Module 5(b): Sensitivity Analysis and Partial Identification

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.
- What if this assumption doesn't hold? Two potential solutions:

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.
- What if this assumption doesn't hold? Two potential solutions:
    1. **Sensitivity analysis**: try to vary the amount of unmeasured confounding to see if it changes the effect.

Where are we? Where are we going?

- Saw how to estimate the ATE with regression under selection on observables.
- What if this assumption doesn't hold? Two potential solutions:
    1. **Sensitivity analysis**: try to vary the amount of unmeasured confounding to see if it changes the effect.
    2. **Partial identification**: abandon point identification and try to find bounds for the ATE under different assumptions.

1/ Sensitivity analysis

Sensitivity analysis for regression

- Standard regression estimator of the ATE:

$$Y_i = \hat{\alpha} + \widehat{\tau} D_i + \mathbf{X}_i' \widehat{\beta} + \hat{\varepsilon}_i$$

Sensitivity analysis for regression

- Standard regression estimator of the ATE:

$$Y_i = \hat{\alpha} + \widehat{\tau} D_i + \mathbf{X}_i' \widehat{\beta} + \widehat{\varepsilon}_i$$

- What if the true regression model contained U_i which we omitted?

$$Y_i = \alpha + \tau D_i + \mathbf{X}_i'\beta + \gamma U_i + \varepsilon_i \qquad \widehat{\tau} \xrightarrow{p} \tau + \gamma \times \underbrace{\frac{\text{cov}(D_i^{\perp \mathbf{X}}, U_i^{\perp \mathbf{X}})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}_{\text{regression of } U_i^{\perp \mathbf{X}} \text{ on } D_i^{\perp \mathbf{X}}}$$

Sensitivity analysis for regression

- Standard regression estimator of the ATE:

$$Y_i = \hat{\alpha} + \widehat{\tau} D_i + \mathbf{X}_i' \widehat{\beta} + \hat{\varepsilon}_i$$

- What if the true regression model contained U_i which we omitted?

$$Y_i = \alpha + \tau D_i + \mathbf{X}_i' \beta + \gamma U_i + \varepsilon_i \qquad \widehat{\tau} \xrightarrow{p} \tau + \gamma \times \underbrace{\frac{\text{cov}(D_i^{\perp \mathbf{X}}, U_i^{\perp \mathbf{X}})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}_{\text{regression of } U_i^{\perp \mathbf{X}} \text{ on } D_i^{\perp \mathbf{X}}}$$

- Standard **omitted variable bias formula**:

$$\widehat{\tau} \xrightarrow{p} \tau + \gamma \times \underbrace{\frac{\text{cov}(D_i^{\perp \mathbf{X}}, U_i^{\perp \mathbf{X}})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}_{\text{regression of } U_i^{\perp \mathbf{X}} \text{ on } D_i^{\perp \mathbf{X}}}$$

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

- **Sensitivity analysis** can then vary two unknown parameters:

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

- **Sensitivity analysis** can then vary two unknown parameters:
  - $R^2_{Y\sim U|D,\mathbf{X}} \in [0,1]$ incremental predictive value of U for the outcome

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

- **Sensitivity analysis** can then vary two unknown parameters:
  - $R^2_{Y\sim U|D,\mathbf{X}} \in [0,1]$ incremental predictive value of U for the outcome
  - $R^2_{D\sim U|\mathbf{X}} \in [0,1]$ incremental predictive value of U for treatment

Partial R-squared interpretations

- Regression coefficients with unknowns are difficult to reason about.
- Easier to reason with partial R^2 version of OVB (Cinelli and Hazlett, JRSSB, 2019):

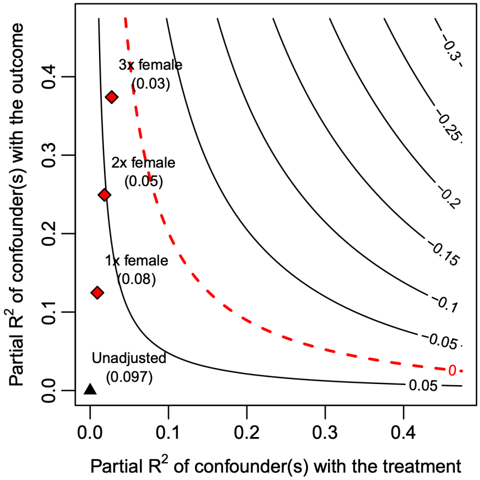
$$|\text{bias}| = \sqrt{\frac{R^2_{Y\sim U|D,\mathbf{X}} R^2_{D\sim U|\mathbf{X}}}{1 - R^2_{D\sim U|\mathbf{X}}} \frac{\mathbb{V}(Y^{\perp \mathbf{X},D})}{\mathbb{V}(D_i^{\perp \mathbf{X}})}}$$

- Partial R^2 is the incremental predictive value of one variable:

$$R^2_{Y\sim U|D,\mathbf{X}} = \frac{R^2_{Y\sim D+\mathbf{X}+U} - R^2_{Y\sim D+\mathbf{X}}}{1 - R^2_{Y\sim D+\mathbf{X}}} = \frac{\text{additional variance explained by } U}{\text{variance unexplained by } D, \mathbf{X}}$$

- **Sensitivity analysis** can then vary two unknown parameters:
  - $R^2_{Y\sim U|D,\mathbf{X}} \in [0,1]$ incremental predictive value of U for the outcome
  - $R^2_{D\sim U|\mathbf{X}} \in [0,1]$ incremental predictive value of U for treatment
  - From these we can determine the bias and thus the true value of τ

Sensitivity analysis example



2/ Partial identification and bounds

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.
    - May not be point identified, but maybe we can bound the effect.

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.
    - May not be point identified, but maybe we can bound the effect.
- If Y is bounded $[y_L, y_U]$, τ logically must be in $[y_L - y_U, y_U - y_L]$.

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.
    - May not be point identified, but maybe we can bound the effect.
- If Y is bounded $[y_L, y_U]$, τ logically must be in $[y_L - y_U, y_U - y_L]$.
- Can we improve using data? Rewrite the ATE with $p = \mathbb{P}(D_i = 1)$:

$$\begin{aligned}\tau = &\ \mathbb{E}[Y_i \mid D_i = 1]p + \mathbb{E}[Y_i(1) \mid D_i = 0](1-p) \\ &- \mathbb{E}[Y_i(0) \mid D_i = 1]p - \mathbb{E}[Y_i \mid D_i = 0](1-p)\end{aligned}$$

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.
    - May not be point identified, but maybe we can bound the effect.
- If Y is bounded $[y_L, y_U]$, τ logically must be in $[y_L - y_U, y_U - y_L]$.
- Can we improve using data? Rewrite the ATE with $p = \mathbb{P}(D_i = 1)$:

$$\begin{aligned}\tau = &\ \mathbb{E}[Y_i \mid D_i = 1]p + \mathbb{E}[Y_i(1) \mid D_i = 0](1-p) \\ &- \mathbb{E}[Y_i(0) \mid D_i = 1]p - \mathbb{E}[Y_i \mid D_i = 0](1-p)\end{aligned}$$

- Plug in y_L and y_U for the counterfactual means to get bounds for τ:

$$\tau \geq \mathbb{E}[Y_i \mid D_i = 1]p + y_L(1-p) - y_U p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$
$$\tau \leq \mathbb{E}[Y_i \mid D_i = 1]p + y_U(1-p) - y_L p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
    - Idea: pick assumptions and then figure out what you can learn.
    - May not be point identified, but maybe we can bound the effect.
- If Y is bounded $[y_L, y_U]$, τ logically must be in $[y_L - y_U, y_U - y_L]$.
- Can we improve using data? Rewrite the ATE with $p = \mathbb{P}(D_i = 1)$:

$$\begin{aligned}\tau = &\ \mathbb{E}[Y_i \mid D_i = 1]p + \mathbb{E}[Y_i(1) \mid D_i = 0](1-p) \\ &- \mathbb{E}[Y_i(0) \mid D_i = 1]p - \mathbb{E}[Y_i \mid D_i = 0](1-p)\end{aligned}$$

- Plug in y_L and y_U for the counterfactual means to get bounds for τ:

$$\tau \geq \mathbb{E}[Y_i \mid D_i = 1]p + y_L(1-p) - y_U p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$
$$\tau \leq \mathbb{E}[Y_i \mid D_i = 1]p + y_U(1-p) - y_L p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$

    - These bounds have width of $|y_U - y_L|$ which is half of the logical bounds.

No assumption bounds

- **Law of decreasing credibility** (Manski): credibility of inferences decreases with strength of assumptions
  - Idea: pick assumptions and then figure out what you can learn.
  - May not be point identified, but maybe we can bound the effect.
- If Y is bounded $[y_L, y_U]$, τ logically must be in $[y_L - y_U, y_U - y_L]$.
- Can we improve using data? Rewrite the ATE with $p = \mathbb{P}(D_i = 1)$:

$$\begin{aligned}\tau &= \mathbb{E}[Y_i \mid D_i = 1]p + \mathbb{E}[Y_i(1) \mid D_i = 0](1-p) \\ &\quad - \mathbb{E}[Y_i(0) \mid D_i = 1]p - \mathbb{E}[Y_i \mid D_i = 0](1-p)\end{aligned}$$

- Plug in y_L and y_U for the counterfactual means to get bounds for τ:

$$\tau \geq \mathbb{E}[Y_i \mid D_i = 1]p + y_L(1-p) - y_U p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$
$$\tau \leq \mathbb{E}[Y_i \mid D_i = 1]p + y_U(1-p) - y_L p - \mathbb{E}[Y_i \mid D_i = 0](1-p)$$

  - These bounds have width of $|y_U - y_L|$ which is half of the logical bounds.
  - But always will contain 0. Weak assumptions $\rightsquigarrow$ weak inferences

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$
  - $\mathbb{E}[Y_i(1) \mid D_i = 0] \leq \mathbb{E}[Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i \mid D_i = 0] = \mu(0)$

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$
  - $\mathbb{E}[Y_i(1) \mid D_i = 0] \leq \mathbb{E}[Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i \mid D_i = 0] = \mu(0)$
- Implies the following bounds for τ:

$$\tau \in \left[(1-p)\left(y_L - \mathbb{E}[Y_i \mid D_i = 0]\right),\ \ p\left(\mathbb{E}[Y_i \mid D_i = 1] - y_L\right)\right]$$

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$
  - $\mathbb{E}[Y_i(1) \mid D_i = 0] \leq \mathbb{E}[Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i \mid D_i = 0] = \mu(0)$
- Implies the following bounds for τ:

$$\tau \in \left[(1-p)\left(y_L - \mathbb{E}[Y_i \mid D_i = 0]\right), \;\; p\left(\mathbb{E}[Y_i \mid D_i = 1] - y_L\right)\right]$$

- Width of these bounds: $\mathbb{E}[Y_i] - y_L$

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
  - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$
  - $\mathbb{E}[Y_i(1) \mid D_i = 0] \leq \mathbb{E}[Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i \mid D_i = 0] = \mu(0)$
- Implies the following bounds for τ:

$$\tau \in \left[(1-p)\left(y_L - \mathbb{E}[Y_i \mid D_i = 0]\right),\ \ p\left(\mathbb{E}[Y_i \mid D_i = 1] - y_L\right)\right]$$

- Width of these bounds: $\mathbb{E}[Y_i] - y_L$
  - Width now depends on the observed data!

Optimized treatment choice

- **Assumptions** can narrow the bounds even further.
- Assumption: people choose the treatment with the highest outcome.
    - $\mathbb{E}[Y_i(0) \mid D_i = 1] \leq \mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1] = \mu(1)$
    - $\mathbb{E}[Y_i(1) \mid D_i = 0] \leq \mathbb{E}[Y_i(0) \mid D_i = 0] = \mathbb{E}[Y_i \mid D_i = 0] = \mu(0)$
- Implies the following bounds for τ:

$$\tau \in \left[(1-p)\left(y_L - \mathbb{E}[Y_i \mid D_i = 0]\right),\;\; p\left(\mathbb{E}[Y_i \mid D_i = 1] - y_L\right)\right]$$

- Width of these bounds: $\mathbb{E}[Y_i] - y_L$
    - Width now depends on the observed data!
    - Interval will still always include zero.

Confidence regions for bounds

- More general setup:

Confidence regions for bounds

- More general setup:
    - True bounds $[\delta_L, \delta_U]$ also called the **identification region**

Confidence regions for bounds

- More general setup:
    - True bounds $[\delta_L, \delta_U]$ also called the **identification region**
    - Estimated bounds $[\hat{\delta}_L, \hat{\delta}_U]$.

Confidence regions for bounds

- More general setup:
  - True bounds $[\delta_L, \delta_U]$ also called the **identification region**
  - Estimated bounds $[\hat{\delta}_L, \hat{\delta}_U]$.
  - $\widehat{\text{se}}(\hat{\delta}_L), \widehat{\text{se}}(\hat{\delta}_U)$ are the standard errors of the estimated bounds

Confidence regions for bounds

- More general setup:
    - True bounds $[\delta_L, \delta_U]$ also called the **identification region**
    - Estimated bounds $[\hat{\delta}_L, \hat{\delta}_U]$.
    - $\widehat{\text{se}}(\hat{\delta}_L), \widehat{\text{se}}(\hat{\delta}_U)$ are the standard errors of the estimated bounds
- Two possible CI approaches that find intervals that...

Confidence regions for bounds

- More general setup:
  - True bounds $[\delta_L, \delta_U]$ also called the **identification region**
  - Estimated bounds $[\hat{\delta}_L, \hat{\delta}_U]$.
  - $\widehat{\text{se}}(\hat{\delta}_L), \widehat{\text{se}}(\hat{\delta}_U)$ are the standard errors of the estimated bounds
- Two possible CI approaches that find intervals that...
  1. Covers the identified region with probability $1 - \alpha$

$$\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$$

Confidence regions for bounds

- More general setup:
  - True bounds $[\delta_L, \delta_U]$ also called the **identification region**
  - Estimated bounds $[\hat{\delta}_L, \hat{\delta}_U]$.
  - $\widehat{\text{se}}(\hat{\delta}_L), \widehat{\text{se}}(\hat{\delta}_U)$ are the standard errors of the estimated bounds
- Two possible CI approaches that find intervals that...
  1. Covers the identified region with probability $1 - \alpha$

     $$\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$$

  2. Covers the true value of the parameter with probability $1 - \alpha$

     $$\mathbb{P}(\tau \in [\hat{\delta}_L, \hat{\delta}_U]) \geq 1 - \alpha$$

Calculating confidence intervals

- Case 1: covering the identified region $\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$

$$[\hat{\delta}_L - z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\delta}_U)]$$

Calculating confidence intervals

- Case 1: covering the identified region $\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1-\alpha$

$$[\hat{\delta}_L - z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_U)]$$

- Works because of **Bonferroni inequality**:

$$\mathbb{P}(\hat{\delta}_L \leq \delta_L \text{ and } \hat{\delta}_U \geq \delta_U) \geq \mathbb{P}(\hat{\delta}_L \leq \delta_L) + \mathbb{P}(\hat{\delta}_U \leq \delta_U) - 1 = 1-\alpha$$

Calculating confidence intervals

- Case 1: covering the identified region $\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$

$$[\hat{\delta}_L - z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_U)]$$

- Works because of **Bonferroni inequality**:

$$\mathbb{P}(\hat{\delta}_L \leq \delta_L \text{ and } \hat{\delta}_U \geq \delta_U) \geq \mathbb{P}(\hat{\delta}_L \leq \delta_L) + \mathbb{P}(\hat{\delta}_U \leq \delta_U) - 1 = 1 - \alpha$$

- Case 2: cover the true parameter, τ.

$$[\hat{\delta}_L - z_{1-\alpha}\widehat{\text{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha}\widehat{\text{se}}(\hat{\delta}_U)]$$

Calculating confidence intervals

- Case 1: covering the identified region $\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$

$$[\hat{\delta}_L - z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\delta}_U)]$$

- Works because of **Bonferroni inequality**:

$$\mathbb{P}(\hat{\delta}_L \leq \delta_L \text{ and } \hat{\delta}_U \geq \delta_U) \geq \mathbb{P}(\hat{\delta}_L \leq \delta_L) + \mathbb{P}(\hat{\delta}_U \leq \delta_U) - 1 = 1 - \alpha$$

- Case 2: cover the true parameter, τ.

$$[\hat{\delta}_L - z_{1-\alpha}\widehat{\mathsf{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha}\widehat{\mathsf{se}}(\hat{\delta}_U)]$$

  - If $\tau = \delta_L$ or $\tau = \delta_U$, then coverage converges to $1 - \alpha$

Calculating confidence intervals

- Case 1: covering the identified region $\mathbb{P}(\hat{\delta}_L \leq \delta_L, \hat{\delta}_U \geq \delta_U) \geq 1 - \alpha$

$$[\hat{\delta}_L - z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha/2}\widehat{\text{se}}(\hat{\delta}_U)]$$

- Works because of **Bonferroni inequality**:

$$\mathbb{P}(\hat{\delta}_L \leq \delta_L \text{ and } \hat{\delta}_U \geq \delta_U) \geq \mathbb{P}(\hat{\delta}_L \leq \delta_L) + \mathbb{P}(\hat{\delta}_U \leq \delta_U) - 1 = 1 - \alpha$$

- Case 2: cover the true parameter, τ.

$$[\hat{\delta}_L - z_{1-\alpha}\widehat{\text{se}}(\hat{\delta}_L), \quad \hat{\delta}_U + z_{1-\alpha}\widehat{\text{se}}(\hat{\delta}_U)]$$

  - If $\tau = \delta_L$ or $\tau = \delta_U$, then coverage converges to $1 - \alpha$
  - If $\delta_L < \tau < \delta_U$, then coverage converges to 1.