

# Unobservable Selection and Coefficient Stability: Theory and Evidence Oster (2019, JBES)

Takaaki Kishida

Kobe University

April 3, 2019

# What we will study?

- Learn a new coefficient stability approach developed by Oster (2019).
- This can be one of robustness checks!
- Do some exercise using stata (`psacalc`).

# What Oster (2019) developed?

- The method for evaluating robustness to omitted variable bias (confounders) under the some assumptions (built on Altonji et al. (2005)).
- Testing coefficient movements in response to the inclusion of observed controls alone can be deceptive regarding the sensitivity of the results to confounders.

Failure of conditional independence assumption (CIA)  
⇒ selection on unobservables

However, CIA isn't sufficient to evaluate the robustness to OVB.

# Example: CIA implies... (from my paper)

Table 1. Effect of Forest Loss on Malaria

Dependent Variable	Malaria		
	(1)	(2)	(3)
Forest Loss (log)			
Last 12 Months	0.0341** (0.0165)	0.0331** (0.0160)	0.0328** (0.0159)
One Year Before	-0.0247 (0.0166)	-0.0254 (0.0164)	-0.0267 (0.0165)
Two Years Before	-0.0133 (0.0156)	-0.0125 (0.0153)	-0.0125 (0.0154)
Subdistrict FE	YES	YES	YES
Year-Month FE	YES	YES	YES
HH and Indvi Controls		YES	YES
Precipitation			YES
Nighttime Lights			YES
Population Density			YES
Observations	20,820	20,820	20,820
R-squared	0.001	0.019	0.020
Number of Subdistrict	1,540	1,540	1,540

Notes : Standard errors clustered at the subdistrict level in parentheses. Statistical significance is denoted as \*\* at 5%.

# How to interpret the results?

- As mentioned, there is still a concern about OVB even the estimated coef(s) is stable to additional observed controls.
- It is necessary to take into account coefficient and R-squared movements.

# Bias-adjusted treatment effect

Oster (2019) defines an approximation of the bias-adjusted treatment effect:

$$\beta^* \approx \tilde{\beta} - \delta[\hat{\beta} - \tilde{\beta}] \frac{R_{\max} - \tilde{R}}{\tilde{R} - \hat{R}}.$$

- $\tilde{\beta}$  and  $\tilde{R}$  are the estimated coefficient and  $R^2$  from a regression with observed controls (column (3))
- $\hat{\beta}$  and  $\hat{R}$  are their equivalents from a regression without observed controls (column (1))
- $\delta$  captures the degree of proportionality, indicating how much of the variation in the outcome explained by the observables and unobservables
- $R_{\max}$  is the  $R^2$  from a hypothetical regression of the outcome on treatment and both observed and unobserved controls
- To identify  $\beta^*$ , we need assumptions regarding  $\delta$  and  $R_{\max}$
- See the original paper for details

# What we need to show in the table?

As far as I know, at least

- $\hat{\beta}$  from the baseline model (w/o controls)
- $\tilde{\beta}$  from the controlled model (w/ full set of controls)
- Identified set  $[\tilde{\beta}, \beta^*(R_{\max}, \delta)]$

Since  $R_{\max}$  and  $\delta$  are unknown, though Oster (2019) suggests useful bounds, calculating  $\beta^*$  with varying  $R_{\max}$  and  $\delta$  would be important.

# Example & interpretation (from my paper)

**Table 2.** Robustness Checks: Selection on Observables and Unobservables

Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Indep. Var. (log)	Baseline Effect $\tilde{\beta}$ , (S.E.), [ $\hat{R}$ ]	Controlled Effect $\tilde{\beta}$ , (S.E.), [ $\hat{R}$ ]	Identified Set $[\tilde{\beta}, \beta^* (2\hat{R}, 1)]$	Exclude Zero?	Within 95% CI?	$\beta^*$ $\delta = 2$
Fever						
Last 12 Months	0.0341** (0.0178)[0.001]	0.0325** (0.0176)[0.02]	[0.0305, 0.0325]	Yes		0.02828
One Year Before	-0.0247 (0.0166)[0.001]	-0.0256 (0.0165)[0.02]	[-0.0266, -0.0256]	Yes		-0.02786
Two Years Before	-0.0133 (0.0156)[0.001]	-0.0125 (0.0153)[0.02]	[-0.0125, -0.0116]	Yes		-0.01073

Note: Column (5) is imcompleted.

Check whether

- identified set exclude zero and
- identified set is within (x% confidence intervals).
- Calculate  $\beta^*$  with different assumptions on  $R_{\max}$  and  $\delta$
- We see that all identified sets do not include zero, and
- the bias-adjusted coefs ( $\beta^*$ ) are very close to actual estimates ( $\tilde{\beta}$ )



Now we  
go to stata!

# References I

- Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy, 113(1):151–184.
- Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. Journal of Business & Economic Statistics, 37(2):187–204.