Causal Inference with Measurement Error:

Nonparametric Identification and Sensitivity Analyses of a Field Experiment on Democratic Deliberations

Kosuke Imai **Teppei Yamamoto**

Princeton University

July 11, 2008

Measurement Error (ME) and Causal Inference

- Political Science:
  - Extensively studied in the context of survey research
  - Achen (1975), Zaller & Feldman (1992), Bartels (1993), etc.
  - Mostly focused on classical ME in the regression framework
  - Rise of randomized survey experiments (e.g., TESS)
  - How does ME affect causal inference?
- Statistics:
  - Long history of research on ME
  - Mostly focused on non-differential ME
  - Fast growing literature on causal inference
  - Little work on the impact of ME on causal estimation

Differential Measurement Error in Political Science

Differential ME:

- Survey respondents' propensity to misreport causal variables is often correlated with the outcome
- Especially common in retrospective studies

Examples:

1. Causal effects of political knowledge on voting behavior
   - Many election surveys (e.g. ANES and BES) ask knowledge questions only after election
   - Voting could affect the level of political knowledge
   - Regressing voting on knowledge will induce bias!
2. Causal effects of implicit cues and racial predispositions
   - Racial attitudes are often measured after experiment
   - Justification: asking attitudes could nullify implicit cues
   - Do implicit cues work only for those with strong racial predispositions?

Nonparametric Identification Analysis

- Advocated by Manski and others
- Few applications in political science
- Question: What can we learn from the observed data alone?
- Different from the identification of parametric models
- Start with no modeling assumption
- Consider additional assumptions
- Bounds rather than point estimates
- Goals:
  1. Establish the domain of consensus among researchers
  2. Highlight the limitations and advantages of research designs
  3. Characterize the roles of additional assumptions

Nonparametric Sensitivity Analysis

- Advocated by Rosenbaum and others
- Few applications in political science

- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Different from sensitivity analysis under parametric models
- IDENTIFICATION ANALYSIS: Bounding conclusions given certain assumptions
- SENSITIVITY ANALYSIS: Bounding the degree of violations of assumptions given certain conclusions

- Goals:
  1. Examine the sensitivity of conclusions to key assumptions
  2. Evaluate the robustness of conclusions

Methodological Contributions

1. Study causal inference with differential ME
2. Derive sharp (best possible) bounds of the average causal effect
3. Incorporate qualitative knowledge into quantitative analysis
4. Exploit auxiliary information
5. Propose a new sensitivity analysis

Motivating and Illustrative Example

Randomized field experiment on democratic deliberations in São Tomé and Príncipe (Humphreys *et al*. 2006):

- A national forum was held in 2004 after discovery of oil
- Citizens deliberated spending priorities in small groups
- Discussions were moderated by randomly assigned leaders
- Units of observation = discussion groups ($n = 148$)
- Group discussion outcomes were then recorded
- Finally, leaders were asked their own preferences

Questions:

- Can a deliberative process lead to better decision outcomes?
- Can discussion leaders manipulate group discussion outcomes?

Causal Quantities of Interest

- Can discussion leaders manipulate group decisions towards their own policy preferences?
- $Z_i^* \in \{0, 1\}$: leaders' (pre-deliberation) preference
- $Z_i \in \{0, 1\}$: leaders' (post-deliberation) preference
- $Y_i \in \{0, 1\}$: Group discussion outcome

- $Y_i(Z_i^*) \in \{0, 1\}$: potential outcomes
- Average treatment effect (ATE): $\tau^* \equiv \mathbb{E}(Y_i(1) - Y_i(0))$

- Does NOT measure the causal effect of leaders' preferences
- The causal effect of having discussions moderated by a leader with particular preferences

Differential ME in the Deliberations Experiment

Problem:

- Leaders' preferences are measured after group discussions

  We cannot discount the possibility that the preferences of the leaders are a result of, rather than a determinant of, the outcomes of the discussions (Humphreys et al., 2006, p.598)

- It may be $Z_i \neq Z_i^*$ for some i

Possible reasons for differential ME:

- Leaders want to appear effective
- Persuasion by groups

Average Treatment Effect (ATE)

Assumption 1 (Strong Ignorability)

$$Z_i^* \perp\!\!\!\perp (Y_i(1), Y_i(0)) \quad \text{and} \quad 0 < \Pr(Z_i^* = 1) < 1.$$

- In observational studies, condition on pre-treatment covariates X_i
- Under A.1, the ATE is:

$$\tau^* = \Pr(Y_i = 1 \mid Z_i^* = 1) - \Pr(Y_i = 1 \mid Z_i^* = 0).$$

- If ignoring the measurement error problem:

$$\tau = \Pr(Y_i = 1 \mid Z_i = 1) - \Pr(Y_i = 1 \mid Z_i = 0).$$

- But, in general, $\tau^* \neq \tau$.

Classical and Nondifferential Measurement Error

- Classical error-in-variables models:
  - ME is independent of the true treatment status, i.e., $Z_i \perp\!\!\!\perp Z_i^*$
  - ME generally leads to attenuation biases
  - e.g. linear least squares regression
  - Necessarily violated for binary variables!
- Non-differential ME:
  - ME is conditionally independent of the outcome given the true value

Assumption 2 (Nondifferential Measurement Error)

$$Z_i \perp\!\!\!\perp Y_i \mid Z_i^*.$$

Assumption 3 (Restriction on the Degree of Measurement Error)

$$\Pr(Z_i = 0 \mid Z_i^* = 1) + \Pr(Z_i = 1 \mid Z_i^* = 0) < 1.$$

Two known identification results under A.1–3:

- Lewbel (2007):

$$\tau \leq \tau^* < \infty$$

- Bollinger (1996):

$$\tau \leq \tau^* < \max\{\kappa \Pr(Z_i = 1) + \tau \Pr(Z_i = 0), \kappa \Pr(Z_i = 0) + \tau \Pr(Z_i = 1)\},$$

  where $\kappa = var(Y_i)/cov(Z_i, Y_i)$.

Limited Informativeness of Assumptions 1 and 3

- ME is likely to be differential in the deliberation experiment
- Under A.1 alone, the sharp bounds are $[-1, 1]$
- Assume A.3 as well as A.1 (but not A.2; i.e., allow differential ME)

Proposition 4

Under A.1 & 3, the sharp bounds $[\alpha, \beta]$ have the following properties:

1. $\alpha = -1$ if and only if $\Pr(Z_i = 1 \mid Y_i = 1) < \Pr(Z_i = 1 \mid Y_i = 0)$,
2. $\beta = 1$ if and only if $\Pr(Z_i = 1 \mid Y_i = 1) > \Pr(Z_i = 1 \mid Y_i = 0)$.

- The bounds on the ATE are always informative, but only on one side (upper or lower).

An Alternative Formulation

Assumption 4 (Positive Correlation between Z_i^* and Z_i)

$$0 < \Pr(Z_i = 1) < 1, \quad \text{and} \quad corr(Z_i, Z_i^*) > 0.$$

Proposition 5

Under A. 1 & 4,

1. $\alpha = -1$ if and only if $\Pr(Y_i = 1 \mid Z_i = 1) < \Pr(Y_i = 1 \mid Z_i = 0)$ or equivalently $\tau < 0$,
2. $\beta = 1$ if and only if $\Pr(Y_i = 1 \mid Z_i = 1) > \Pr(Y_i = 1 \mid Z_i = 0)$ or equivalently $\tau > 0$.

- Under a minimal set of assumptions, the bounds are informative but wide when differential ME exists

Incorporating Additional Assumptions

- Additional assumptions for more informative inference
- Weaker assumptions than what is necessary for point identification

- Qualitative knowledge about the source of measurement error
- Leaders could be persuaded by groups
- Leaders might have an incentive to misreport

Principal Stratification

- $S_i \in \{c, a, n, d\}$: Group i's "type"
  - $S_i = c$: compliant groups, same discussion outcomes as leader's (pre-deliberation) preference
  - $S_i = a$: always prefers a given policy
  - $S_i = n$: never prefers a given policy
  - $S_i = d$: defiers, outcomes always opposite to leader's preference
- Often called "principal strata" (Frangakis & Rubin 2002)

Observed Str. Y_i	True TTT Z_i^*	Principal Str. S_i
0	0	c, n
0	1	n, d
1	0	a, d
1	1	c, a

Possible Assumptions

Assumption 5 (No persuasion by compliant groups)

$$\Pr(Z_i = z \mid S_i = c, Z_i^* = z) = 1 \quad \text{for} \quad z \in \{0, 1\}.$$

Assumption 6 (Leaders' incentives)

$$\Pr(Z_i = z \mid Y_i = z, Z_i^* = z) = 1 \quad \text{for} \quad z \in \{0, 1\}.$$

- Leaders do not misreport if the actual group decision outcome agrees with their pre-deliberation preference
- Mathematically, A.6 implies A.5

Pooling multiple outcome variables:

- Multiple measures of the same concept are often used
- Might be reasonable to assume that respondents have the same *types* in those questions

Nonparametric Sensitivity Analysis

- Without assumptions, bounds are too wide
- But, additional assumptions may be too controversial
- Leaders may have an incentive to conceal their influence
- Can the study be saved?

- How good does the measurement have to be in order for the study's conclusions to hold? (Recall A.4)

Sensitivity Analysis

$$0 < \Pr(Z_i = 1) < 1, \text{ and } corr(Z_i, Z_i^*) \geq \rho.$$

- Find the minimum value of ρ such that the results hold

How to Derive the Sharp Bounds

Setup:

- Formulate the problem as that of constrained linear optimization
- Use the standard linear programming algorithm

Notation:

- $P_{yz} = \Pr(Y_i = y, Z_i = z)$: observable joint probability
- $Q = \Pr(Z_i^* = 1)$: Treatment assignment probability
- $\psi_{yz} = \Pr(Y_i = y, Z_i = z \mid Z_i^* = 1)$
- $\phi_{yz} = \Pr(Y_i = y, Z_i = z \mid Z_i^* = 0)$

Example: Under A. 1, 4 & 6,

- Objective function: $\tau^* = \sum_{z=0}^{1} \psi_{1z} - \sum_{z=0}^{1} \phi_{1z}$
- Constraints:
  - $P_{yz} = (1-Q)\phi_{yz} + Q\psi_{yz}, \; y, z \in \{0, 1\}$
  - A.4 $\Leftrightarrow \frac{\phi_{01}+\phi_{11}}{P_{01}+P_{11}}(1-Q) + \frac{\psi_{00}+\psi_{10}}{P_{00}+P_{10}}Q < 1$
  - A.6 $\Leftrightarrow \phi_{01} = \psi_{10} = 0$

Sharp Bounds under the Incentive Assumption

Proposition 6 (Sharp Bounds under A.1, 4 & 6)

1. The identification region of τ^* can be expressed as

$$\max\left(-\frac{P_{10}+P_{11}}{1-Q}, \; -\frac{P_{01}}{Q} - \frac{P_{10}}{1-Q}, \; -\frac{P_{00}+P_{01}}{Q}\right) \leq \tau^* \leq \min\left(\frac{P_{00}}{1-Q} - \frac{P_{01}}{Q}, \frac{P_{11}}{Q} - \frac{P_{10}}{1-Q}\right).$$

2. The sharp upper and lower bounds are given by,

$$\max\left\{-1, \; \min\left(P_{00} - \frac{P_{01}P_{10}}{P_{11}} - 1, P_{11} - \frac{P_{01}P_{10}}{P_{00}} - 1\right)\right\} \leq \tau^* \leq \tau.$$

- The naïve estimator τ always leads to overestimation (contrary to nondifferential measurement error)
- The sharp lower bound never exceeds zero
- Auxiliary information about Q

Analytical Strategies in the Other Cases

Setup under the Persuasion Assumption (A. 1, 4 & 5):

- Need to introduce the principal strata probabilities:
  - $\pi_{sz} = \Pr(S_i = s, Z_i = z \mid Z_i^* = 1)$
  - $\eta_{sz} = \Pr(S_i = s, Z_i = z \mid Z_i^* = 0)$ for $s \in \{c, a, n, d\}$ and $z \in \{0, 1\}$
- Objective function: $\tau^* = \pi_{c1} + \pi_{a1} - (\eta_{a1} + \eta_{d1} + \eta_{a0} + \eta_{d0})$
- Constraints:
  - $P_{0z} = (1 - Q)(\eta_{cz} + \eta_{nz}) + Q(\pi_{nz} + \pi_{dz})$
  - $P_{1z} = (1 - Q)(\eta_{az} + \eta_{dz}) + Q(\pi_{cz} + \pi_{az})$
  - A.4 $\Leftrightarrow \sum_{j \in \{c,a,n,d\}} \left\{ \frac{\eta_{j1}}{P_{01}+P_{11}}(1 - Q) + \frac{\pi_{j0}}{P_{00}+P_{10}} Q \right\} < 1$
  - A.5 $\Leftrightarrow \pi_{c0} = \pi_{a0} = \eta_{c1} = \eta_{n1} = 0$
- Now can solve *numerically*
- Similar analysis for different sets of assumptions

Sensitivity Analysis:

- A linear inequality constraint: $\frac{\phi_{01}+\phi_{11}}{P_{01}+P_{11}}(1 - Q) + \frac{\psi_{00}+\psi_{10}}{P_{00}+P_{10}} Q \leq 1 - \rho$
- Plot the bounds against ρ

Data

Questions:

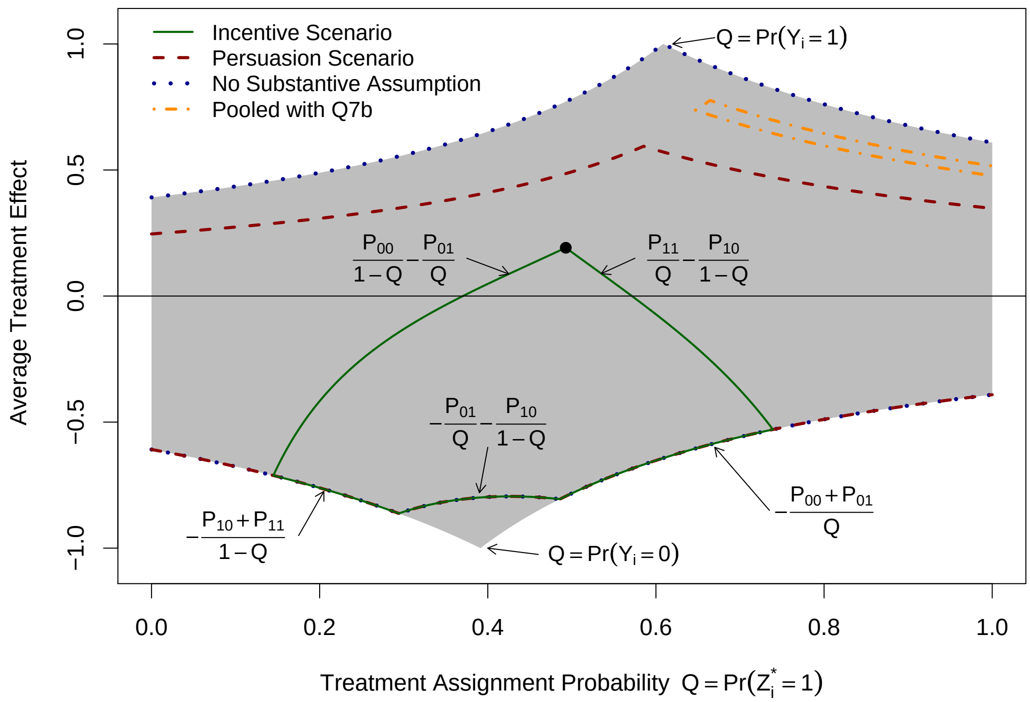
- Q3: local clinics (0) vs. hospitals (1)
- Q4c: advanced education (0) vs. basic education (1)
- Q7b: improving roads (0) vs. public transportation (1)
- Q7c: building village roads (0) vs. roads between centers (1)
- Q11a: consuming (0) vs. investing (1) windfall money

Descriptive Statistics:

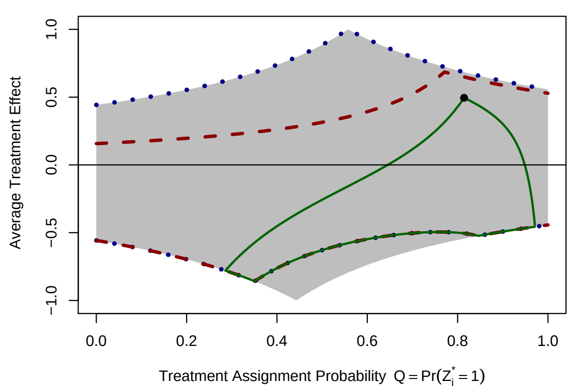
	P_{00}	P_{01}	P_{10}	P_{11}	τ	$\hat{Q}$
Q3	0.157	0.286	0.029	0.529	0.495	0.58
Q4c	0.213	0.025	0.175	0.588	0.508	–
Q7b	0.697	0.171	0.105	0.026	0.002	0.15
Q7c	0.246	0.145	0.261	0.348	0.192	0.19
Q11a	0.176	0.352	0.121	0.352	0.093	0.46

Estimated Sharp Bounds on the ATE

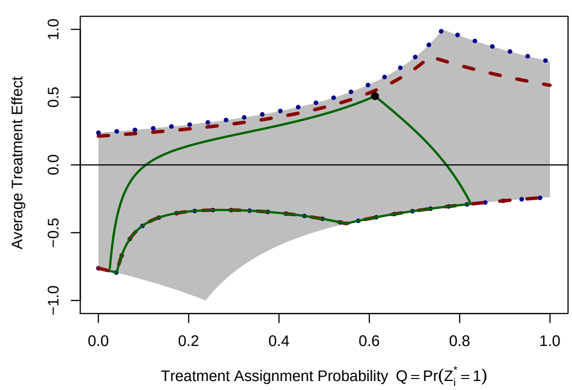
Q7c: Villages (0) or Major Centers (1)?



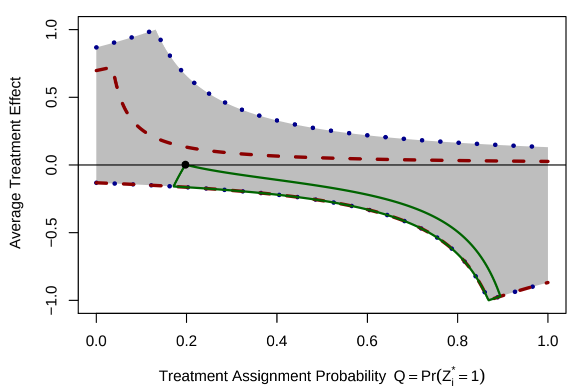
Q3: Clinics (0) or Hospitals (1)?



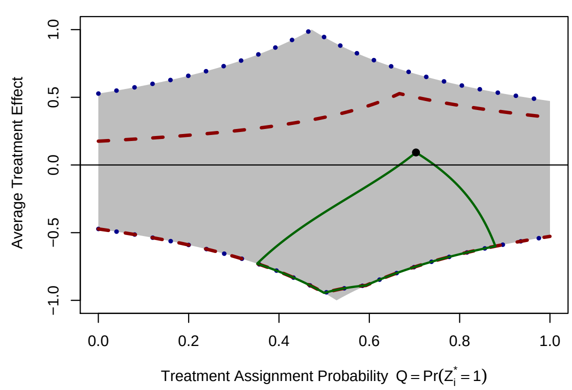
Q4c: Advanced (0) or Basic (1) Education?



Q7b: Roads (0) or Public Transportation (1)?



Q11a: Consume (0) or Invest (1) Windfalls?



Auxiliary Information from the Pre-Forum Survey Data

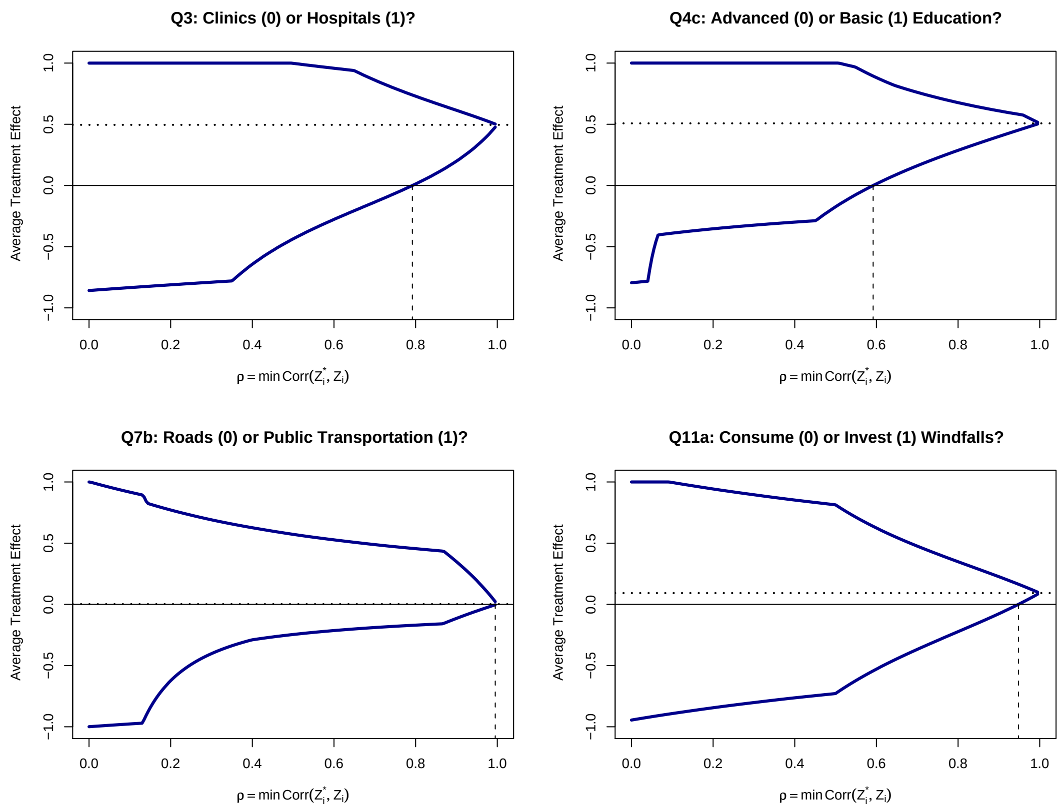
- In the pre-forum survey, 19% preferred major roads to village roads
- Using this as an estimate of Q in Q7c, the sharp bounds on τ^* become $[-0.862, 0.192] \rightarrow [-0.751, -0.459]$.
- Similar results for other questions:
  - Q3 ($\hat{Q} = 58\%$): $[-0.858, 0.495] \rightarrow [-0.561, -0.118]$
  - Q11a ($\hat{Q} = 46\%$): $[-0.945, 0.092] \rightarrow [-0.875, -0.439]$

- Can also use interval estimates (e.g. Q is in $\pm 5\%$ of $\hat{Q}$)
- For Q7b, $\hat{Q}$ is not contained in the possible range of Q $\Longrightarrow$ A.1, 4 & 6 are unlikely to be satisfied or $\hat{Q}$ is a bad estimate
- Impossible to distinguish the two scenarios

Pooling Multiple Discussion Questions

- Questions 7b and 7c asked about similar policy issues (transportation)
- Assume that groups had the same misclassification types in these Qs
- Derivation of the bounds:
  - Use linear programming algorithm as before
  - Add additional constraints on the principal strata probabilities (π_{yz}, η_{yz})
- Results:
  - Strong identification power: Sharp bounds tightened to $[0.477, 0.776]$
  - Barely overlap with the bounds under A.1, 4 & 6

Sensitivity Analysis



Concluding Remarks

- Causal inference is difficult when differential ME is present
- Bounds are informative but wide
- No definitive conclusion about the influence of leaders
- Avoid differential ME if possible!

- Sensitivity analysis: Can the study be saved?
- Additional assumptions based on qualitative knowledge

- Nonparametric identification analysis as a starting point
- Nonparametric sensitivity analysis for evaluating robustness

- To what degree do debates and disagreements in the discipline depend on assumptions rather than empirical data?