_i} and $\hat{\beta}_{T_i}$ values.)
- 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 treatment effect, 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

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)

Summary: When to Use a Continuous Measure of Treatment Intensity

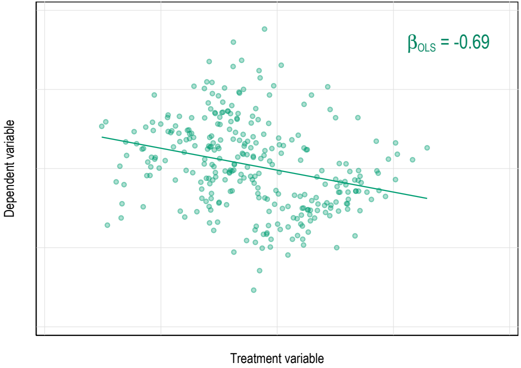
- A single treatment dummy: $Y_i = \alpha + \beta T_i$
- Mutually exclusive treatment dummies: $Y_i = \alpha + \beta_1 T_{1,i} + \beta_2 T_{2,i} + \beta_3 T_{3,i}$
- A continuous measure of treatment intensity: $Y_i = \alpha + \beta X_i$

Fixed Effects

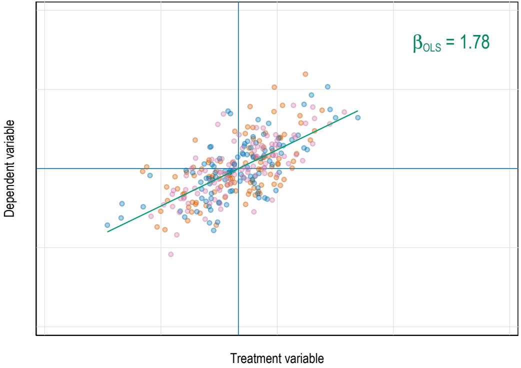
What Are Fixed Effects?

- Individual dummy variables for (all but one of the) 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

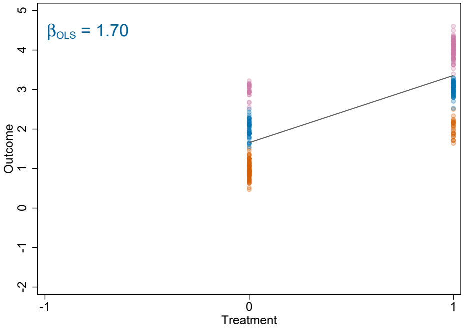
- Consider the following setup:
  - $N = 300$
  - Three groups: $N_1 = N_2 = N_3 = 100$
  - Treatment probability: 25, 50, 75 percent in Groups 1, 2, and 3, respectively
  - Data-generating process: $Y_i = G_i + D_i + \varepsilon_i$
- In this setting?
  - What is the (real) treatment effect?
  - What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

Binary Treatment, Homogeneous Treatment Effects: Example

G_i	p	N	N_T	N_C	$E\left[\bar{Y}_T\right]$	$E\left[\bar{Y}_C\right]$
1	0.25	100				
2	0.50	100				
3	0.75	100				

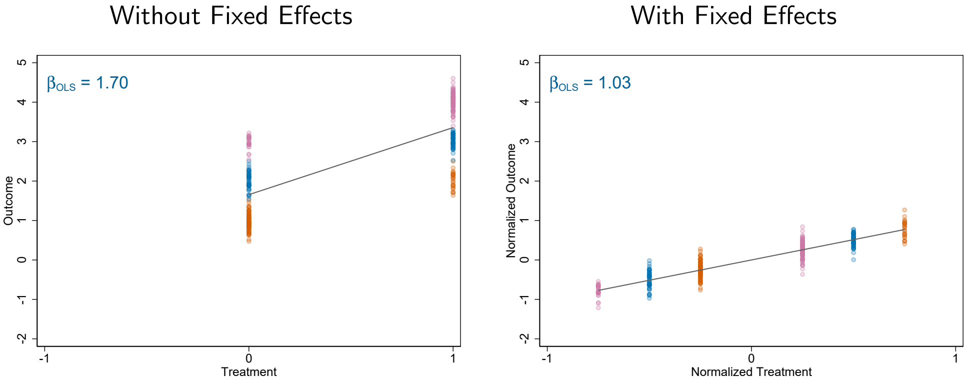
Binary Treatment, Homogeneous Treatment Effects: Example

Without Fixed Effects



- 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

The OLS regression specification:

$Y_i = \alpha + \beta D_i + \theta_{G2} G2_i + \theta_{G3} G3_i$

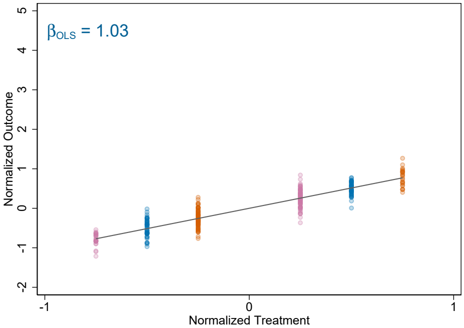
is equivalent to

$\tilde{Y}_i = \alpha + \beta \tilde{D}_i$

where

- $\tilde{Y}_i = Y_i - \bar{Y}_g$
- $\tilde{D}_i = D_i - \bar{D}_g$

With Fixed Effects



Binary Treatment, Homogeneous Treatment Effects: Example

ℓ	G_i	D_i	$\bar{D}_g = p$	$\tilde{D}_i = D_i - \bar{D}_g$	N_ℓ
1	1	0	0.25		
2	1	1	0.25		
3	2	0	0.50		
4	2	1	0.50		
5	3	0	0.75		
6	3	1	0.75		

Binary Treatment, Homogeneous Treatment Effects: Example

The OLS regression specification:

$$Y_i = \alpha + \beta D_i + \theta_{G2} G2_i + \theta_{G3} G3_i$$

is equivalent to

$$\tilde{Y}_i = \alpha + \beta \tilde{D}_i$$

$$\Rightarrow \hat{\beta} = \left(\sum_i \tilde{Y}_i \tilde{D}_i \right) / \sum_i \tilde{D}_i^2$$

$$= \left(\sum_\ell N_\ell \bar{\tilde{Y}}_\ell \tilde{D}_\ell \right) / \sum_i \tilde{D}_i^2$$

$$= \left[\sum_g N_g p_g (1 - p_g) \left(\bar{Y}_{g,T} - \bar{Y}_{g,C} \right) \right] / \sum_i \tilde{D}_i^2$$

Notice:

1. $\bar{\tilde{Y}}_\ell = \bar{Y}_\ell - \bar{Y}_g$
2. $N_\ell \tilde{D}_\ell = N_g p_g (1 - p_g)$ when $D_\ell = 1$
3. $N_\ell \tilde{D}_\ell = -N_g p_g (1 - p_g)$ when $D_\ell = 0$

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 group-specific weights
  3. Groups 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

Things We Know About Fixed Effects

1. When you regress Y on a set of fixed effects: $Y = \alpha + \beta T_1 + \beta T_2 + \ldots + \beta_{k-1} T_{k-1}$
   - OLS coefficients $\beta_1, \ldots, \beta_{k-1}$ are...
   - Predicted values $\hat{Y}_i$ are...
   - The residuals from this regression are...
2. When you regress Y on a treatment dummy plus fixed effects, the coefficient on D_i is a weighted-average of the within-group estimates of the treatment effect $(\bar{Y}_{g,T} - \bar{Y}_{g,C})$
   - Weights are proportional to $N_G p_G (1 - p_G)$

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 $Z_1, \ldots, Z_K$

The Frisch-Waugh-Lovell Theorem Explains Fixed Effects

- In a regression of Y on mutually exclusive dummies T_1 and T_2 ($Y = \alpha + \beta T_1 + \beta T_2$), we can use FWL to prove that the OLS coefficients on T_1 and T_2 are the differences in means
- FWL can be used iteratively, for example when $Z_1, \ldots, Z_K$ are mutually exclusive dummies
- When $Z_1, \ldots, Z_K$ are mutually exclusive dummies (fixed effects), FWL $\Rightarrow$ coefficient $\hat{\beta}$ from the OLS regression $Y = \alpha + \beta X + \gamma_1 Z_1 + \ldots \gamma_K Z_K$ (of Y on X plus fixed effects) is identical to the coefficient from a bivariate regression of normalized Y on normalized X
    - Normalizing $\leftrightarrow$ subtracting off group-level means $\leftrightarrow$ residualizing Y and X using FEs
- FWL can be applied iteratively with multiple sets of fixed effects ("two-way fixed effects")
- FWL provides a clear intuition for the mechanics of multivariate regression, when we regress Y on multiple continuous X variables (e.g. in observational data, or in macro)

Summary: Treatment Effects in a Regression Framework

- 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
  - Observations with above-mean treatment intensity receive positive weight in OLS
  - Common with fixed effects in experimental and non-experimental (e.g. country FEs) settings
    - When randomized/exogenous treatment probability/intensity varies across groups, you must include fixed effects or control directly for probability of treatment (the propensity score)

Empirical Exercise

The Graduation Approach

THE GRADUATION APPROACH

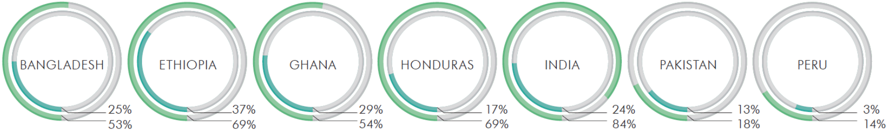
The Graduation approach consists of six complementary components, each designed to address specific constraints facing ultra-poor households.

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

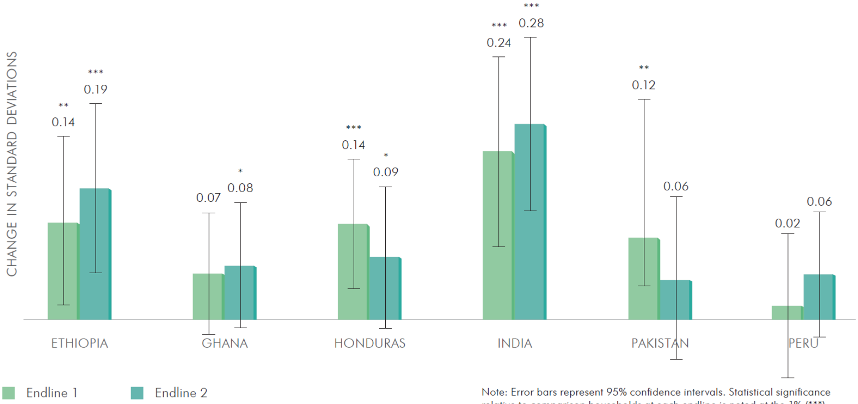


The program successfully targeted ultra-poor households. The proportion of households living below US$1.25 per day that were identified as eligible for the program and included in the study sample exceeded—often by a substantial margin—the proportion of the population living below US$1.25 per day in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