March 2011

LABOUR LECTURES

Selection on Observables to Address Bias from Unobservables, with Applications to the School Choice Literature

Joseph G. Altonji

Yale University

1 Overview of the Lectures

1.1 Two Objectives

- Provide an overview of empirical evidence on two key issues that determine the merits school choice programs.

1. (a) Are private schools more effective than public schools?

2. Will school choice lead the most advantaged students to exit regular public schools, leading to a harmful "cream skimming effect" of school choice on students who remain in the existing schools.

- Discuss new estimators of treatment effects based on formally modeling how observed variable and unobserved variables are related.
  - The new approaches will be placed in the context of informal checks for selection and omitted variables bias that researchers routinely use.

1.2 Lecture 1. Sample Selection Bias and Studies of the Effectiveness of Private schools

Part A

- Regression, Instrumental Variables, and Regression Discontinuity Approaches to Measuring School Effectiveness

- Basic Results

- The Informal Use of Selection on the Observables to Assess Bias from Omitted Variables, Selection, and Invalid Exclusion Restrictions

Part B

- The Theoretical Foundation for Using the Degree of Selection on Observables to Assess Bias from Unobservables

1.3 Lecture 2. Estimation Methods and Applications

- The "Observables-Unobservables" (OU) Estimator
- Application of OU to Catholic School Effect, Swan Ganz procedure
- Sensitivity analysis related to the OU Estimator, with applications to Catholic school effect and Swan-Ganz
- brief discussion of heterogenous treatment effects case (very preliminary)
- The OU-Factor Estimator
- Consistency of OU-Factor
- Constructing Confidence Intervals

- Monte Carlo Evidence

- Conclusion

Lectures 1a, 1b, and Lecture 2 draw on

- "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, with Todd Elder and Christopher Taber, *Journal of Political Economy*, Vol. 113, February 2005.

- "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling, with Todd Elder and Christopher Taber, *Journal of Human Resources*, Fall 2005.

- Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization Selection on Observables and Unobservables, with-Todd Elder and Christopher Taber, *American Economic Review Papers and Proceeding, May 2008*

- “Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables,” with Timothy Conley, Todd Elder and Christopher Taber, September 2011 (on my web page).

1.4 Lecture 3 : Issues in Evaluating of School Choice Reform

- The Effects of Competition on Public School Performance

- A Theoretical Framework for Assessing the Cream Skimming Effect of School Choice
  - How heterogenous are schools
  - Which Students will leave
  - How large are peer effects

- Econometric Issues
  - Estimating a School Choice Model When Peers Effect Demand
  - Allowing peer effects to depend on unobserved variables

- Evidence on the Cream-Skimming Effect.

Key Reference for Lecture: "Estimating the Cream Skimming Effect of School Choice," with Ching-I Huang and Christopher R. Taber, NBER Working Paper No 16579, December 2010, http://www.nber.org/papers/w16579

Lecture 1a:

Sample Selection Bias and Studies of the Effectiveness of Private schools

2 Research on the Effectiveness of Private Schools

- Dissatisfaction with state operated schools has lead to long running debate in the U.S. and in other countries over whether private schools would do a better job

- Many studies examines effects of private Schools on test scores

- A smaller set looks at effects educational attainment and wages

- Why?
  - Assess merits of private schooling
  - lessons for public schools?
  - consequences of expansion of school choice, vouchers, alternative publicly funded school (Charter Schools).
  - Court decisions have expanded the scope for religion schools in the U.S.

2.1 Previous Literature

- Much of the literature is on Catholic Schools
  - Historically, about 2/3 of all private schools
  - More homogenous
- Literature is very large.

2.2 Early Work

- Simple cross tabulations or multivariate regressions of outcomes such as test scores and post secondary educational attainment typically show a substantial positive effect of Catholic school attendance.

2.2.1 Regression approach.

Estimate

$$Y_i = X_i\gamma + \alpha CH_i + \varepsilon_i$$

- Y_i is an outcome
- CH_i is a dummy variable for Catholic high school attendance,
- X_i are background, perhaps prior achievement, test scores
- ε_i captures other factors that affect test scores, dropping out, etc. that we fail to control for.
- α is the average effect of Catholic school.

- Estimate of α is biased if CH if related to ε conditional on X.

$$CH_i = 1(X_i\beta_1 + u_i > 0)$$

We can't measure all of the characteristics of the home environment and the student body that influence performance.

Likely that some of these unobserved characteristics affect Catholic school attendance.

- Strategy: **If**
  - control variables explain a lot
  - control variables cover many dimensions relevant for outcome
  - differences in mean of the control variables by CH_i modest
  - and the estimate of α is large

  then conclude that part of the Catholic school effect is real.

- Leading examples: Coleman, Hoffer, and Kilgore (1982) and Coleman and Hoffer (1987) using High School and Beyond,

- Positive effects can be found in many data sets.

2.3 Is Positive Effect Plausible ($\alpha > 0$)?

- Coleman and Hoffer (1987) and Bryk, Lee and Holland (1993): offer a coherent theory for why Catholic schools are more successful that is consistent with the pattern of results they obtain.
  - Catholic schools emphasize core academic subjects and academic achievement.
  - More "academic press" on students by maximizing school time spent in instruction, by assigning more homework, by grading more strictly, and by holding all students to high standards.
  - Catholic schools are more likely to operate as communities
    - parents and staff agree on achievement and good behavior as school goals
    - teachers are more personally connected to their students.
  - The "social capital" provided in Catholic schools can substitute for weaknesses in the home environment and in the community.
  - Proponents argue that many strengths of Catholic schools can be replicated in public school settings.

- Successful charter schools (KIP) seem to have some similar features, but even more academic press, emphasis on behavior

- The CHK book ignited a firestorm of debate and criticism that is still being played out today.

- Many prominent social scientists argued that the positive effects of Catholic school attendance in ordinary least squares (OLS) regressions are due to spurious correlations between Catholic school attendance and family characteristics that are favorable to education.

  – (e.g., Goldberger and Cain (1982)), ...

2.3.1 Argument for Positive Bias from Selection and Omitted Variables:

-
  - It costs money and time to send children to a private school. Public school is the default, low cost decision.
  - Parents who are highly motivated or who have substantial resources will be more likely to choose private schools for their children.
    * Result: Family backgrounds of the children who attend Catholic schools will differ systematically from the backgrounds of those in public schools.
  - Some of these differences will be hard to measure, such as how much parents care about their children's behavior and education.
  - Since the other students have also selected into the school, the peer group influences will also be more favorable to education.
  - Some of the selection may reflect parental preferences for religious instruction, which may or may not be correlated with background factors favorable to achievement in school and the labor market.
  - But about sixty-three percent of Catholic school parents cite academic standards and courses rather than religious instruction as the primary reason for sending their children to Catholic schools. (Hoxby 1995).Coleman and Hoffer (1987)

- Counterargument:
  - Parents may opt out of public schools if their kids are not doing well.
  - Might lead to negative selection on unobservable attributes of the child.

2.4 IV Strategies

Identification Strategies are nested by the following model:

$$
\begin{aligned}
Y_i^* &= aCH_i + X_i\gamma + \varepsilon_i && \text{(1a)} \\
Y_i &= Y_i^* \text{ or } Y_i = 1(Y_i^* > 0) \\
CH_i &= 1(X_i\beta_1 + Z_i\beta_2 + u_i > 0) && \text{(1b)}
\end{aligned}
$$

- Choice of what to exclude from X_i and include in Z_i is not obvious.

- **Student background characteristics?**

  – variables that affect choice, such as income, attitudes, and education of the parents, are likely to influence outcomes independently of the school.

- **School Characteristics?**

  – Characteristics of the private and public schools that influence choice, such as student body characteristics and school policies, are likely to be related to the effectiveness of the schools.

- **Tuition?**
  - average tuition will influence the resources available to the school and so may affect outcomes.
  - average tuition will be correlated with average family income in a school for two reasons.
    - tuition choice reflects "ability to pay" of the potential student body of the school in mind. (Not a problem if can control for family income.)
    - variation in tuition for reasons unrelated to demand for the school (such as resources in the Diocese the high school is in) will influence the average motivation of parents who send children to the school.

- **Religion (**C_i)?
  - Several IV type studies use religion of the student or the student's parent.
    - Evans and Schwab (1995), main source of ID in Neal (1997)

- **Proximity to Catholic Schools** (D_i)**?**
  - Neal (1997) uses C and proxies for distance from nearest Catholic school
    - Fraction of Catholics in the county population, and the number of Catholic schools in the county
    - Most of his results are based on bivariate probit models of Catholic high school attendance and two different outcomes (high school graduation and college graduation) in which Catholic school effects are identified by excluding whether the person is Catholic.
    - AET (2005b) show when C_i is excluded, the probit functional form is the main source of identification, not D_i

- $C_i \cdot D_i$? I started with a strong prior that $C_i \cdot D_i$ would be a valid instrument. Altonji, Elder and Taber (2005b) consider the use interaction between religion and distance.
  - Can control for C_i, D_i

3 How Do We Evaluate the Strategies?

- If one has very strong apriori info that one or more instruments are valid, perform Hausman tests of others.
  - Does not fit the "effects of Catholic schools" case.

- Researchers typically consider:

1. Face plausibility of the instruments
2. Pattern of correlation between Z_i and X_i
3. Sensitivity of Results to Additional Controls
4. Whether estimates of α make sense
5. Find subpopulation for which is $CH_i = 0$ and use to test exogeneity of Z_i

- Additional Issue: Is first stage sufficiently powerful?
  - Is identification coming from the exclusion? Or from functional form.

Let's explore in Catholic School Case, drawing on AET (2005a, b)

3.1 Data

- National Longitudinal Survey of the High School Class of 1972 (NLS72)
  - follows students who in last year of high school in 1972
- National Education Longitudinal Survey: 1988 (NELS:88)
  - follows students who were in 8th grade in 1988

Probit and Regression Estimates for Catholic eighth grade sample.

- Treat CH_i as exogenous: $cov(\varepsilon_i, v_i) = 0$. Don't need any exclusions
  - Many studies used this strategy
  - Use rich control set??. Pick homogenous sample???

TABLE 3

OLS and Probit Estimates of Catholic High School Effe

	Full Sample: 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)
				A. High School
Probit	.97	.57	.48	.41
	(.17)	(.19)	(.22)	(.21)
	[.123]	[.081]	[.068]	[.052]
Pseudo R^{2c}	.01	.16	.21	.34
				B. College
Probit	.73	.37	.33	.32
	(.08)	(.09)	(.09)	(.09)
	[.283]	[.106]	[.084]	[.074]
Pseudo R^2	.02	.19	.29	.34

CTS 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)
Graduation				
Probit	.99	.88	.95	1.27
	(.24)	(.25)	(.27)	(.29)
	[.105]	[.084]	[.081]	[.088]
Pseudo R^{2c}	.11	.35	.44	.58
in 1994				
Probit	.60	.48	.56	.60
	(.13)	(.15)	(.15)	(.15)
	[.236]	[.154]	[.154]	[.149]
Pseudo R^2	.04	.18	.29	.36

- Find a strong positive effect of CH on high school graduation (.07 or .08) and college attendance (.14)

3.1.1 Results for Test Scores

- 12th grade math: .13 standard deviations

- 12th grade reading: near 0.

- Studies of Voucher programs in elementary school in US show mixed evidence on test score gains)

- Studies of Privately funded vouchers based on field experiments show 0 or small gains over all. Some evidence for a positive effect on African Americans, but result not robust.

(See Rouse and Barrow (2009) for a review and references)

3.1.2 Results for Urban Minorities:

- Urban Minorities: Effects are stronger for HS grad, less precise: HS Grad .191. College .144

TABLE 7

OLS, Fixed-Effect, and Probit Estimates of Catholic High School Effects by Race and Urban Residence: Full Set of Controls

	Sample			
	Urban and Suburban White Only (1)	Urban and Suburban Minorities Only (2)	Urban White Only (3)	Urban Minorities Only (4)
	A. High School Graduate			
Sample mean	.88	.80	.88	.80
	(N=3,799)	(N=1,308)	(N=1,002)	(N=697)
Probit	.443	.524	1.176	1.592
	(.279)	(.338)	(.417)	(.673)
	[.046]	[.085]	[.091]	[.191]
	B. College in 1994			
Sample mean	.37	.26	.32	.26
	(N=3,695)	(N=1,258)	(N=981)	(N=666)
Probit	.354	.697	.506	.677
	(.107)	(.201)	(.167)	(.303)
	[.087]	[.158]	[.110]	[.144]

Key question: How much of the estimated high school effect on educational attainment is real, and how much is due to selection bias?

3.2 Are Instruments Needed? Can CH be Treated as Exogenous?

3.2.1 Face Plausibility?

- Highly Questionable.

3.2.2 Pattern of Correlation between CH and X

AET (2005a), Table 1.

TABLE 1

Means of Key Variables in High School and Eighth Grade Sector

	Full Sample			Catholic 8th Grade		
Variable	Public 10th (N=11,167)	Catholic 10th (N=672)	Difference	Public 10th (N=366)	Catholic 10th (N=640)	Difference
Demographics:						
Female	.52	.45	−.07**	.61	.50	−.11**
Asian	.03	.04	.01	.05	.05	.00
Hispanic	.09	.09	.00	.08	.09	.01
Black	.10	.09	−.01	.07	.11	.04
White	.78	.78	.00	.80	.74	−.06
Family background:						
Mother's education (years)	13.21	13.96	.75***	13.34	13.88	.54***
Father's education (years)	13.49	14.51	1.01***	13.39	14.38	.99***
Log of family income	10.23	10.72	.49***	10.47	10.66	.19***
Mother only in house	.14	.09	−.05***	.07	.09	.02***
Parent married	.79	.89	.10***	.90	.88	−.02
Parents Catholic	.28	.82	.54***	.84	.84	.00
Geography:						
Rural	.36	.03	−.33***	.13	.01	−.12**
Suburban	.45	.51	.06*	.40	.48	.08
Urban	.19	.46	.27***	.47	.51	.04
Distance to closest Catholic high school (miles)	22.16	2.97	−19.19***	6.91	2.37	−4.53***

TABLE 1

MEANS OF KEY VARIABLES IN HIGH SCHOOL AND EIGHTH GRADE SECTOR

VARIABLE	Full Sample: Public 10th (N=11,167)	Full Sample: Catholic 10th (N=672)	Full Sample: Difference	Catholic 8th Grade: Public 10th (N=366)	Catholic 8th Grade: Catholic 10th (N=640)	Catholic 8th Grade: Difference
Expectations:[a]						
Schooling expectations (years)	15.25	15.97	.72***	15.52	15.92	.40***
Very sure to graduate from high school	.84	.89	.05***	.84	.90	.06*
Parents expect at least some college	.89	.98	.09***	.94	.98	.04
Parents expect at least college graduation	.79	.92	.13***	.88	.91	.03
Student expects white-collar job	.47	.61	.14***	.55	.59	.04
8th grade variables:						
Delinquency index (range 0–4)	.64	.53	−.11*	.54	.46	−.08
Student got into fight	.24	.23	−.02	.20	.19	−.01
Student rarely completes homework	.19	.08	−.11***	.08	.06	−.01
Student frequently disruptive	.12	.08	−.05***	.08	.08	.00
Student repeated grade 4–8	.06	.02	−.05***	.03	.02	−.01
Risk index (range 0–4)	.69	.35	−.34***	.39	.39	.00
Grades composite	2.94	3.16	.22***	3.09	3.20	.11**
Unpreparedness index (0–25)	10.77	11.08	.31***	10.84	11.02	.17
8th grade reading score	51.19	55.05	3.86***	54.12	55.59	1.47
8th grade mathematics score	51.13	54.57	3.44***	52.89	53.98	1.09
Outcomes:						
12th grade reading standardized score	51.20	54.60	3.40***	53.25	54.70	1.45
12th grade math standardized score	51.20	55.54	4.34***	53.13	55.63	2.49***
Enrolled in four-year college in 1994	.31	.59	.28***	.38	.61	.23***
High school graduate	.85	.98	.13***	.88	.98	.10***

- Huge difference in HS grad rates (.104), college attendance (.23)

- smaller differences in 12th grade test scores

- 8th grade outcomes, family background more favorable for kids in Catholic high schools.

- Difference in observables is much smaller for Catholic 8th grade sample.

- **Those who attend CH are clearly advantaged.**

- Difference is smaller among those who attend Catholic 8th grade.

- In the NLS72, the differences are smaller.

- Will return to the CH exogenous case later.

3.3 Is Religion a Valid Instrument?

3.3.1 Face Plausibility:

- Not much discussion of why Catholic religion per se would influence educational attainment and wages in contemporary U.S..

- Evan and Schwab, others have pointed out that Catholics aren't that far from the median on most socieconomic meausures. Look at Difference in Means of Covariates by C

	(1)	(2)	(3)
	Overall Mean	Difference by C_i	Difference by D_i
Demographics			
Female	0.50	0.01	0.00
Asian	0.04	0.01	0.04
Hispanic	0.10	0.19	0.08
Black	0.13	–0.15	0.08
White	0.73	–0.05	–0.20
Family background			
Mother's education	13.14	–0.26	0.17
Father's education	13.42	–0.07	0.17
Log of family income	10.20	0.11	0.12
Mother only in house	0.15	–0.04	0.02
Parent married	0.78	0.06	–0.02
Geography			
Rural	0.32	–0.15	–0.44
Suburban	0.44	0.06	0.08
Urban	0.24	0.09	0.36
Expectations			
Schooling	15.17	0.15	0.31
Very sure to graduate high school	0.83	–0.01	0.00
Parents: some college	0.88	0.04	0.05
Parents: college graduates	0.88	0.03	0.06
White collar job	0.46	0.03	0.06

Demographics			
Female	0.50	0.01	0.00
Asian	0.04	0.01	0.04
Hispanic	0.10	0.19	0.08
Black	0.13	−0.15	0.08
White	0.73	−0.05	−0.20
Family background			
Mother's education	13.14	−0.26	0.17
Father's education	13.42	−0.07	0.17
Log of family income	10.20	0.11	0.12
Mother only in house	0.15	−0.04	0.02
Parent married	0.78	0.06	−0.02
Geography			
Rural	0.32	−0.15	−0.44
Suburban	0.44	0.06	0.08
Urban	0.24	0.09	0.36
Expectations			
Schooling	15.17	0.15	0.31
Very sure to graduate high school	0.83	−0.01	0.00
Parents: some college	0.88	0.04	0.05
Parents: college graduates	0.88	0.03	0.06
White collar job	0.46	0.03	0.06

- **Conclusion: Catholics look advantaged.**

- Less so in earlier cohorts

Hard to judge whether the links are large or small—this is what AET address

3.4 Do the Estimates Using $Catholic$ as the Excluded Instrument Make Sense?

	Excluded Instruments		
	(1) Catholic (C_i)	(2) Distance (D_i)	(3) $C_i \times D_i$
High school graduation (NELS:88)			
Probit (controls exclude "instruments")	0.065	0.047	0.052
	(0.025)	(0.025)	(0.026)
Bivariate probit	0.128	−0.007	−0.022
	(0.032)	(0.085)	(0.119)
OLS	0.041	0.021	0.023
	(0.014)	(0.014)	(0.015)
2SLS	0.34	−0.04	0.09
	(0.08)	(0.10)	(0.11)
College in 1994(NELS:88)			
Probit (controls exclude "instruments")	0.094	0.085	0.077
	(0.022)	(0.022)	(0.022)
Bivariate probit	0.170	0.103	−0.043
	(0.055)	(0.062)	(0.070)
OLS	0.128	0.119	0.111
	(0.026)	(0.026)	(0.026)
2SLS	0.40	0.31	−0.11
	(0.10)	(0.11)	(0.12)

2SLS estimate of .34 is unreasonably large given that the mean of the graduation rate is .84.

Bivariate probit estimate is .128—double the univariate probit estimate.

2SLS estimate of.40 for College attendance is larger than the mean of is .29. *This is unreasonable.*

- AET reject Catholic as instrument based on above evidence rather than argue that the large 2SLS and bivariate probit estimates bolster evidence for a positive effect based on OLS and probit.

3.5 Evaluating C_i by Examining Effect of Catholic Religion for Students from Public Eighth Grades

- For persons for whom Catholic high school is not an option, the coefficient on C_i in a single equation model for Y_i is an estimate of the *direct effect* of Catholic religion on the outcome.

- Only 0.3% of public school eighth graders go to Catholic HS. Only 0.7% among public eighth grade attendees whose parents are Catholic.

$$Y_i = \alpha CH_i + X_i'\gamma + \varepsilon_i, \tag{2}$$

where γ is defined so that cov$(\varepsilon_i, X_i) = 0$.

- CH_i is potentially endogenous and thus correlated with ε_i.

- Assume C_i does not influence Y_i directly but is correlated with CH_i.

- Is C_i is correlated with ε_i?

- Define β, π, and λ to be the coefficients of the least squares projections

$$\text{Proj}\,(C_i \mid X_i) = X_i'\pi, \tag{3}$$
$$\text{Proj}\,(CH_i \mid X_i, C_i) = X_i'\beta + \lambda C_i. \tag{4}$$

- Let $\widetilde{C}_i$ be the residual of the projection of C_i on X_i :

$$\widetilde{C}_i \equiv C_i - X_i' \pi. \tag{5}$$

- $\hat{\alpha}_{IV}$ converges to

$$\widehat{\alpha}_{IV} \xrightarrow{p} \alpha + \frac{1}{\lambda} \frac{cov\left(\widetilde{C}_i, \varepsilon_i\right)}{var(\widetilde{C}_i)}. \tag{6}$$

- Let p_i be an event such that

$$\Pr(CH_i = 1 \mid p_i) = 0.$$

- In our application p_i is attendance of a public eighth grade by individual i.

- Assume for now that $(X_i, C_i, \varepsilon_i) \perp p_i$

- Consider regression Y_i on X_i and C_i conditional on p_i. Coef. converges to

$$\frac{cov\left(\widetilde{C}_i, \varepsilon_i\right)}{var(\widetilde{C}_i)}$$

- Have a consistent estimate of λ from the first stage regression, so can obtain a consistent estimate of the bias

$$\hat{\psi} = \frac{1}{\lambda}\frac{cov\left(\widetilde{C}_i, \varepsilon_i\right)}{var(\widetilde{C}_i)}$$

by using 8th grade public school sample to estimate

$$Y_i = X_i'\Lambda + [C_i\widehat{\lambda}]\psi + \omega_i \tag{7}$$

Bias estimate $\hat{\psi}$ is huge when C is used as an instrument.

Enough to rule out use of C.

Table 4

Comparison of 2SLS Estimates[a] and Bias Implied by OLS Estimation of $Y_i = X_i'\gamma + [Z_i' \hat{\lambda}]\psi + \omega_i$, on the Public Eighth Grade Subsample.[b] Various Outcomes and instruments. NELS:88 Sample Weighted, (Huber-White Standard Errors in Parentheses)

Outcome (Y)	Instruments (Z_i) (1) Catholic	(2) Distance	(3) Catholic × D
High school graduation			
Implied bias in 2SLS (ψ)[c]	0.34 (0.08)	–0.05 (0.12)	0.15 (0.
2SLS coefficient	0.34 (0.08)	–0.04 (0.10)	0.09 (0.
College attendance			
Implied bias in 2SLS (ψ)	0.29 (0.11)	0.37 (0.12)	–0.23 (0.
2SLS coefficient	0.40 (0.10)	0.31 (0.11)	–0.11 (0.
Twelfth Grade Reading Score			
Implied bias in 2SLS (ψ)	0.54 (1.68)	–0.51 (2.08)	–0.50 (1.
2SLS Coefficient	1.40 (1.54)	–1.09 (1.84)	1.24 (1.
Twelfth grade math score			
Implied bias in 2SLS (ψ)	1.85 (1.41)	1.83 (1.69)	–4.37 (2.
2SLS Coefficient	2.64 (1.41)	2.43 (1.45)	–2.63 (1.

- Paper argues that conclusion is robust to accounting for selection induced by restricting analysis to those who chose Catholic 8th grades.

3.6 VI. Are Exclusion Restrictions or Nonlinearity the Main Source of Identification? Bivariate Probit Versus 2SLS

- The specification used in Neal (1997), Evans and Schwab (1995), and AET (2005a, b) is

$$\begin{aligned} CH_i &= 1(X_i'\beta + Z_i'\lambda + u_i > 0) \\ Y_i &= 1(\alpha CH_i + X_i'\gamma + \varepsilon_i > 0), \end{aligned}$$

(u_i, ε_i) are standard normals with an unknown correlation.

- Many studies work with a similar model, perhaps with continuous Y.

- Exclusion restrictions are useful for semiparametric identification in limited dependent variable models (see, e.g., Heckman, 1990 or Taber, 2000).

- However, the linearity and normality assumptions of the model are technically sufficient. An exclusion restriction is not necessary.

- When one imposes both exclusion restrictions and functional form restrictions, both contribute to identification of α.

- Should explore whether identification is primarily coming from the exclusion restrictions or primarily coming from the functional form restrictions

- Our analysis was motivated by large differences in some cases between 2SLS and bivariate probit estimates.

- Follow Neal (1997) focussed on people from urban areas with separate effects for minorities and whites.

To isolate role of nonlinearity, consider two stage probit.

Usual First stage:

$$\Pr(CH_i = 1 \mid X_i, Z_i) = \Phi(X_i'\beta + Z_i'\lambda)$$

Second Stage:

$$\Pr(COLLEGE_i = 1 \mid X_i, Z_i) = \Phi\left[X_i'\gamma + \alpha\Phi(X_i'\widehat{\beta} + Z_i'\widehat{\lambda})\right]$$

- This yields point estimates and standard errors for α that are similar to bivariate probit

$\Phi(X_i'\widehat{\beta} + Z_i'\widehat{\lambda})$

- Instead of including $\Phi(X_i'\widehat{\beta} + Z_i'\widehat{\lambda})$ as the key variable in the second stage, include separate predicted probabilities holding X_i and Z_i constant at their sample means, respectively.

$$\Pr(COLL_i = 1 \mid X_i, Z_i, \overline{X}_i, \overline{Z}_i) = \Phi\left[X_i'\gamma + \alpha_1\Phi(\overline{X}_i'\widehat{\beta} + Z_i'\widehat{\lambda}) + \alpha_2\Phi(X_i'\widehat{\beta} + \overline{Z}_i'\widehat{\lambda})\right] \tag{8}$$

- Idea: isolate the effects of variation in Z_i on $\widehat{CH}_i$, given by $\alpha_1\Phi(\overline{X}_i'\widehat{\beta} + Z_i'\widehat{\lambda})$, from the effects of variation in X_i, which is given by $\alpha_2\Phi(X_i'\widehat{\beta} + \overline{Z}_i'\widehat{\lambda})$.

- $\hat{\alpha}_1$ measures the extent to which variation in the excluded instruments are influencing college attendance, rather than just nonlinearity in X_i in the function $\Phi(X_i'\widehat{\beta} + \overline{Z}_i'\widehat{\lambda})$.

- (Note: This is only an informal exercise to explore the extent of the identifying power of Z_i. $\hat{\alpha}_1$ is not a consistent estimator of α in most circumstances.)

Column 1, 4: α based on $\Pr(COLLEGE_i = 1 \mid X_i, Z_i) = \Phi\left[X_i'\gamma + \alpha\Phi(X_i'\widehat{\beta} + Z_i'\widehat{\lambda})\right]$

Column 3, 6: α_1 based on $\Pr(COLL_i = 1 \mid X_i, Z_i, \overline{X}_i, \overline{Z}_i)$

$$= \Phi\left[X_i'\gamma + \alpha_1\Phi(\overline{X}_i'\widehat{\beta} + Z_i'\widehat{\lambda}) + \alpha_2\Phi(X_i'\widehat{\beta} + \overline{Z}_i'\widehat{\lambda})\right]$$

	Sample					
	Nonwhites in cities (N = 1,532)			Whites in Cities (N = 5,326)		
	Nonlinear Models (Probit) (1)	Linear Models (OLS/2SLS) (2)	Nonlinear Models Holding X_i Constant[4] (3)	Nonlinear Models (Probit) (4)	Linear Models (OLS/2SLS) (5)	Nonlinear Models Holding X_i Constant[4] (6)
Single equation (OLS/Probit)	0.640	0.239		0.253	0.093	
	(0.198)	(0.070)		(0.062)	(0.022)	
	[0.239]			[0.093]		
Two equation models						
Excluded Instruments						
$\%CCH_i$ and CH/M_i	1.471	1.375	5.541	0.048	0.115	0.084
	(0.442)	(0.583)	(2.082)	(0.250)	(0.158)	(0.783)
	[0.517]		[0.706]	[0.018]		[0.031]
C_i and $\%CCH_i$	0.879	0.054	0.012	–0.090	–0.036	–0.084
	(0.523)	(0.309)	(1.443)	(0.121)	(0.050)	(0.148)
	[0.329]		[0.004]	[–0.033]		[–0.031]
C_i, $\%CCH_i$ and CH/M_i	1.106	0.331	1.302	–0.085	–0.034	–0.069
	(0.460)	(0.254)	(0.706)	(0.118)	(0.048)	(0.125)
	[0.409]		[0.471]	[–0.031]		[–0.025]

	Sample					
	Nonwhites in cities (N = 1,532)			Whites in Cities (N = 5,326)		
	Nonlinear Models (Probit) (1)	Linear Models (OLS/2SLS) (2)	Nonlinear Models Holding X_i Constant[4] (3)	Nonlinear Models (Probit) (4)	Linear Models (OLS/2SLS) (5)	Nonlinear Models Holding X_i Constant[4] (6)
C_i only	0.761	–0.093	–0.505	–0.133	–0.056	–0.149
	(0.543)	(0.324)	(1.638)	(0.130)	(0.054)	(0.151)
	[0.285]		[–0.148]	[–0.049]		[–0.054]
$C_i \times D_i$	1.333	2.572	1.409	–0.121	–0.395	2.624
	(0.516)	(2.442)	(1.276)	(0.262)	(0.169)	(5.173)
	[0.478]		[0.497]	[–0.044]		[0.559]
None	1.224			–0.094		
	(0.542)			(0.301)		
	[0.446]			[–0.034]		

- Consider one extreme case, one in which C_i, $\%CCH_i$, and CH/P_i are all used as excluded instruments for urban whites.

- Estimate of α_1 of -0.069 (0.125) is similar to the corresponding 2 stage probit coefficient of -0.085 (0.118) in terms of both magnitude and precision. No evidence of a problem.

- Consider second extreme case in which only $\%CCH_i$, and CH/P_i are used to identify the model for urban minorities.

- $\hat{\alpha}_1$ is 5.541 (2.082). Very different from 2 stage probit estimate (se) of 1.471 (0.442) .

- Variation in the instruments contributes substantially to identification in the first case, but not in the second.

- Estimates for urban minorities in columns (1) and (3) indicate that in every case the point estimates and standard errors differ dramatically.

- This implies that no combination of instruments drives all (or nearly all) of the identification of these models for urban minorities. Even C_i doesn't play much of a role.

- For urban whites, the exclusion restrictions show substantially more power, but only when Catholic religion (C_i) is used as an instrument - the models using $\%CCH_i$ and CH/P_i or $C_i \times D_i$ still exhibit large discrepancies between columns (4) and (6).
  - The implication is that Catholic religion drives identification in models for urban whites, but none of the other candidate instruments are effective for this sample
  - No combination of instruments appears to be powerful for urban minorities.

- AET (2005b) argue that the analysis explains some of the discrepancies among the results in papers on the Catholic school effect that have used these various exclusion restrictions.

- See also Grogger and Neal (2000), who conclude that county level instruments for Catholic school availability are not very informative in NELS:88.

3.6.1 Conclusion About Role of Nonlinearity and Lesson for Applied Researchers

- Bivariate probit can *sometimes* produce misleading results which are consistent with a powerful instrumental variable, when in fact identification is stemming from a weak instrument in combination with functional form assumptions.

- To isolate the role of each of these factors, one should experiment with specifications that rely solely on exclusion restrictions for identification.

4 A Few Words About Regression Discontinuity Design

$$\begin{aligned} Y_i^* &= aCH_i + X_i\gamma_1 + f(Z_i) + \varepsilon_i \qquad (9a) \\ Y_i &= Y_I^* \text{ or } Y_i = 1(Y_I^* > 0) \end{aligned}$$

$$CH_i = 1(Z_i > \bar{Z}) \qquad \text{strict discontinuity}$$

$$CH_i = 1(X_i\beta_1 + 1(Z_i > \bar{Z})\beta_2 + u_i > 0) \quad \text{fuzzy discontinuity}$$

- $f(Z_i)$ is a smooth function in neighborhood of $\bar{Z}$. Z is called the running variable
- References: Theory: Hahn, J., P. Todd, W. Van der Klaauw, (Econometrica Vol 69), D. Lee and T. Lemieux, (JEL 2010)

- In school choice context, Z is a test score or distance from an attendance boundary.
  - Campbell and Thistelwaite (1960) introduced the idea. Studied the effect of merit awards on future academic outcomes. Allocation of the awards was based a test score.
  - Example: Bayer, Black (QJE, 1999), Bayer, Ferreira, Robert McMillan (JPE, 2007) (distance from attendance boundary)
- Other examples:
  - In social transfer study, Z might be earnings and $\bar{Z}$ the earnings cutoff for eligibility (Lemieux and Milligan)
  - In medical study, Z might be some indicator of the condition of the patient and $\bar{Z}$ the cutoff for eligibility to a treatment. Example: (Almond, Doyle, Kowalski, and Williams (QJE 2010) study of value of intensive care for low birth weight babies. Z is birth weight and $\bar{Z}$ =1.5 Kilo.
- With strict discontinuity, just estimate the Y_i equation.

- With fuzzy discontinuity, one works with the equations for Y and CH. The excluded variable is $1(Z_i > \bar{Z})$. If Y is continous, just 2SLS.

- The assumption is the $1(Z_i > \bar{Z})$ is exogenous conditional on X and $f(Z)$.

4.1 Using the Observables to Check on $1(Z_i > \bar{Z})$

- Researchers graph X_i and $X_i\gamma_i$ again Z_i and look for a break around $1(Z_i > \bar{Z})$

- Worry if:

1. Find systematic relationship between how the X_i are related to Y_i and how they are related to Z_i and

2. There is a break in the relationship between X_i around Z_i around $\bar{Z}$. (Figure 17 from David Lee's (2008) study of on incumbent advantage shows no break in Democratic vote share in last election at $\bar{Z} = 0$.)

3. There is a discontinuity in the distribution of Z_i around $\bar{Z}$. (Perfect ability to manipulate Z)

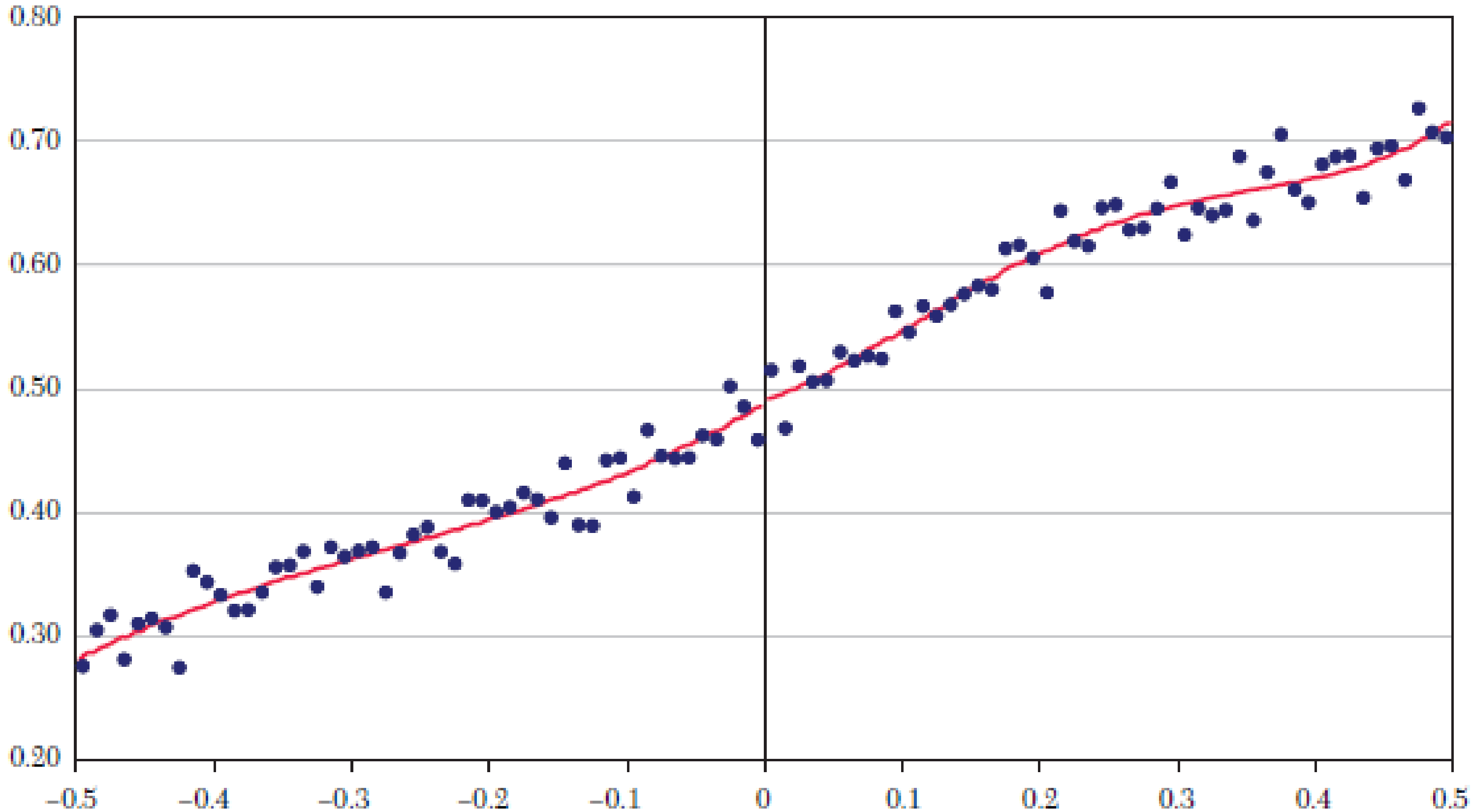

Figure 17. Discontinuity in Baseline Covariate (Share of Vote in Prior Election)

Example: Lee 2008, study of effect of incumbency on vote share in next election (Z_t). Graph shows no break in distribution of prior Democratic vote share (Z_{t-2}) on distribution of running variable Z_{t-1} at $Z_{t-1} = \bar{Z} = 0$

5 Next: Foundation for Using Observed Variables to Assess Selection on Unobserved Variables