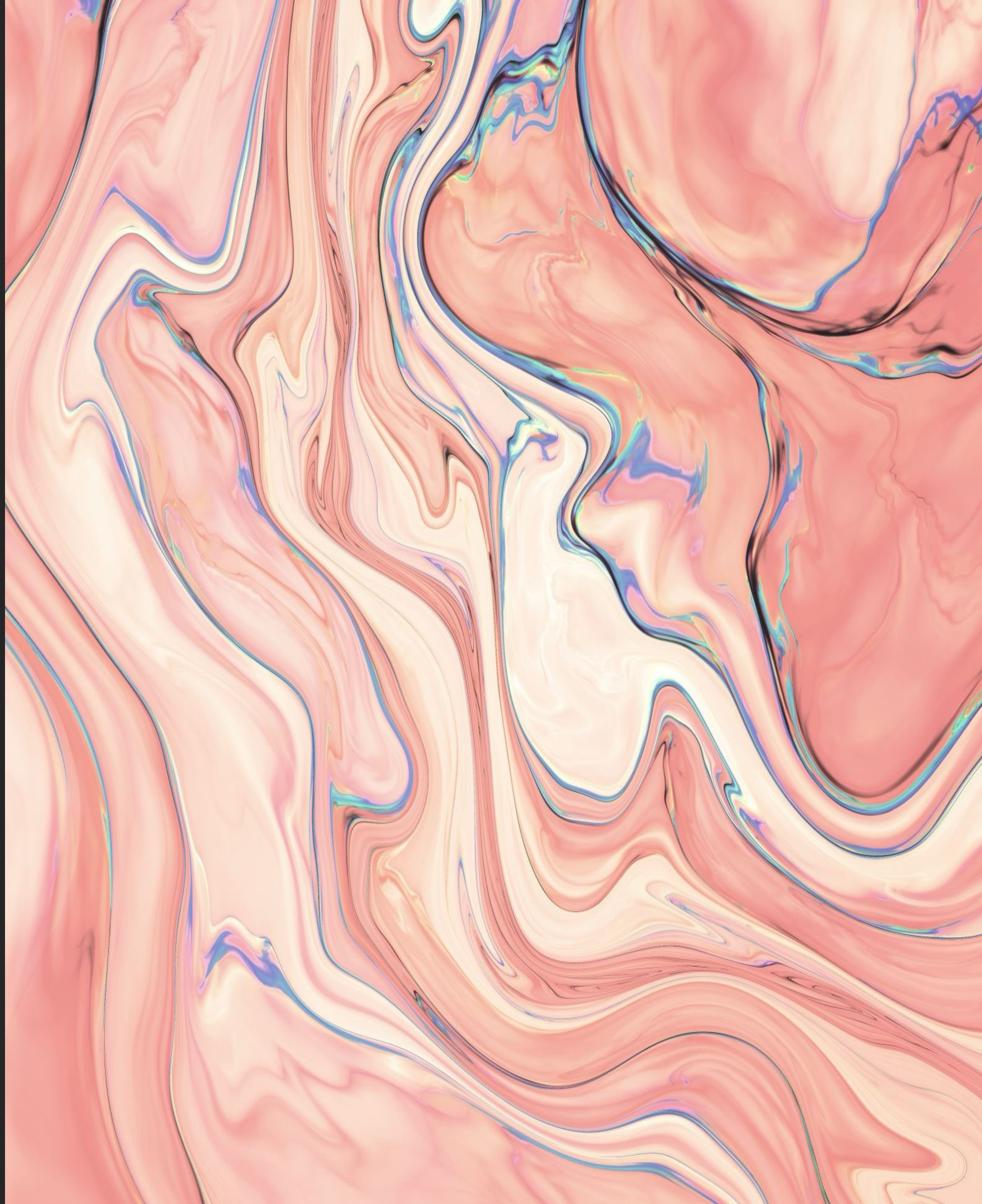
EMPIRICAL ESTIMATION IN MICROECONOMICS:

INTRODUCTION TO ECONOMETRICS & EXAMPLES OF REAL-WORLD APPLICATIONS

ANDREW GATES

2024



INTRODUCTION TO EMPIRICAL ESTIMATION METHODS

- Ordinary Least Squares estimation
  - Motivation and intuitive explanation
  - Formal mathematical derivation
  - Examples of applications and interpreting results
  - Basic examples of different functional forms

- Differences in Differences estimation
  - Card & Krueger (AER 1994) paper: minimum wages & employment

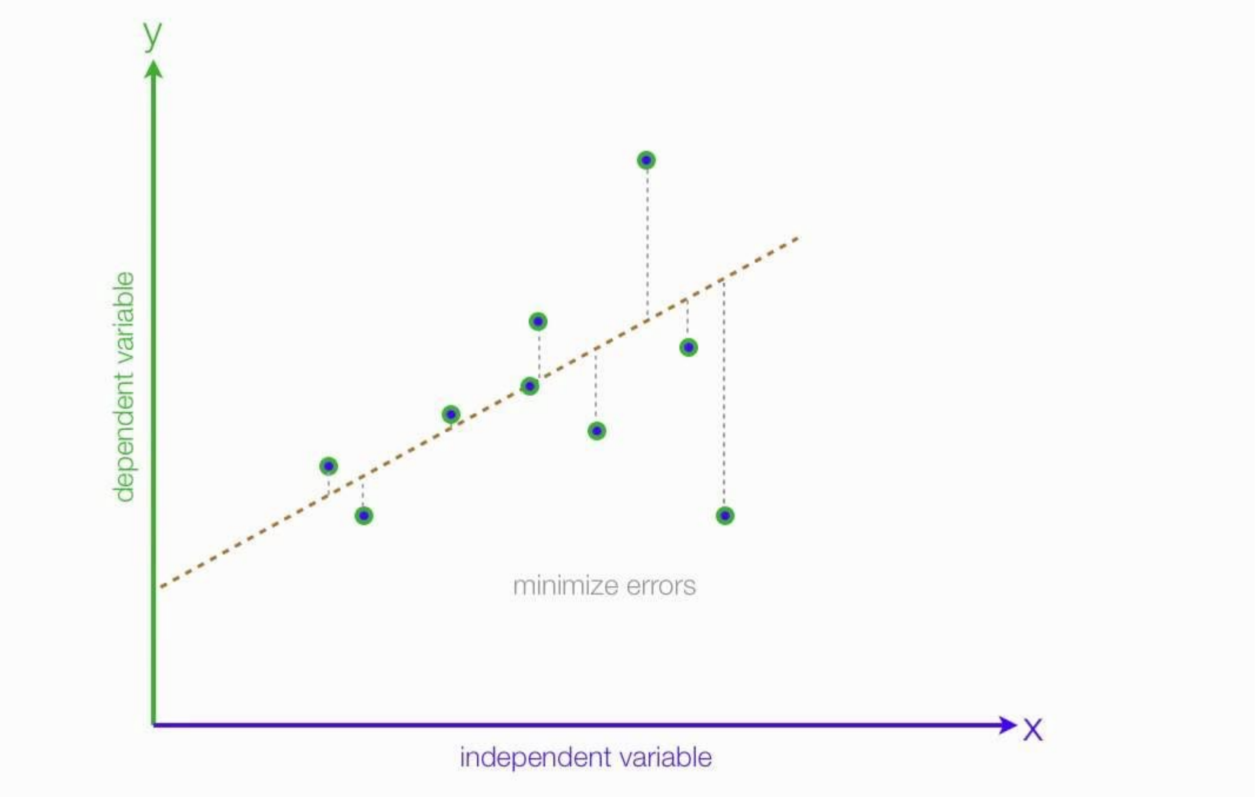
    - *Future Topic:* Instrumental Variables approach
      - Acemoglu, Johnson, Robinson (2001) paper: settler mortality, institutions, econ growth
      - Other examples…

OLS: ORDINARY LEAST SQUARES ESTIMATION

- To estimate how an "independent variable" (***x***) affects a "dependent variable" (***y***) we can use a minimization problem with derivatives to find a line that best fits a plot of the data points
  - We have N observations: each point represents a measured value of outcome variable Y and exogenous variable X
- This (linear regression) is an extremely common technique to do empirical estimations in research to understand relationships between variables

<u>Examples:</u>

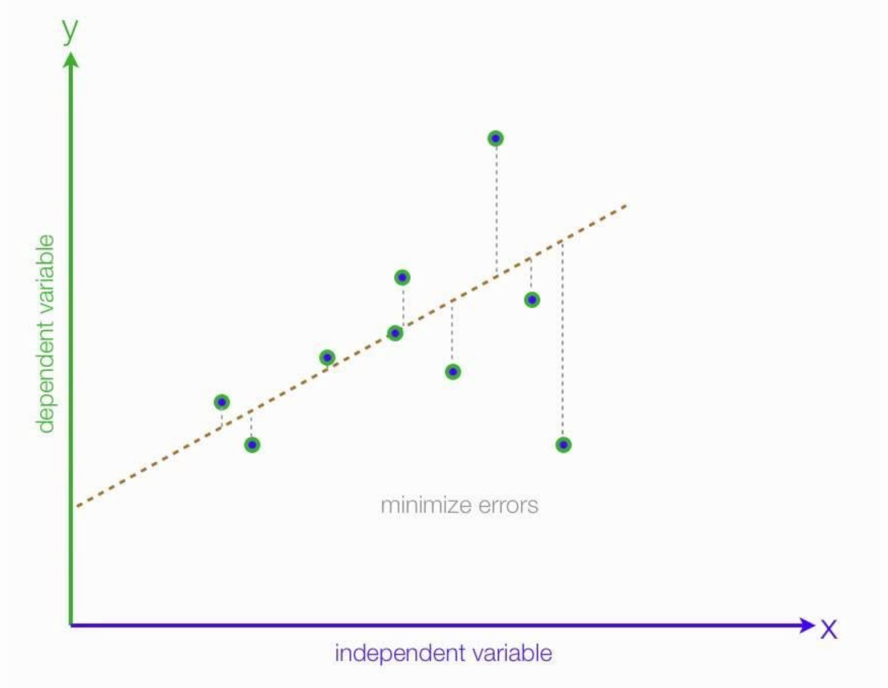
- wage and years of education
- college GPA and high school SAT score
- stock price and sales volume
- pollution levels and factory regulations
- automobile weight and crash fatality rate



OLS ESTIMATION: EQUATION AND INTUITION

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

- Y_i = estimated outcome value for observation i
- β_0 = intercept of the best fit line
- β_1 = slope of the best fit line
- X_i = value of independent variable X for observation i
- ε_i = random error



- The intuition is to fit a line that minimizes the **"sum of squared residuals"** (RSS): the total amount of all vertical distances between the *N* data points and this line of "best fit" for the given set of data
  - Changing the intercept (starting point on the y-axis) shifts the line up or down
  - Changing the slope rotates the line's orientation / direction
- Squaring each of the vertical distances before adding them together achieves two important things:
  - Mathematically ensures all distances are positive: otherwise points above and below the line would cancel each other out
  - Exponentially increases the penalty for being farther away from the best fit line: outlier points have more "weight"

Deriving the formulas for the OLS intercept and slope coefficient starts with minimizing the sum of squared residuals (RSS) - the distance between the line of best fit and the data points. Mathematically, this problem is the following:

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 = RSS$$

This minimization problem above is solved by taking the partial derivatives and setting them equal to 0. We take the derivative of RSS with respect to $\hat{\beta}_0$ and set it equal to 0 and solve, and perform the same process with respect to $\hat{\beta}_1$. This obtains the following FOCs:

$$\frac{\partial(RSS)}{\partial\hat{\beta}_0} = \sum_{i=1}^{N} -2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

and

$$\frac{\partial(RSS)}{\partial\hat{\beta}_1} = \sum_{i=1}^{N} -2x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

Now we can solve these first derivative equations, or "first order conditions", using algebra and basic properties of summations. Dividing both sides by -2 we can re-write the first FOC as $\sum_{i=1}^{N} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = 0$.

Using the fact that $\sum_{i=1}^{N} y_i = N\bar{y}$ we have $N\hat{\beta}_0 = N\bar{y} - N\hat{\beta}_1\bar{x}$

Finally, we can divide everything by N to obtain

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

To solve for $\hat{\beta}_1$ we can divide both sides by -2 again and rearrange to obtain $\sum_{i=1}^{N} x_i y_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2 = 0$. Substituting the result for $\hat{\beta}_0$ into this yields

$$\sum_{i=1}^{N} x_i y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x}\right) x_i - \hat{\beta}_1 x_i^2 = 0$$

The summation applies to everything in the equation above. Using the property that a constant term can be removed from a summation and distributing the sum to each term, we obtain:

$$\sum_{i=1}^{N} x_i y_i - \bar{y} \sum_{i=1}^{N} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

From this same summation property, $\sum_{i=1}^{N} y_i = N\bar{y}$ and $\sum_{i=1}^{N} x_i = N\bar{x}$.

Using these for $\hat{\beta}_1$ we obtain:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}$$

Finally, we can use the facts that $\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}$ and $\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - N\bar{x}^2$, which stem from basic algebra, to substitute these two properties into the above equation and obtain the standard notation for the OLS slope estimator:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}.$$

This is the “best fit" explanation of the how the dependent variable changes with respect to the independent variable.

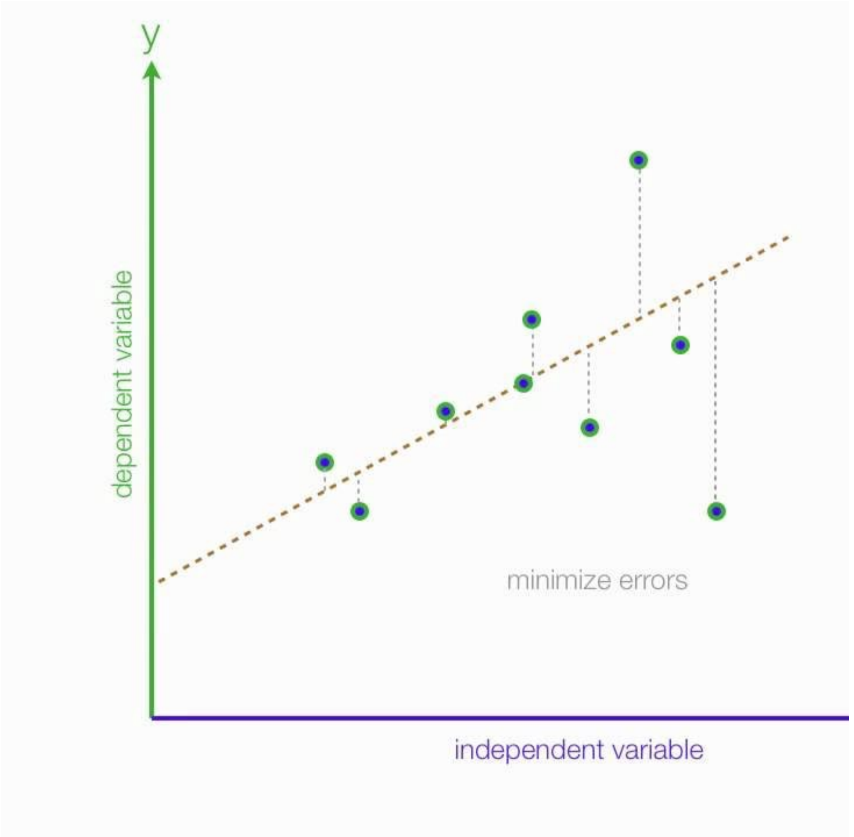
LINEAR REGRESSION: INPUTS & OUTPUTS

- We can enter a set of data, containing ***N*** observations of a Y value and an X value, and an equation that we believe is a good explanation of the relationship between Y and X for the sample

- In Stata we could enter the following for a linear regression:

  regress Y X

  - The coefficient (β_1) that will be returned from this input command represents the slope of the best fit line, and this indicates how a one unit change in X will affect Y
  - We will also obtain an intercept for the vertical axis: graphically this is the value of the line of best fit when X=0



LINEAR REGRESSION: MULTIPLE VARIABLES

- A set of data containing ***N*** observations of a Y value and an X_1 value and an X_2 value can also be used, or potentially a data structure with many independent variables

- In Stata we could enter the following for this linear regression with multiple variables:

  regress Y X_1 X_2

  - The coefficients (β_1, β_2) that will be returned from this input command indicate how a one unit change in each of these two X variables will separately affect Y

REAL-WORLD APPLICATION 1: WAGE EQUATION

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

- Y_i = wage for individual i
- β_0 = intercept
- β_1 = effect on Y of a 1 unit increase in X
- X_i = years of education for individual i
- ε_i = error term

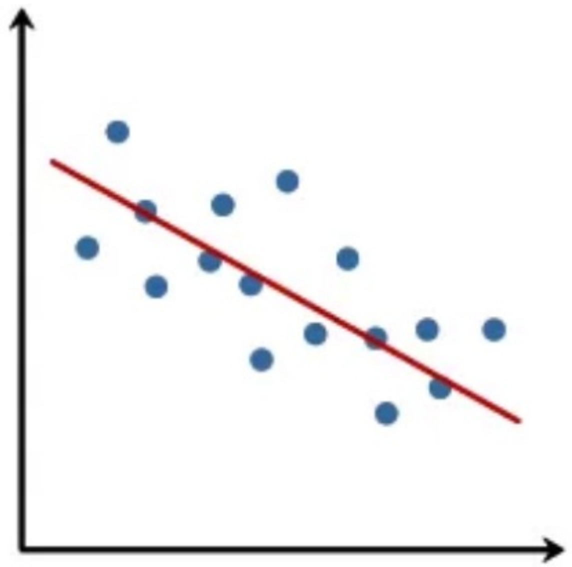
- Suppose β_0 = 10,000 and β_1 = 4000 : the basic interpretation is that an average random individual from the sample is predicted to earn $10,000 plus $4000 per year of education
- Someone with 12 total years of schooling is predicted to earn: 10,000+12(4000) = $58,000
- Someone with 16 total years of schooling is predicted to earn: 10,000+16(4000) = $74,000

Question: *Do you think anything is missing with this model?*

REAL-WORLD APPLICATION 2: REGULATIONS & BREWERIES

Linear regression equation: $Y_s = \beta_0 + \beta_1 X_s + \varepsilon_s$

- Y_s = brewery count in state s
- β_0 = intercept
- β_1 = effect on Y of a 1 unit increase in X
- X_s = "entry barrier" regulations in state s
- ε_s = error term



- Suppose $\beta_0 = 600$ and $\beta_1 = -10$: the basic interpretation is that each additional regulation in a state is results in a decrease of 10 breweries in that state
- A state with 40 entry barrier regulations is expected to have 600 + 40(-10) = 200 breweries

REAL-WORLD APPLICATION 3: ATTENDANCE & EXAM SCORES

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

- Y_i = score for individual i
- β_0 = intercept
- β_1 = effect on Y of a 1 unit increase in X_1
- X_{1i} = classes attended for individual i
- ε_i = error term

- Suppose $\beta_0 = 78$ and $\beta_1 = 1.3$: the basic interpretation for this intercept and coefficient is that attending zero classes obtains a 78 on average and each class attended increases exam score by 1.3 points on average

- Someone with 10 classes attended is predicted to score 78 + 10(1.3) = 91

- What may be missing? Is it correlated with the included "exogenous" X-variable in this specification?

MULTIPLE VARIABLES EXAMPLE: EXAM SCORES BY ATTENDANCE & STUDY HOURS

Linear regression equation with multiple explanatory variables: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^{0.5} + \varepsilon_i$

- Y_i = score for individual i
- β_0 = intercept
- β_1 = effect on Y of a 1 unit increase in X_1
- X_{1i} = classes attended for individual i
- β_2 = effect on Y of a 1 unit increase in X_2
- X_{2i} = total hours of work and study for individual i
- ε_i = error term

- Suppose β_0 = 75 and b_1 = 0.7 and β_2 = 1.5 : now we can see that there is a positive correlation between attendance and total hours of work, which obviously matters but was not included before (omitted variable bias)
- Including this new exogenous variable decreases the extent to which attendance explains the outcome, but this probably increases the overall accuracy and explanatory power (which is called R^2) of the regression
- Also note the functional form of this new variable: hours studied has diminishing marginal returns
- Someone with 10 sessions attended and 64 hours of work is predicted to score 75 + 10(0.7) + 1.5(8) = 94.0
- Someone with 3 sessions attended and 36 hours of work is predicted to score 75 + 3(0.7) + 1.5(6) = 86.1

WAGE EQUATIONS WITH SEVERAL VARIABLES & DIFFERENT FUNCTIONAL FORMS

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^{0.5} + \beta_3 X_{3i}^{0.5} + \beta_4 X_{4i} + \varepsilon_i$

- Y_i = annual income for individual i
- β_0 = intercept
- $\beta_1, \beta_2, \beta_3, \beta_4$: slope coefficients
- X_{1i} = total years of education for individual i
- X_{2i} = total years of work experience for individual i
- X_{3i} = total number of years that individual i has spent in prison
- X_{4i} = total number of years that individual i has been married
- ε_i = error term

- Suppose that $\beta_0 = 15000$, $\beta_1 = 2000$, $\beta_2 = 7000$, $\beta_3 = -8000$, $\beta_4 = 1000$
- A person with 16 years of education, 4 years of work experience, and zero criminal convictions who has never been married is expected to have annual income of 15000 + 2000(16) + 7000(2) – 8000(0) + 1000(0) = $61,000
- A person with 18 years of education, 25 years of work experience, and zero criminal convictions who has been married for 16 years is expected to have annual income of 15000 + 2000(18) + 7000(5) – 8000(0) + 1000(4) = $90,000

UNDERSTANDING CAUSATION & ENDOGENEITY CONCERNS

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^{0.5} + \beta_3 X_{3i}^{0.5} + \beta_4 X_{4i} + \varepsilon_i$

- Y_i = annual income for individual i
- β_0 = intercept
- β_1 , β_2 , β_3 , β_4 : slope coefficients
- X_{1i} = total years of education for individual i
- X_{2i} = total years of work experience for individual i
- X_{3i} = total number of years that individual i has spent in prison
- X_{4i} = total number of years that individual i has been married
- ε_i = error term

- Suppose that $\beta_0 = 15000$, $\beta_1 = 2000$, $\beta_2 = 7000$, $\beta_3 = -8000$, $\beta_4 = 1000$
- A person with 12 years of education, 4 years of work experience, 9 years spent in prison, and who has never been married is expected to have annual income of $15000 + 2000(12) + 7000(2) - 8000(3) + 1000(0) = \$29,000$
- Is it reasonable to assume that spending time in prison is the reason for a low income instead of the possibility that having a low income might have been a significant factor affecting that person's initial decision to get involved in criminal activity?
  - The data often does not conclusively tell us the "causal flow direction" - what happened first - making it impossible to know which was the cause and which was the effect: this ambiguity is usually called ***endogeneity*** and it creates major limitations and accuracy issues in statistical models
  - The easiest way to deal with this, if possible, is to choose outcome variables measured at a later point in time than the explanatory variables

USING BINARY INDICATOR "DUMMY" VARIABLES & MULTICOLLINEARITY ISSUES

- Sometimes it is either more accurate or more convenient to represent variables as a binary measure
  - An "indicator variable" equal to either 1 or 0 that represents whether someone is or is not a convicted felon is usually more accurate and efficient for statistical modeling than some continuous measure of years in jail
  - The effect of a binary variable in a standard model is always either zero or the value of its slope estimate

- Consider a variable like marriage: if someone is married then they are statistically more likely to have certain personality, lifestyle, and other attributes that positively correlate with being a good employee
  - Years of marriage is generally not as good of a measure as a simple binary variable representing whether the person is currently married
  - Potential concern: marriage could increase or potentially decrease a person's incentives to earn lots of money
  - Marriage status is correlated with age, education, etc… and potentially affected by prior income or lack of income

<u>It is very complicated to find the most accurate and efficient combination and format of variables in empirical analysis!</u>

USING BINARY INDICATOR "DUMMY" VARIABLES

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^{0.5} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$

- Y_i = annual income for individual i
- β_0 = intercept
- $\beta_1, \beta_2, \beta_3, \beta_4$: slope coefficients
- X_{1i} = total years of education for individual i
- X_{2i} = total years of work experience for individual i
- X_{3i} = binary indicator set equal to 1 if individual i has ever been in prison
- X_{4i} = binary indicator set equal to 1 if individual i has ever been married
- ε_i = error term

- Suppose that $\beta_0 = 15000$, $\beta_1 = 2000$, $\beta_2 = 7000$, $\beta_3 = -15000$, $\beta_4 = 13000$
- A person with 16 years of education, 4 years of work experience, and zero criminal convictions who has never been married is expected to have annual income of 15000 + 2000(16) + 7000(2) − 15000(0) + 13000(0) = $61,000
- A person with 16 years of education and 4 years of work experience who has been to prison and has never been married is expected to have annual income of 15000 + 2000(16) + 7000(2) − 15000(1) + 13000(0) = $46,000
- A person with 16 years of education, 4 years of work experience, and zero criminal convictions who has been married is expected to have annual income of 15000 + 2000(16) + 7000(2) − 15000(0) + 13000(1) = $74,000

USING BINARY INDICATORS AND CONTINUOUS VARIABLES TOGETHER

Linear regression equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^{0.5} + \beta_3 X_{3i} + \beta_4 X_{4i}^{0.5} + \beta_5 X_{5i} + \beta_6 X_{6i}^{0.5} + \varepsilon_i$

- Y_i = annual income for individual i
- β_0 = intercept
- β_1 , β_2 , β_3 , β_4 , β_5 , β_6 : slope coefficients
- X_{1i} = total years of education for individual i
- X_{2i} = total years of work experience for individual i
- X_{3i} = binary indicator set equal to 1 if individual i has ever been in prison
- X_{4i} = total number of years that individual i has spent in prison
- X_{5i} = binary indicator set equal to 1 if individual i has ever been married
- X_{6i} = total number of years that individual i has been married
- ε_i = error term

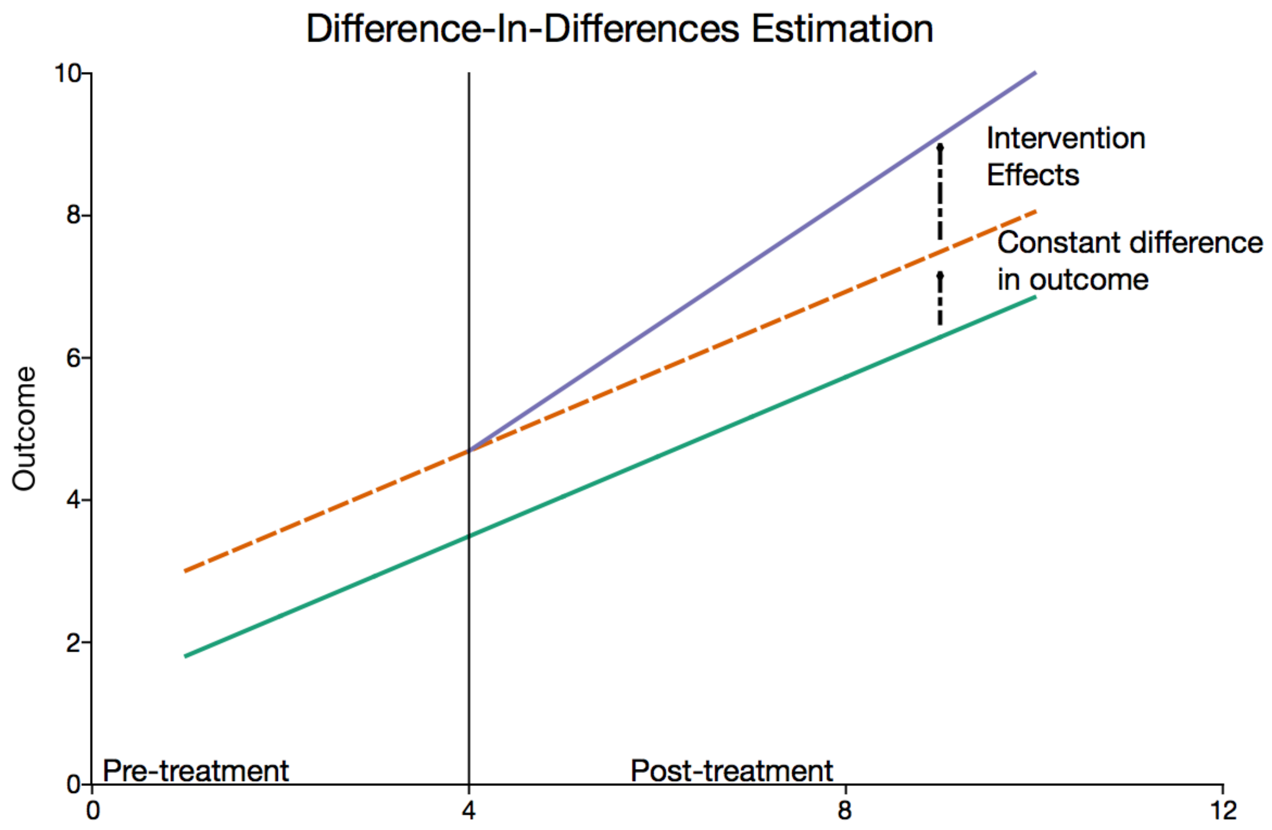
- Suppose that $\beta_0 = 15000$, $\beta_1 = 2000$, $\beta_2 = 7000$, $\beta_3 = -8000$, $\beta_4 = -3000$, $\beta_5 = 6000$, $\beta_6 = 2000$
- A person with 16 years of education, 9 years of work experience, and zero criminal convictions who has been married 4 years is expected to have annual income of $15000 + 2000(16) + 7000(3) - 8000(0) - 3000(0) + 6000(1) + 2000(2) = \$78,000$
- A person with 16 years of education, 9 years of work experience, who spent 4 years in prison and has been married for 1 year is expected to have annual income of $15000 + 2000(16) + 7000(3) - 8000(1) - 3000(2) + 6000(1) + 2000(1) = \$62,000$

DIFFERENCES IN DIFFERENCES ESTIMATION

- Differences in Differences (often abbreviated DD or DiD) is a way to analyze the effect of a “treatment” by comparing trends in outcomes while taking pre-existing trajectories of the outcomes into account

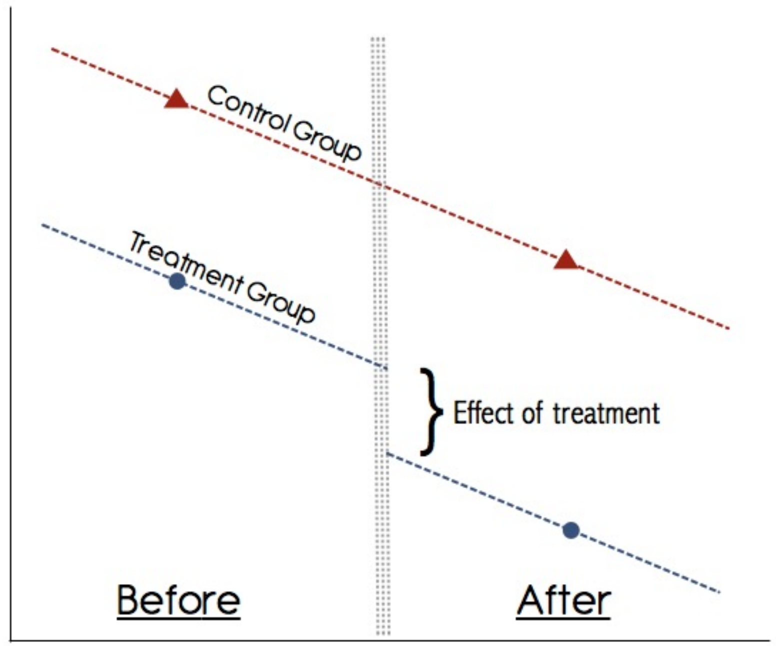
- Example: Imagine two hospital patients who are both recovering from an illness at the same rate and receiving medicine. One patient receives an additional treatment T and then her rate of improvement increases afterwards. Using a DD approach can estimate the isolated effect of T on whatever quantitative measure of health is being observed as the outcome.

DIFFERENCES IN DIFFERENCES ESTIMATION



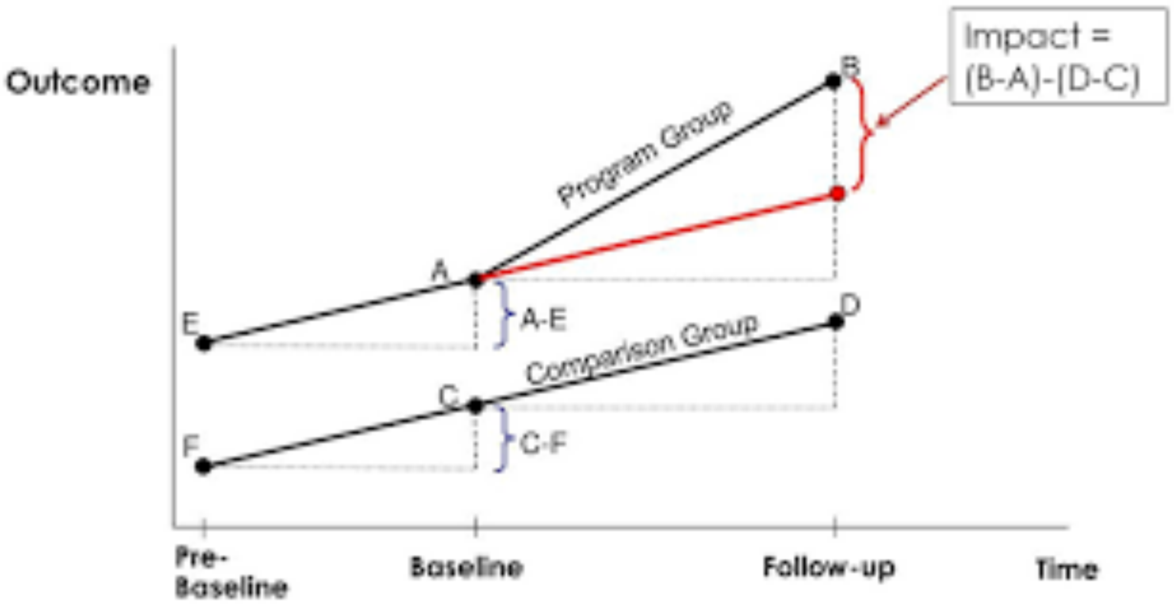
- DiD approach estimates the *change in the trend* compared to the pre-existing trend
- The green line above would represent a control group with the orange representing a treatment group

DIFFERENCES IN DIFFERENCES ESTIMATION



- DiD approach can capture both immediate and delayed effects compared to the counterfactual trend
- There can be immediate shifts, changes in the slope, or a combination of both, or of course possibly neither

DIFFERENCES IN DIFFERENCES ESTIMATION



- Intuitively, (D-C) is the change that would have happened anyway and (B-A) is this change plus the treatment effect, so we subtract the measured change from a "control group" to isolate the treatment effect

- It is extremely important to have a properly comparable control group: if there are differences between the treatment group observations and control group observations besides the treatment then that could bias the results

DIFFERENCES IN DIFFERENCES ESTIMATION: POLICY EXAMPLE

- Howard Bodenhorn (2016) “Blind tigers and red-tape cocktails”: Found that early local Prohibition policies in the late 1800s increased homicides by 50% in dry counties
    - Used DiD approach to compare outcomes across counties in South Carolina
    - After a county outlawed alcohol there was a massive increase in violence compared to overall trends
    - Logic: buyers and sellers resorting to “black markets” did not have any government to resolve commercial disputes
    - Government could not play the role of “monopolist of violence” so market participants took matters into their own hands
    - If homicide rates were not constant over time then a DiD approach is necessary to disentangle ”treatment effect”

MINIMUM WAGES & EMPLOYMENT:
A CASE STUDY OF THE FAST-FOOD INDUSTRY IN NEW JERSEY AND PENNSYLVANIA

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)

- NJ minimum wage significantly increased in 1992
  - Presumably exogenous minimum wage law change: “treatment” in this model

- Classical economic theory predicts a decrease in employment and production
  - Input cost of labor increases: firms expected to respond by shifting towards capital and decreasing output
  - Fast food industry is considered highly competitive

- Compared employment in 410 fast-food restaurants in NJ and eastern PA before and after the change
  - Conducted multiple waves of phone interviews with extremely high response rate (87% overall; 91% NJ, 73% PA)
  - Follow-ups and physical visits to verify closed stores
  - No substantial evidence of selection bias / non-response bias

MINIMUM WAGES & EMPLOYMENT: IMPLEMENTATION

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)

- Initially very similar wages ($4.61 NJ and $4.63 PA) across the two groups but higher prices in NJ stores

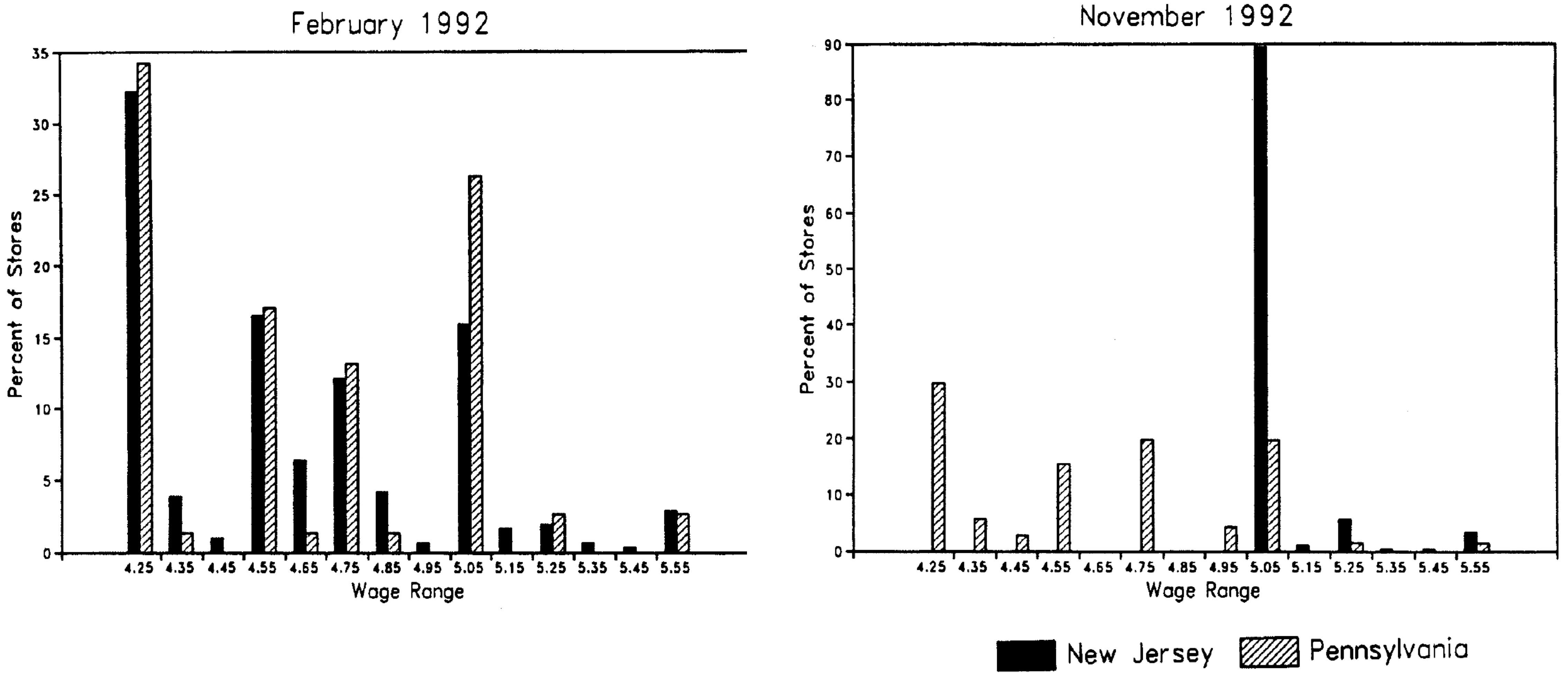
- Two year delay in New Jersey's implementation of the higher minimum wage
  - Policy change was made in good economic conditions before a mild to moderate economic recession occurred
  - Challenges to stop the policy change before implementation failed: these might have caused firms to delay taking potential action in response because there was uncertainty about whether the minimum wage increase would actually happen
  - Wage increased in multiple steps months apart: likely a policy design to minimize negative response from firms

- Initially **23.3** full-time equivalent (FTE) workers in PA and **20.4** in NJ

- From Feb 1992 to Nov 1992 the distribution of wages shifted substantially – especially in NJ
  - NJ initially had clusters around $4.25 and other values below the post-hike minimum and shifted to one huge cluster just barely above the new minimum around $5.05: result was almost uniformity in wages after this policy change took effect

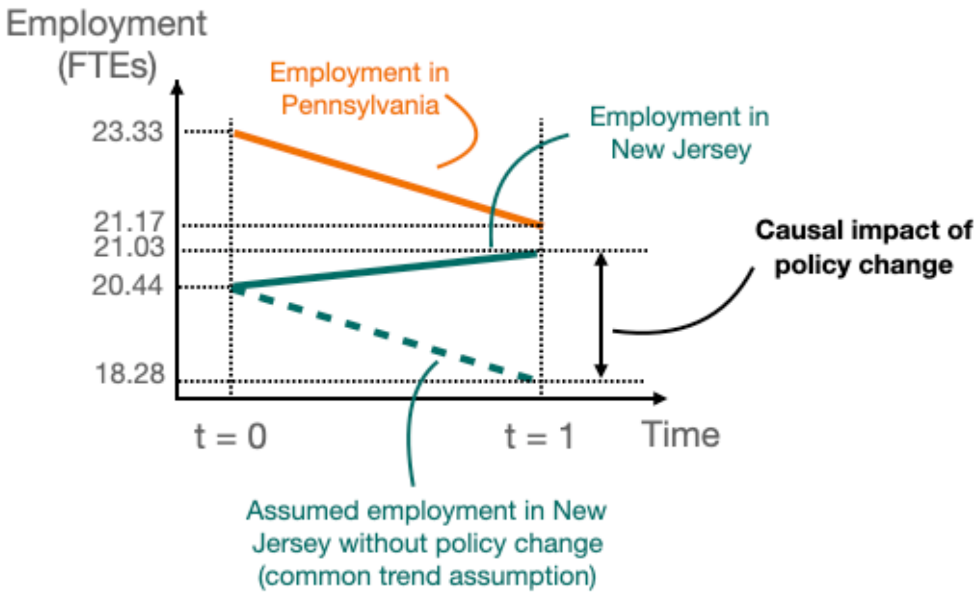
MINIMUM WAGES & EMPLOYMENT: RESULTS

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)



MINIMUM WAGES & EMPLOYMENT: MAIN RESULTS

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)



- Main DiD estimation result: **relative *gain* of +2.76 FTE employees in NJ** fast food restaurants over the time period when NJ minimum wage increases were implemented

- In NJ, stores were initially smaller on average but more increased in size, and employment expanded at lower-wage stores while employment contracted at higher-wage stores
  - These results contradict the predictions of traditional economic theory about how firms behave

- High-wage NJ stores had a change in FTE employees similar to comparable PA stores (-2.16 vs. -2.28)

MINIMUM WAGES & EMPLOYMENT: ROBUSTNESS CHECKS & OTHER FACTORS

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)

- Tested for reduction in employee benefits as a way for firms to "make up" additional costs of wage increase: found that fringe benefits / other compensation forms matched changes in wages
  - Starting bonuses, employee discounts, training did not appear to decline to offset higher labor costs
  - This suggests that firms did not find other ways to reduce total compensation for workers

- Effects of economic recession on consumer choices?
  - Demand might actually have increased for fast food if consumers shift spending towards cheaper options
  - Employment declines are obviously common during recessions, hence the decreases observed in this study
  - These two effects predicted by standard economic reasoning work in opposite directions

- One potential explanation for the unexpected findings: "unobserved demand shocks" in NJ could have outweighed negative employment effects
  - Tested for this using subsamples (Newark area and Camden area) and did not find evidence of demand shocks

DISCUSSION: MINIMUM WAGES & EMPLOYMENT

CARD & KRUEGER 1994 (AMERICAN ECONOMIC REVIEW)

- Are their experimental design and results convincing?
  - Could there have been other simultaneous economic or policy change "treatment effects" which are not accounted for?
  - Is the sample in eastern PA a good "control group" to compare to the "treatment group" restaurants in NJ?
  - Do fast-food restaurants choose output quantity in a way that is similar to classical industrial production examples?
  - To what extent are labor and capital substitute inputs in fast food and how has this changed over time?

- In what ways might responses to minimum wage hikes and ability to substitute inputs vary across industries?
  - Healthcare (or anything with inelastic demand and specialized labor on the supply side) – probably less flexibility on quantity and quality changes for firms

- What are the limitations or modern economic and policy implications of this analysis?
  - How high could minimum wages be raised before these findings no longer hold?
  - What effects to we expect to see now that many CA fast food workers just unionized a few days ago?