Selection on Observables

Hasin Yousaf (UC3M)

9th November

Summary

- Altonji, Elder and Taber, JPE, 2005
- Bellows and Miguel, JPubE, 2009
- Oster, J Bus & Econ Stat, 2017

Motivation

- To start lets think about a standard (instrumental variables) model

$$Y_i = \alpha T_i + W_i'\Gamma + u_i, \tag{1}$$

- The key assumption is that we have some instrument Z_i which is correlated with T_i, but

$$Cov(Z_i, u_i) = 0$$

- One can never verify this assumption but must take it on face value
- Of course a special case of this model is OLS in which $Z_i = T_i$

Motivation

- The best justification for the instrument is random assignment
- If Z_i are truly randomly assigned, it should not be correlated with the observable covariates
  - Balancing tests are standard in randomized control trials
  - It is common to run a regression of Z_i on W_i and test whether these are related
- Problems:
  - Just because we don't reject the null does not mean that the assumption is right (low power)
  - If we do reject the null that doesn't mean the assumption is not approximately true
- In other words what we really care about is the magnitude of the relationship not just the F-statistic

Characterize Observables to be like the Unobservables

$$Y_i = \alpha T_i + X_i' \Gamma_X + W_i^* \Gamma^*, \qquad (2)$$

where W_i^* contains all possible co-variates

$$W_i^* \Gamma^* = \sum_{j=1}^{K^*} S_j W_{ij} \Gamma_j + \sum_{j=1}^{K^*} (1 - S_j) W_{ij} \Gamma_j = W_i^{'} \Gamma + u_i$$

where S_j is an indicator for whether W_{ij} is contained in the data set

This motivates the main idea: **If Selection on the unobservables is the same as selection on the observables how large would the bias be?**

Characterize Observables to be like the Unobservables

- Think about running a regression of Z_i on the observable index and unobservable:

$$Proj(Z_i|X_i, W_i^{'}\Gamma, u_i) = \phi_0 + \phi_X^{'} X_i + \phi(W_i^{'}\Gamma) + \phi_u u_i,$$

- Imagine two types of data collectors:
- An incompetent data collector would have no idea what he was doing and choose S_j at random. It turns out (asymptotically) $\phi_u = \phi$
- By contrast suppose we had a perfect data collector. That person would collect all of the variables that were correlated with Z_i so that the only unobservables left would be uncorrelated with Z_i. In that case $\phi_u = 0$
- The truth is probably somewhere in between
- Throughout the remaining part, assumption is incompetent data collector

Bias Formula

Consider the regression model:

$$Y = \beta X + \Psi\omega_i^o + W_2 + \epsilon, \tag{3}$$

- X is the treatment
- Index: $W_1 = \Psi\omega^o$ is vector of observables
- W_2 is not observed
- Proportional selection relation:

$$\delta\frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}$$

- where $Cov(W_i, X) = \sigma_{iX}$ and $Var(W_i) = \sigma_i^2$
- $W_1 \perp W_2$

Bias Formula

- Regression of Y on X gives $\dot{\beta}$ and $\dot{R}$
- Regression of Y on X and ω^o gives $\tilde{\beta}$ and $\tilde{R}$
- Regression (hypothetical) of Y on X, ω^o and W_2 gives β and R_{max}
- Omitted Variable Bias from the first two regressions can be fully determined by following regressions:
  - Each component of ω^o on $X \rightarrow$ Coeff. on X: $\hat{\lambda}_{\omega_i^o|X}$
  - W_2 on $X \rightarrow$ Coeff. on X: $\hat{\lambda}_{W_2|X}$
  - W_2 on X and $\omega^o \rightarrow$ Coeff. on X: $\hat{\lambda}_{W_2|X,\omega^o}$

Bias Formula

Assumption 1: $\delta = 1$

Regression of X on $\omega^o \rightarrow (\mu_1, \cdots, \mu_J)$

Regression of Y on X and ω^o are φ_i

Assumption 2: $\frac{\mu_i}{\mu_j} = \frac{\varphi_i}{\varphi_j}$

A1: Equal Selection Relation (Relaxed later)

A2: Coeff. on X in intermediate reg: Y on X and controls is same with the observed control set ω^o as it would be if we could control for the index W_1

A2: Relative contribution of each var to X is same as their contribution to Y

Bias Formula

$$\dot{\beta} \xrightarrow{p} \beta + \sum_{i=1}^{J} \varphi_i^o \lambda_{\omega_i^o|X} + \lambda_{W_2|X}$$

$$\tilde{\beta} \xrightarrow{p} \beta + \lambda_{W_2|X,\omega^o}$$

$$Then : Bias = \Pi = \frac{\sigma_2^2 \sigma_{1X}}{\sigma_1^2 (\sigma_X^2 - \frac{\sigma_{1X}^2}{\sigma_1^2})}$$

$$Let : \beta^* = \tilde{\beta} - (\dot{\beta} - \tilde{\beta})(\frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}})$$

Bias Formula

Proposition 1: $\beta^* \xrightarrow{p} \beta$

$$(\mathring{\beta} - \tilde{\beta}) \xrightarrow{p} \left(\frac{\sigma_{1X}}{\sigma_X^2}\right)\left(1 - \frac{\sigma_{1X}}{\sigma_1^2}\Pi\right)$$

$$\left(\tilde{R} - \mathring{R}\right)\hat{\sigma}_y^2 \xrightarrow{p} \sigma_1^2 + \Pi^2(\sigma_X^2 - \frac{\sigma_{1X}^2}{\sigma_1^2}) - \frac{1}{\sigma_X^2}\left(\sigma_{1X} + \Pi(\sigma_X^2 - \frac{\sigma_{1X}^2}{\sigma_1^2})\right)^2$$

$$(R_{max} - \tilde{R})\hat{\sigma}_y^2 \xrightarrow{p} \Pi\left(\frac{\sigma_1^2\left(\sigma_X^2 - \frac{\sigma_{1X}^2}{\sigma_1^2}\right)}{\sigma_{1X}} - \Pi\left(\sigma_X^2 - \frac{\sigma_{1X}^2}{\sigma_1^2}\right)\right)$$

3 equations and 3 unknowns $(\sigma_1^2, \sigma_{1X}, \Pi)$

$$\Pi = (\dot{\beta} - \tilde{\beta})(\frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}})$$

Significance

- "It is reassuring that the estimates are very similar in the standard and the augmented specifications, indicating that our results are unlikely to be driven by omitted variables bias." (Chiappori et al., JPE, 2012)
- "These controls do not change the coefficient estimates meaningfully, and the stability of the estimates from columns 4 through 7 suggests that controlling for the model and age of the car accounts for most of the relevant selection."(Lacetra et al., AER, 2012)
- However, even under the most optimistic assumption, coefficient movements alone are not a sufficient statistic to calculate bias

Example

$$Y = \beta X + W + C$$

- X is education and Y is wage
- W and C are two orthogonal components of ability
- Assume $\sigma_W^2 >> \sigma_C^2$
- Assume both W and C relate to X in the same way
- Different results if researcher observes W or C Table
- Coeff. changes very little if C is observed because it is less important in explaining wages

Comparison with Altonji, Elder and Taber, JPE, 2005

- Estimate is consistent only under the null of zero treatment effect ($\beta = 0$)
  - Cannot do under some other null
  - Oster not only finds δ for $\beta = 0$, but also
  - allows to estimate β for each assumed δ

- Do not provide explicit form for bias-adjusted treatment effect
  - Can only calculate statistic for controls moving coeff. *towards* zero
  - Inclusion of controls moving coeff. away from zero is "robust" by definition
  - Oster allows to calculate bias-adjusted β for each assumed δ

- Assume $R_{max} = 1$
  - Unlikely to hold in many cases Figure
  - Oster allows any potential value of R_{max}
  - Oster suggests $R_{max} = 1.3\tilde{R}$ as a reasonable approx.

Comparison with Bellows and Miguel, JPubE, 2009

- Follow Altonji, Elder and Taber, JPE, 2005
- **Additional Assumption:** Observed and unobserved controls are equally important in explaining the outcome
- Technically, this implies $R_{max} = \tilde{R} + (\tilde{R} - \dot{R})$
- Yields a nice form: $\frac{\beta_F}{\beta_R - \beta_F}$
- **Problem:** Ignore that coefficient movements should be scaled by R^2

Rules of Thumb

- Both Oster and AET suggest $\delta = 1$ as an upper bound
- Note: $\delta = 0 \implies$ "No role of Unobservables"
- Bounding Set:

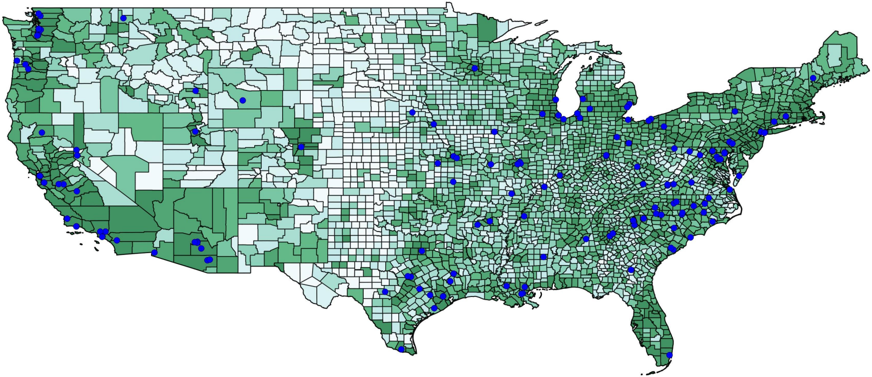
$$\Delta_s = [\tilde{\beta}, \beta^*(R_{max}, 1)]$$

- Easily implementable in Stata with command: **psacalc**

Application

- In my research I am interested in gauging the effect of mass shootings *MS* on political outcomes
- I have MS and electoral data at county level
- I am interested in estimating following DiD setup:

Locations of Mass Shootings



Application

$$repshare_{it} = \alpha_i + \alpha_t + \beta(MSE * Post)_{it} + X'_{it}\Gamma + u_{it}, \quad (4)$$

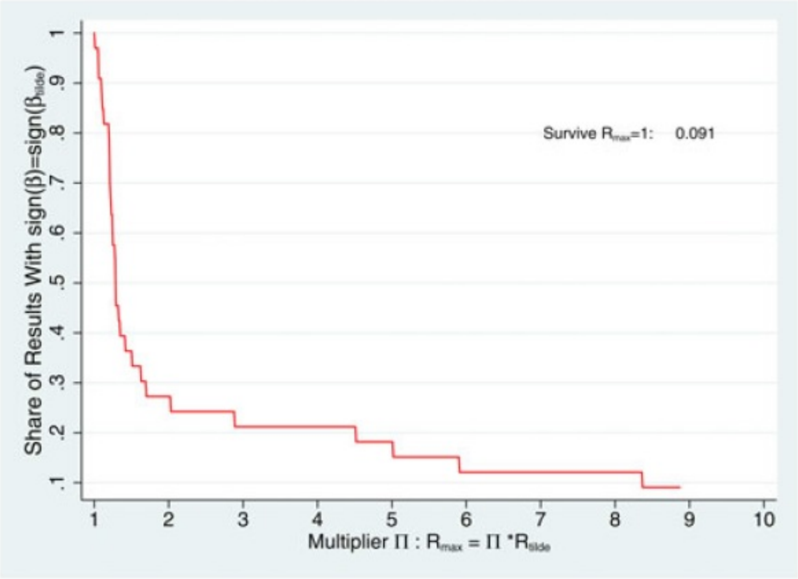
- $repshare_{it}$: Republican vote share in county c, election t
- MSE: Dummy equal to 1 if a county has MS during sample period, 0 otherwise
- $Post$: Dummy equal to 1 for all periods after a county has MS, 0 otherwise
- X'_{ct}: Vector of controls
    - Demographic, Economic, Gun related, Crime related, Health related

VARIABLES	(1) REP	(2) REP	(3) REP	(4) REP	(5) REP	(6) REP	(7) REP	(8) REP
MS * Post	-0.044***	-0.043***	-0.042***	-0.042***	-0.040***	-0.045***	-0.040***	-0.036***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
N	15,658	15,658	15,658	15,657	15,654	15,653	15,658	15,649
R^2	0.378	0.406	0.383	0.380	0.382	0.388	0.410	0.424
Controls	None	Demog.	Econ.	Gun	Crime	Health	Pref.	All

	Restricted	Full	BM	$R_{max} = 1.2\tilde{R}$	$R_{max} = 1.3\tilde{R}$	$R_{max} = 1.5\tilde{R}$
Presidential	-0.044	-0.040	11.49	3.87	2.58	1.55
R	0.357	0.410				
SE	0.011	0.011				
House	-0.024	-0.028	-7.79	-23.14	-15.44	-9.28
R	0.089	0.171	*			
SE	0.010	0.011				

	Baseline	Controlled	Identified Set	Exclude Zero	Within Coef Interval	δ for $\beta = 0$	BM δ for $\beta = 0$
Presidential	-0.044	-0.040	[-0.026, -0.040]	Yes	Yes	2.58	11.49
House	-0.024	-0.028	[-0.028, -0.041]	Yes	Yes	-15.44	-7.79

(a) Rejection of Zero, $R_{max} = \Pi \tilde{R}$.



Return

Example

Table 1. Calibrated example

High versus low variance control			
Quality of observed control	Uncontrolled coefficient [R^2]	Controlled coefficient [R^2]	True effect
High variance control observed	0.202 [0.004]	0.002 [0.990]	0
Low variance control observed	0.202 [0.004]	0.200 [0.013]	0

NOTES: This table shows calculations based on simulated data. The true model is $Y = \beta X + W + C$, with $\beta = 0$. The data are constructed so the high variance control is W and $\text{var}(W) = 10$ and the low variance controls is C and $\text{var}(C) = 0.1$. $\text{var}(X) = 1$, $\text{cov}(X, W) = 0.2$ and $\text{cov}(X, C) = 0.002$. Note that $\text{cov}(X, C)$ is implied by the equal selection assumption, $\text{cov}(X, W)$, $\text{var}(C)$, and $\text{var}(W)$.

Return