Correcting Measurement Error Bias in Conjoint Survey Experiments[1]

Gary King[2]

Institute for Quantitative Social Science, Harvard University

Harvard Experiments Working Group, 2/9/2024

[1] Paper, software, slides, data: GaryKing.org/conjointE

[2] Based on work with Katherine Clayton, Yusaku Horiuchi, Aaron Kaufman, and Mayya Komisarchik

This talk

This talk

- What's conjoint? (with easier, more applicable methods)

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
    - How much? Lots!

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
    - How much? Lots!
    - Why so much? Approximating real world decisions
    - Ignore it? Bias!

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - Hard to fix? A few lines of code & estimate IRR

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - Hard to fix? A few lines of code & estimate IRR
    - Planning a survey: add one question

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
    - How much? Lots!
    - Why so much? Approximating real world decisions
    - Ignore it? Bias!
    - Hard to fix? A few lines of code & estimate IRR
        - Planning a survey: add one question
        - Reanalyzing existing data: extrapolate

This talk

- What's conjoint? (with easier, more applicable methods)
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - Hard to fix? A few lines of code & estimate IRR
    - Planning a survey: add one question
    - Reanalyzing existing data: extrapolate
- Evidence: 13+ surveys, 9,472 respondents, 137,785 questions

Conjoint Questions: Complicated Real World Trade Offs

Conjoint Questions: Complicated Real World Trade Offs

(Ono and Burden, 2018)

Conjoint Questions: Complicated Real World Trade Offs

(Ono and Burden, 2018)

Please carefully review the two potential candidates running for election to the U.S. House of Representatives, detailed below.

	Candidate 0	Candidate 1
Race/Ethnicity	Hispanic	Asian American
Age	52	60
Favorability rating among the public	70%	34%
Position on immigrants	Favors giving citizenship or guest worker status to undocumented immigrants	Opposes giving citizenship or guest worker status to undocumented immigrants
Party affiliation	Republican Party	Democratic Party
Position on abortion	Abortion is not a private matter (pro-life)	Abortion is a private matter (pro-choice)
Position on government deficit	Wants to reduce the deficit through tax increase	Wants to reduce the deficit through tax increase
Salient personal characteristics	Really cares about people like you	Really cares about people like you
Position on national security	Wants to cut military budget and keep the U.S. out of war	Wants to maintain strong defense and increase U.S. influence
Gender	Female	Female
Policy area of expertise	Education	Foreign policy
Family	Single (divorced)	Married (no child)
Experience in public office	12 years	4 years

If you had to choose between them, which of these candidates would you vote to be a member of the U.S. House of Representatives?

Simplified (but more general) Data Structure

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)
- Better than profile as unit of analysis (legacy from marketing & psychology)

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)
- Better than profile as unit of analysis (legacy from marketing & psychology)
  - Less: Intuitive (2 rows in data per question)

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)
- Better than profile as unit of analysis (legacy from marketing & psychology)
  - Less: Intuitive (2 rows in data per question)
  - More: complicated (clustering) methods required

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)
- Better than profile as unit of analysis (legacy from marketing & psychology)
  - Less: Intuitive (2 rows in data per question)
  - More: complicated (clustering) methods required
  - Fewer: substantive questions can be asked

Simplified (but more general) Data Structure

- Unit of Analysis: the binary choice; N rows
- Dependent var: Candidate 0 vs 1
- Covariates: Attributes (randomized)
  - Age: "78, 60", "60, 60", "52, 43"
  - Incumbency: "Incumbent, nonincumbent", or "open seat"
- Mean, SE: as usual (see intro stat class)
- Better than profile as unit of analysis (legacy from marketing & psychology)
  - Less: Intuitive (2 rows in data per question)
  - More: complicated (clustering) methods required
  - Fewer: substantive questions can be asked
  - Same: measurement error corrections

Quantities of Interest

Quantities of Interest

- Notation

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}} \left[\rho_i(a_\ell) \right].$$

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}}\left[\rho_i(a_\ell, a_{-\ell})\right].$$

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}}\left[\rho_i(a_\ell, a_{-\ell})\right].$$

  - A little wacky, but convenient: Each attribute can of interest

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \operatorname*{mean}_i \left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell})\right].$$

  - A little wacky, but convenient: Each attribute can of interest
  - Attribute selection: defines QOI

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \operatorname*{mean}_i \left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell})\right].$$

  - A little wacky, but convenient: Each attribute can of interest
  - Attribute selection: defines QOI
  - Weighting over other attributes: Why uniform?

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}}\left[\rho_i(a_\ell, a_{-\ell})\right].$$

  - A little wacky, but convenient: Each attribute can of interest
  - Attribute selection: defines QOI
  - Weighting over other attributes: Why uniform?
  - Other conjoint QOIs: linear functions of this mean

Quantities of Interest

- Notation
  - Attributes: $a = \{a_\ell, a_{-\ell}\} = \{$Attribute of interest, Others$\}$
  - Preference: $\rho_i(a_\ell, a_{-\ell}) \in \{0, 1\}$
- Marginal Mean in traditional (non-conjoint!) survey

$$\rho(a_\ell) = \underset{i}{\text{mean}}\left[\rho_i(a_\ell)\right].$$

  - E.g.: Mean preferring Candidate 0 (Female) to 1 (Male)
  - Implicit: Other attributes set to observed (or like "Clarify")
- Marginal Mean in conjoint survey

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}}\left[\rho_i(a_\ell, a_{-\ell})\right].$$

  - A little wacky, but convenient: Each attribute can of interest
  - Attribute selection: defines QOI
  - Weighting over other attributes: Why uniform?
  - Other conjoint QOIs: linear functions of this mean
- Casual effect: Average Marginal Component Effect (AMCE)

$$\theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

Measurement Error in Binary Choice

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed,

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed, preferences $\{0,1\}$ unobserved

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed, preferences $\{0,1\}$ unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \qquad \leftarrow \text{Pr(swapping error)} \end{cases}$$

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed, preferences $\{0,1\}$ unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \quad \leftarrow \text{Pr(swapping error)} \end{cases}$$

- Estimate τ: Ask 1 extra question

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed, preferences $\{0,1\}$ unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \qquad \leftarrow \text{Pr(swapping error)} \end{cases}$$

- Estimate τ: Ask 1 extra question

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices {0,1} observed, preferences {0,1} unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \end{cases} \quad \leftarrow \text{Pr(swapping error)}$$

- Estimate τ: Ask 1 extra question
    1. Estimate Intra-Respondent Reliability (IRR)

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices $\{0,1\}$ observed, preferences $\{0,1\}$ unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \end{cases} \quad \leftarrow \text{Pr(swapping error)}$$

- Estimate τ: Ask 1 extra question
  1. Estimate Intra-Respondent Reliability (IRR)
     - Add one question: Q1, Q2, Q3, Q4, Q5, Q1

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices {0,1} observed, preferences {0,1} unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \qquad \leftarrow \text{Pr(swapping error)} \end{cases}$$

- Estimate τ: Ask 1 extra question
  1. Estimate Intra-Respondent Reliability (IRR)
     - Add one question: Q1, Q2, Q3, Q4, Q5, Q1
     - IRR = % agreement on 2 identical questions asked moments apart

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices {0,1} observed, preferences {0,1} unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \end{cases} \quad \leftarrow \text{Pr(swapping error)}$$

- Estimate τ: Ask 1 extra question
  1. Estimate Intra-Respondent Reliability (IRR)
     - Add one question: Q1, Q2, Q3, Q4, Q5, Q1
     - IRR = % agreement on 2 identical questions asked moments apart
     - No one remembers the duplicate Q

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices {0,1} observed, preferences {0,1} unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \qquad \leftarrow \text{Pr(swapping error)} \end{cases}$$

- Estimate τ: Ask 1 extra question
  1. Estimate Intra-Respondent Reliability (IRR)
     - Add one question: Q1, Q2, Q3, Q4, Q5, Q1
     - IRR = % agreement on 2 identical questions asked moments apart
     - No one remembers the duplicate Q
     - Different "considerations" (confounders) unlikely

- Extrapolate τ: no new data (to come)

Measurement Error in Binary Choice

(Observation mechanism assumed; evidence coming later)

- Choices {0,1} observed, preferences {0,1} unobserved

$$C_i(a) = \begin{cases} \rho_i(a) & \text{w.p. } 1-\tau \\ 1-\rho_i(a) & \text{w.p. } \tau \end{cases} \quad \leftarrow \text{Pr(swapping error)}$$

- Estimate τ: Ask 1 extra question
  1. Estimate Intra-Respondent Reliability (IRR)
     - Add one question: Q1, Q2, Q3, Q4, Q5, Q1
     - IRR = % agreement on 2 identical questions asked moments apart
     - No one remembers the duplicate Q
     - Different "considerations" (confounders) unlikely
  2. Calculate: $\hat{\tau} = \dfrac{1-\sqrt{1-2(1-\text{IRR})}}{2}$
- Extrapolate τ: no new data (to come)

What changed from the last time you saw this?

What changed from the last time you saw this?

(Ono and Burden, 2018)

What changed from the last time you saw this?

(Ono and Burden, 2018)

Please carefully review the two potential candidates running for election to the U.S. House of Representatives, detailed below.

	Candidate 0	Candidate 1
Race/Ethnicity	Hispanic	Asian American
Age	52	60
Favorability rating among the public	70%	34%
Position on immigrants	Favors giving citizenship or guest worker status to undocumented immigrants	Opposes giving citizenship or guest worker status to undocumented immigrants
Party affiliation	Republican Party	Democratic Party
Position on abortion	Abortion is not a private matter (pro-life)	Abortion is a private matter (pro-choice)
Position on government deficit	Wants to reduce the deficit through tax increase	Wants to reduce the deficit through tax increase
Salient personal characteristics	Really cares about people like you	Really cares about people like you
Position on national security	Wants to cut military budget and keep the U.S. out of war	Wants to maintain strong defense and increase U.S. influence
Gender	Female	Female
Policy area of expertise	Education	Foreign policy
Family	Single (divorced)	Married (no child)
Experience in public office	12 years	4 years

If you had to choose between them, which of these candidates would you vote to be a member of the U.S. House of Representatives?

Bias Due to Swapping Error

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell}) \right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \operatorname*{mean}_{i : A_i = a} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell}) \right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \operatorname*{mean}_{i : A_i = a} \left[C_i(a) \right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell}) \right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i : A_i = a}{\text{mean}} \left[C_i(a) \right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i\,:\,A_i = a}{\text{mean}} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \operatorname*{mean}_{i\,:\,A_i = a} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias
- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}}\left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i : A_i = a}{\text{mean}}\left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias

- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

  - with τ known: unbiased

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i:A_i = a}{\text{mean}} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias

- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

  - with τ known: unbiased
  - with consistent $\hat{\tau}$: Consistent, approximately unbiased

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i : A_i = a}{\text{mean}} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias
- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

  - with τ known: unbiased
  - with consistent $\hat{\tau}$: Consistent, approximately unbiased
  - SE estimators: Fast, accurate, easy

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \operatorname*{mean}_{i, a_{-\ell}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \operatorname*{mean}_{i : A_i = a} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\leadsto$ bias
- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

  - with τ known: unbiased
  - with consistent $\hat{\tau}$: Consistent, approximately unbiased
  - SE estimators: Fast, accurate, easy
- What we need to show about τ

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i : A_i = a}{\text{mean}} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias
- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

  - with τ known: unbiased
  - with consistent $\hat{\tau}$: Consistent, approximately unbiased
  - SE estimators: Fast, accurate, easy
- What we need to show about τ
  - IRR small: τ large $\rightsquigarrow$ bias large enough to matter

Bias Due to Swapping Error

- Quantities of Interest (from earlier)

$$\rho(a_\ell) = \underset{i, a_{-\ell}}{\text{mean}} \left[\rho_i(a_\ell, a_{-\ell})\right], \qquad \theta(a_\ell, a'_\ell) = \rho(a_\ell) - \rho(a'_\ell).$$

- Standard estimators

$$\hat{\rho}(a) = \underset{i: A_i = a}{\text{mean}} \left[C_i(a)\right], \qquad \hat{\theta}(a, a') = \hat{\rho}(a) - \hat{\rho}(a')$$

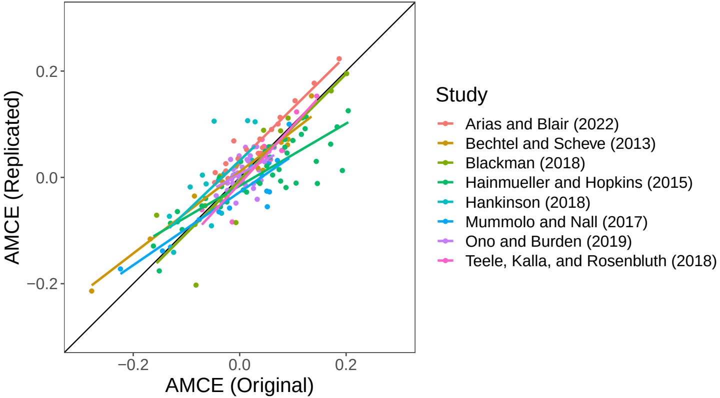
  - If $\tau = 0$: both unbiased (identified by randomization)
  - If $\tau > 0$: both biased (i.e., most prior research)
  - Not like regression: error in outcome variable $\rightsquigarrow$ bias
- Alternative estimators

$$\tilde{\rho}(a) = \frac{\hat{\rho}(a) - \tau}{1 - 2\tau}, \qquad \tilde{\theta}(a, a') = \frac{\hat{\theta}(a, a')}{1 - 2\tau},$$

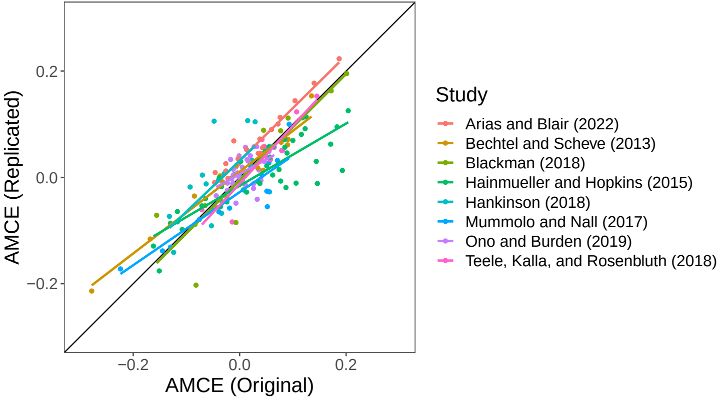
  - with τ known: unbiased
  - with consistent $\hat{\tau}$: Consistent, approximately unbiased
  - SE estimators: Fast, accurate, easy
- What we need to show about τ
  - IRR small: τ large $\rightsquigarrow$ bias large enough to matter
  - $\tau(a) \approx \tau$: correction is easy

8 Replications of Data Collection & Analysis

8 Replications of Data Collection & Analysis

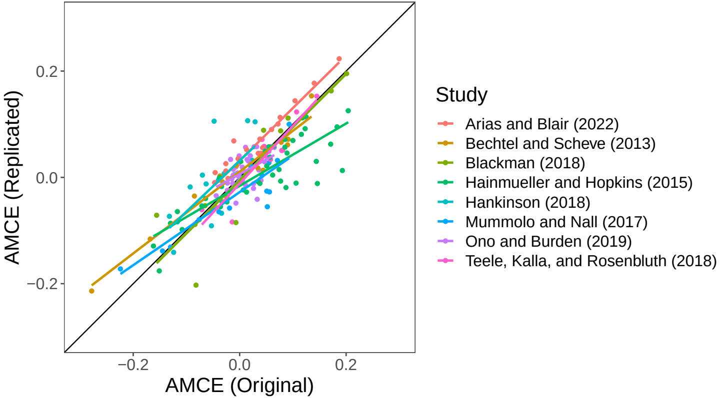


8 Replications of Data Collection & Analysis



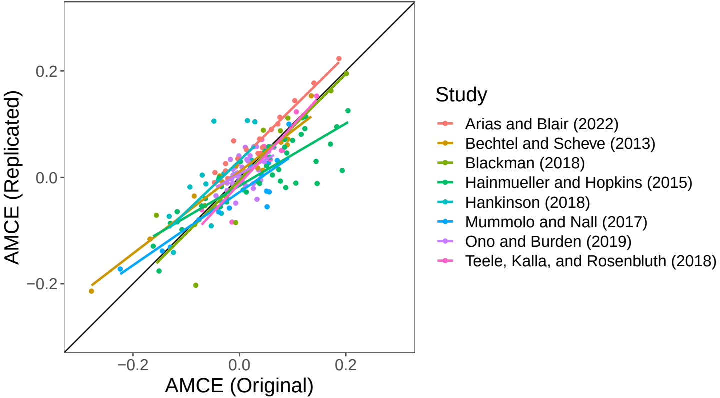
- Replicated all data collection and analysis from scratch

8 Replications of Data Collection & Analysis



- Replicated all data collection and analysis from scratch
- Every study replicates! (Median correlation 0.9)

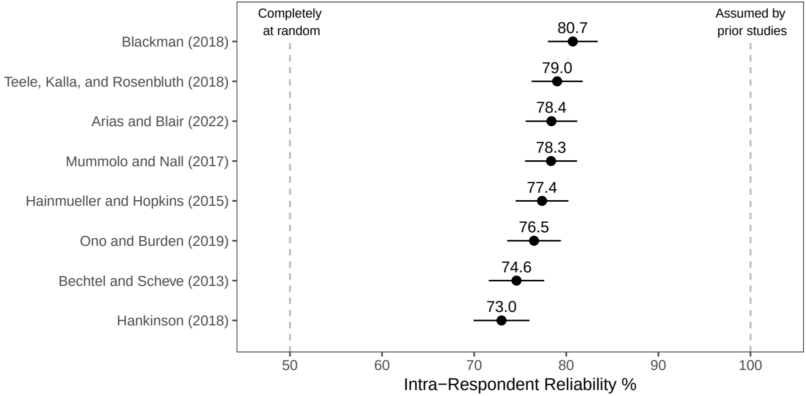
8 Replications of Data Collection & Analysis



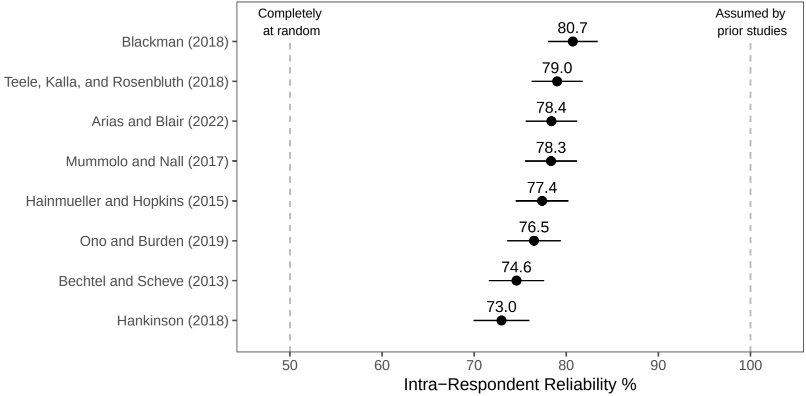
- Replicated all data collection and analysis from scratch
- Every study replicates! (Median correlation 0.9)
- Impressive literature, especially given crises in other fields

Low Intra-Respondent Reliability!

Low Intra-Respondent Reliability!

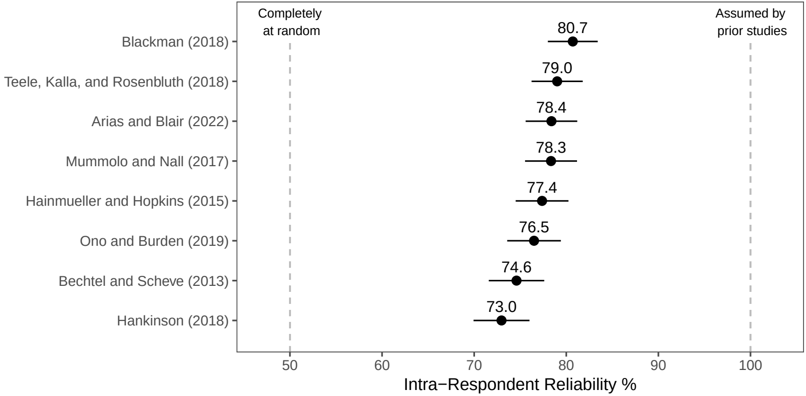


Low Intra-Respondent Reliability!



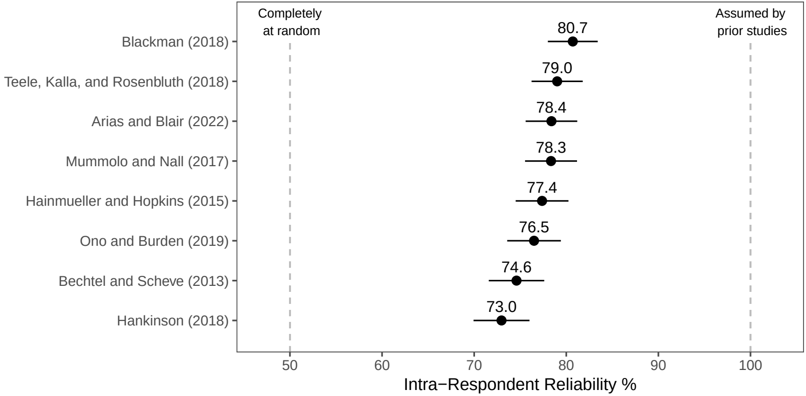
- Randomize 2 of 8 studies & 5 Qs per resp.; $n = 3,289$

Low Intra-Respondent Reliability!



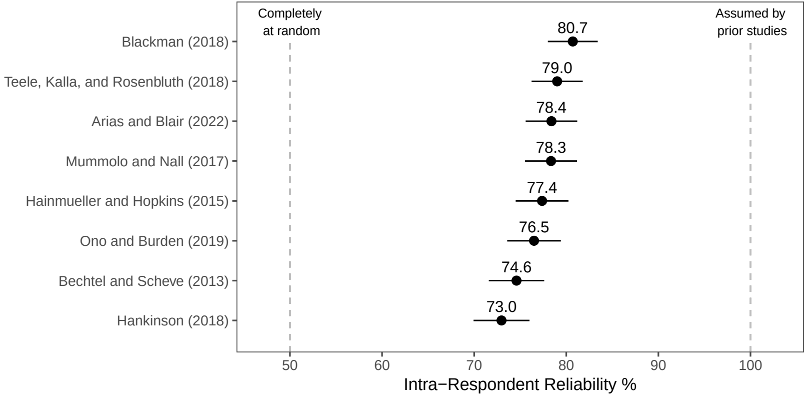
- Randomize 2 of 8 studies & 5 Qs per resp.; $n = 3,289$
- Add one Q: Q1, Q2, Q3, Q4, Q5, Q1; Compute IRR.

Low Intra-Respondent Reliability!



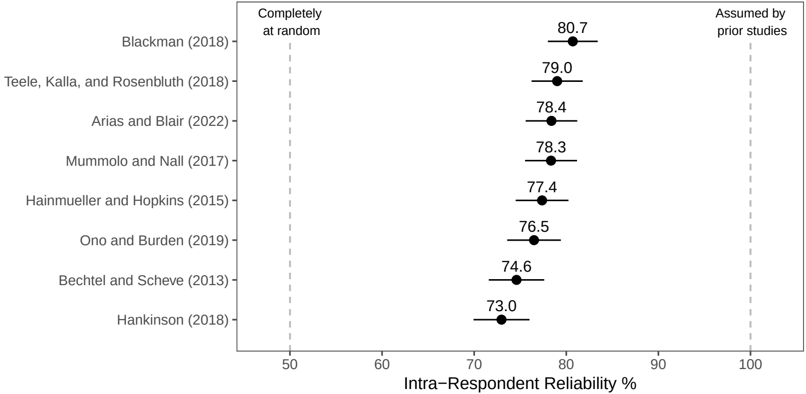
- Randomize 2 of 8 studies & 5 Qs per resp.; $n = 3,289$
- Add one Q: Q1, Q2, Q3, Q4, Q5, Q1; Compute IRR.
- Measurement error: IRR ≈ 0.75

Low Intra-Respondent Reliability!



- Randomize 2 of 8 studies & 5 Qs per resp.; $n = 3,289$
- Add one Q: Q1, Q2, Q3, Q4, Q5, Q1; Compute IRR.
- Measurement error: IRR $\approx 0.75 \rightsquigarrow \tau \approx 0.15$

Low Intra-Respondent Reliability!



- Randomize 2 of 8 studies & 5 Qs per resp.; $n = 3,289$
- Add one Q: Q1, Q2, Q3, Q4, Q5, Q1; Compute IRR.
- Measurement error: IRR $\approx 0.75 \rightsquigarrow \tau \approx 0.15$
- 15% of 0s should be 1s or 1s should be 0s

IRR Doesn't Vary by Attribute Levels

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists

IRR Doesn’t Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs
  - Collect $n = 50$ for each of the 48 pairs: estimate IRR

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs
  - Collect $n = 50$ for each of the 48 pairs: estimate IRR
  - Repeat a second time with $n = 100$, estimate IRR again

IRR Doesn't Vary by Attribute Levels

- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs
  - Collect $n = 50$ for each of the 48 pairs: estimate IRR
  - Repeat a second time with $n = 100$, estimate IRR again
- Conclusions:

IRR Doesn't Vary by Attribute Levels

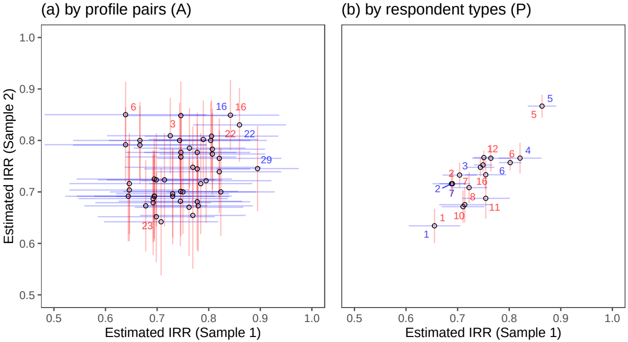
- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs
  - Collect $n = 50$ for each of the 48 pairs: estimate IRR
  - Repeat a second time with $n = 100$, estimate IRR again
- Conclusions:
  - Surveys + real world trade offs: IRR plummets

IRR Doesn't Vary by Attribute Levels

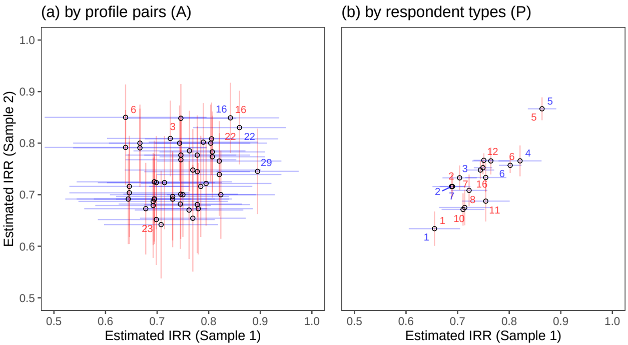
- Top-down approach: Theories tested, rejected by evidence
  - Inconsistency among attribute values
  - Complexity of question wording, attribute lists
  - Divergence between attribute values of the two profiles
- Bottom-up approach
  - Replicate experiment with media articles (Mummolo, 2016)
  - 48 possible sets of attributes for the profile pairs
  - Collect $n = 50$ for each of the 48 pairs: estimate IRR
  - Repeat a second time with $n = 100$, estimate IRR again
- Conclusions:
  - Surveys + real world trade offs: IRR plummets
  - $\text{IRR}(a) \approx$ IRR constant

IRR Varies by Personal Characteristics, not Attributes

IRR Varies by Personal Characteristics, not Attributes

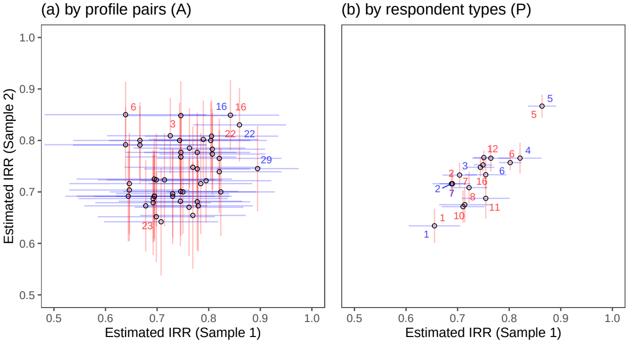


IRR Varies by Personal Characteristics, not Attributes



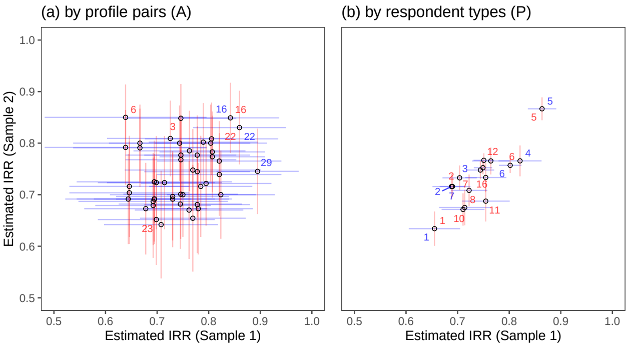
- Left Panel: IRR by Attribute

IRR Varies by Personal Characteristics, not Attributes



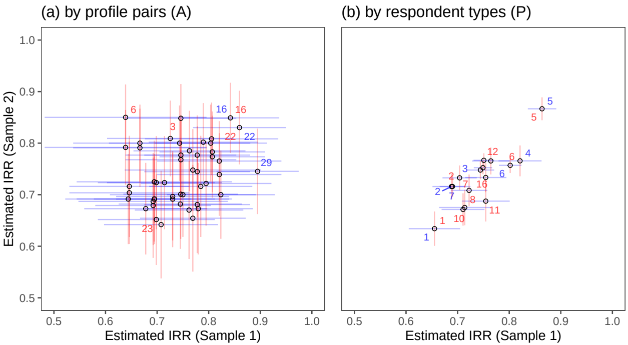
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation

IRR Varies by Personal Characteristics, not Attributes



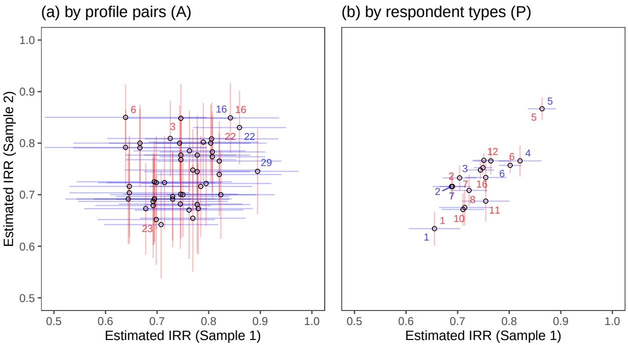
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)

IRR Varies by Personal Characteristics, not Attributes



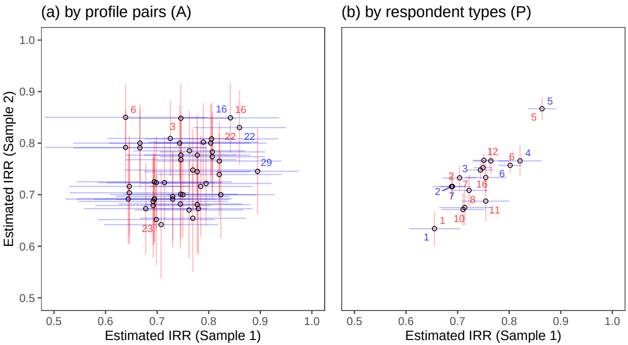
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)
  - IRR(a) ≈ IRR

IRR Varies by Personal Characteristics, not Attributes



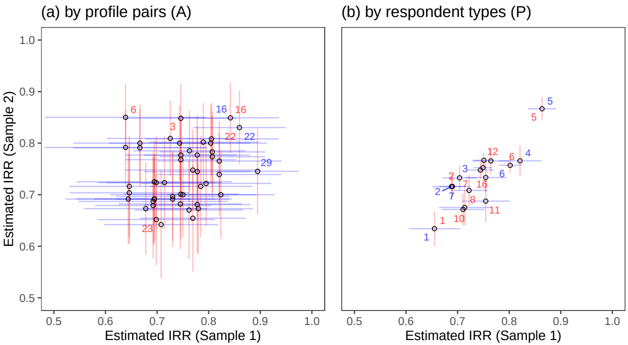
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)
  - $\text{IRR}(a) \approx \text{IRR}$
- Right Panel: IRR by Personal Characteristics

IRR Varies by Personal Characteristics, not Attributes



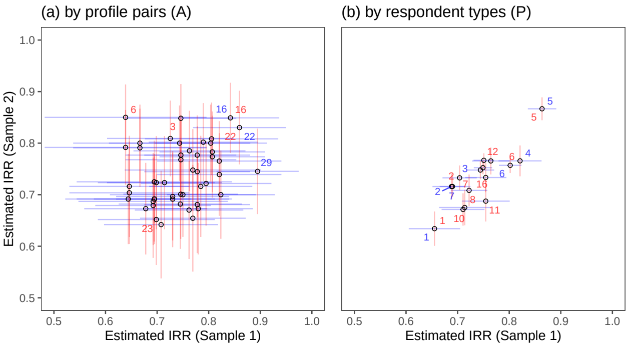
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)
  - $\text{IRR}(a) \approx \text{IRR}$
- Right Panel: IRR by Personal Characteristics
  - Most pairs differ significantly from mean; high correlation (0.85)

IRR Varies by Personal Characteristics, not Attributes



- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)
  - $\text{IRR}(a) \approx \text{IRR}$
- Right Panel: IRR by Personal Characteristics
  - Most pairs differ significantly from mean; high correlation (0.85)
  - Substance: younger, minority, male respondents have lower IRR

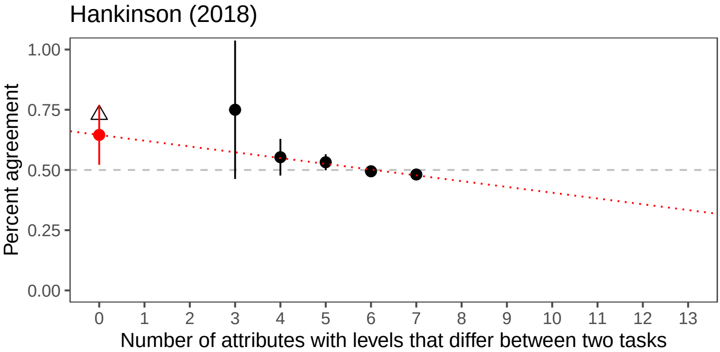
IRR Varies by Personal Characteristics, not Attributes



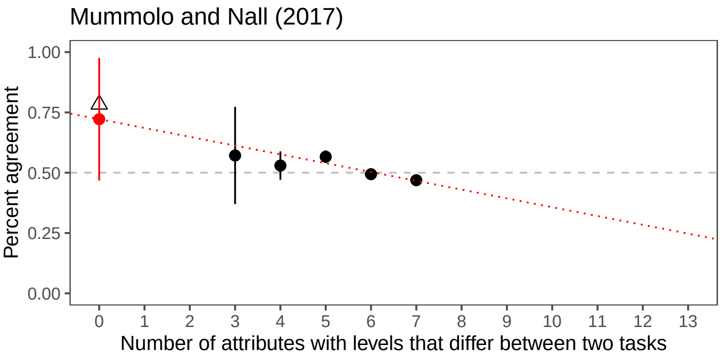
- Left Panel: IRR by Attribute
  - Mean IRR ≈ 0.75 (again); little correlation
  - $E(\text{sig.}|\text{null}) = 2.4$. Actual: 3 (Sample 1) and 5 (Sample 2)
  - IRR(a) ≈ IRR
- Right Panel: IRR by Personal Characteristics
  - Most pairs differ significantly from mean; high correlation (0.85)
  - Substance: younger, minority, male respondents have lower IRR
  - Estimate IRR by subgroups, not by attribute values

Estimating IRR Without New Data

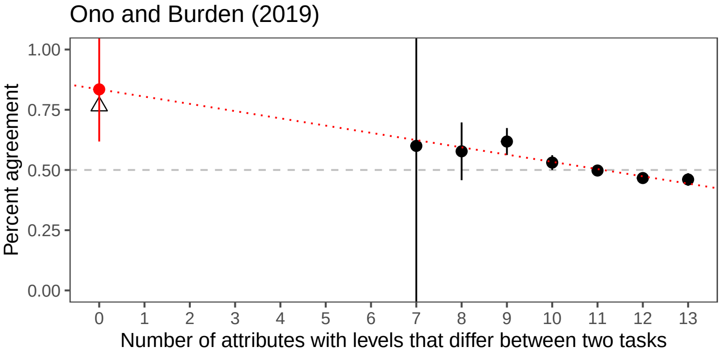
Estimating IRR Without New Data



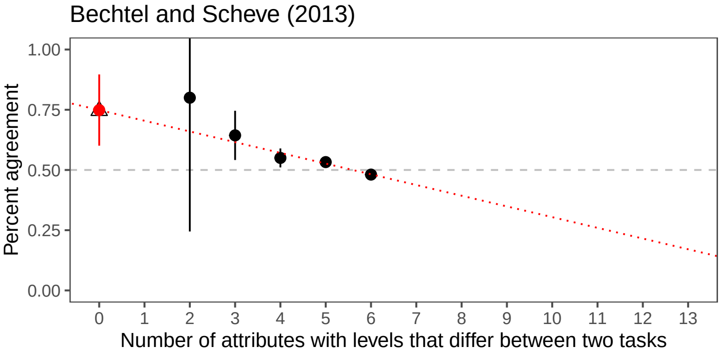
Estimating IRR Without New Data



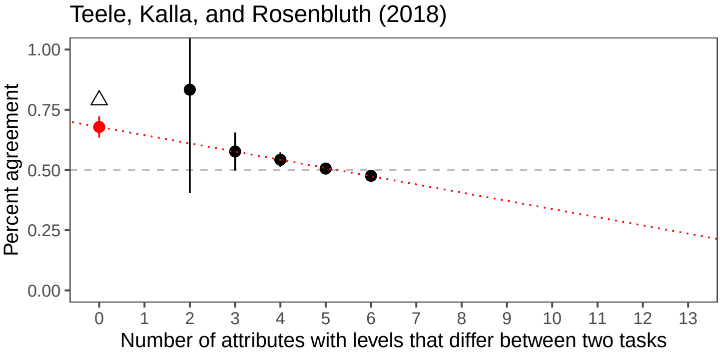
Estimating IRR Without New Data



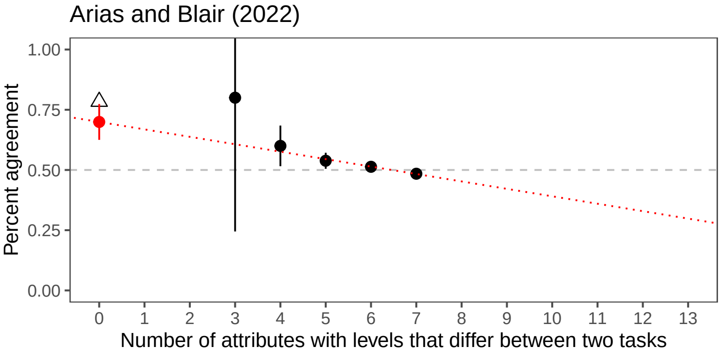
Estimating IRR Without New Data



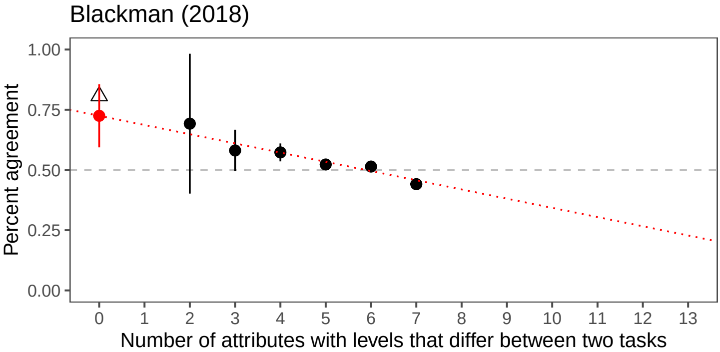
Estimating IRR Without New Data



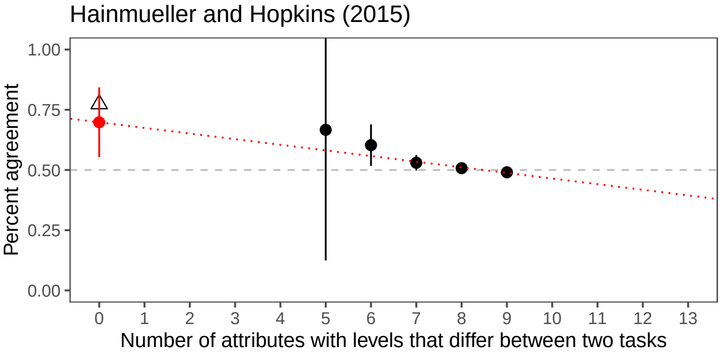
Estimating IRR Without New Data



Estimating IRR Without New Data

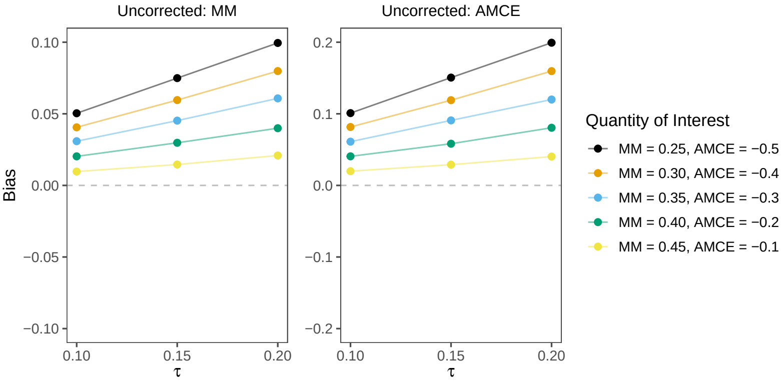


Estimating IRR Without New Data

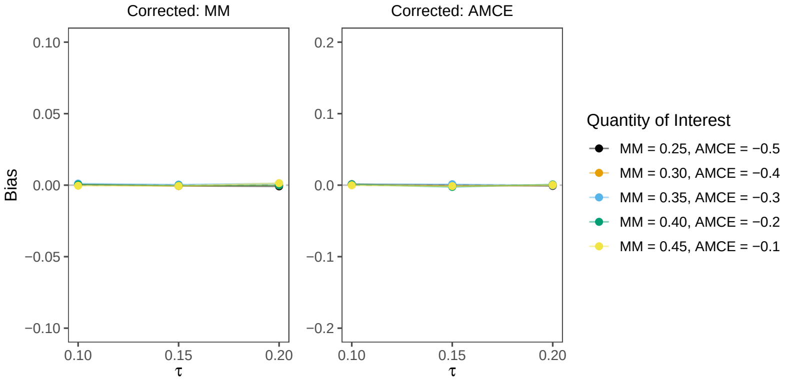


Reducing Bias and Mean Square Error

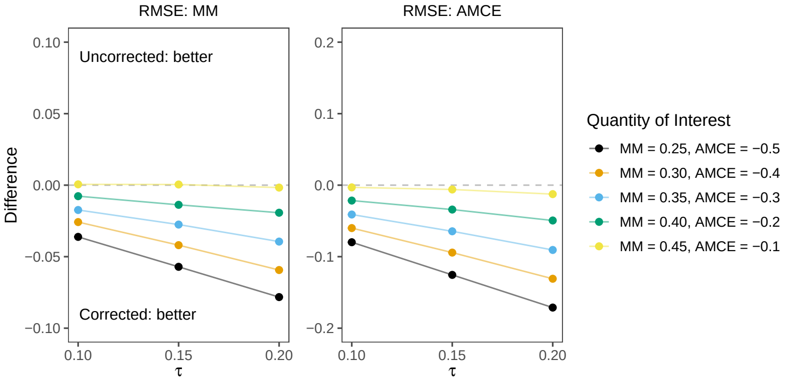
Reducing Bias and Mean Square Error



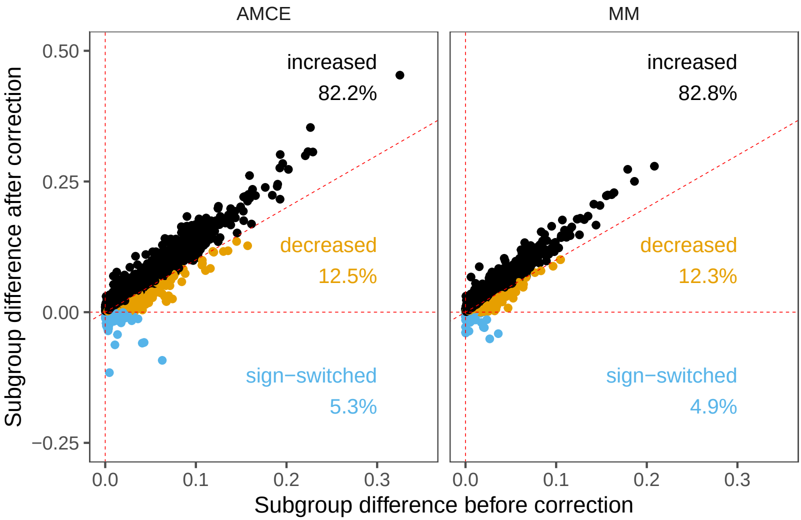
Reducing Bias and Mean Square Error



Reducing Bias and Mean Square Error

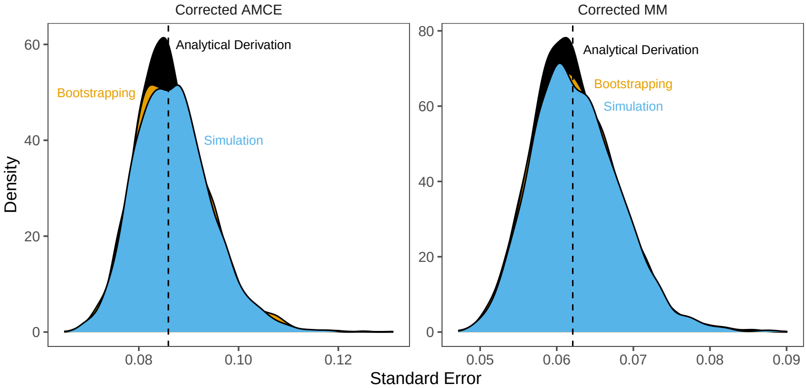


Consequences of Bias Correction in 8 Studies

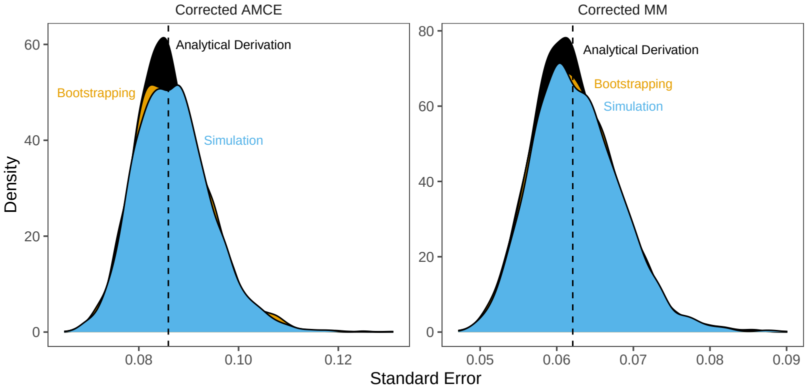


Standard Errors

Standard Errors

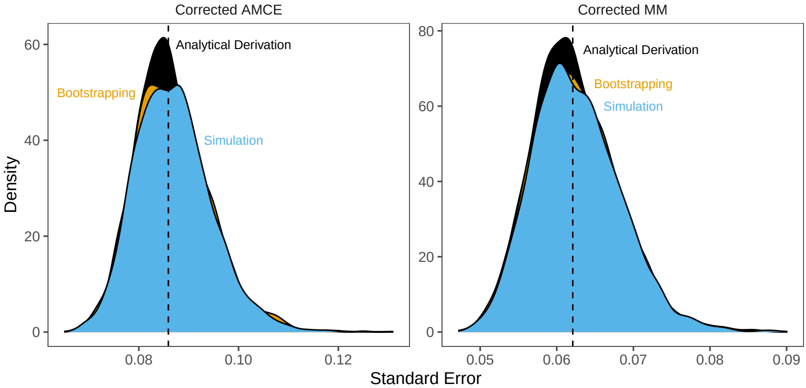


Standard Errors



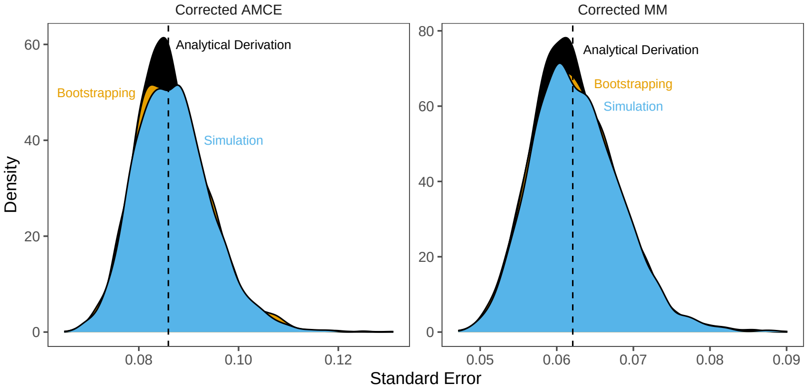
- Convenient: Bootstrapping

Standard Errors



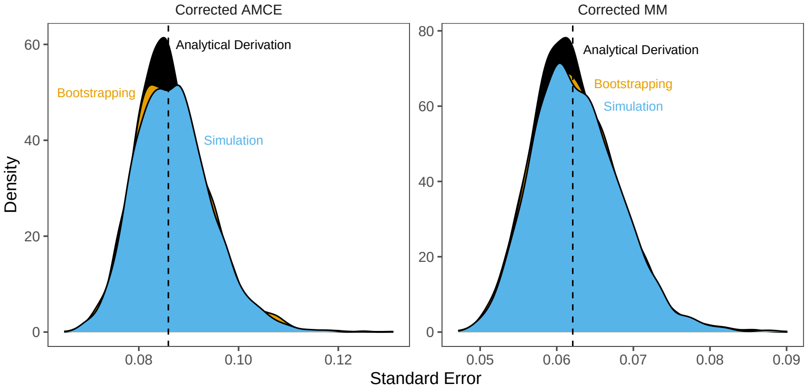
- Convenient: Bootstrapping
- Familiar: Clarify-like simulation

Standard Errors



- Convenient: Bootstrapping
- Familiar: Clarify-like simulation
- Fast: Analytical derivation (790x faster)

Standard Errors



- Convenient: Bootstrapping
- Familiar: Clarify-like simulation
- Fast: Analytical derivation (790x faster)
- Our "Projoint" software: Implements them all

What did he say?

What did he say?

- Simplify methods, generalize substance: use choice-level UOA

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - How Hard to Fix? A few lines of code & estimate IRR

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - How Hard to Fix? A few lines of code & estimate IRR
    - Planning a survey: add one question

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - How Hard to Fix? A few lines of code & estimate IRR
    - Planning a survey: add one question
    - Reanalyzing existing data: extrapolate

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - How Hard to Fix? A few lines of code & estimate IRR
    - Planning a survey: add one question
    - Reanalyzing existing data: extrapolate
- Evidence: 13+ surveys, 9,472 respondents, 137,785 questions

What did he say?

- Simplify methods, generalize substance: use choice-level UOA
- Measurement error in conjoint
  - How much? Lots!
  - Why so much? Approximating real world decisions
  - Ignore it? Bias!
  - How Hard to Fix? A few lines of code & estimate IRR
    - Planning a survey: add one question
    - Reanalyzing existing data: extrapolate
- Evidence: 13+ surveys, 9,472 respondents, 137,785 questions

---

Paper, slides, software, data GaryKing.org/conjointE