Policy Evaluation - Regression

March 14, 2024

Estimating Treatment Effects Review

- $ATE =$

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$
- $ATE_{est} =$

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$
- $ATE_{est} = Avg_n[Y_i^1|D_i = 1] - Avg_n[Y_i^0|D_i = 0]$

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$
- $ATE_{est} = Avg_n[Y_i^1 | D_i = 1] - Avg_n[Y_i^0 | D_i = 0]$
- When $(Y^1, Y^0) \not\perp D$:

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$
- $ATE_{est} = Avg_n[Y_i^1 | D_i = 1] - Avg_n[Y_i^0 | D_i = 0]$
- When $(Y^1, Y^0) \not\perp D$:

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0 | D_i = 1] - Avg_n[Y_i^0 | D_i = 0]\}}_{\text{Selection Bias}}$$

$$+ \underbrace{(1 - \pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias}}$$

Estimating Treatment Effects Review

- $ATE = Avg_n[Y_i^1 - Y_i^0]$
- $ATE_{est} = Avg_n[Y_i^1 | D_i = 1] - Avg_n[Y_i^0 | D_i = 0]$
- When $(Y^1, Y^0) \not\!\perp\!\!\!\perp D$:

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0 | D_i = 1] - Avg_n[Y_i^0 | D_i = 0]\}}_{\text{Selection Bias}} + \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias}}$$

- Solution?

Estimating Treatment Effects Review

- Randomize treatment exposure so that $(Y^1, Y^0) \perp\!\!\!\perp D$.

Estimating Treatment Effects Review

- Randomize treatment exposure so that $(Y^1, Y^0) \perp\!\!\!\perp D$.

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0|D_i = 1] - Avg_n[Y_i^0|D_i = 0]\}}_{\text{Selection Bias} = 0}$$

$$+ \quad \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias} = 0}$$

Estimating Treatment Effects Review

- Randomize treatment exposure so that $(Y^1, Y^0) \perp\!\!\!\perp D$.

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0|D_i = 1] - Avg_n[Y_i^0|D_i = 0]\}}_{\text{Selection Bias} = 0}$$

$$+ \quad \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias} = 0}$$

- Added bonus when LLNs applies:

Estimating Treatment Effects Review

- Randomize treatment exposure so that $(Y^1, Y^0) \perp\!\!\!\perp D$.

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0|D_i = 1] - Avg_n[Y_i^0|D_i = 0]\}}_{\text{Selection Bias} = 0}$$

$$+ \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias} = 0}$$

- Added bonus when LLNs applies: Sample ATE = Population ATE

Estimating Treatment Effects Review

- Randomize treatment exposure so that $(Y^1, Y^0) \perp\!\!\!\perp D$.

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y_i^0|D_i = 1] - Avg_n[Y_i^0|D_i = 0]\}}_{\text{Selection Bias} = 0}$$
$$+ \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias} = 0}$$

- Added bonus when LLNs applies: Sample ATE = Population ATE

$$\{Avg_n[Y_i^1|D_i = 1] - Avg_n[Y_i^0|D_i = 0]\} = \{E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 0]\}$$

Policy Evaluation - Randomization

- So randomization is the ideal and the "gold standard" of research methodologies.

Policy Evaluation - Randomization

- So randomization is the ideal and the “gold standard” of research methodologies.
- But...in policy analysis it’s often (though not always) impossible to randomize individuals in our sample into treatment?

- So randomization is the ideal and the "gold standard" of research methodologies.
- But...in policy analysis it's often (though not always) impossible to randomize individuals in our sample into treatment?
- How can we attempt to identify causal effects of policy when randomization is infeasible?

Policy Evaluation - Regression Analysis

- **Regression Analysis**

- **Regression Analysis**
  - Assume that when key observed variables have been "made equal" across treatment and control groups, omitted variable bias from unobserved confounders will also be eliminated.

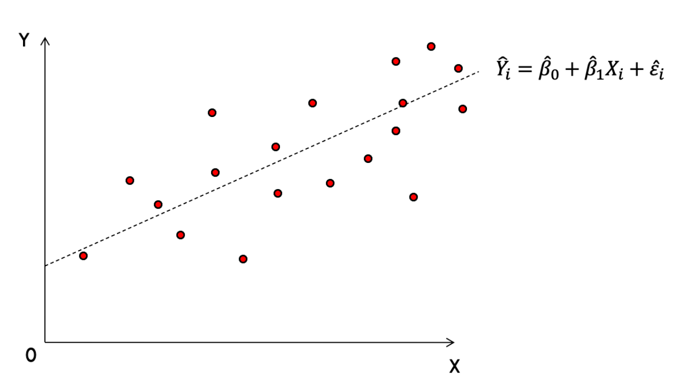
- **Regression Analysis**
  - Assume that when key observed variables have been "made equal" across treatment and control groups, omitted variable bias from unobserved confounders will also be eliminated.
  - Even if D is not randomly assigned, we can still *hope* that $(Y^1, Y^0) \perp\!\!\!\perp D$.

- **Regression Analysis**
  - Assume that when key observed variables have been "made equal" across treatment and control groups, omitted variable bias from unobserved confounders will also be eliminated.
  - Even if D is not randomly assigned, we can still *hope* that $(Y^1, Y^0) \perp\!\!\!\perp D$.
  - And that $ATE_{est} = ATE$.

- **Regression Analysis**
  - Assume that when key observed variables have been “made equal” across treatment and control groups, omitted variable bias from unobserved confounders will also be eliminated.
  - Even if D is not randomly assigned, we can still *hope* that $(Y^1, Y^0) \perp\!\!\!\perp D$.
  - And that $ATE_{est} = ATE$.
  - BUT THIS NOT “GUARANTEED” LIKE IT IS WITH RANDOMIZATION AND DEPENDS ON A STRONG (UNTESTABLE) ASSUMPTION.

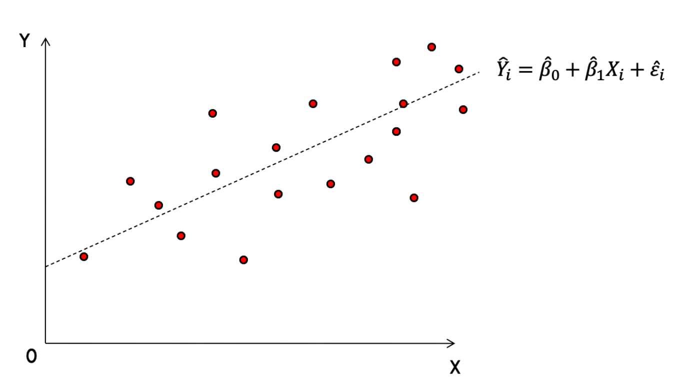
Policy Evaluation - Regression

- So how does regression actually work?



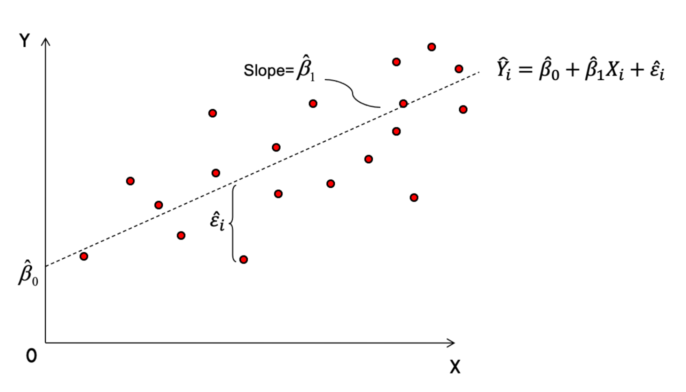
Policy Evaluation - Regression

- So how does regression actually work?

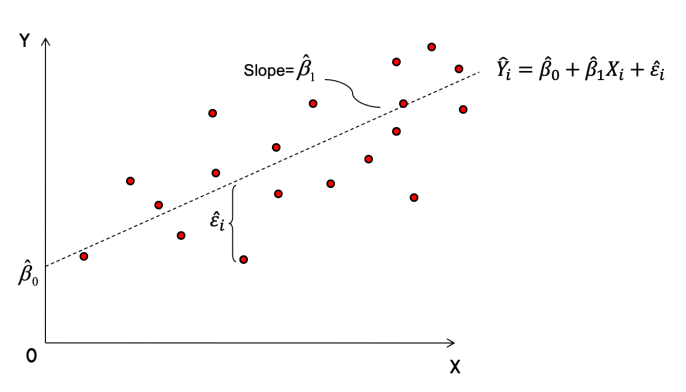


- Note that we have not randomized people into X in this case. We simply observe an individual's value of X in our sample.

Policy Evaluation - Regression



Policy Evaluation - Regression



- How does OLS use the regression residuals to fit a trend line to a series of data points?

- Suppose that we estimate a linear regression model of the following form: $y_{it} = \beta_0 + \beta_1 Edu_{it} + \varepsilon_{it}$

  Where y is the number of times that individual *i* visits the emergency department in year *t*

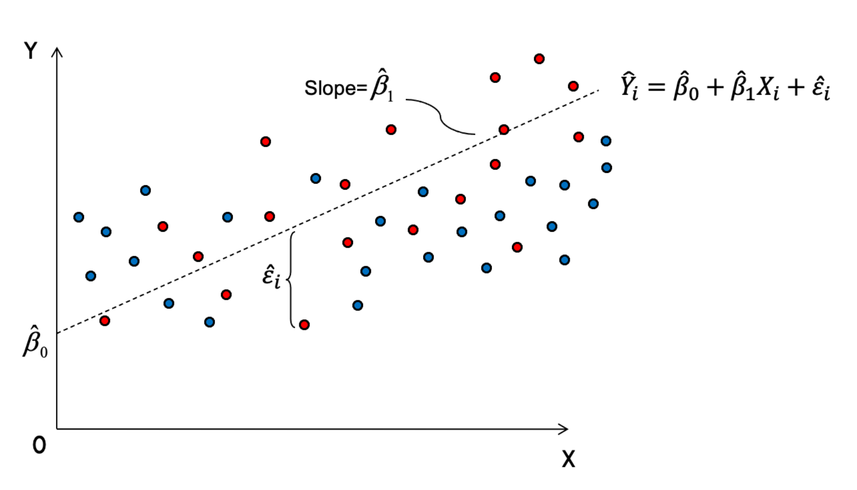
  Edu is individual *i*'s education in years (e.g. a value of 12 represents a high school education, a value of 14 represents a college education, etc.).

  Our coefficient estimates are as follows: $\hat{\beta}_0$=4.0 and $\hat{\beta}_1$=-0.25.

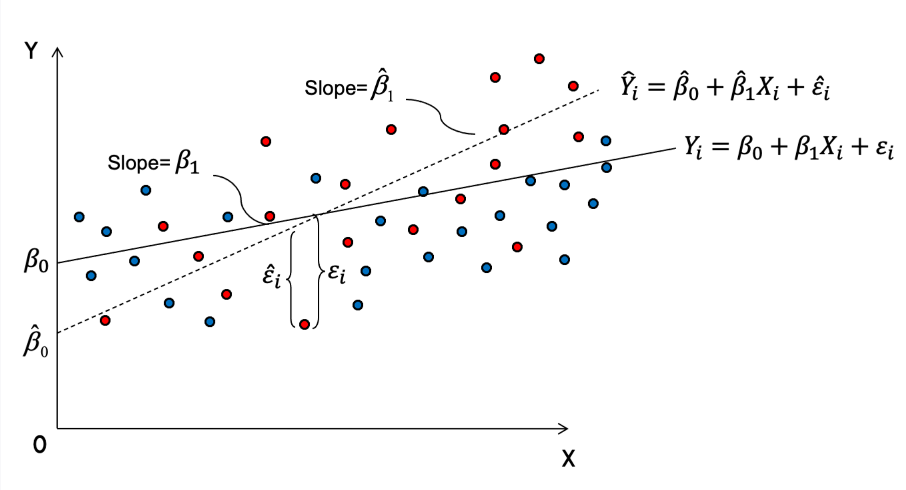
  How do we interpret these coefficient estimates?

Policy Evaluation - Regression

- But suppose the *population* distribution is...



Policy Evaluation - Regression



Policy Evaluation - Regression

- $\hat{\beta}_1$ is an estimate of the true β_1, the relationship between X (the independent variable) and Y (the dependent variable)

Policy Evaluation - Regression

- $\hat{\beta}_1$ is an estimate of the true β_1, the relationship between X (the independent variable) and Y (the dependent variable)
- How reliable is our estimate?

Policy Evaluation - Regression

- $\hat{\beta}_1$ is an estimate of the true β_1, the relationship between X (the independent variable) and Y (the dependent variable)
- How reliable is our estimate?
  - Variance of the sampling distribution of the sample mean
    - Variance of the distribution of β's from independent samples of the population
    - Equal to the variance of the population distribution divided by n (number of samples)
    - Square root of this variance is the standard error

Policy Evaluation - Regression

- $\hat{\beta}_1$ is an estimate of the true β_1, the relationship between X (the independent variable) and Y (the dependent variable)
- How reliable is our estimate?
  - Variance of the sampling distribution of the sample mean
    - Variance of the distribution of β's from independent samples of the population
    - Equal to the variance of the population distribution divided by n (number of samples)
    - Square root of this variance is the standard error
- Inference
  - Calculate a t-statistic and corresponding p-value
  - $C.I._{\hat{\beta}} = \hat{\beta} \pm t^* \times SE(\hat{\beta})$

Policy Evaluation - Regression

- $\hat{\beta}_1$ is an estimate of the true β_1, the relationship between X (the independent variable) and Y (the dependent variable)
- How reliable is our estimate?
  - Variance of the sampling distribution of the sample mean
    - Variance of the distribution of β's from independent samples of the population
    - Equal to the variance of the population distribution divided by n (number of samples)
    - Square root of this variance is the standard error
- Inference
  - Calculate a t-statistic and corresponding *p*-value
  - $C.I._{\hat{\beta}} = \hat{\beta} \pm t^* \times SE(\hat{\beta})$
  - Suppose you estimate the regression model above and find that the coefficient estimate of β_1 has a p-value of 0.115 from a two-sided t-test. What does that mean and what would you conclude about the statistical significance of the relationship between education and emergency department use?

Policy Evaluation - Regression

- Note that at this point, regression doesn't look all that different from randomization.

Policy Evaluation - Regression

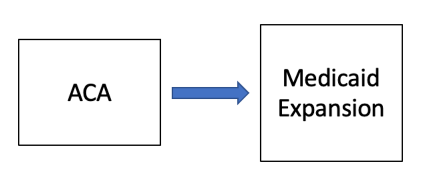
- Note that at this point, regression doesn't look all that different from randomization.
  - In both cases we're just comparing means of treated and control groups from a sample.

Policy Evaluation - Regression

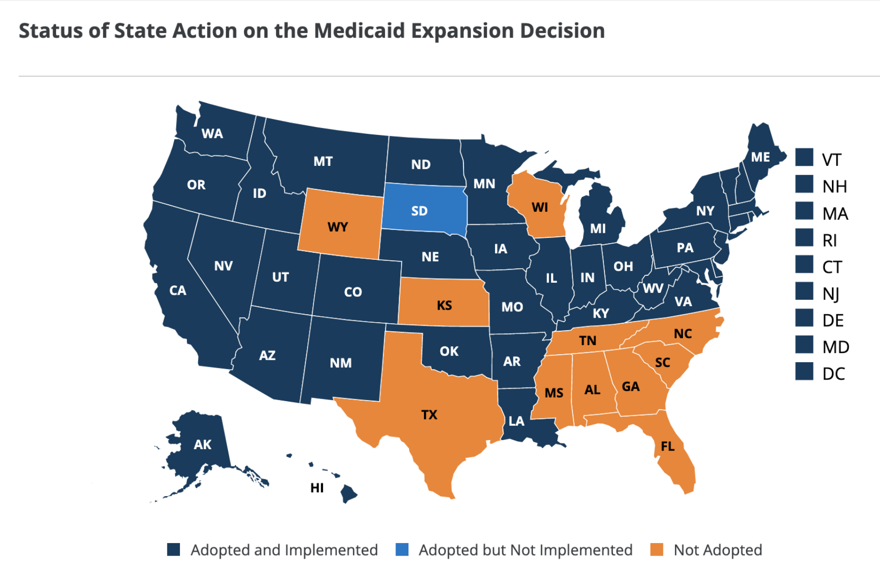
- Note that at this point, regression doesn't look all that different from randomization.
  - In both cases we're just comparing means of treated and control groups from a sample.
- Where these methods begin to diverge is when we add additional covariates to our regression model to "control" for potental confounders.

Policy Evaluation - Regression

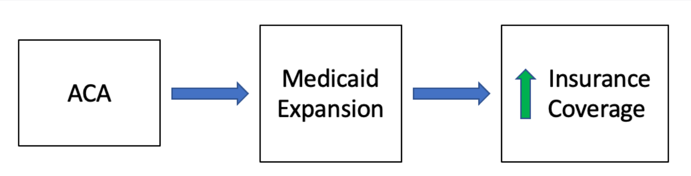
- Medicaid Expansion and Mortality



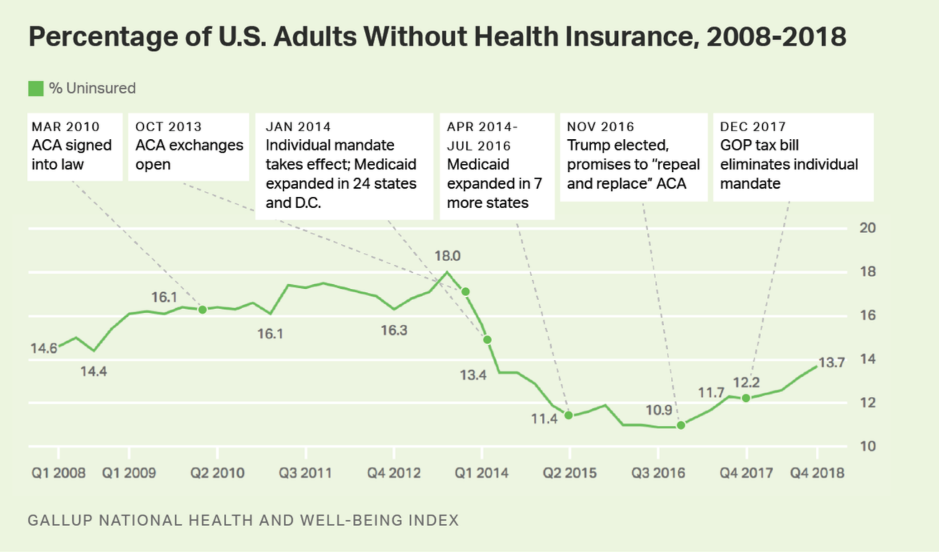
Policy Evaluation - Regression



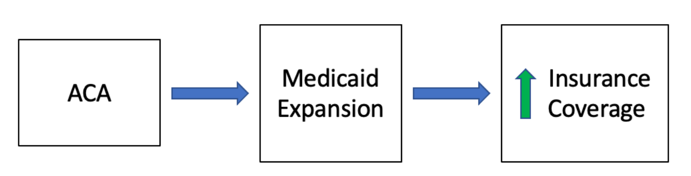
Policy Evaluation - Regression



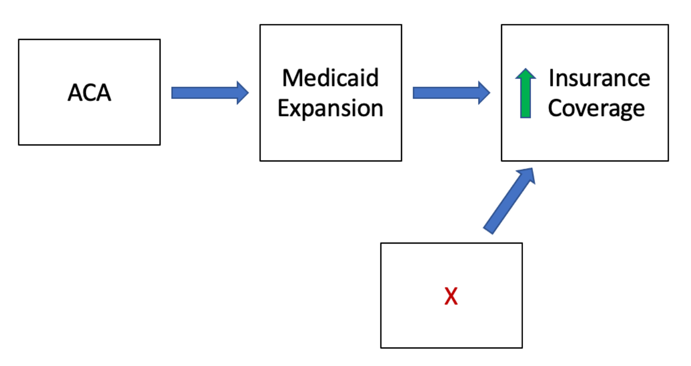
Policy Evaluation - Regression



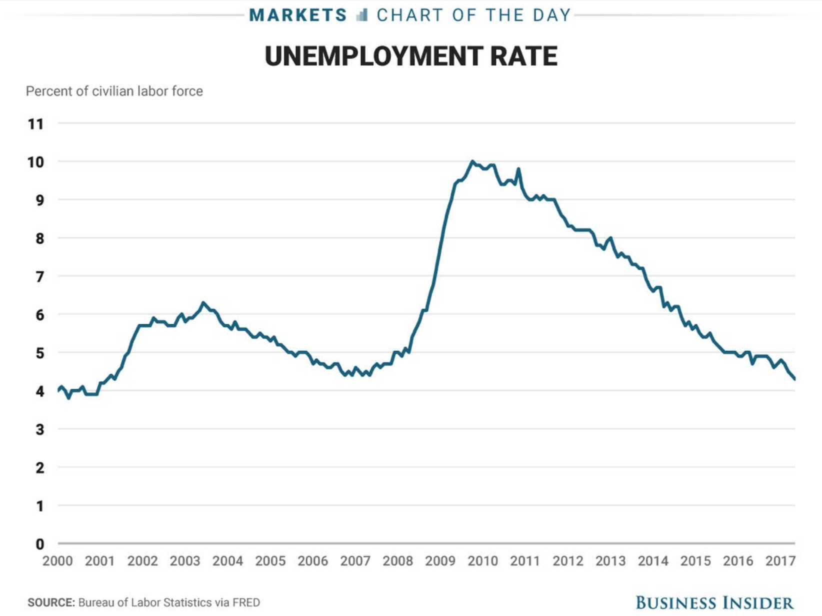
Policy Evaluation - Regression



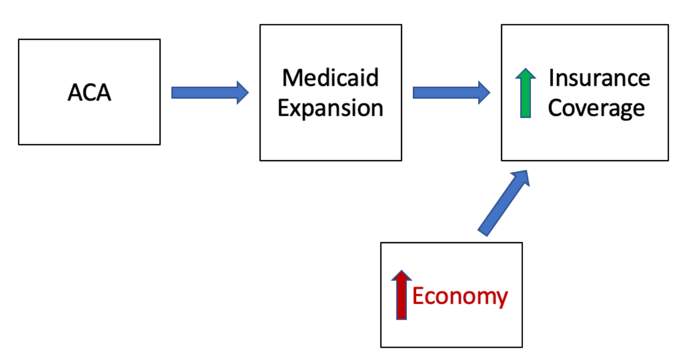
Policy Evaluation - Regression



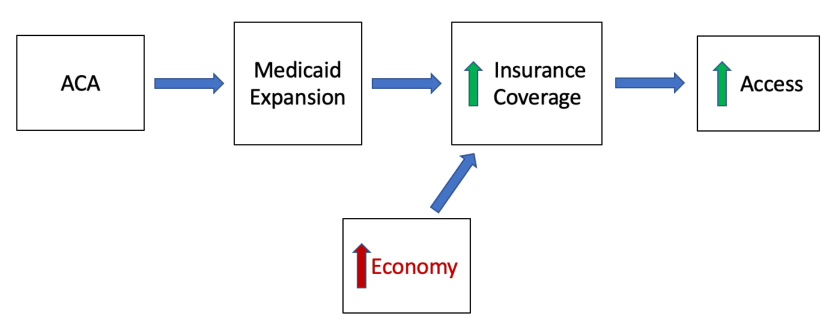
Policy Evaluation - Regression



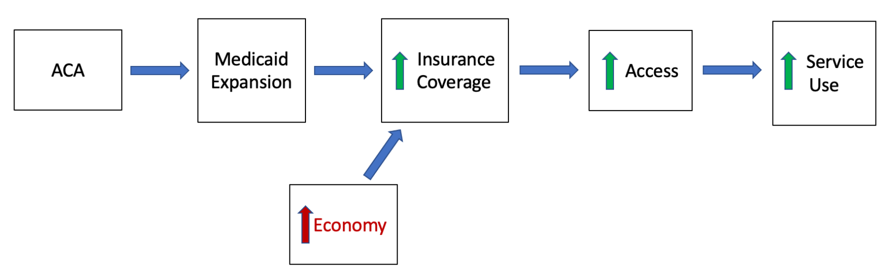
Policy Evaluation - Regression



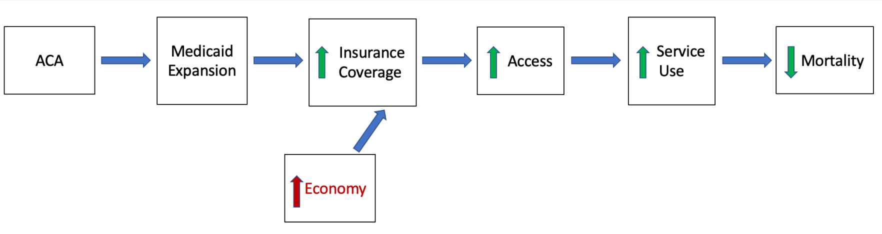
Policy Evaluation - Regression



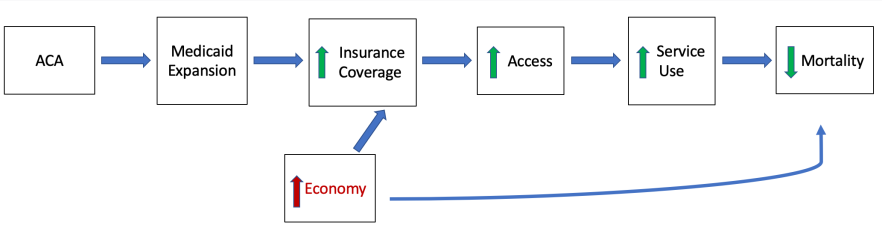
Policy Evaluation - Regression



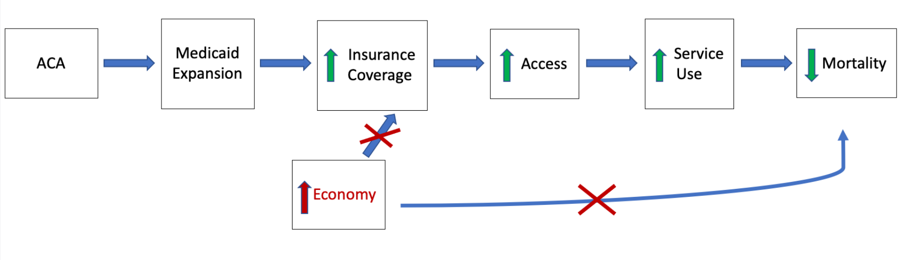
Policy Evaluation - Regression



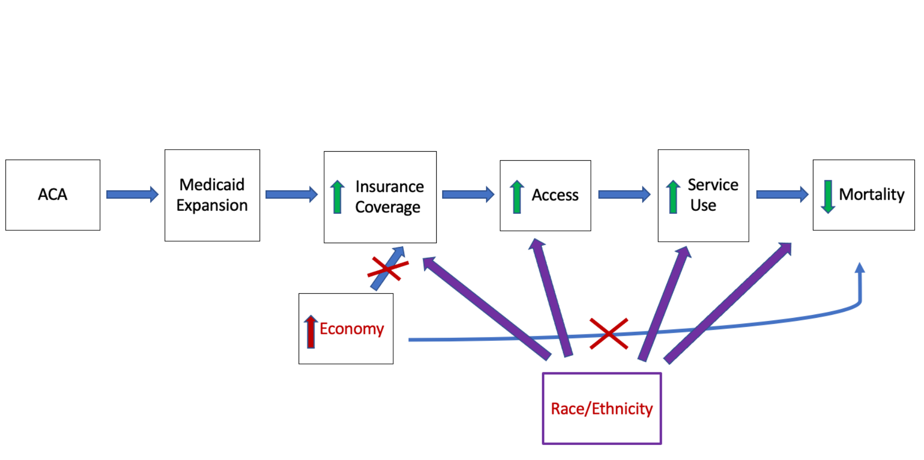
Policy Evaluation - Regression



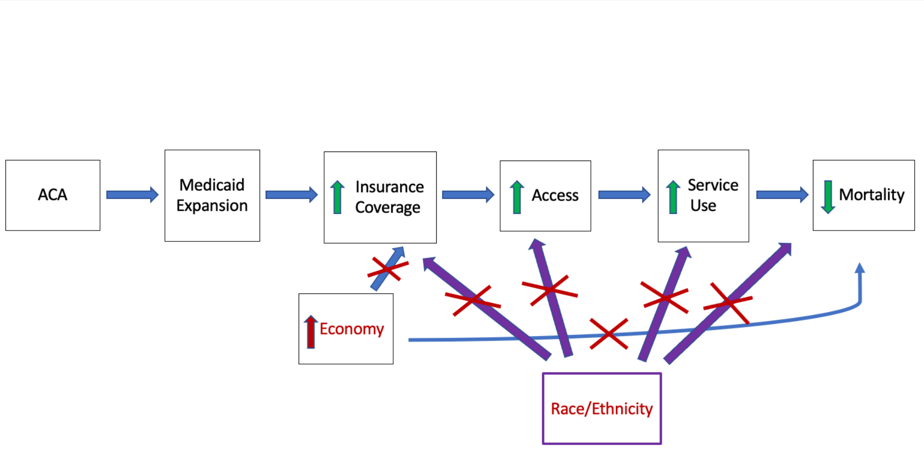
Policy Evaluation - Regression



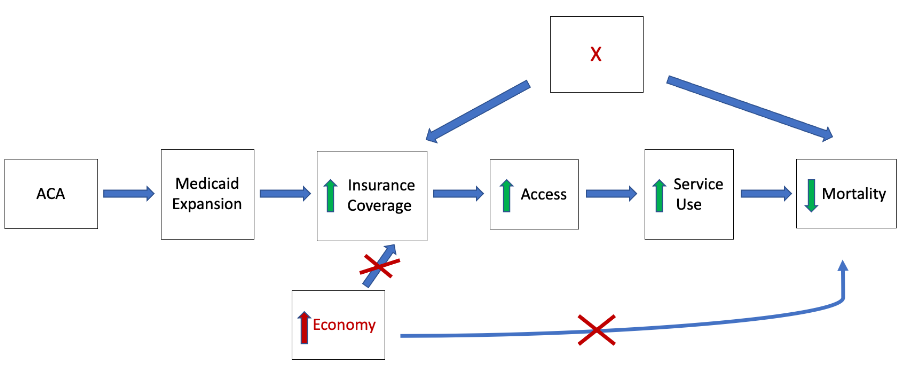
Policy Evaluation - Regression



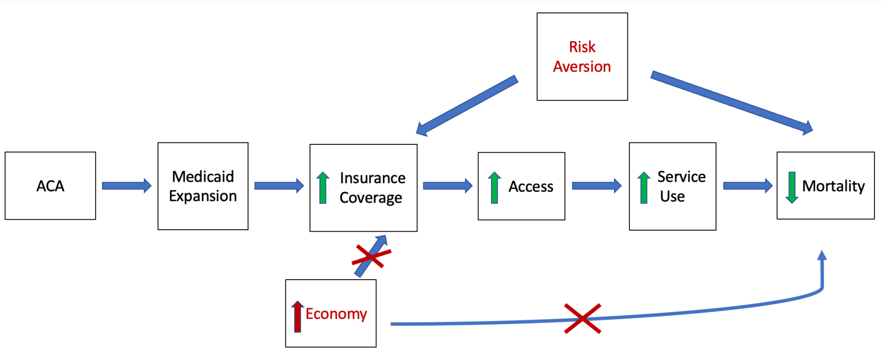
Policy Evaluation - Regression



Policy Evaluation - Regression



Policy Evaluation - Regression



Policy Evaluation - Regression

- Regression model estimating the effect of Medicaid coverage on mortality:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \varepsilon_{it}$$

Policy Evaluation - Regression

- Regression model estimating the effect of Medicaid coverage on mortality:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \varepsilon_{it}$$

Policy Evaluation - Regression

- Regression model estimating the effect of Medicaid coverage on mortality:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + \varepsilon_{it}$$

- Regression model estimating the effect of Medicaid coverage on mortality:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + ...\beta_k X_{it} + \varepsilon_{it}$$

Policy Evaluation - Regression

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + ... \beta_k X_{it} + \varepsilon_{it}$$

- Our goal is to "control" for as many characteristics (X) as we can that could be related to insurance coverage and morality (i.e., confounders).

Policy Evaluation - Regression

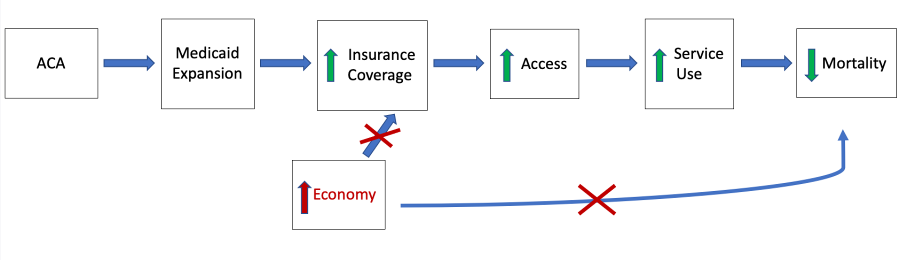
$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + ...\beta_k X_{it} + \varepsilon_{it}$$

- Our goal is to "control" for as many characteristics (X) as we can that could be related to insurance coverage and morality (i.e., confounders).
- Estimate the effect of Medicaid coverage on mortality *conditional* on education, age, etc.

Policy Evaluation - Regression

$$y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 Edu_{it} + \beta_3 Age_{it} + ...\beta_k X_{it} + \varepsilon_{it}$$

- Our goal is to "control" for as many characteristics (X) as we can that could be related to insurance coverage and morality (i.e., confounders).
- Estimate the effect of Medicaid coverage on mortality *conditional* on education, age, etc.



Policy Evaluation - Regression

- Remember, as long as we have eliminated all relevant confounders, our regression estimate will be equal to the average treatment effect:

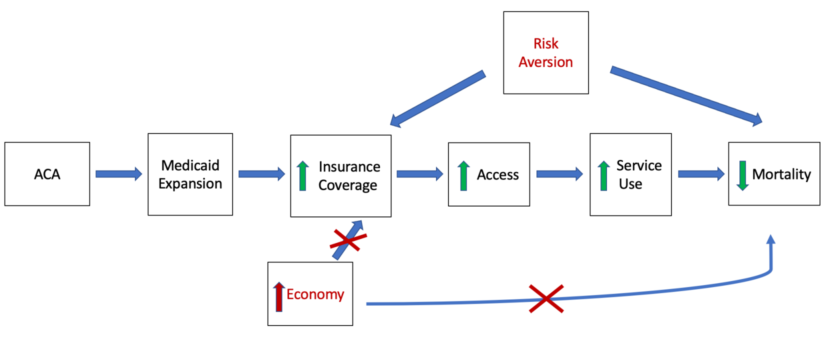
$$ATE_{est} = ATE$$

Policy Evaluation - Regression

- Remember, as long as we have eliminated all relevant confounders, our regression estimate will be equal to the average treatment effect:

$$ATE_{est} = ATE$$

- But then there's this...



Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?

Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?
- Our estimate of the effect of Medicaid coverage on mortality will be *biased*.

Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?
- Our estimate of the effect of Medicaid coverage on mortality will be *biased*.
- We think we're estimating the effect of Medicaid coverage on mortality, but we're actually estimating the combined effect of Medicaid coverage and risk aversion on mortality.

Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?
- Our estimate of the effect of Medicaid coverage on mortality will be *biased*.
- We think we're estimating the effect of Medicaid coverage on mortality, but we're actually estimating the combined effect of Medicaid coverage and risk aversion on mortality.
- Sometimes we can determine which direction that bias is likely to take.

Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?
- Our estimate of the effect of Medicaid coverage on mortality will be *biased*.
- We think we're estimating the effect of Medicaid coverage on mortality, but we're actually estimating the combined effect of Medicaid coverage and risk aversion on mortality.
- Sometimes we can determine which direction that bias is likely to take.
    - Suppose that those who are more risk averse are more likely to sign up for Medicaid coverage when eligible and are more likely to be in better health.

Policy Evaluation - Regression

- What happens if we fail to account for an observed or unobserved confounder?
- Our estimate of the effect of Medicaid coverage on mortality will be *biased*.
- We think we're estimating the effect of Medicaid coverage on mortality, but we're actually estimating the combined effect of Medicaid coverage and risk aversion on mortality.
- Sometimes we can determine which direction that bias is likely to take.
  - Suppose that those who are more risk averse are more likely to sign up for Medicaid coverage when eligible and are more likely to be in better health.
    - Then our estimate of the effect of Medicaid coverage on mortality will be biased *away from zero*.

Policy Evaluation - Regression

- Moral of the story:

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.
  - The goal is to compare statistically identical people who differ only in their exposure to treatment.

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.
  - The goal is to compare statistically identical people who differ only in their exposure to treatment.
  - Under certain conditions, regression can return causal estimates of our relationship of interest.

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.
  - The goal is to compare statistically identical people who differ only in their exposure to treatment.
  - Under certain conditions, regression can return causal estimates of our relationship of interest.
  - But omitted variable bias is a serious threat when exposure to treatment is non-random.

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.
  - The goal is to compare statistically identical people who differ only in their exposure to treatment.
  - Under certain conditions, regression can return causal estimates of our relationship of interest.
  - But omitted variable bias is a serious threat when exposure to treatment is non-random.
- So what do we do if randomization is infeasible and we're not confident that regression estimates will be unbiased?

Policy Evaluation - Regression

- Moral of the story:
  - Regression analysis is often used in place of randomization when randomization is infeasible.
  - The goal is to compare statistically identical people who differ only in their exposure to treatment.
  - Under certain conditions, regression can return causal estimates of our relationship of interest.
  - But omitted variable bias is a serious threat when exposure to treatment is non-random.
- So what do we do if randomization is infeasible and we're not confident that regression estimates will be unbiased?
  - Natural experiments and quasi-experimental research design.