http://www.econ.ucsd.edu/~elib/250A/

Causal Inference with and without Experiments I

1. The Big Picture: Data, Statistics, Economic Theory and Applied Econometrics
2. Causal Inference: The SAT Prep example
3. What can we learn from these data?
4. Causal inference with experiments: NSW evaluation
5. Causal inf. without experiments: Adding omitted variables
6. Causal inference without experiments: Heckman 2 step

1. The Big Picture

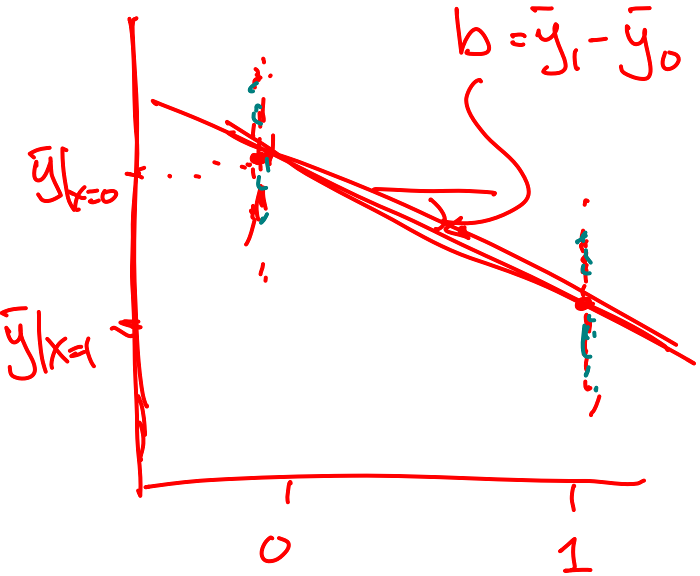
Data

Economic
Theory

Statistics
& Probability

2. Causal Inference: The SAT Prep course example

Nonexperimental

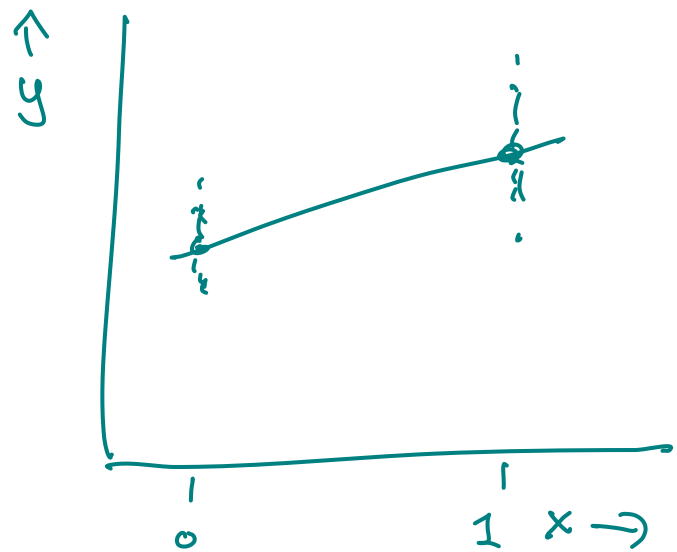


y - SAT score

$X = \begin{cases} 0 & \text{NOT} \\ 1 & \text{SAT PREP COURSE} \end{cases}$

b could be a useful predictor.

Experimental



$X_i = \begin{cases} 0 & \text{NOT} \\ 1 & \text{SAT PREP COURSE} \end{cases}$

RANDOMLY ASSIGNED. $\Rightarrow Cov(x_i, a_i) = 0$

Linear Causal Effect

$\beta^c = V(x)^{-1} Cov(x, y)$ in Population where X R.A.

- $Y = X\beta^c + \varepsilon$, $\varepsilon = Y - X\beta^c$, $Cov(X, \varepsilon) \equiv 0 \Leftarrow$ X R.A.,
  $\Leftarrow$ (OLS $\Rightarrow X \perp \varepsilon$)

Distinguish from

$\beta = V(x)^{-1} Cov(x, y)$ when X not R.A. (or maybe it is)

$Y = X\beta + v$, $Cov(v, X) = 0 \Leftarrow$ (OLS)

Short & Long Regressions - SAT Prep Example, X_i not R.A.

(L) $Y_i = X_{1i} b_1^L + a_i + u_i$; $b_1^L \xrightarrow{P} \beta^c$ (assumed)

(S) $Y_i = X_{1i} b_1^S + e_i^S$

$b_1^S = b_1^L + b_{21} \overset{1}{\overset{\shortparallel}{b_2^L}}$ $\quad b_{21} = (x_1' x_1)^{-1} x_1' x_2$, $(x_2 = a)$

3. What can we learn from these data?

- Best linear predictor
  $V(x)^{-1}Cov(x,y)$ in population
- Linear Causal effect
  $V(x)^{-1}Cov(x,y)$ in population with RA of x
- Why call it "causal"?
  - "OVB" justification: imagine listing the omitted variables in residual
  - Full derivative vs. partial derivative analogy
- Selection –Cov(x,a)
  [appeals to theory for interpretation]
- There is no "wrong" estimate, just different applications. Each estimate has a customer, in this case the covariance term has one as well.
- Why linear predictors?

4. Causal Inference with experimental data

- The gold standard of inference because “OVB”=0 (e.g., Cov(x,a) = 0) by construction with random assignment.
- Other design issues:
  - proper randomization of assignment
  - representativeness (random sampling among the appropriate population)
- Problems: Cost, Ethical Issues
- The Great Society, Patrick Moynihan and Experimental Evaluations

Causal Inference with Experimental Data: Selection into Treatment

- Selection by Individuals – e.g., charter schools, GRE prep courses
- Selection by Institutions – e.g., job training programs, military draft

- Selection by Individuals and Institutions:
  - e.g. The "Ashenfelter Dip"

Causal Inf. with experiments: Lalonde (86)

Lalonde checks randomization

TABLE 1—THE SAMPLE MEANS AND STANDARD DEVIATIONS OF PRE-TRAINING EARNINGS AND OTHER CHARACTERISTICS FOR THE NSW AFDC AND MALE PARTICIPANTS

	Full National Supported Work Sample			
	AFDC Participants		Male Participants	
Variable	Treatments	Controls	Treatments	Controls
Age	33.37	33.63	24.49	23.99
	(7.43)	(7.18)	(6.58)	(6.54)
Years of School	10.30	10.27	10.17	10.17
	(1.92)	(2.00)	(1.75)	(1.76)
Proportion	.70	.69	.79	.80
High School Dropouts	(.46)	(.46)	(.41)	(.40)
Proportion Married	.02	.04	.14	.13
	(.15)	(.20)	(.35)	(.35)
Proportion Black	.84	.82	.76	.75
	(.37)	(.39)	(.43)	(.43)
Proportion Hispanic	.12	.13	.12	.14
	(.32)	(.33)	(.33)	(.35)
Real Earnings	$393	$395	1472	1558
1 year Before	(1,203)	(1,149)	(2656)	(2961)
Training	[43]	[41]	[58]	[63]
Real Earnings	$854	$894	2860	3030
2 years Before	(2,087)	(2,240)	(4729)	(5293)
Training	[74]	[79]	[104]	[113]
Hours Worked	90	92	278	274
1 year Before	(251)	(253)	(466)	(458)
Training	[9]	[9]	[10]	[10]
Hours Worked	186	188	458	469
2 years Before	(434)	(450)	(654)	(689)
Training	[15]	[16]	[14]	[15]
Month of Assignment	−12.26	−12.30	−16.08	−15.91
(Jan. 78 = 0)	(4.30)	(4.23)	(5.97)	(5.89)
Number of Observations	800	802	2083	2193

Note: The numbers shown in parentheses are the standard deviations and those in the square brackets are the standard errors.

Causal Inf. w/ expmnts: Lalonde (86)

TABLE 2—ANNUAL EARNINGS OF NSW TREATMENTS, CONTROLS, AND EIGHT CANDIDATE COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*

Year	Treat-ments	Controls	Comparison Group[a,b] *PSID*-1	*PSID*-2	*PSID*-3	*PSID*-4	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3	*CPS-SSA*-4
1975	$895	$877	7,303	2,327	937	6,654	7,788	3,748	4,575	2,049
	(81)	(90)	(317)	(286)	(189)	(428)	(63)	(250)	(135)	(333)
1976	$1,794	$646	7,442	2,697	665	6,770	8,547	4,774	3,800	2,036
	(99)	(63)	(327)	(317)	(157)	(463)	(65)	(302)	(128)	(337)
1977	$6,143	$1,518	7,983	3,219	891	7,213	8,562	4,851	5,277	2,844
	(140)	(112)	(335)	(376)	(229)	(484)	(68)	(317)	(153)	(450)
1978	$4,526	$2,885	8,146	3,636	1,631	7,564	8,518	5,343	5,665	3,700
	(270)	(244)	(339)	(421)	(381)	(480)	(72)	(365)	(166)	(593)
1979	$4,670	$3,819	8,016	3,569	1,602	7,482	8,023	5,343	5,782	3,733
	(226)	(208)	(334)	(381)	(334)	(462)	(73)	(371)	(170)	(543)
Number of Observations	600	585	595	173	118	255	11,132	241	1,594	87

[a] The Comparison Groups are defined as follows: *PSID*-1: All female household heads continuously from 1975 through 1979, who were between 20 and 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all women who received AFDC in 1975; *PSID*-3: Selects from the *PSID*-2 all women who were not working when surveyed in 1976; *PSID*-4: Selects from the *PSID*-1 group all women with children, none of whom are less than 5-years-old; *CPS-SSA* − 1: All females from Westat *CPS-SSA* sample; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all females who received AFDC in 1975; *CPS-SSA*-3: Selects from *CPS-SSA*-1 all females who were not working in the spring of 1976; *CPS-SSA*-4: Selects from *CPS-SSA*-2 all females who were not working in the spring of 1976.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. For the NSW treatments and controls, the number of observations refer only to 1975 and 1979. In the other years there are fewer observations, especially in 1978. At the time of the resurvey in 1979, treatments had been out of Supported Work for an average of 20 months.

Population: AFDC women, ex-addicts, ex-cons; Treatment: 9-18 mo. guaranteed work beginning in 1976.

Note the “Ashenfelter Dip” (mean reversion in income) in “means-based” programs.

5. Causal Inference Without Experiments

- One Approach: Include the omitted variables in the hope of reducing OVB
  - imagine measuring “a”, ability in the SAT example
  - perhaps there’s a proxy for “a”
- Formally: assume (hopefully) that Cov $(x_1, \varepsilon \mid \beta_2' x_2) = 0$
  - e.g., finding some more x’s, fixed effects, differences in differences, selection correction
- “Perfectly specified equation including all relevant variables”

- Another approach: Matching (or Propensity Score) estimator: assumes (hopefully) that Cov $(x_1, \varepsilon \mid x_2) = 0$
  note: without assuming a linear function $\beta_2' x_2$ for influence of “omitted” variables
  note: If values of x_2 in treatment and comparison observations were identical across paired observations this assumption is sufficient. This is generally impractical.
- Problem: See Tables 2 & 4 in Lalonde

Causal Inf. w/out expmnts: Lalonde (86)

TABLE 2—ANNUAL EARNINGS OF NSW TREATMENTS, CONTROLS, AND EIGHT CANDIDATE COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*

			Comparison Group[a,b]							
Year	Treatments	Controls	*PSID*-1	*PSID*-2	*PSID*-3	*PSID*-4	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3	*CPS-SSA*-4
1975	$895	$877	7,303	2,327	937	6,654	7,788	3,748	4,575	2,049
	(81)	(90)	(317)	(286)	(189)	(428)	(63)	(250)	(135)	(333)
1976	$1,794	$646	7,442	2,697	665	6,770	8,547	4,774	3,800	2,036
	(99)	(63)	(327)	(317)	(157)	(463)	(65)	(302)	(128)	(337)
1977	$6,143	$1,518	7,983	3,219	891	7,213	8,562	4,851	5,277	2,844
	(140)	(112)	(335)	(376)	(229)	(484)	(68)	(317)	(153)	(450)
1978	$4,526	$2,885	8,146	3,636	1,631	7,564	8,518	5,343	5,665	3,700
	(270)	(244)	(339)	(421)	(381)	(480)	(72)	(365)	(166)	(593)
1979	$4,670	$3,819	8,016	3,569	1,602	7,482	8,023	5,343	5,782	3,733
	(226)	(208)	(334)	(381)	(334)	(462)	(73)	(371)	(170)	(543)
Number of Observations	600	585	595	173	118	255	11,132	241	1,594	87

[a] The Comparison Groups are defined as follows: *PSID*-1: All female household heads continuously from 1975 through 1979, who were between 20 and 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all women who received AFDC in 1975; *PSID*-3: Selects from the *PSID*-2 all women who were not working when surveyed in 1976; *PSID*-4: Selects from the *PSID*-1 group all women with children, none of whom are less than 5-years-old; *CPS-SSA* − 1: All females from Westat *CPS-SSA* sample; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all females who received AFDC in 1975; *CPS-SSA*-3: Selects from *CPS-SSA*-1 all females who were not working in the spring of 1976; *CPS-SSA*-4: Selects from *CPS-SSA*-2 all females who were not working in the spring of 1976.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. For the NSW treatments and controls, the number of observations refer only to 1975 and 1979. In the other years there are fewer observations, especially in 1978. At the time of the resurvey in 1979, treatments had been out of Supported Work for an average of 20 months.

Causal Inf. w/out expmnts: Lalonde '86

TABLE 4—EARNINGS COMPARISONS AND ESTIMATED TRAINING EFFECTS FOR THE NSW AFDC PARTICIPANTS USING COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*[a,b]

		NSW Treatment Earnings Less Comparison Group Earnings				Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons		Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79		Controlling for All Observed Variables and Pre-Training Earnings	
		Pre-Training Year, 1975		Post-Training Year, 1979							
Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–79 (1)	Unadjusted (2)	Adjusted[c] (3)	Unadjusted (4)	Adjusted[c] (5)	Without Age (6)	With Age (7)	Unadjusted (8)	Adjusted[c] (9)	Without AFDC (10)	With AFDC (11)
Controls	2,942	−17	−22	851	861	833	883	843	864	854	–
	(220)	(122)	(122)	(307)	(306)	(323)	(323)	(308)	(306)	(312)	
PSID-1	713	−6,443	−4,882	−3,357	−2,143	3,097	2,657	1746	1,354	1664	2,097
	(210)	(326)	(336)	(403)	(425)	(317)	(333)	(357)	(380)	(409)	(491)
PSID-2	1,242	−1,467	−1,515	1,090	870	2,568	2,392	1,764	1,535	1,826	–
	(314)	(216)	(224)	(468)	(484)	(473)	(481)	(472)	(487)	(537)	
PSID-3	665	−77	−100	3,057	2,915	3,145	3,020	3,070	2,930	2,919	–
	(351)	(202)	(208)	(532)	(543)	(557)	(563)	(531)	(543)	(592)	
PSID-4	928	−5,694	−4,976	−2,822	−2,268	2,883	2,655	1,184	950	1,406	2,146
	(311)	(306)	(323)	(460)	(491)	(417)	(434)	(483)	(503)	(542)	(652)
CPS-SSA-1	233	−6,928	−5,813	−3,363	−2,650	3,578	3,501	1,214	1,127	536	1,041
	(64)	(272)	(309)	(320)	(365)	(280)	(282)	(272)	(309)	(349)	(503)
CPS-SSA-2	1,595	−2,888	−2,332	−683	−240	2,215	2,068	447	620	665	–
	(360)	(204)	(256)	(428)	(536)	(438)	(446)	(468)	(554)	(651)	
CPS-SSA-3	1,207	−3,715	−3,150	−1,122	−812	2,603	2,615	814	784	−99	1,246
	(166)	(226)	(325)	(311)	(452)	(307)	(328)	(305)	(429)	(481)	(720)
CPS-SSA-4	1,684	−1,189	−780	926	756	2,126	1,833	1,222	952	827	–
	(524)	(249)	(283)	(630)	(716)	(654)	(663)	(637)	(717)	(814)	

[a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1979. Based on the experimental data, an unbiased estimate of the impact of training presented in col. 4 is $851. The first three columns present the difference between each comparison group's 1975 and 1979 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.

[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.

[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

[d] See Table 2 for definitions of the comparison groups.

$y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2' x_{2it} + \varepsilon_{it}$, and difference, or even "quasi-difference" (include time invariant vars.)

Bad news: adding the omitted vars. misses positive selection

Good news: including more covariates reduces OVB, though only sometimes

Bad news: no way of knowing which comparison group to choose to start with, so no way to know a priori which estimate is consistent, if any!

6. Causal inference w/out experiments: Heckman 2 step selection correction (1979)

(1) $$y_{it} = \delta D_i + \beta X_{it} + b_i + n_t + \varepsilon_{it}$$

(2) $$\varepsilon_{it} - \rho \varepsilon_{it-1} = \nu_{it}$$

(3) $$d_{is} = y_{is} + \gamma Z_{is} + \eta_{is}$$

(4) $$D_i = 1 \quad \text{if } d_{is} > 0; \quad D_i = 0 \quad \text{if } d_{is} < 0.$$

(6) $$E(b_i + \varepsilon_{it} | Z_i, D_i) = \rho \sigma_\varepsilon \left[D_i \frac{\phi(\gamma Z_i)}{1 - \Phi(\gamma Z_i)} - (1 - D_i) \frac{\phi(\gamma Z_i)}{\Phi(\gamma Z_i)} \right] = r H_i,$$

..if errors epsilon and eta are joint normal with correlation rho, both equations are perfectly specified including all relevant variables.

This allows the expected selection bias to be treated as an omitted variable in (7).

(7) $$Y_{it} = \delta D_i + \beta X_{it} + r H_i + v_i^*,$$

Causal Inf. w/out expts: Heckman 2 step

TABLE 6—ESTIMATED TRAINING EFFECTS USING TWO-STAGE ESTIMATOR

Variables Excluded from the Earnings Equation, but Included in the Participation Equation	Comparison Group	NSW AFDC Females: Training Dummy	NSW AFDC Females: Estimate of Expectation	NSW Males: Training Dummy	NSW Males: Estimate of Expectation
		Heckman Correction for Program Participation Bias, Using Estimate of Conditional Expectation of Earnings Error as Regressor in Earnings Equation — Estimate of Coefficient for			
Marital Status, Residency in an SMSA, Employment Status in 1976, AFDC Status in 1975, Number of Children	*PSID*-1	1,129 (385)	−894 (396)	−1,333 (820)	−2,357 (781)
	CPS-SSA-1	1,102 (323)	−606 (480)	−22 (584)	−1,437 (449)
	NSW Controls	837 (317)	−18 (2376)	899 (840)	−835 (2601)
Employment Status in 1976, AFDC Status in 1975, Number of Children	*PSID*-1	1,256 (405)	−823 (410)	–	–
	CPS-SSA-1	439 (333)	−979 (481)	–	–
	NSW Controls	–	–	–	–
Employment Status in 1976, Number of Children	*PSID*-1	1,564 (604)	−552 (569)	−1,161 (864)	−2,655 (799)
	CPS-SSA-1	552 (514)	−902 (551)	13 (584)	−1,484 (450)
	NSW Controls	851 (318)	147 (2385)	889 (841)	−808 (2603)
No Exclusion Restrictions	*PSID*-1	1,747 (620)	−526 (568)	−667 (905)	−2,446 (806)
	CPS-SSA-1	805 (523)	−908 (548)	213 (588)	−1,364 (452)
	NSW Controls	861 (318)	284 (2385)	889 (840)	−876 (2601)

Notes: The estimated training effects are in 1982 dollars. For the females, the experimental estimate of impact of the supported work program was \$851 with a standard error of \$317. The one-step estimates from col. 11 of Table 4 were \$2,097 with a standard error of \$491 using the *PSID*-1 as a comparison group, \$1,041 with a standard error of \$503 using the *CPS-SSA*-1 as a comparison group, and \$854 with a standard error of \$312 using the NSW controls as a comparison group. Estimates are missing for the case of three exclusions using the NSW controls since AFDC status in 1975 cannot be used as an instrument for the NSW females. For the males, the experimental estimate of impact of the supported work program was \$886 with a standard error of \$476. The one-step estimates from col. 10 of Table 5 were \$−1,228 with a standard error of \$896 using the *PSID*-1 as a comparison group, \$−805 with a standard error of \$484 using the *CPS-SSA*-1 as a comparison group, and \$662 with a standard error of \$506 using the NSW controls as a comparison group. Estimates are missing for the case of three exclusions for the NSW males as AFDC status is not used as an instrument in the analysis of the male trainees.

Table 6: Evaluating the “2 step”

- Using experimental controls Heckman “2 step” provides a close estimate of treatment effect and a reassuring zero coefficient on selection effect (since selection was random).
- Using nonexperimental comparisons the “2 step” estimates show *negative* selection of varying amounts across comparison groups so they generally overestimate the treatment effect.

The “2 step” vs. more regression controls

- Comparing Tables 4 and 6, the “2 step” estimates of treatment effects are no more stable than the regression controls. Both vary across comparison groups, undermining confidence in the method.

  - Notes 1: the only difference between using regression controls and using the selection correction is the functional form in (7) in this case
  - 2: if one had a variable that belonged in the selection equation (D) but not in the earning equation (y) then the selection term could be estimated without relying only on the functional form information.
  .. but then why not just use that variable as an instrument?

  - 3: Even if correction doesn’t work, Heckman (1979) taught us that selection bias can be treated as an OVB.

- Conclusions: Either the functional forms are incorrect or relevant omitted variables remain, or both.

  - i.e., Selection is a pretty devious form of OVB, at least with strict eligibility criteria.

Bonus: Causal Inf. w/out exp - Propensity Scores and Matching

- Wahba & Dehejia (1998) claim that you can reproduce the selection process using the propensity score method for the NSW sample.
  Note: Lalonde is doing some of this already.
- Smith & Todd (1998 or so) dispute the claim.
- Journal of Econometrics (2005) contains responses and rejoinders.
- Nobody claims that matching on propensity scores is a panacea (i.e., a general solution).

Bonus: Rubin's causality definition

Assume a binary RHS variable, D є (0,1).

Assume that for each i, $Y_i(D)$ has two *potential outcomes*, $Y_i(0)$ and $Y_i(1)$ (only one of which we can ever observe).

Assume that if $D_i = D'_i$ then $Y_i(D) = Y_i(D')$, i.e., regardless of the values of $D_{k \neq i}$.
(SUTVA: Stable unit treatment value assumption)

The *causal effect* is then defined as $Y_i(1) - Y_i(0)$.
Note: the notion of ceteris paribus is implicit in the potential outcomes.

OLS regression of Y on D is an unbiased estimate of $E(Y_i(1) - Y_i(0))$, the *average causal effect*, if D is randomly assigned.

This approach dates back to R.A. Fisher and John Neyman in the 1910s.

Statisticians prefer this "ε-free" notation.

Next time.. Problems with "Perfectly Specified" equations and how to solve them.

- Reading: Could be anything applied, so you might as well read the classics..

  Griliches, Z. (1986) "Economic Data Issues," in *Handbook of Econometrics, Volume III*, (Z. Griliches and M.D. Intriligator eds.) Elsevier Science.

  or

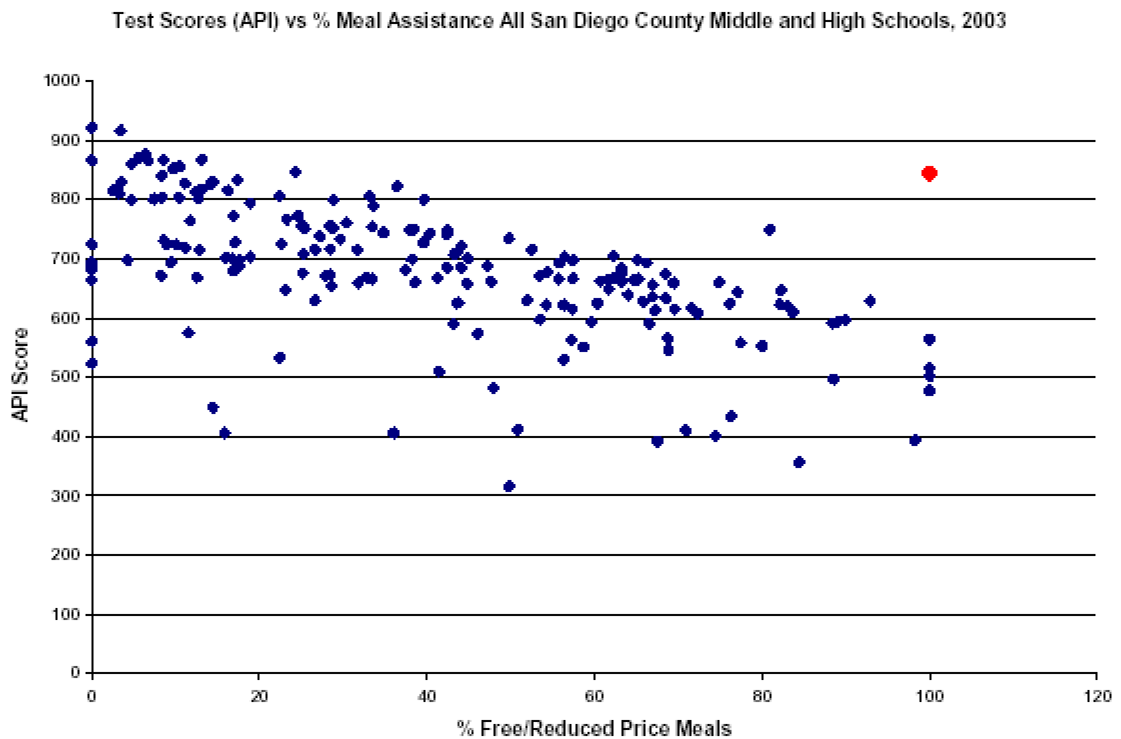
  Smith, Adam (1981 [1776]). *An inquiry into the nature and causes of the wealth of nations.* Indianapolis, Liberty Classics.

Causal Inference with and without Experiments II

0. Examples of Experiments
1. Population Parameters: What did your sample regression aspire to estimate?
2. Flavors of Omitted Variable Bias
3. Measurement Error
4. Misspecification of Functional Form
5. Heterogeneity
6. Endogeneity
7. Problems and Solutions

Preuss School: Selection or Treatment? (Betts, 2006)

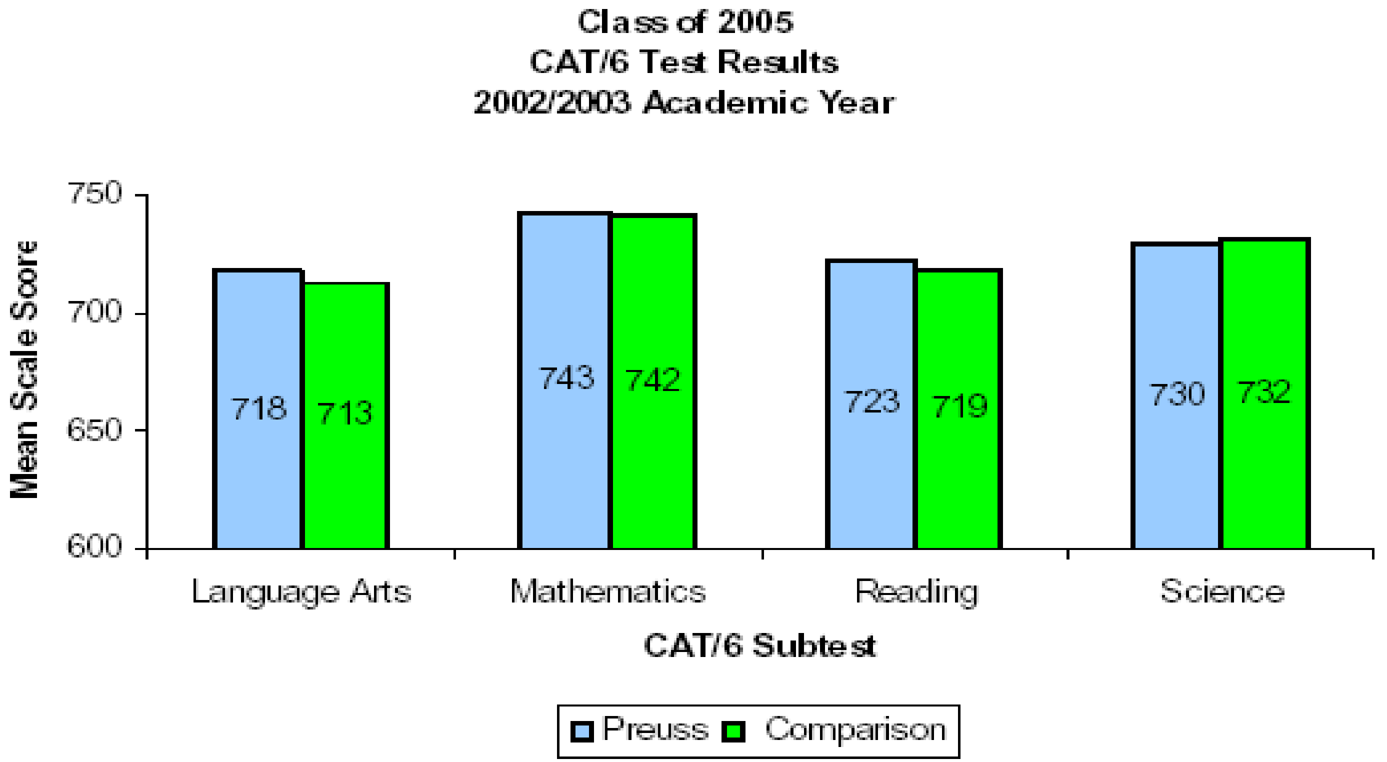
Figure 1.5.1



SOURCES: California Department of Education at http://api.cde.ca.gov/datafiles.html and http://www.cde.ca.gov/demographics/files/afdc.htm were the sources for figure 1.5.1

Experimental Evaluation of Preuss Students

Figure 3.3.1 – Cohort 2005



There were no statistically significant or marginal differences between Preuss and Comparison Groups on these measures.

Other Examples: Corruption in Indonesia

Monitoring Corruption:* *Evidence from a Field Experiment in Indonesia* Benjamin A. Olken, NBER, November 2004

- about half of road construction projects chosen for audit

Table 3: Audits – main theft results

			No Fixed Effects		Engineer Fixed Effects		Stratum Fixed Effects		
Percent missing: Log reported value – Log actual value	Control Mean	Treatment Mean: Audits	Audit Effect	P-Value	Audit Effect	P-Value	Audit Effect	P-Value	Num Obs
Major items in roads	0.296 (0.036)	0.211 (0.033)	-0.086 (0.049)	0.083	-0.090 (0.044)	0.041	-0.056 (0.038)	0.140	490
Major items in roads and ancillary projects	0.296 (0.031)	0.218 (0.034)	-0.077 (0.046)	0.098	-0.079 (0.042)	0.061	-0.072 (0.037)	0.051	556
Breakdown of roads:									
Materials	0.223 (0.043)	0.175 (0.040)	-0.049 (0.058)	0.404	-0.031 (0.054)	0.561	0.004 (0.050)	0.934	486
Unskilled labor	0.333 (0.086)	0.263 (0.079)	-0.066 (0.117)	0.573	-0.102 (0.096)	0.293	0.001 (0.086)	0.990	434

Notes: Robust standard errors in parentheses. Audit effect and p-values are computed from a regression of the dependent variable on a dummy for audit treatment, invitations treatment and invitations + comment forms treatments, allowing for robust standard errors clustered by subdistrict to account for clustering of treatment by subdistrict. All dependent variables are the log of the value reported by the village less the log of the estimated actual value, which is approximately equal to the percent of corruption. Villages are included in each row only if there was positive reported expenditures for the dependent variable listed in that row.

Callen and Long: Election Monitoring in Afghanistan (Sept., 2010)

Table 1: Summary Statistics and Randomization Verification

Variable	Mean	Std. Dev.	N	Mean Control	Mean Treatment	p-value (Control = Treatment)
Plans to turnout during election (=1)	0.794	0.234	449	0.789	0.798	0.732
Believes vote is secret (=1)	0.659	0.26	449	0.668	0.650	0.473
People in precinct will vote for same cand. (=1)	0.241	0.255	444	0.234	0.249	0.552
Problems with ballot transport are likely (=1)	0.535	0.303	443	0.536	0.534	0.939
Police in PC help security (=1)	1.326	0.324	449	1.332	1.319	0.676
District Elect. Officer keeps elect. fair (=1)	0.55	0.289	447	0.555	0.545	0.716
District Governor keeps elect. fair (=1)	0.112	0.169	447	0.110	0.115	0.801
People like you are threatened to vote one way (=1)	0.208	0.227	448	0.214	0.202	0.576
Local violence likely on elect. day (=1)	0.492	0.332	444	0.501	0.483	0.560
Current MP from same Qawm (=1)	0.186	0.202	449	0.181	0.190	0.650
MP Candidate from same Qawm (=1)	0.233	0.223	449	0.235	0.232	0.900
Current MP from same Qawm (=1)	0.186	0.202	449	0.181	0.190	0.650
Electrified (=1)	0.714	0.312	449	0.723	0.706	0.572
Local violence likely on elect. day (=1)	0.492	0.332	444	0.501	0.483	0.560
Trad. auth. helps settle disputes (=1)	1.86	1.639	449	1.860	1.859	0.993
Pashtun (=1)	2.098	2.597	449	2.144	2.053	0.710
Tajik (=1)	2.797	2.49	449	2.797	2.797	0.999
Income generating activity (=1)	3.913	1.392	449	3.932	3.894	0.772
Monthly Income	56.465	32.496	449	56.583	56.351	0.940

Table 2: Treatment Effects

	Removed Tally Sheet (=1)				Number of ECC Complaints			
Received Letter	0.093**	0.086**	0.087**	0.085**	0.099	0.106*	0.106*	0.108*
	(0.047)	(0.039)	(0.039)	(0.040)	(0.063)	(0.057)	(0.057)	(0.058)
Electrified (=1)			0.120	0.116			0.048	0.026
			(0.089)	(0.093)			(0.054)	(0.058)
District Governor keeps elect. fair (=1)			-0.079	-0.097			0.198	0.237
			(0.121)	(0.131)			(0.266)	(0.254)
Province Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Full covariates	No	No	No	Yes	No	No	No	Yes
R-squared	0.009	0.336	0.339	0.331	0.005	0.232	0.234	0.252
N	450	450	450	440	450	450	450	440

Notes: *Level of significance*: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors reported in parentheses. ECC complaints are the sum of complaints made to the Electoral Complaints Commission by candidates about polling center officials at a given polling center. The full covariates are the share of Pashtun respondents, the share of Tajik respondents, income, employment, and expectations about violence on election day.


08/13/2010 04:50

Costs

Financial

- Baseline Survey – $84,632
- Election Day – $71,550
- Travel etc. – $11,000
- Research Asst. – $2,440
- Endline Survey –$84,632
- Overhead (26%) –$43,223
- Total $297,477

Human

- Mike – 60 Days
- James – 75 Days
- Shahim – 30 Days
- Mohammad – 30 Days
- UCSD Support –10 Days

More Examples: Battles of Sexes over Budgets

Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines. *Nava Ashraf, Harvard University*

Table 2: Main Experimental Outcomes

	Private		Public		Negotiation	
	Male N=48 (1)	Female N=48 (2)	Male N=48 (3)	Female N=48 (4)	Male N=50 (5)	Female N=50 (6)
Prefer Direct Deposit over Self[a]	75.0%	62.5%	41.7%***	60.4%	71.4%+++	70.0%
Prefer Direct Deposit over Food[b]	60.4%	52.1%	21.7%***	41.7%	54.0%+++	56.0%

***Significant at 1%, when compared to Private condition

**Significant at 5%, when compared to Private condition

*Significant at 10%, when compared to Private condition

+++Significant at 1%, when compared to Public condition

Notes: This table is a comparison of means across treatment groups for males and females.

[a]Percent of individuals who preferred direct deposit for 200 into savings account over gift certificate for self worth 200 pesos

[b]Percent of individuals who preferred direct deposit for 200 into savings accout over gift certificate for food worth 200 pesos

1. Population Parameters: What did your sample regression aspire to estimate?

<u>Sample</u>	<u>Population</u>
	1. CEF ("Best" Predictor)
y = Xb + e, x'e=0	2. BLP
	3. Causal Effect
	4. Linear Causal Effect
	5. Perfectly specified equation including all relevant variables

- In principle #4 and #5 yield identical population parameters for β_1 if Cov $(x_1, \varepsilon \mid \beta_2'x_2) = 0$, where $\varepsilon = y - \beta_1'x_1 - \beta_2'x_2$ (and β_1 is L.C.E.).

2. Flavors of Omitted Variable "Bias"

- If the sample regression aspires to estimate (5) a perfectly specified equation including all relevant variables then your biggest worry is: Did I include all the relevant variables?
- Lots of problems can be though of as OVB:
- 1. Forgot X_2 (e.g., reconstruction spending)
- 2. Selection bias (Heckman '79)
- 3. Measurement error
- 4. Misspecification of functional form
- 5. Heterogeneity bias (Mundlak '61)
- 6. Endogeneity bias

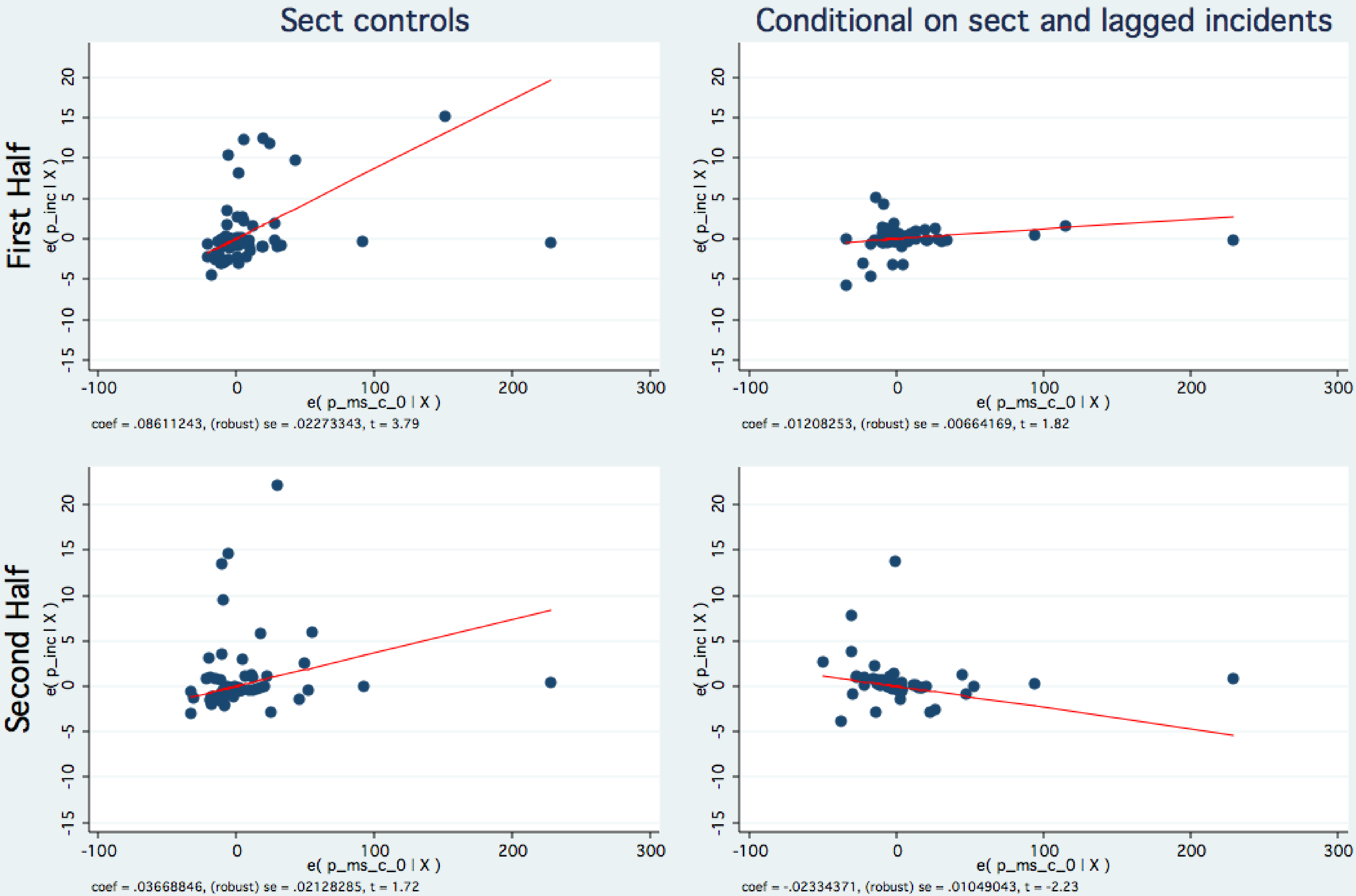
Our agenda: Problems and Solutions

Solution Problem	Add the omitted var.	Experiment	instrument
1. Forgot X_2	X?	X	
2. Selection	X?	X	
3. Meas. Err.	X?	-	x
4. Misspecification	X?	-	
5. Heterogeneity	X (fe) ?	X	
6. Endogeneity	-	limited?	X

Note: Matching is an alternative solution for binary X in selection problems.
Bottom Line: Dogmatism is fun but appropriate technology is more productive.

Nonexperimental Example from Iraq:

Incidents vs. Spending, 2007



3. Measurement Error and OVB

- $y_i = \beta_0 + \beta_1 x^*_i + \varepsilon_i$
- x^*_i not observed. The best we can do is observe a noisy measure of x*..

  $x_i = x^*_i + v_i$, Cov (v,x*)=0, Cov(ε,x*)=0, Cov(v, ε)=0
  "classical" measurement error assumptions
- Rewrite as an omitted variable..
  $y_i = \beta_0 + \beta_1 (x_i - v) + \varepsilon_i$
  $= \beta_0 + \beta_1 x_i - \beta_1 v + \varepsilon_{i\,,}$ (L)
  $y_i = \beta_0 + \beta_1 x_i + u_i$ (S), $Cov(x,u) \neq 0$

- ..and use OVB formula to solve

$b_1^s = b_1^L + b_{21}\, b_2^L$

3. Measurement Error (cont.)

- How big might a signal to noise ratio be?
- Data dependent. Survey data tends to be noisy. Bound & Krueger (1991) ask income of both employers and employees. Noise is 10-15% of variance in survey income in CPS, implying a 10-15% attenuation bias (i.e., bias toward zero) in the coeff. of a simple regression.
- Quiz: What happens to that attenuation bias as you add covariates to the RHS under classical assumptions?

Aside: Frisch-Waugh Theorem

- Y and X both have a trend. You want to estimate a regression of Y on X.
- Should you detrend before regressing or detrend as part of the regress?
- Frisch & Waugh, Econometrica 1(4), (1933)

4. Misspecification of functional form as OVB

- Typical misspecification of functional form is to oversimplify and assume that E(y|x) is linear.
- You can always approximate a nonlinear function with a power series, turning it into a "long"(er) regression.
- Often the gain is accuracy will cost precision (larger std. errors).
- E.g., Probit vs. "Linear Probability"

5. Heterogeneity Bias

- Mundlak (1961) was concerned with returns to scale in farming.
  Journal of Farm Economics, Vol. 43, No. 1 (Feb., 1961)
- Estimated a log-linear production function (i.e., Cobb-Douglas)
- If inputs and ability are complements (as implied by Cobb-Douglas), then high ability farmers use more inputs so that OVB is positive on input coefficients.
- So he used the fixed-effects regression to deal with OVB.

5. Heterogeneity Bias

- Mundlak (1961) was concerned with returns to scale in farming.
  Journal of Farm Economics, Vol. 43, No. 1 (Feb., 1961)
- Estimated a log-linear production function (i.e., Cobb-Douglas)
- If inputs and ability are complements (as implied by Cobb-Douglas), then high ability farmers use more inputs so that OVB is positive on input coefficients.
- So he invented the fixed-effects regression to deal with OVB.

Heterogeneity bias graphically

PRODUCTION FUNCTION FREE OF MANAGEMENT BIAS

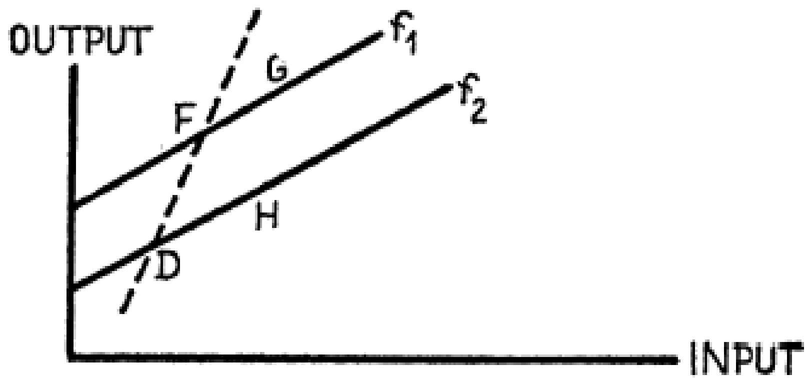


FIG. 1

The variables included in the regression are:

Y = value of product
X_1 = number of labor days
X_2 = variable expense
X_3 = value of livestock at the beginning of the year
X_4 = value of livestock and poultry barns measured in I£. The value was derived by first measuring the capacity of the barns disregarding age differences. The capacity values were then multiplied by the market price.
X_i = amount of land on irrigated area basis (1 irrigated dunam = 4 dunams of dry land; 1 dunam = $\frac{1}{4}$ acre).

Heterogeneity bias – Managerial ability

TABLE 1

Item	X_1	X_2	X_3	X_4	X_5	Output
1. Geometric means	539	9900	1792	6621	255	18973
2. Estimated elasticities						Σb
a. Data pooled	.130	.692	.0043**	.103	.037	.967
b. Allowing for a year effect	.153	.679	.0042**	.101	.032*	.969
c. Allowing for a firm effect	.083*	.635	.0021**	.156	.002**	.878
d. Allowing for both year & firm effects	.115	.582	.005**	.100*	−.007**	.795
3. Absolute bias						
(a) Assumption of no year effect	.047	.057	.0022	−.053	.035	
(b) Allowing for year effect	.038	.097	−.0008	.001	.039	

Evidence of heterogeneity/managerial/ability bias.
Or not. What about measurement error?

6. Endogeneity Bias

- We seek an estimate of the effect of x on y, but y affects x.
- E.g., p and q in markets.
- Can long regressions generate consistent estimates in principle?
- Can experiments ?

Endogeneity: Supply-Demand example

Instrumental Variables

- Validity
- Relevance

- ..but a good instrument is hard to find

Instrumental Variables

- IV vs. Measurement Error
- IV vs. Selection Bias

  Angrist, Joshua (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Records," *American Economic Review*, 80:3 (June).

7. Problems and Solutions

Solution Problem	Add the omitted var.	experiment	instrument
1. Forgot X_2			
2. Selection			
3. Meas. Err.			
4. Misspecification			
5. Heterogeneity			
6. Endogeneity			

Note: Matching is an alternative solution for binary X in selection problems.

Bottom Line: Dogmatism is fun but appropriate technology is more productive.