ECON 626: Applied Microeconomics

Lecture 6:

Selection on Observables

Professors: Pamela Jakiela and Owen Ozier

Department of Economics
University of Maryland, College Park

Altonji, Elder, Taber (2005)

$$Y^* = \alpha CH + \boldsymbol{W}'\boldsymbol{\Gamma}$$

$$Y^* = \alpha CH + \boldsymbol{X}'\boldsymbol{\Gamma}_X + \xi$$

$$Y^* = \alpha CH + \boldsymbol{X}'\gamma + \epsilon$$

Condition:

$$\frac{E[\epsilon|CH=1] - E[\epsilon|CH=0]}{Var(\epsilon)} = \frac{E[\boldsymbol{X}'\gamma|CH=1] - E[\boldsymbol{X}'\gamma|CH=0]}{Var(\boldsymbol{X}'\gamma)}$$

Let $CH = \boldsymbol{X}'\beta + \widetilde{CH}$.

$$Y^* = \alpha\widetilde{CH} + \boldsymbol{X}'(\gamma + \alpha\beta) + \epsilon$$

$$plim\ \hat{\alpha} = \alpha + \frac{Var(CH)}{Var(\widetilde{CH})}\left(E[\epsilon|CH=1] - E[\epsilon|CH=0]\right)$$

Altonji, Elder, Taber (2005)

"1. the elements of X are chosen at random from the full set of factors W that determine Y;
2. the numbers of elements in X and W are large, and none of the elements dominates the distribution of CH or the outcome Y; and
3. the relationship between the observable elements X and the unobservables obeys an assumption that is very strong but weaker than the standard assumption Cov (X, ξ) = 0. Roughly speaking, the assumption is that the regression of CH^* on Y^* - αCH is equal to the regression of the part of CH^* that is orthogonal to X on the corresponding part of Y^* - αCH."

Altonji, Elder, Taber (2005)

"... as a result of the limits on the number of the factors that we know matter and that we know how to collect and can afford to collect, many elements of W are left out."
"...it is better to think of the elements of X as a ... random subset of ... W rather than a set ... systematically chosen to eliminate bias."
"...The relatively large number and wide variety of observables that enter into our problem suggest that the observables may provide a useful guide to the unobservables."
"...[we may] expect the relationship between the unobservables and CH (or, more generally, any potentially endogenous treatment) to be weaker than the relationship between the observables and CH. First, X often has been selected with an eye toward reducing bias in single-equation estimates rather than at random. ... Second, in the case of the twelfth grade test scores, ϵ will also reflect the substantial variability in test performance on a particular day, which presumably has nothing to do with the decision to start Catholic high school. Finally, and most importantly, shocks that occur after eighth grade are excluded from X."

Altonji, Elder, Taber (2005)

"(Note that when $Var(\epsilon)$ is very large relative to $Var(\boldsymbol{X}'\gamma)$, what one can learn is limited, because even a small shift in $(E[\epsilon|CH=1] - E[\epsilon|CH=0])\,/\,Var(\epsilon)$ is consistent with a large bias in α.)"

Altonji, Elder, Taber (2005)

TABLE 3
OLS and Probit Estimates of Catholic High School Effects in Subsamples of NELS:88 (Weighted)

	Full Sample: Controls				Catholic 8th Grade Attendees: Controls			
	None (1)	Family Background, City Size, and Region[a] (2)	Col. 2 Plus 8th Grade Tests (3)	Col. 3 Plus Other 8th Grade Measures[b] (4)	None (5)	Family Background, City Size, and Region[a] (6)	Col. 2 Plus 8th Grade Tests (7)	Col. 3 Plus Other 8th Grade Measures[b] (8)
A. High School Graduation								
Probit	.97	.57	.48	.41	.99	.88	.95	1.27
	(.17)	(.19)	(.22)	(.21)	(.24)	(.25)	(.27)	(.29)
	[.123]	[.081]	[.068]	[.052]	[.105]	[.084]	[.081]	[.088]
Pseudo R^{2}[c]	.01	.16	.21	.34	.11	.35	.44	.58
B. College in 1994								
Probit	.73	.37	.33	.32	.60	.48	.56	.60
	(.08)	(.09)	(.09)	(.09)	(.13)	(.15)	(.15)	(.15)
	[.283]	[.106]	[.084]	[.074]	[.236]	[.154]	[.154]	[.149]
Pseudo R^2	.02	.19	.29	.34	.04	.18	.29	.36
C. 12th Grade Reading Score								
OLS	4.28	2.08	1.18	1.14	1.92	.17	.37	.33
	(.47)	(.54)	(.38)	(.38)	(.82)	(.98)	(.63)	(.62)
R^2	.01	.19	.60	.60	.01	.19	.59	.62
D. 12th Grade Math Score								
OLS	4.86	1.98	1.07	.92	2.79	1.10	1.46	1.14
	(.44)	(.54)	(.34)	(.32)	(.77)	(1.00)	(.53)	(.46)
R^2	.01	.26	.72	.74	.02	.26	.73	.77

Altonji, Elder, Taber (2005)

TABLE 3
OLS and Probit Estimates of Catholic High School Effects in Subsamples of NELS:88 (Weighted)

		Catholic 8th Grade Attendees: Controls			
		None (5)	Family Background, City Size, and Region[a] (6)	Col. 2 Plus 8th Grade Tests (7)	Col. 3 Plus Other 8th Grade Measures[b] (8)
A. High School Graduation	Probit	.99	.88	.95	1.27
		(.24)	(.25)	(.27)	(.29)
		[.105]	[.084]	[.081]	[.088]
	Pseudo R^{2c}	.11	.35	.44	.58
B. College in 1994	Probit	.60	.48	.56	.60
		(.13)	(.15)	(.15)	(.15)
		[.236]	[.154]	[.154]	[.149]
	Pseudo R^2	.04	.18	.29	.36
C. 12th Grade Reading Score	OLS	1.92	.17	.37	.33
		(.82)	(.98)	(.63)	(.62)
	R^2	.01	.19	.59	.62
D. 12th Grade Math Score	OLS	2.79	1.10	1.46	1.14
		(.77)	(1.00)	(.53)	(.46)
	R^2	.02	.26	.73	.77

Altonji, Elder, Taber (2005)

TABLE 6
Amount of Selection on Unobservables Relative to Selection on Observables Required to Attribute the Entire Catholic School Effect to Selection Bias

Outcome	$[\hat{E}(\boldsymbol{X}'\hat{\gamma}\mid CH = 1) - \hat{E}(\boldsymbol{X}'\hat{\gamma}\mid CH = 0)] \div \widehat{\text{Var}}(\boldsymbol{X}'\hat{\gamma})$ (1)	$\widehat{\text{Var}}(\hat{\epsilon})$ (2)	$E(\epsilon\mid CH = 1) - E(\epsilon\mid CH = 0)$[a] (3)	$\text{Cov}(\epsilon, \widetilde{CH}) \div \text{Var}(\widetilde{CH})$ (4)	$\hat{\alpha}$ (5)	Implied Ratio[b] (6)
	A. $\hat{\alpha}$ Estimated from the Catholic Eighth Grade Subsample, Full Set of Controls[c]					
High school graduation (N=859)	.24	1.00	.24	.29	1.03 (.31)	3.55
College attendance (N=834)	.39	1.00	.39	.47	.67 (.16)	1.43
12th grade reading (N=739)	.091	36.00	3.28	3.94	.33 (.62)	.08
12th grade math (N=739)	.038	24.01	.91	1.09	1.14 (.46)	1.04

Bellows and Miguel (2009)

J. Bellows, E. Miguel / Journal of Public Economics 93 (2009) 1144–1157 1151

Table 3
Community meetings and conflict victimization.

	Dependent variable: did you attend any community meetings in the past year?			
	IRCBP			
	2005 and 2007		2007	
Explanatory variables	(1)	(2)	(3)	(4)
Conflict victimization index	0.0704***	0.0652***	0.0775***	0.0686***
	(0.0164)	(0.0165)	(0.0253)	(0.0246)
Respondent is female		−0.1300***		−0.1276***
		(0.0084)		(0.0126)
Respondent age		0.0003		0.0002
		(0.0003)		(0.0005)
Respondent has any education		0.0590***		0.0466**
		(0.0108)		(0.0194)
Traditional authority household		0.0928***		0.0647***
		(0.0128)		(0.0194)
1990 Household head had any education				0.0205
				(0.0199)
1990 Household had a traditional leader				0.1054***
				(0.0217)
1990 Household had a community leader				−0.0067
				(0.0169)
R-squared	0.361	0.391	0.267	0.298
Observations	10,471	10,471	5193	5193
Enumeration area/Year fixed effects	X	X	X	X

Oster (2016)

"A common approach to evaluating robustness to omitted variable bias is to observe coefficient movements after inclusion of controls. This is informative only if selection on observables is informative about selection on unobservables. Although this link is known in theory (i.e. Altonji, Elder and Taber (2005)), very few empirical papers approach this formally. I develop an extension of the theory which connects bias explicitly to coefficient stability. I show that it is necessary to take into account coefficient and R-squared movements. I develop a formal bounding argument. I show two validation exercises and discuss application to the economics literature."

Oster (2016)

$$Y = \beta X + \psi\omega^o + W_2 + \epsilon$$

Define:

$\overset{o}{\beta}$ to be the coefficient from regressing Y on X;
$\overset{o}{R}$ to be the R^2 from this regression;

$\tilde{\beta}$ to be the coefficient (on X) from regressing Y on X and ω^o;
$\tilde{R}$ to be the R^2 from this regression;

R_{max} to be the R^2 from regressing Y on X, ω^o, and W_2;

$$\delta \frac{Cov(\psi\omega^o, X)}{Var(\psi\omega^o)} = \frac{Cov(W_2, X)}{Var(W_2)}$$

Then:

$$\beta^* \approx \tilde{\beta} - \delta \left[\overset{o}{\beta} - \tilde{\beta} \right] \frac{R_{max} - \tilde{R}}{\tilde{R} - \overset{o}{R}} \xrightarrow{p} \beta$$