Extended Sensitivity Analysis for Heterogeneous Unmeasured Confounding with an Application to Sibling Studies of Returns to Education

Raiden B. Hasegawa   Colin B. Fogarty

Heterogeneous Patterns of Confounding

- A **sensitivity analysis** assesses how robust causal conclusions from an observational study are to departures from random assignment due to unmeasured confounding.
- In a stratified study, the **sensitivity value** is the smallest bound on the maximal bias present in any stratum, while allowing for *arbitrarily unfavorable patterns of confounding*, such that the qualitative conclusions of the study change.
- If the investigator suspects hidden confounding is **present but heterogeneous** across strata *the least favorable patterns of confounding may be very unlikely*, leading the sensitivity analysis to be **overly pessimistic**.

Goal of the Paper: Extend sensitivity analysis to place bounds on both maximal and typical biases present in strata, to account for the presence of heterogeneous confounding.

Example: Returns to Education

Ashenfelter and Rouse (1998):

- Collected survey data on 340 monozygotic twin pairs from the Twinsburg Twins Festival in Twinsburg, Ohio over three summers (1991-1993).
- $I = 40$ pairs where one twin had at least two years of college education ($Z_{ij} = 1$) and the other no more than a HS degree ($Z_{ij'} = 0$).
- Included twin pairs where both twins held a job at the time of survey collection; log wage data reflecting the 1989-1993 period ($R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$).
- Retrospective survey $\implies$ no measures of baseline ability, such as IQ recorded earlier in life (**a potential unobserved confounder** u_{ij}).

Naive analysis assuming no unmeasured confounding:

- p-value ≈ 0.0001 (Fisher's Sharp Null – $H_0 : \mathbf{r}_T = \mathbf{r}_C$)
- 95% CI: $[16\%, 43\%]$

A Model for Unmeasured Confounding

- Assuming **no unmeasured confounding** between twins, treatment assignment in a twin study looks like a randomized experiment:

$$\mathbb{P}(\mathbf{Z} = \mathbf{z} | \, \mathbf{z} \in \Omega) = 1/2^I, \quad \Omega = \{\mathbf{z} : z_{i1} + z_{i2} = 1, \forall\, i = 1, \ldots, I\}.$$

- In the **presence of unmeasured confounding** ($u_{i1} \neq u_{i2}$), distribution of treatment allocations is **unknown**:

$$\mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Z} = \mathbf{z} | \, \mathbf{z} \in \Omega) = \prod_{i=1}^{I} \pi_{i1}^{Z_{i1}} (1 - \pi_{i1})^{1 - Z_{i1}}$$

because $\pi_{ij} = \mathbb{P}(Z_{ij} = 1 \,|\, \mathbf{Z} \in \Omega)$ are **unobserved**.

- Departures from randomization parameterized by $\Gamma > 1$, a **bound on the maximal odds of treatment assignment** in any twin pair:

$$\pi_i^*/(1 - \pi_i^*) \leq \Gamma, \quad \pi_i^* = \max\{\pi_{i1}, \pi_{i2}\}, \quad \forall\, i = 1, \ldots, I.$$

- **Conventional Sensitivity Analysis under this model:** For a test statistic $t = \mathbf{Z}^T \mathbf{q}(\mathbf{R})$ and given maximal bias bound Γ, find the pattern of $\boldsymbol{\pi}$ that maximizes p-value $P(\boldsymbol{\pi})$ under reference distribution $\mathbb{P}_{\boldsymbol{\pi}}$,

$$P_\Gamma^* = \max_{\boldsymbol{\pi} \in \mathcal{P}_\Gamma} P(\boldsymbol{\pi})$$

where $\mathcal{P}_\Gamma = \{\boldsymbol{\pi} : 1/2 \leq \pi_i^* \leq \Gamma/(1 + \Gamma), \forall\, i = 1, \ldots, I\}$

Conventional Sensitivity Analysis of Example

- **Worst case pattern of confounding:** Amounts to assigning twin with higher hourly wage the largest allowable probability of attending college $\implies$ **average and maximal π_i^* coincide.**
- **Sensitivity value:** $\Gamma^* \approx 2.36$ where $\Gamma^* = \inf\{\Gamma : \, P_\Gamma^* \geq 0.05/2\}$.

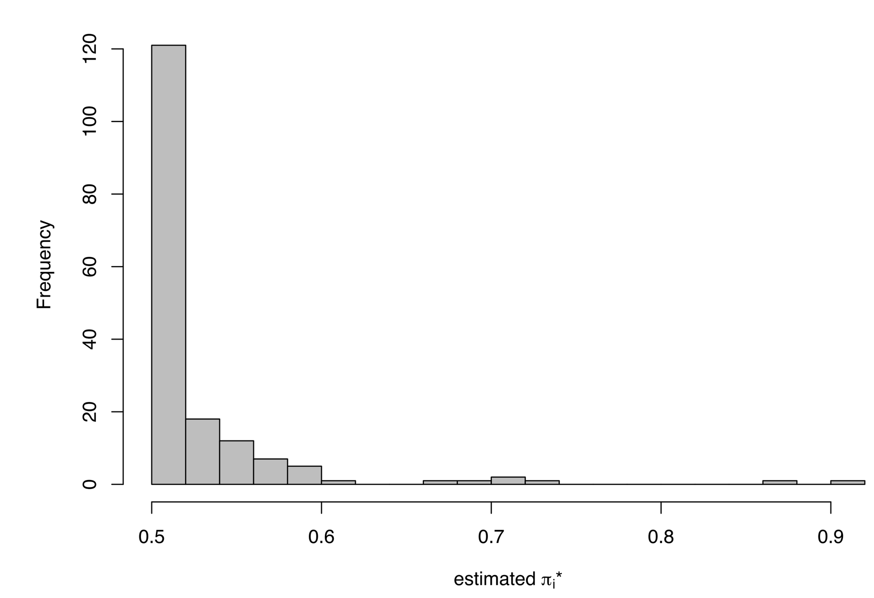
Is 2.36 a plausible bound on the maximal bias among twins due to unmeasured confounding?

- **Example: Ability Bias** – *Could disparities in ability between twins in our study increase the odds that one of the twins went to college by more than a factor of 2.36?*

Cross-Study Calibration

A partial solution to the above open question: Calibrate Γ^* to an estimate of maximal bias due to disparities in ability from a **calibration study** that has comparable design and information on baseline IQ [using calibration methods from Hsu and Small (2013)]. For example,

Estimated π_i^* from 171 same-sex sibling pairs in the Wisconsin Longitudinal Study (WLS) where $Z_{i1} + Z_{i2} = 1$:



- $\max \pi_i^*/(1 - \max \pi_i^*) = 9.3$ and $\bar{\pi}^*/(1 - \bar{\pi}^*) = 1.1 \implies$ mostly modest biases, some moderate and few very large biases.
- $\Gamma^* < 9.3 \implies$ Ashenfelter study **likely sensitive** to plausible levels of ability bias but also suggests *maximal and typical biases due to IQ disparities are quite different*.

Extended Sensitivity Analysis

Our Contribution: Adapt the conventional sensitivity analysis to provide a less pessimistic view of the study's robustness to hidden bias when heterogeneous confounding leads to differing bounds on maximal and typical bias.

Extended Sensitivity Analysis (ESA): A two-parameter sensitivity analysis that simultaneously bounds the maximal and typical biases by Γ and $\bar{\Gamma}$, respectively:

$$P_{\Gamma,\bar{\Gamma}}^* = \max_{\boldsymbol{\pi} \in \mathcal{P}_{\Gamma,\bar{\Gamma}}} P(\boldsymbol{\pi})$$

where $\mathcal{P}_{\Gamma,\bar{\Gamma}}$ is the set of $\boldsymbol{\pi}$ that respect both bounds on maximal and typical biases.

Properties:

- Returns a curve of $(\Gamma, \bar{\Gamma})$ changepoints, analagous to Γ^*.
- If desired, bias parameters can be calibrated to population-level bias information (important in cross-study calibration).
- Can recover conventional sensitivity analysis when $\bar{\Gamma} = \Gamma$.
- Computationally feasible – can be expressed as a standard QP [Fogarty and Small (2016)].

Superpopulation Model for Paired Studies

- Sample a treated subject from an infinite population of treated subjects and record $X_{i1} = x_i$ then sample a control subject from an infinite population of controls conditional on $X_{i2} = x_i$.
- Under this model, π_{ij} is a realization of random variable Π_{ij} $\implies$ π_i^* a realization of random variable $\Pi_i^* = \max\{\Pi_{i1}, \Pi_{i2}\}$.

Why assume a population when inferential framework is conditional on sample?

Why Population-Level Bias Calibration?

Bias estimates from calibration sample are relevant to study sample only insofar as the two samples believably came from some common superpopulation!

A difficulty with constructing a set $\mathcal{P}_{\Gamma,\bar{\Gamma}}$ in terms of bounds on population-level bias:

- Although $1/2 \leq \Pi_i^* \leq \Gamma/(1+\Gamma) \implies 1/2 \leq \max \pi_i^* \leq \Gamma/(1 + \Gamma)$, the same correspondence does not hold for typical biases,

$$1/2 \leq \mathbb{E}[\Pi_i^*] \leq \bar{\Gamma}/(1 + \bar{\Gamma}) \;\not\Longrightarrow\; 1/2 \leq \bar{\pi}_i^* \leq \bar{\Gamma}/(1 + \bar{\Gamma}).$$

Implication for constructing $\mathcal{P}_{\Gamma,\bar{\Gamma}}$: A deterministic set $\mathcal{P}_{\Gamma,\bar{\Gamma}}$ respecting population-level bias bounds would allow for an allocation $\boldsymbol{\pi}$ such that $\bar{\pi}^*$ and $\max \pi_i^*$ coincide, however unlikely $\implies$ **back in the conventional sensitivity analysis.**

How can we remedy this?

Making $\mathcal{P}_{\Gamma,\bar{\Gamma}}$ Stochastic

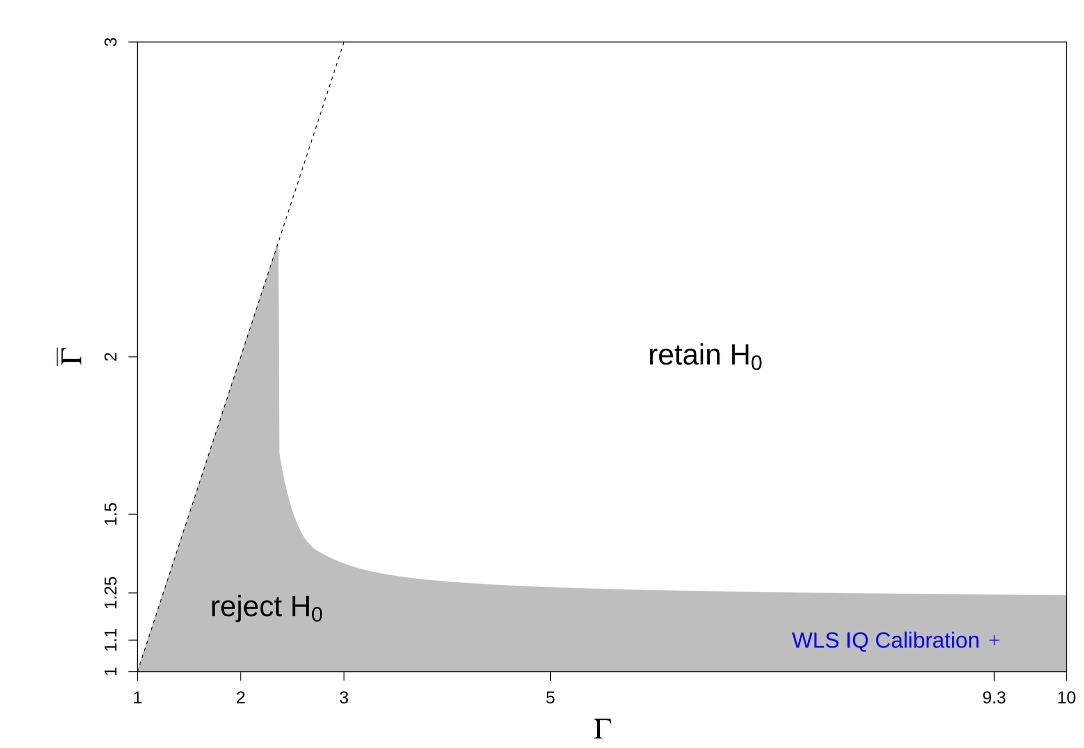
A strategy from Berger and Boos (1994): Maximize p-value over $1 - \beta$ confidence set for a nuisance parameter and add β to get a valid p-value for tests based on non-pivotal quantities.

Modification for ESA: $\boldsymbol{\pi}$ takes the place of the nuisance parameter, need $\boldsymbol{\Pi} \in \mathcal{P}_{\Gamma,\bar{\Gamma}}$ with probability at least $1 - \beta$, and β is added to the maximal p-value to account for stochastic nature of set.

Ensuring $\mathbb{P}(\boldsymbol{\Pi} \in \mathcal{P}_{\Gamma,\bar{\Gamma}}) \geq 1 - \beta$: Can use fact that $\bar{\Pi}^*$ is a sample average, conservative variance estimate for Π_i^* that depends only on $\bar{\Gamma}$ and Γ (Bhatia-Davis Inequality), and apply CLT.

Calibrated Sensitivity Curve

Rather than a single sensitivity parameter, ESA returns a **sensitivity curve** of $(\Gamma, \bar{\Gamma})$ below which we cannot reject H_0, calibrated to estimates of **both** typical and maximal biases due to disparities in IQ:



Takeaway: Unlike the conventional analysis, the calibrated ESA suggests that the Ashenfelter study **is not sensitive** to plausible patterns of ability bias!

Special Cases:

- At (Γ, Γ) we recover the conventional sensitivity analysis $\implies \Gamma^* \approx 2.36$.
- If no plausible bound on maximal bias, taking $\lim_{\Gamma_n \to \infty} (\Gamma_n, \bar{\Gamma})$ recovers single-parameter sensitivity analysis bounding only the typical bias $\implies \bar{\Gamma}^* \approx 1.22$.

Concluding Remarks

- In summary, the extended sensitivity analysis generalizes the conventional sensitivity analysis by allowing the researcher to place bounds on both the maximal and typical bias.
- As illustrated with the Ashenfelter study, this will generally result in less conservative conclusions when there is reason to believe the bounds on maximal and typical bias do not coincide.

Paper pre-print: https://arxiv.org/abs/1711.05570 (*In Revision*)