Sensitivity Analysis for Outcome Tests

Elisha Cohen Emory University
Email: eacohe3@emory.edu Web: http://elishacohen.rbind.io/

1. Introduction: Methodological Contribution

Outcome tests, a method that can be used for evaluating bias in selection making processes, are especially useful when using administrative datasets that contain only observations after the selection process has occurred. I show the outcome test lower bound derived by Knox, Lowe, and Mummolo (2019), can be adapted for use with risk ratios and hence extended to include an E-value style (Ding and VanderWeele, 2016) sensitivity analysis adjustment.

2. Introduction: Application

A very small proportion, 12.5% nationally, of all sworn law enforcement positions are held by women. The status of women in policing has not demonstrably changed in the last two decades and while there is a growing focus in political science on race and ethnicity in policing, little attention has been given to gender. Using data on all settlements paid by the Chicago Police Department from 1985-2015 I estimate a robust lower bound on gender bias with an outcome test. My parameter of interest is the lower bound on the proportion of men who would not have been hired had they been women.

3. Notation and Assumptions

Given two groups, men and women, the selection process to hire a person i is defined as $S_i \in \{0, 1\}$

- $S_i = 1$ hired (observed); $S_i = 0$ not hired (unobserved)
- $Y_i \in \{0, 1\}$ is good performance (no settlement)
- $S_i(m) = 1$ Person i would be hired if a man; $S_i(w) = 1$ Person i would be hired if a woman

Assumption: Monotonicity

The probability that a women who was hired, would not have been hired had they been a man is zero.

$$Pr[S(w) = 1, S(m) = 0)] = 0$$

Women are only "Always" hired strata

		Women	
		$S_i(w) = 0$	$S_i(w) = 1$
Men	$S_i(m) = 0$	Never	~~Hurt~~ ~~$S(m) < S(w)$~~
	$S_i(m) = 1$	Helped $S(m) > S(w)$	Always $S(m) = S(w) = 1$

- "Never" group is undefined since they are not in the data
- $S_i = 1$ for all in the dataset because everyone observed has been hired
- Let $\mathcal{G}_m$ be the probability of no settlement among men
- $G_{m,s(m)>s(w)}$: Probability of no settlement among "Helped" men
- $G_{m,s(m)=s(w)=1}$: Probability of no settlement among "Always" men

Assumption: Comparability

The "Always" hire men are comparable to the "Always" hire women conditional on covariates

$$\mathcal{G}_{m,s(m)=s(w)=1,\boldsymbol{X}} = \mathcal{G}_{w,s(m)=s(w)=1,\boldsymbol{X}}$$

4. Parameter of Interest

Parameter of Interest is p: The proportion of men who would not been hired had they been women

$$\mathcal{G}_m = \underbrace{p}_{\text{Proportion Helped}} \cdot \overbrace{\mathcal{G}_{m,s(m)>s(w)}}^{\text{Mean Helped}} + \underbrace{(1-p)}_{\text{Proportion Always}} \cdot \overbrace{\mathcal{G}_{m,s(m)=s(w)=1}}^{\text{Mean Always}}$$

$$p = \frac{\mathcal{G}_{m,s(m)=s(w)=1} - \mathcal{G}_m}{\mathcal{G}_{m,s(m)=s(w)=1} - \mathcal{G}_{m,s(m)>s(w)}} \Rightarrow p \geq \frac{\mathcal{G}_w - \mathcal{G}_m}{\mathcal{G}_w}$$

For a binary model the group means can be replaced with relative risks R_w and R_m

$$p \geq \frac{\mathcal{R}_w - \mathcal{R}_m}{\mathcal{R}_w} = 1 - \frac{\mathcal{R}_m}{\mathcal{R}_w} = 1 - \frac{1}{RR}$$

Where RR is the risk ratio of women to men. Using a logistic model to estimate the following specification:

$$P[\hat{Y} = 1|X] = \frac{exp^{(\hat{\beta}_w D + \boldsymbol{X}\hat{\gamma} + \hat{\epsilon})}}{1 + exp^{(\hat{\beta}_w D + \boldsymbol{X}\hat{\gamma} + \hat{\epsilon})}}$$

and

$$\hat{RR} = \frac{E[P(\hat{Y} = 1|w = 1, \boldsymbol{X})]}{E[P(\hat{Y} = 1|w = 0, \boldsymbol{X})]}$$

I therefore estimate the lower bound on the proportion of men who would not have been hired had they been women: $\hat{p} \geq 1 - \frac{1}{\hat{RR}}$

Results 1

	Model 1		Model 2	
	Est.	Lower CI	Est.	Lower CI
$\hat{RR}$	1.108	1.097	1.097	1.085
$\hat{OT}$	9.8%	8.8%	8.9%	7.9%

Table 1:Estimated RR and outcome test lower bound conditional on covariates. All confidence interval estimates are bootstrapped.

Two model specifications are used: Model 1 controls for gender, race and officer appointed year and Model 2 additionally controls for current unit and current rank fixed effects. **If Monotonicity and Comparability hold, I estimate at least 9.8%(8.9%) of men would not have been hired had they been women.**

Sensitivity Analysis

The sensitivity analysis is necessary to evaluate how strong an unobserved confounder U must be to potentially reduce the estimated lower bound on the parameter of interest to zero. Adapting the approach by Ding and VanderWeele (2016) I estimate an outcome test lower bound robust to measured and unmeasured confounders.

In order to adjust my estimates to be robust to possible unmeasured confounders I choose an observed covariate with a strong association with the outcome to use as a benchmark. I use the cohort of officers appointed in 2000-2005, a time period with heavily enforced broken windows policing. This cohort has by far the most settlements making it a reasonable choice. I separately estimate the RR for the association of this cohort with the outcome and with gender. Then using the bias factor equation in Ding and VanderWeele (2016) scale the overall RR and lower bound accordingly. This gives me an RR and outcome test estimate robust to possible unmeasured confounders of the same strength as the observed cohort covariate. (This is a "naive" benchmark as discussed in Cinelli and Hazlett (2020) and does not adjust for potential collider bias and therefore could be underestimating the bias. However, it is useful to give some context to scale how much an unmeasured confounder, if measured, could alter the estimated RR.)

Results 2

	Model 1		Model 2	
	Est.	Lower CI	Est.	Lower CI
$\hat{RR}$	1.085	1.066	1.081	1.061
$\hat{OT}$	7.9%	6.2%	7.5%	5.8%

Table 2:Sensitivity adjusted RR and outcome test lower bound estimates conditional on covariates. All confidence interval estimates are bootstrapped.

Using my sensitivity adjusted outcome test, I estimate at least 7.9%(7.5%) of men would not have been hired had they been women.

References

[1] Carlos Cinelli and Chad Hazlett.
Making sense of sensitivity: Extending omitted variable bias.
Journal of Royal Statistical Society Series B, 82(1):39–67, 2020.

[2] Dean Knox, Will Lowe, and Jonathan Mummolo.
Administrative records mask racially biased policing.
American Political Science Review, pages 1–19, 2019.

[3] Peng Ding and Tyler J VanderWeele.
Sensitivity analysis without assumptions.
Epidemiology, 27(3):368, 2016.