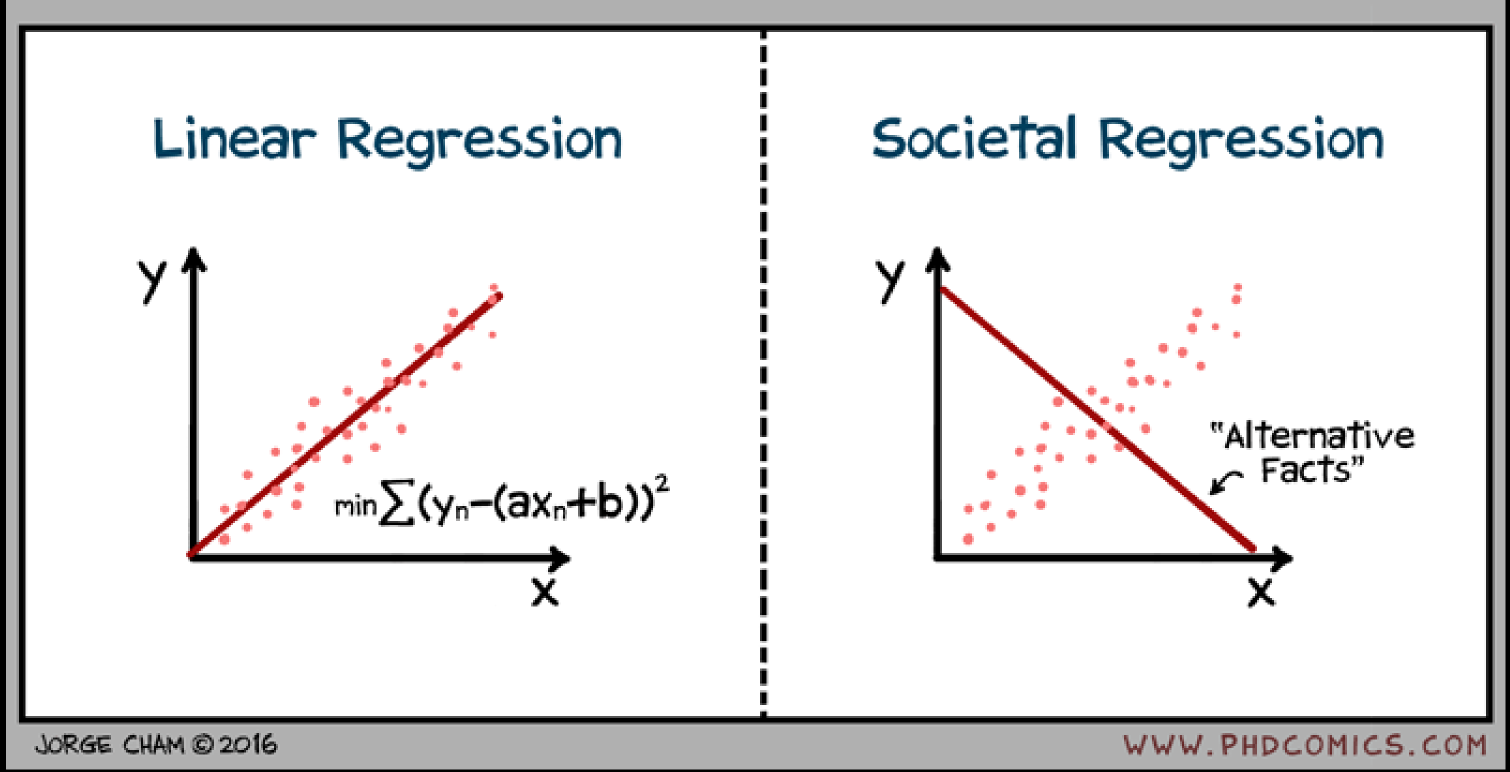


INFO 251: Applied Machine Learning

Regression and Impact Evaluation

Announcements

- PS2 posted
- Enrollment updates
- Note: please have a pen and paper handy for today's lecture, in order to do some basic calculations

About the class

What field/discipline is your degree in?

MIMS 2022	3 respondents	3 %
MIMS 2023	29 respondents	29 %
Public Policy	4 respondents	4 %
Social Science	5 respondents	5 %
Natural Science / Engineering	24 respondents	24 %
Haas	6 respondents	6 %
Statistics	12 respondents	12 %
Other	26 respondents	26 %

What was your undergraduate major?

Math / Statistics	27 respondents	27 %
Computer Science	20 respondents	20 %
Engineering	28 respondents	28 %
Social Science	17 respondents	17 %
Humanities	3 respondents	3 %
Natural Sciences	9 respondents	9 %
Other	15 respondents	15 %

About the class

Which of the following topics are you most interested in? Choose exactly 2:

Experimental methods for causal inference (impact evaluation, A-B testing)	31 respondents	31 %
Non-experimental methods for causal inference (instrumental variables, regression discontinuity)	11 respondents	11 %
Nearest neighbors and instance-based learning	9 respondents	9 %
Gradient descent	12 respondents	12 %
Regularization and linear models	6 respondents	6 %
Naive Bayes	11 respondents	11 %
Decision trees and random forests	16 respondents	16 %
Neural networks	38 respondents	38 %
"Deep" learning	33 respondents	33 %
Reinforcement learning	25 respondents	25 %
Practical issues with machine learning (features, missing/imbalanced data, multi-label classification)	36 respondents	36 %
Cluster analysis	13 respondents	13 %
Dimensionality reduction	11 respondents	11 %
Recommender systems	24 respondents	24 %
Natural language processing	14 respondents	14 %
Map-Reduce and Hadoop	10 respondents	10 %
Machine learning for causal inference	26 respondents	26 %

Random breakout session #2

- You will be randomly divided into groups of 3-4
  - This breakout room will only last a few minutes
  - Introduce yourselves
    - Name, program, something you look forward to in 2022
    - Are you potentially looking for study partners?
    - Consider exchanging contact information!

Course Outline

- Causal Inference and Research Design
  - **Experimental methods**
  - Non-experiment methods
- Machine Learning
  - Design of Machine Learning Experiments
  - Linear Models and Gradient Descent
  - Non-linear models
  - Neural models
  - Unsupervised Learning
  - Practicalities, Fairness, Bias
- Special topics

Key Concepts (last lecture)

- Random selection and assignment
- Internal and external validity
- Counterfactuals
- Identifying assumptions
- Control groups
- Statistical Power
- Single difference design
- Pre vs. Post research design
- Difference-in-Difference (Double Difference) design
- Differential Trends
- Multiple hypothesis testing, and adjustments

Outline

- **Lecture 1 wrap-up: Multiple hypothesis testing**
- A complete impact evaluation example: Progresa
- Regression recap
- Regression and Impact Evaluation
- Heterogeneous treatment effects
- Double-Difference via Regression
- (Fixed effects and Normalization)

Key Concepts (today's lecture)

- Progresa
- Interpreting regression coefficients
- Dummy variables, "one-hot" vectors
- Heterogeneous treatment effects
- Regression and impact evaluation
  - Estimating treat vs. control
  - Interaction variables
  - Estimating difference-in-difference
- Cross-sectional vs. panel data
- Between vs. within variation
- Difference regressions
- Normalization
- Fixed effects

Outline

- **A complete impact evaluation example: Progresa**
- Regression recap
- Regression and Impact Evaluation
- Heterogeneous treatment effects
- Double-Difference via Regression
- Fixed effects and Normalization

Progresa

- Goals of Progresa?
  - Increase education of the poor
  - Improve living conditions of the poor
- How to do it?
  - "Conditional cash transfers"
  - Provide subsidies to poor households **if they send their children to school**
- Why would this work?
  - Credit constraints
  - Reducing the cost of schooling is fundamental to many policies designed to improve educational outcomes

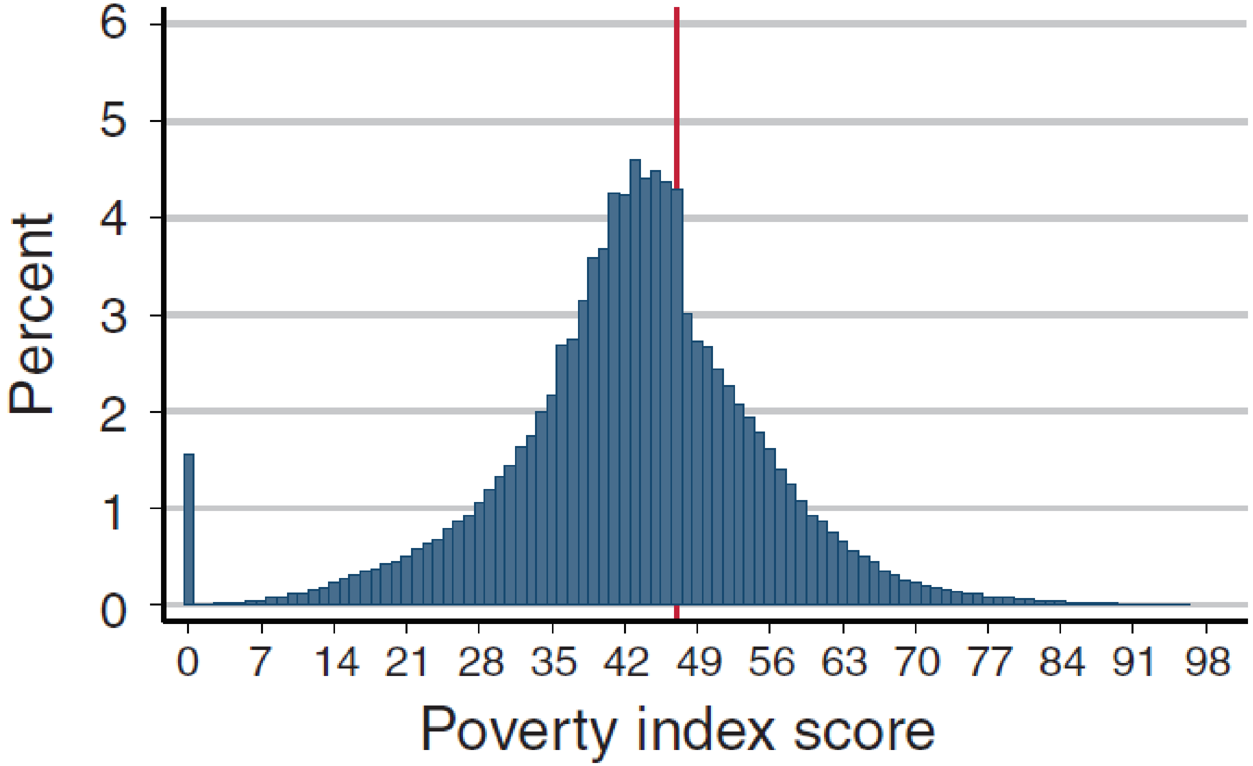
Progresa: Details

- Distinguishing features of Progresa
  - Targeted to poor households with primary school children
  - Money given to women in household
  - Randomized evaluation built in
    - 506 villages randomly assigned to treatment and control
- Scale and scope
  - Started in 1997, continues today (first as Oportunidades, now as Prospera)
  - Now covers 3 million people, 0.3% of Mexico GDP
  - More broadly, CCT's have now been used in dozens of countries
    - Fiszbein, A.S., Norbert R., 2009. Conditional Cash Transfers: Reducing Present and Future Poverty.
    - Baird et al. 2013. Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review.

Progresa: Eligibility

- Eligibility for Progresa was established at the household level, following a two-step targeting procedure
  1. Must live in a "poor rural village"
     - These communities were selected on the basis of a marginality index, established in 1995 using information from the Census
     - For the impact evaluation, a subset of these communities were selected to be eligible for Progresa *before* other communities (these are the "treatment villages")
  2. Household must be a "poor household" (within a poor village)
     - In treatment villages, a household census was conducted prior to program launch
     - The census collected a wealth index, which was used to identify poor households
     - The details of this welfare index were kept secret. Why?

Progresa: Eligibility

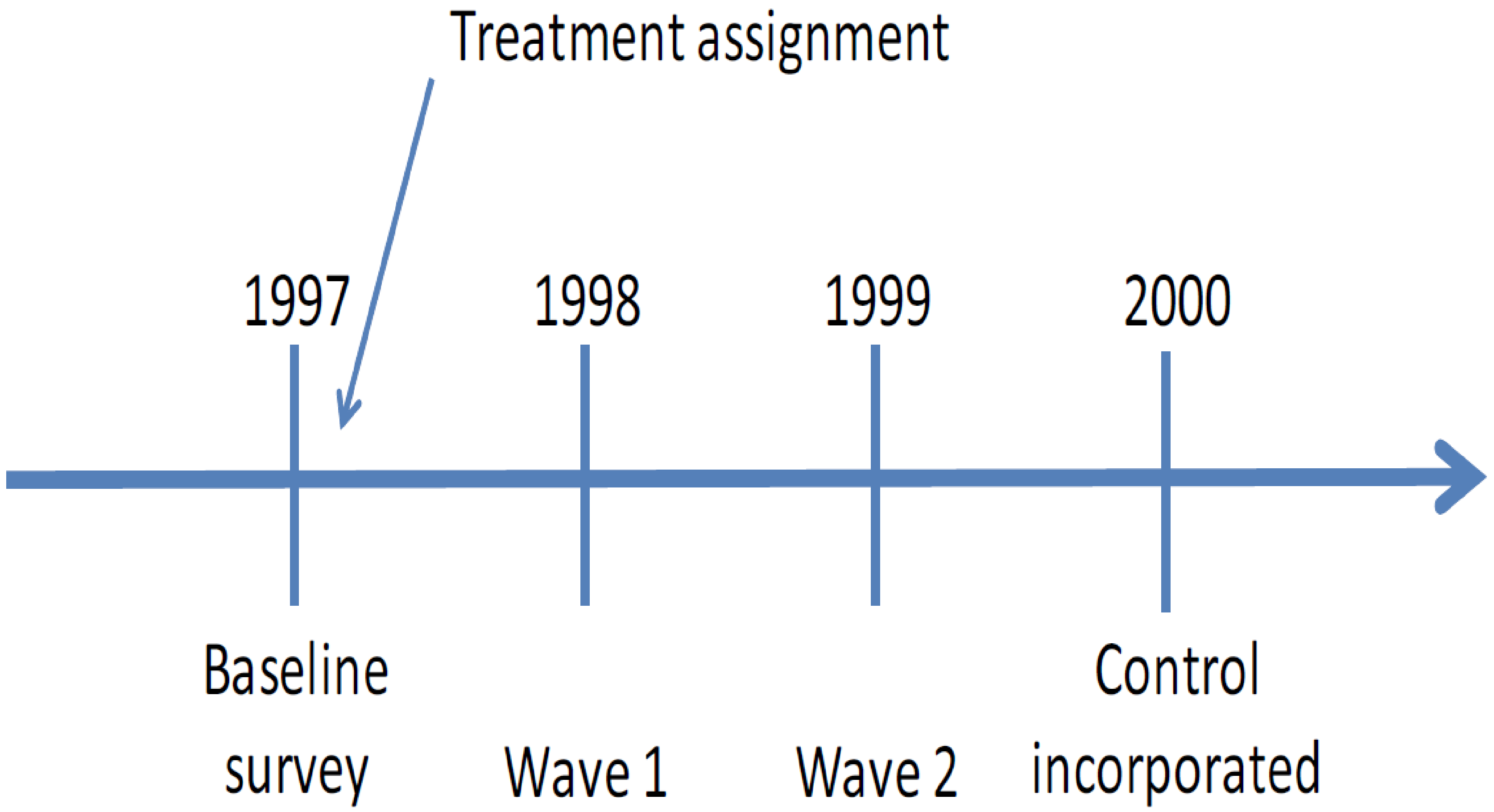


- Camacho, A., Conover, E., 2011. Manipulation of Social Program Eligibility. American Economic Journal: Economic Policy 3, 41–65.

Progresa: Implementation

- Nationwide, 50,000 communities were selected to receive Progresa
  - 78% of the households in the selected communities were deemed eligible
- For the impact evaluation, which occurred prior to the national roll-out, a subset of 506 communities in 7 states were selected to participate (i.e., ~1% of all communities)
  - Program was rolled out to 320 ("treatment") communities in May 1998
  - Program was rolled out to remaining 186 ("control") communities in late 1999
  - This is a classic example of a "staggered roll-out design"
  - 97 percent take-up among households

Progresa: Implementation



Progresa: Implementation

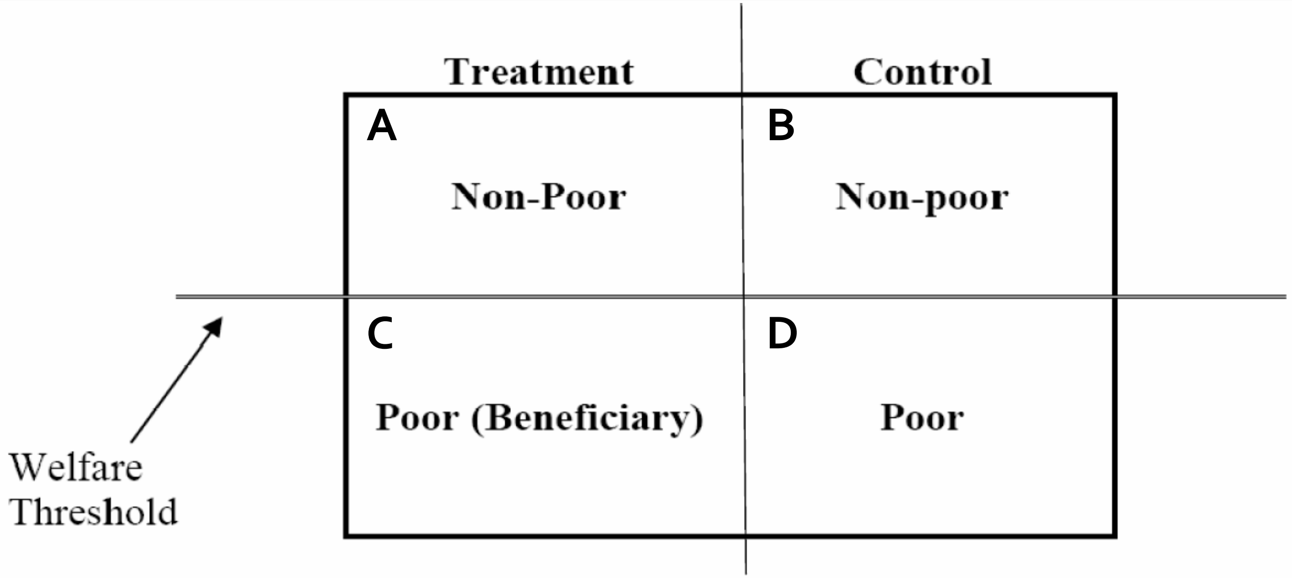


Figure 1: Program Evaluation Design

- How to measure effect of Progresa on enrollment, assuming we only have access to (post-intervention) data from 1998?
  - Compare poor in Treatment to poor in Control (i.e., C - D)

Progresa: Implementation

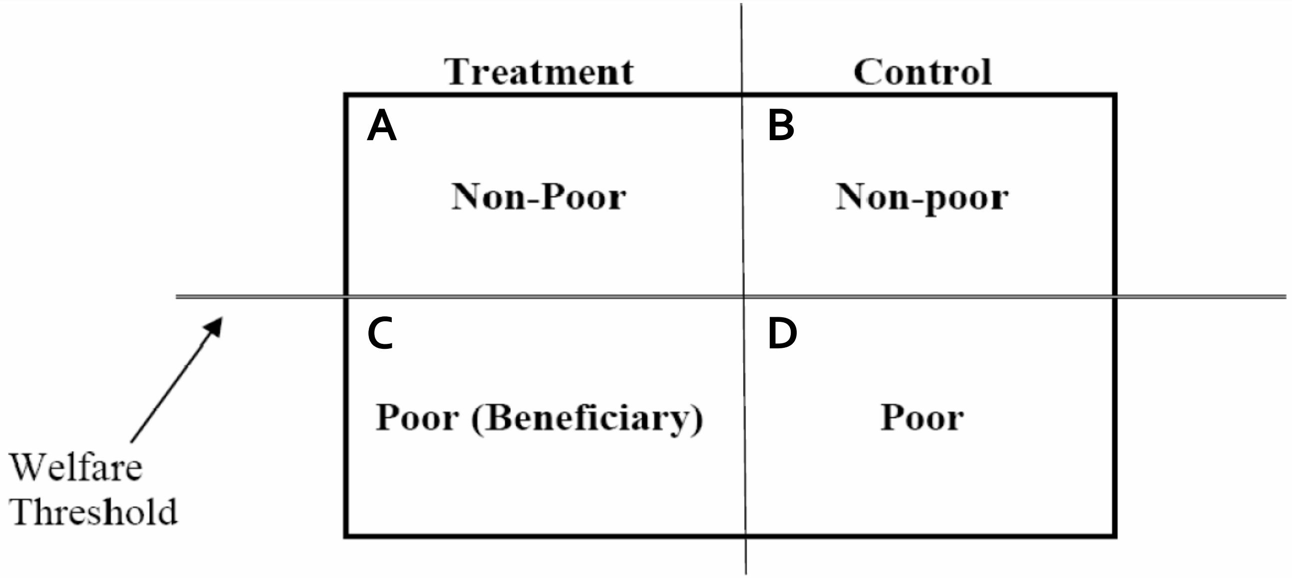


Figure 1: Program Evaluation Design

- The counterfactual (for enrollment of poor in treated villages) is…
  - Enrollment of poor in control villages (in absence of Progresa)
- The key identifying assumption is…
  - In the absence of treatment, enrollment in C would have been same as enrollment in D
- Does this assumption reasonable in this context?
- What evidence might we provide to support using this identifying assumption?

Progresa: Externalities

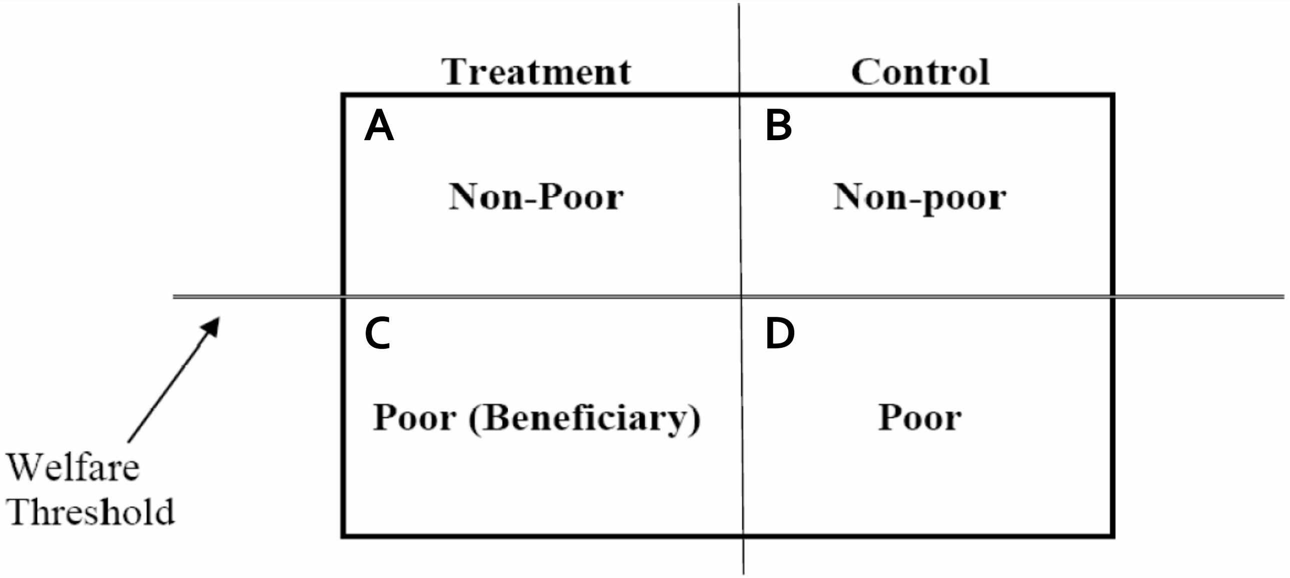


Figure 1: Program Evaluation Design

- The “basic” treat-vs-control design measures the treatment effect by looking at outcomes among the poor, and ignores the non-poor
- What if we observe differences in the outcomes of the non-poor?
  1. Perhaps treatment and control villages were different to begin with (i.e., randomization failed)
  2. Perhaps there were spillover effects from poor to non-poor in treated villages

Progresa: Diff-in-Diff

- What if we have two rounds of survey data: baseline data (before intervention) and endline data (after intervention)?
  - We can assess whether treatment and control villages differed before Progresa
- Difference-in-difference design for evaluating Progresa
  1. Focusing on poor, how do changes between 97 and 98 compare between T and C villages?
  2. Focusing on endline, how do differences between Poor and Non-Poor compare between T and C?
  - What are the identifying assumptions of these two designs?

1997 ("Baseline")

Treatment	Control
E Non-Poor	F Non-poor
G Poor (Beneficiary)	H Poor

1998 ("Endline")

Treatment	Control
A Non-Poor	B Non-poor
C Poor (Beneficiary)	D Poor

Outline

- A complete impact evaluation example: Progresa
- **Regression recap**
- Regression and Impact Evaluation
- Heterogeneous treatment effects
- Double-Difference via Regression
- Fixed effects and Normalization

Regression and Impact Evaluation

- Thus far we've used cross-tabs to "eyeball" the impact of a treatment
- Advantages
  - Very simple to compute
  - Easily interpretable
- Disadvantages
  - How to measure statistical precision?
  - How to deal with known confounds (e.g., differential trends)?
  - How to estimate treatment effect heterogeneity?
  - May be unbiased, but may not be precise - controlling for additional factors may increase precision

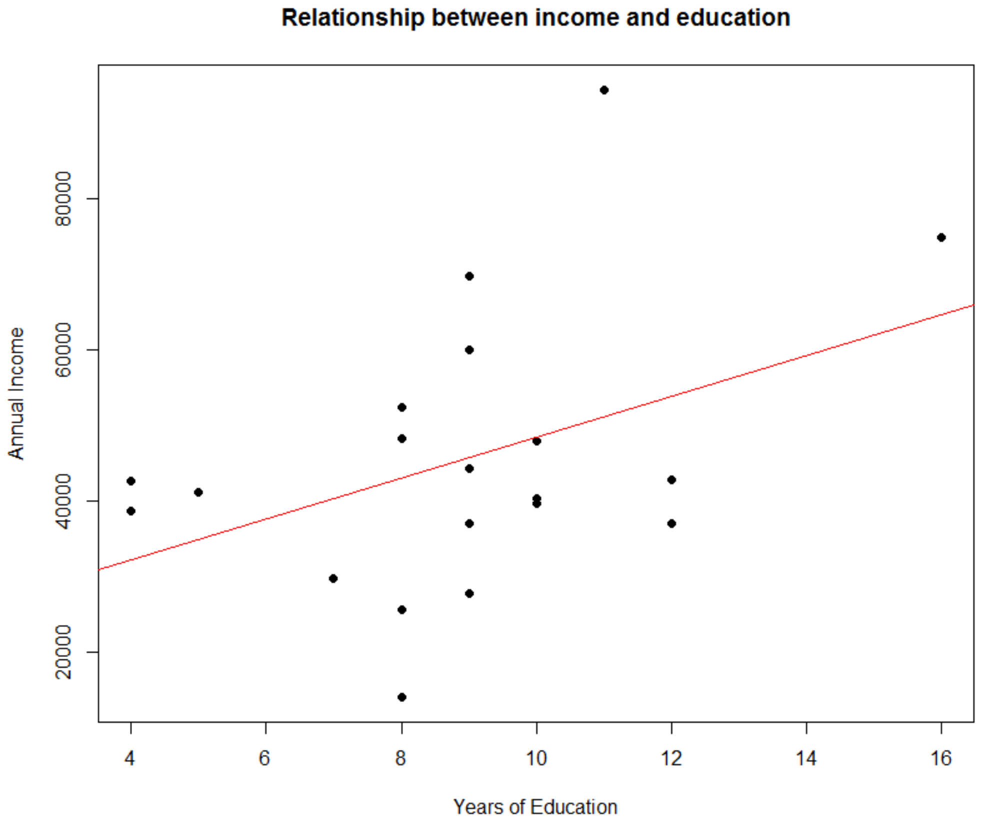
Random breakout session #1

- You will be randomly divided into groups of 3-4
  - This breakout room will only last a few minutes
  - It will be awkward, but that's okay
  - Introduce yourselves
    - Name, program, something you look forward to in 2022
    - Are you potentially looking for study partners?
    - Consider exchanging contact information!
  - This will just last a few minutes…

Regression: Quick Recap

- Linear regression offers a concise summary of the mean of one variable as a function of the other variable through two parameters: the slope and the intercept of the regression line
  - Causality often *implied*, rarely *justified*

```
      Education Age    Income
[1,]          8  35  30942.35
[2,]          8  23  37323.89
[3,]          8  58  49381.84
[4,]          5  41  31680.86
[5,]         13  35  81147.84
[6,]          9  43  38682.86
[7,]          8  35  34632.30
[8,]          7  56  14394.98
[9,]         11  62  22243.85
[10,]        14  24  51831.79
[11,]        12  25  23963.90
[12,]        12  32  66780.27
[13,]         4  41  26979.73
[14,]         8  49  38837.48
[15,]        10  21  40726.37
[16,]         8  33  40269.51
[17,]         4  36  34293.32
[18,]        10  38  61158.98
[19,]        11  36  64329.59
[20,]         9  48  51069.77
```



Regression: Quick Recap

- Simple bivariate (linear) regression
  - The regression model

$$wages_i = \alpha + \beta * education_i + error_i$$

  - The fitted model

$$wages_i = 12409 + 3310 * education_i + error_i$$

- Intuition check
  - What does β tell us?
  - What is *12409*?
  - What are the expected wages be for someone with 14 years of education?
    - 12409 + 14*3310= 58,749

Regression: Quick Recap

- Regression with binary predictor/independent variables
  - The regression model

$$wages_i = \alpha + \beta * isForeign_i + error_i$$

  - The fitted model

$$wages_i = 54212 - 2710 * isForeign_i + error_i$$

- Multiple (linear) regression

$$wages_i = \alpha + \beta * education_i + \gamma * isForeign_i + error_i$$

Regression: Categorical variables

- What if our control variables are categorical?
  - Example: We want to study the relationship between wages and education, controlling for country
  - $Wages_i = \alpha + \beta Education_i + \gamma Country_i + \epsilon_i$

- How to deal with a categorical predictor?
  - Convert to a single binary variable:
    - $Wages_i = \alpha + \beta Education_i + \gamma USA_i + \epsilon_i$
    - $USA_i = 1$ iff worker *i* is from USA, $USA_i = 0$ otherwise
    - Makes sense if we care about the effect of one category relative to others
  - Convert to a set of binary variables:
    - $Wages_i = \beta Education_i + \gamma_1 USA_i + \gamma_2 CHINA_i + \cdots + \gamma_M Country_m + \epsilon_i$
    - $Wages_i = \beta Education_i + \sum_{c=1}^{M} \gamma_c \, \mathbf{1}(Country_i = c) + \epsilon_i$
    - $Wages_i = \beta Education_i + Country_i + \epsilon_i$ ← this is an abuse of notation, but it is very common

Regression: "Dummy" variables

- Converting levels to a series of binary variables
  - $Y_i = \alpha + \beta Education_i + Country_i + \epsilon_i$
  - $Country_i$ is a "dummy variable" or "one-hot vector" or "fixed effect"

- Interpretation
  - Equivalent to creating a country-specific intercept
  - Each of the $Country$ coefficients indicates average wage for workers from that particular country (when $Education_i = 0$), *relative to the reference country*
  - With M countries, we have (M-1) coefficients and an intercept α
  - Alternatively, estimate with no intercept and M coefficients
    - $Y_i = \beta Education_i + Country_i + \epsilon_i$
    - Intuition check: Will the coefficients for "country" dummies be the same in both cases?

Outline

- A complete impact evaluation example: Progresa
- Regression recap
- **Regression and Impact Evaluation**
- Heterogeneous treatment effects
- Double-Difference via Regression
- Fixed effects and Normalization

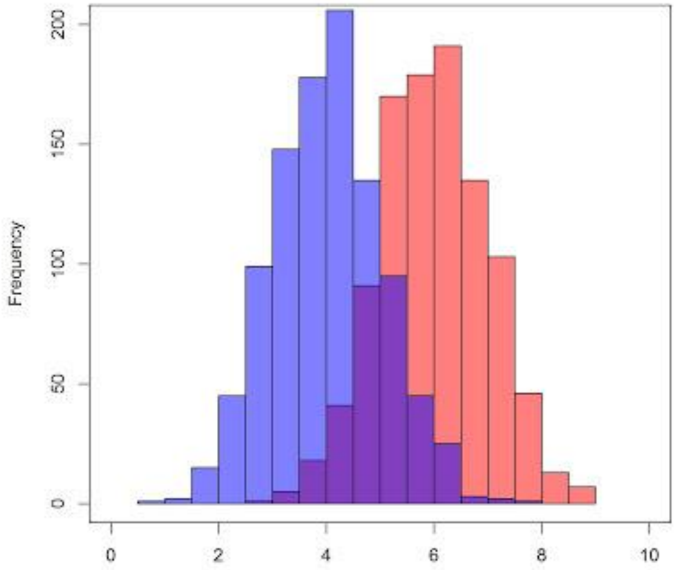
Regression and Impact: Basics

- How to measure the effect of treatment T on outcome Y in a regression?
  - The regression equation:

$$Y_i = \alpha + \beta T_i + \epsilon_i$$

  - Example: We estimate the effect of eating a cookie on happiness on a scale of 1-10. We estimate $\hat{\alpha} = 4.1, \hat{\beta} = 1.3$. What does this mean?

- If T is randomly assigned, $\hat{\beta}$ is an estimate of the ***causal impact*** of T on Y



Regression: "Control" variables

- How to simultaneously measure the effect of a treatment T and a non-experimental control variable X on an outcome Y in a regression setting?

$$Y_i = \alpha_1 + \beta_1 T_i + \gamma X_i + e_i$$

- How is this different from a version without control variables?

$$Y_i = \alpha_2 + \beta_2 T_i + \epsilon_i$$

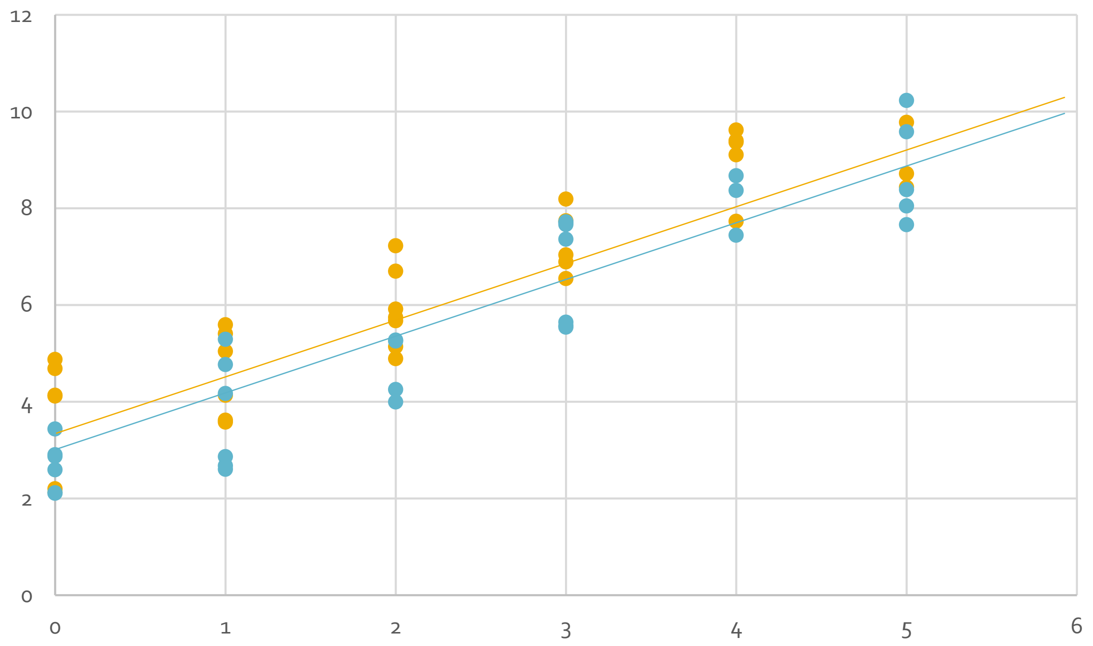
- In a perfectly randomized experiment…
  - What, if anything, can we say about $Cor(T_i, X_i)$?
  - What, if anything, can we say about our estimates of β_1 and β_2?
  - What, if anything, can we say about γ?

Control variables: Example

- Example: We are estimating the effect of eating a cookie (T_i) on happiness (Y_i) on a scale of 1-10, while controlling for years in grad school (X_i)

- Regression equation?
  - $Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i$

- Coefficient estimates
  - $\hat{\alpha} = 3.4, \hat{\beta} = -0.5, \hat{\gamma} = 1.2$
  - What do these results mean?



Outline

- Regression recap
- Regression and Impact Evaluation
- **Heterogeneous treatment effects**
- Double-Difference via Regression
- Fixed effects and Normalization

Treatment effect heterogeneity

- How to simultaneously measure the effect on an outcome *Y* of: (i) a treatment *T*; and (ii) a non-experimental control variable *X*; and (iii) allow for the treatment effect to vary with the control variable?
  - Example: Tall students may respond differently to cookies than short students
  - The treatment effect is different for different types of people (where "type" is measured with X_i)

- We can estimate this with a regression:

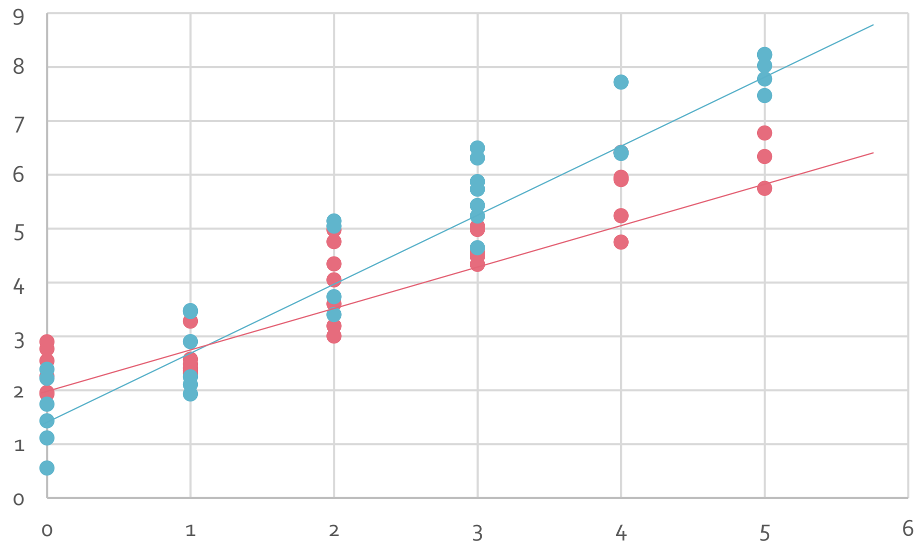
$$Y_i = \alpha + \beta T_i + \gamma X_i + \delta(T_i * X_i) + \epsilon_i$$

Treatment effect heterogeneity: Binary X

- Example 1: We want to estimate the effect of eating a cookie (T) on happiness (Y), while controlling for height (X) and the interaction (T*X)
  - $Y_i = \alpha + \beta T_i + \gamma X_i + \delta(T_i * X_i) + \epsilon_i$
  - $\hat{\alpha} = 1.4, \hat{\beta} = -1.0, \hat{\gamma} = 0.8, \hat{\delta} = 0.6$
    - What do these results mean?
  - What is the expected happiness for:
    - A short student with a cookie
    - A tall student with a cookie
    - A short student without a cookie
    - A tall student without a cookie

Treatment effect heterogeneity: Continuous X

- Example 2: We want to estimate the effect of eating a cookie on happiness, while controlling for years in grad school (a continuous variable) and the interaction effect.
    - Note: this is identical to the last example, except now our X variable is continuous, not binary
  - $Y_i = \alpha + \beta T_i + \gamma X_i + \delta(T_i * X_i) + \epsilon_i$
  - $\hat{\alpha} = 1.4, \hat{\beta} = -1.0, \hat{\gamma} = 0.8, \hat{\delta} = 0.6$
  - What do these results mean?
  - How to visualize?



Outline

- Regression recap
- Regression and Impact Evaluation
- Heterogeneous treatment effects
- **Double-Difference via Regression**
- Fixed effects and Normalization

Regression and Impact: Diff-in-Diff

- In a typical double-difference setting, data is collected at baseline (pre-treatment) and at endline (post-treatment)
- Very similar to a situation where you have a treatment *T* **and** a non-experimental binary control variable *X* **and** a differential effect of treatment by a binary control variable
- In this case, the control variable is time!
  - Instead of a (continuous) X_i we have a binary $Post_i$

Regression and Impact: Diff-in-Diff

- For example, our data look like this:

ID	Y	Post?	Treatment?	Post * Treat	…
1	12	0 (PRE)	1 (Treat)	0	…
1	14	1 (POST)	1 (Treat)	1	…
2	11	0 (PRE)	0 (Control)	0	…
2	11	1 (POST)	0 (Control)	0	…
3	24	0 (PRE)	0 (Control)	0	…
…	…	…	…	…	…
12419	…	…	…	…	…

	Control	Treat
Pre	A	B
Post	C	D

- Write down the regression equation:

$$Y_i = \alpha + \beta T_i + \gamma Post_i + \delta(T_i * Post_i) + \epsilon_i$$

Regression and Impact: Diff-in-Diff

- Let's return to Progresa (Shultz, 2004: page 209)

$$S_i = \alpha_0 + \alpha_1 P_i + \alpha_2 E_i + \alpha_3 P_i E_i + \sum_{k=1}^{K} \gamma_{ki} C_{ki} + \sum_{j=1}^{J} \beta_j X_{ji} + e_i \quad i = 1, 2 \ldots, n \qquad (1)$$

- P_i = Progresa village ("T")
- E_i = Eligible (poor) household ("X")
- P_iE_i = Difference-in-difference estimator
- C_{ki} = Dummy variable for grade of child
  - Controls for fact that enrollment rates vary across grades
- X_{ji} = Other control variables (I think?)

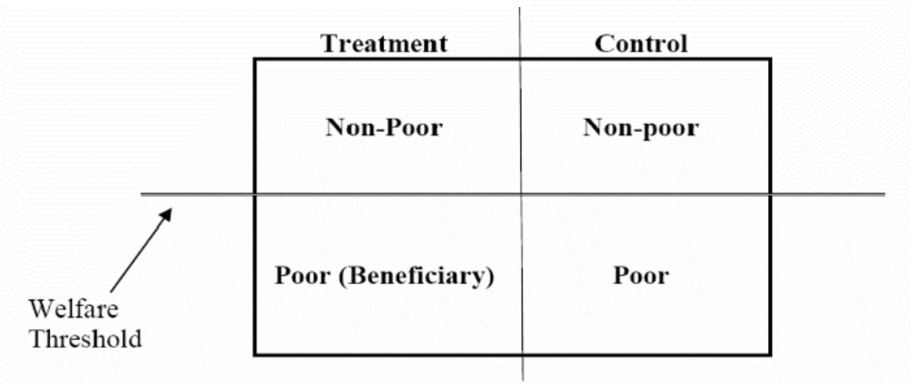


Figure 1: Program Evaluation Design

Regression and Impact: Summary

- Double Difference

$$Y_i = \alpha + \beta T_i + \gamma Post_i + \delta(T_i * Post_i) + \epsilon_i$$

- Simple difference

$$Y_i = \alpha + \beta T_i + \epsilon_i$$

  - (Estimated only in Post period)

- Pre vs. Post

$$Y_i = \alpha + \gamma Post_i + \epsilon_i$$

  - (Estimated only on treatment group)

Outline

- Regression recap
- Regression and Impact Evaluation
- Heterogeneous treatment effects
- Double-Difference via Regression
- **Fixed effects and Normalization**

Fixed Effects and Normalization

- **Cross-sectional data:** Data from several units at a single point in time
  - Schooling outcomes in 1998
  - Happiness of students on Jan 20
  - Number of logins per person in March
- **Time series (panel) data:** Multiple observations for each unit over time
  - Schooling outcomes in 1997 and 1998
  - Happiness of students on Jan 20 and Jun 20
  - Number of logins per person in each month of 2010

Between and Within Variation

- **Cross-sectional data:**
  - Variation is between/across units

Location	Year	Price	Quantity sold (per capita)
Chicago	2003	$75	2.0
Seattle	2003	$50	1.0
Milwaukee	2003	$60	1.5
Madison	2003	$55	0.8

- What do you notice about relationship between price and quantity?
  - Across the four cities, price and quantity are positively correlated
  - This is not what we would expect – omitted variables likely matter

Between and Within Variation

- **Panel data:**
  - Variation is between/across units *and* **within units over time**

Location	Year	Price	Quantity
Chicago	2003	$75	2.0
Chicago	2004	$85	1.8
Seattle	2003	$50	1.0
Seattle	2004	$48	1.1
Milwaukee	2003	$60	1.5
Milwaukee	2004	$65	1.4
Madison	2003	$55	0.8
Madison	2004	$60	0.7

- What do you notice about relationship between price and quantity?
  - Within each of the four cities, price and quantity are inversely correlated (as expected with downward sloping demand)

Exploiting Within-Unit Variation

- How to isolate changes in outcomes correlated with changes *within* a unit ?
  - e.g., Changes in demand caused by changes in price *within* a given city (over time)

Exploiting Within-Unit Variation

- One (familiar?) approach: “difference” regressions
  - Isolate differences over time in X and Y
  - Instead of: $Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}$
  - $(Y_{it} - Y_{i(t-1)}) = \alpha + \beta(X_{it} - X_{i(t-1)}) + \epsilon_{it}$
    - Are *changes* in *Y* related to *changes* in *X*?

- Another approach: normalization
  - Instead of: $Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}$
  - $(Y_{it} - \overline{Y_i}) = \alpha + \beta(X_{it} - \overline{X_i}) + \epsilon_{it}$

Exploiting Within-Unit Variation

- A third (closely related) approach: **Fixed effects**
- Basic idea (refer to lecture notes for details)
  - Instead of: $(Y_{it} - \bar{Y}_i) = \alpha + \beta(X_{it} - \bar{X}_i) + \epsilon_{it}$
  - Add "dummy" variables for each *i*: $Y_{it} = \alpha + \beta X_{it} + (\mu_i + \cdots + \mu_N) + \epsilon_{it}$
    - Conceptually the same as the "country fixed effect" example from slide 27
    - More formally: $Y_{it} = \alpha + \beta X_{it} + \sum_{j=1}^{N} \mu_j \, \mathbf{1}(ID_i = j) + \epsilon_i$
    - Equivalent to adding an intercept for each *i*
  - Shorthand: $Y_{it} = \alpha + \beta X_{it} + \mu_i + \epsilon_{it}$
- Note: we can also "normalize" for time FE's
  - $Y_{it} = \alpha + \beta X_{it} + \mu_i + \pi_t + \epsilon_{it}$

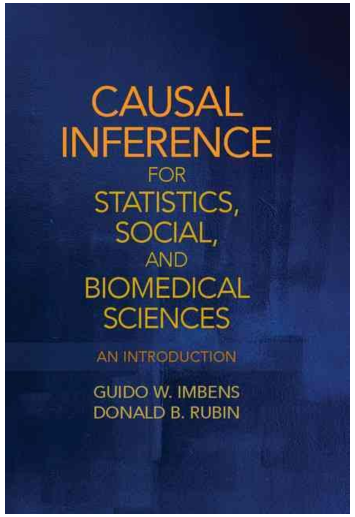
Fixed Effects: Digging Deeper

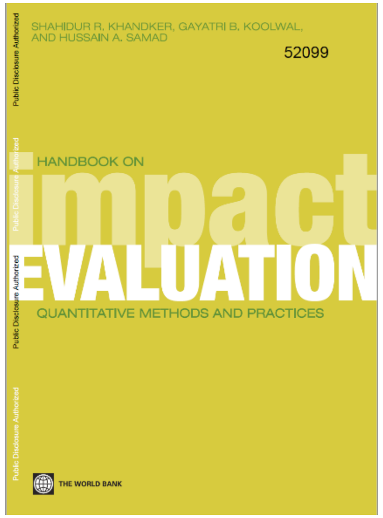
- Key advantage of Fixed Effects
  - Fixed effects control for *unobserved heterogeneity*
  - They remove the effect of time-invariant characteristics to assess the net effect of the predictors on the outcome

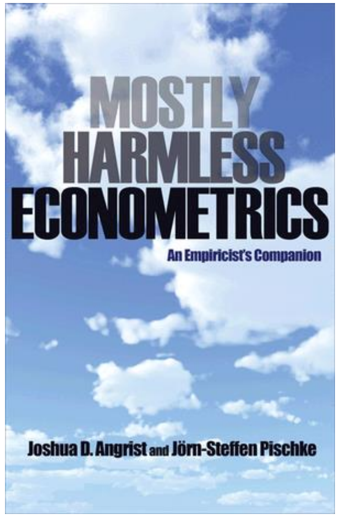
- Extensions
  - Time trends
  - Region-specific slopes

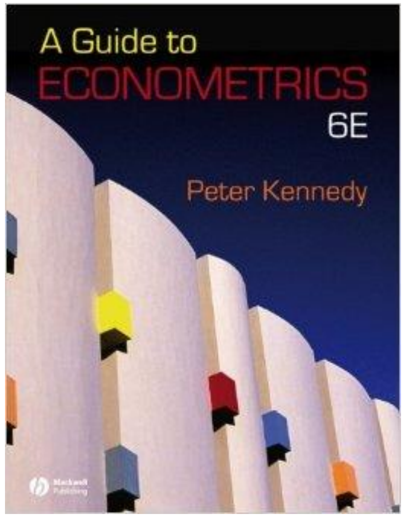
Additional Resources

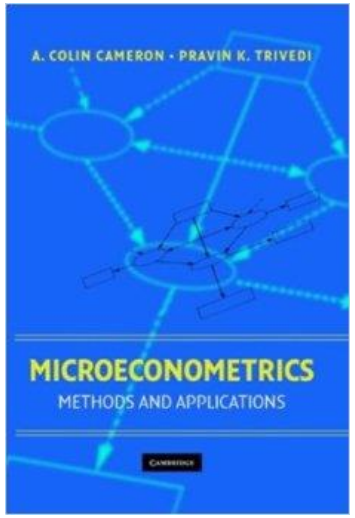
Beginner → Advanced





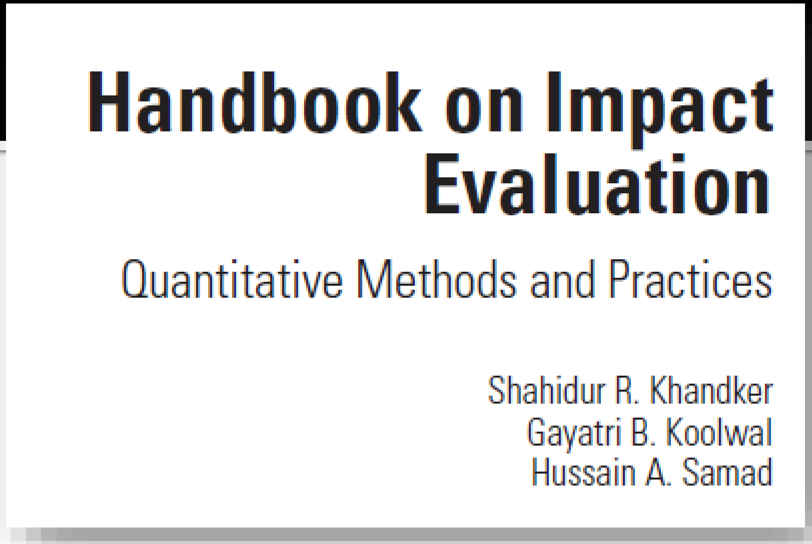






For Next Class:

- Read Chapters 6 and 7:



Handbook on Impact Evaluation

Quantitative Methods and Practices

Shahidur R. Khandker
Gayatri B. Koolwal
Hussain A. Samad

- Get started on problem set 2!