Econometrics 2

Sergey Lychagin

CEU

Winter 2020

Evaluating Training Programs

- What is the causal effect of participating in a training program on future earnings?
- Government programs for unemployed or disadvantaged workers lacking basic skills
- Training: counseling, work experience, courses
- Evaluation strategies
  - Experimental evaluations: randomized assignment
  - Non-experimental methods relying on CIA: simple regression, diff-in-diff, Heckman's correction, propensity score matching.
- Do these methods find similar results?
- LaLonde (1986), Dehejia and Wahba JASA (1999)

National Supported Work Demonstration

- National Supported Work Demonstration (NSW): temporary employment program to help disadvantaged workers lacking basic skills to move in to labor market.
- Separate programs for males and females (AFDC, mostly single unemployed mothers)
- Experiment: Random assignment of qualified applicants to *treatment* groups and *control* groups in 10 sites across the US.
- Guaranteed 9-18 months job plus counseling.
- Data: Survey with baseline information 4 follow up interviews.
- Individual level panel data, male sample $t = 1975, ..., 1979$
- Training participation in $1976 - 1977$
- Baseline period: 1975, After period: $t = 1979$
- Experimental Results

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.

Non-experimental Evaluation

Keep observations from the treatment group

1. Select a control group from publicly available data
2. Choose an econometric model to control for selection into training

Choice of control groups:

- PSID: nationally representative panel surveys of households
- PSID-1: male household heads continuously observed 1975 - 1978
- PSID-2: men not working in spring 1976
- PSID-3: men not working in spring 1975 and 1976
- CPS-3: men unemployed in 1976 with income below poverty level in 1975

Ideally — make the control and the treatment group look similar in observables.

TABLE 3—ANNUAL EARNINGS OF NSW MALE TREATMENTS, CONTROLS, AND SIX CANDIDATE COMPARISON GROUPS FROM THE *PSID* AND *CPS-SSA*

			Comparison Group[a,b]					
Year	Treatments	Controls	*PSID*-1	*PSID*-2	*PSID*-3	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3
1975	$3,066	$3,027	19,056[a]	7,569	2,611	13,650	7,387	2,729
	(283)	(252)	(272)	(568)	(492)	(73)	(206)	(197)
1976	$4,035	$2,121	20,267	6,152	3,191	14,579	6,390	3,863
	(215)	(163)	(296)	(601)	(609)	(75)	(187)	(267)
1977	$6,335	$3,403	20,898	7,985	3,981	15,046	9,305	6,399
	(376)	(228)	(296)	(621)	(594)	(76)	(225)	(398)
1978	$5,976	$5,090	21,542	9,996	5,279	14,846	10,071	7,277
	(402)	(227)	(311)	(703)	(686)	(76)	(241)	(431)
Number of Observations	297	425	2,493	253	128	15,992	1,283	305

[a] The Comparison Groups are defined as follows: *PSID*-1: All male household heads continuously from 1975 through 1978, who were less than 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all men who were not working when surveyed in the spring of 1976; *PSID*-3: Selects from the *PSID*-1 group all men who were not working when surveyed in either spring of 1975 or 1976; *CPS-SSA*-1: All males based on Westat's criteria, except those over 55-years-old; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all males who were not working when surveyed in March 1976; *CPS-SSA*-3: Selects from the *CPS-SSA*-1 unemployed males in 1976 whose income in 1975 was below the poverty level.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. The number of observations refer only to 1975 and 1978. In the other years there are fewer observations. The sample of treatments is smaller than the sample of controls because treatments still in Supported Work as of January 1978 are excluded from the sample, and in the young high school target group there were by design more controls than treatments.

Econometric Evaluation Methods

Main threat to identification — selection on observables and unobservables. Treatment is administered to a very special group!
Regression based methods

$$
\begin{aligned}
Y_{it} &= \delta D_i + \beta X_{it} + a_i + b_t + \epsilon_{it} \\
D_i &= \begin{cases} 1 & \text{if treated} \\ 0 & \text{if control} \end{cases}
\end{aligned}
$$

- Unadjusted $Y_{it} = \delta D_i + \epsilon_{it}, \quad t = 1978$
- Adjusted $Y_{it} = \delta D_i + \beta X_{it} + \epsilon_{it}, \quad t = 1978$
- FD: $Y_{it} = \delta D_i + \beta X_{it} + a_i + b_t + \epsilon_{it}, \; t = 1975, 1978$ (difference-in-differences)
- FD controlling for pre-training earnings ("quasi difference").

Female participants (AFDC)

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]

Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–79 (1)	NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2)	Pre-Training Year, 1975 Adjusted[c] (3)	Post-Training Year, 1979 Unadjusted (4)	Post-Training Year, 1979 Adjusted[c] (5)	Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons Without Age (6)	With Age (7)	Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79 Unadjusted (8)	Adjusted[c] (9)	Controlling for All Observed Variables and Pre-Training Earnings Without AFDC (10)	With AFDC (11)
	2,942	−17	−22	851	861	833	883	843	864	854	–
Controls	(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)	

Female participants (AFDC)

- Columns (2-3): cross-sectional regressions aren't performing well. Selection on unobservables? Rule out (4-5)
- Ashenfelter's dip (see Table 1). Simple diff-in-diff is not a good idea. Rule out (6-7).
- Specifications (8-11) look okay, but:
  - the estimates are sensitive to sample choice
  - better balance $\rightarrow$ smaller sample $\rightarrow$ large std. errors

Male participants

TABLE 5—EARNINGS COMPARISONS AND ESTIMATED TRAINING EFFECTS FOR THE NSW MALE PARTICIPANTS USING COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*[a,b]

Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–78 (1)	NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2)	NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Adjusted[c] (3)	NSW Treatment Earnings Less Comparison Group Earnings: Post-Training Year, 1978 Unadjusted (4)	NSW Treatment Earnings Less Comparison Group Earnings: Post-Training Year, 1978 Adjusted[c] (5)	Difference in Differences: Difference in Earnings Growth 1975–78 Treatments Less Comparisons: Without Age (6)	Difference in Differences: Difference in Earnings Growth 1975–78 Treatments Less Comparisons: With Age (7)	Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–78: Unadjusted (8)	Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–78: Adjusted[c] (9)	Controlling for All Observed Variables and Pre-Training Earnings (10)
Controls	$2,063	$39	$−21	$886	$798	$847	$856	$897	$802	$662
	(325)	(383)	(378)	(476)	(472)	(560)	(558)	(467)	(467)	(506)
PSID-1	$2,043	−$15,997	−$7,624	−$15,578	−$8,067	$425	−$749	−$2,380	−$2,119	−$1,228
	(237)	(795)	(851)	(913)	(990)	(650)	(692)	(680)	(746)	(896)
PSID-2	$6,071	−$4,503	−$3,669	−$4,020	−$3,482	$484	−$650	−$1,364	−$1,694	−$792
	(637)	(608)	(757)	(781)	(935)	(738)	(850)	(729)	(878)	(1024)
PSID-3	($3,322	($455	$455	$697	−$509	$242	−$1,325	$629	−$552	$397
	(780)	(539)	(704)	(760)	(967)	(884)	(1078)	(757)	(967)	(1103)
CPS-SSA-1	$1,196	−$10,585	−$4,654	−$8,870	−$4,416	$1,714	$195	−$1,543	−$1,102	−$805
	(61)	(539)	(509)	(562)	(557)	(452)	(441)	(426)	(450)	(484)
CPS-SSA-2	$2,684	−$4,321	−$1,824	−$4,095	−$1,675	$226	−$488	−$1,850	−$782	−$319
	(229)	(450)	(535)	(537)	(672)	(539)	(530)	(497)	(621)	(761)
CPS-SSA-3	$4,548	$337	$878	−$1,300	$224	−$1,637	−$1,388	−$1,396	$17	$1,466
	(409)	(343)	(447)	(590)	(766)	(631)	(655)	(582)	(761)	(984)

Heckman correction

In theory, this takes care of selection on unobservables

1. Run probit: D_i on pre-training demographics, Z_i, compute the correction term ("estimate of expectation").
2. Run earnings post-treatment on controls, X_i and the correction term from step 1.

One extra dimension of modeling choice: exclusion restriction: what controls are included in Z_i, but not X_i?

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

Heckman correction

- Estimates for women sensitive to the choice of sample and exclusion restriction.
- Non-experimental and experimental estimates for men are not even close.
- Using better matched control groups results in huge std. errors (not reported in table 6)

LaLonde's conclusion

- Results are very sensitive to sample specification and econometric method
- Virtually as many estimates of training effects can be obtained as there are data sets and methods
- Randomized experiments are necessary to evaluate the causal training effects

Dehejia and Wahba JASA (1999)

- Regression based methods cannot recover the experimental extimate (LaLonde 1986)
- Are propensity score methods more successful?

Problems with NSW training evaluation

- Small homogeneous treatment group
- Large, very heterogeneous control groups in non-experimental setting
- Big differences in observable characteristics $\rightarrow$ "limited overlap"
- Search for the "best" subset of the control group to match with treatment group

Samples

- Male Participants
- Lalonde: training participants in 1976 and 1977: 297T, 425C
- Dehejia and Wahba, participants in 1977: 185T, 260C
- Additional year of pre-treatment earnings: important control variable
- Treatment group from experimental data
- Control groups: based on survey data
- Replicate Lalonde's results for linear regression and fixed effects

Table 1. Sample Means of Characteristics for NSW and Comparison Samples

	No. of observations	*Age*	*Education*	*Black*	*Hispanic*	*No degree*	*Married*	*RE74 (U.S. $)*	*RE75 (U.S. $)*
NSW/Lalonde:[a]									
Treated	297	24.63	10.38	.80	.09	.73	.17		3,066
		(.32)	(.09)	(.02)	(.01)	(.02)	(.02)		(236)
Control	425	24.45	10.19	.80	.11	.81	.16		3,026
		(.32)	(.08)	(.02)	(.02)	(.02)	(.02)		(252)
RE74 subset:[b]									
Treated	185	25.81	10.35	.84	.059	.71	.19	2,096	1,532
		(.35)	(.10)	(.02)	(.01)	(.02)	(.02)	(237)	(156)
Control	260	25.05	10.09	.83	.1	.83	.15	2,107	1,267
		(.34)	(.08)	(.02)	(.02)	(.02)	(.02)	(276)	(151)
Comparison groups:[c]									
PSID-1	2,490	34.85	12.11	.25	.032	.31	.87	19,429	19,063
		[.78]	[.23]	[.03]	[.01]	[.04]	[.03]	[991]	[1,002]
PSID-2	253	36.10	10.77	.39	.067	.49	.74	11,027	7,569
		[1.00]	[.27]	[.04]	[.02]	[.05]	[.04]	[853]	[695]
PSID-3	128	38.25	10.30	.45	.18	.51	.70	5,566	2,611
		[1.17]	[.29]	[.05]	[.03]	[.05]	[.05]	(686)	[499]
CPS-1	15,992	33.22	12.02	.07	.07	.29	.71	14,016	13,650
		[.81]	[.21]	[.02]	[.02]	[.03]	[.03]	[705]	[682]
CPS-2	2,369	28.25	11.24	.11	.08	.45	.46	8,728	7,397
		[.87]	[.19]	[.02]	[.02]	[.04]	[.04]	[667]	[600]
CPS-3	429	28.03	10.23	.21	.14	.60	.51	5,619	2,467
		[.87]	[.23]	[.03]	[.03]	[.04]	[.04]	[552]	[288]

NOTE: Standard errors are in parentheses. Standard error on difference in means with RE74 subset/treated is given in brackets. Age = age in years; Education = number of years of schooling; Black = 1 if black, 0 otherwise; Hispanic = 1 if Hispanic, 0 otherwise; No degree = 1 if no high school degree, 0 otherwise; Married = 1 if married, 0 otherwise; RE*x* = earnings in calendar year

Propensity score methods

Average treatment effect on the treated

$$\delta_{TOT} = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)$$

Identification assumption

$$(Y_{1i}, Y_{0i}) \amalg D_i|X_i$$

Conditional Independence Assumption or "Selection on Observables"

Propensity Score Theorem:

$$\{Y_{1i}, Y_{0i}\} \amalg D_i|p(X_i)$$

Propensity score $p(X_i) = Prob(D_i = 1|X_i)$

Estimation steps

1. Estimate the propensity score $\hat{p}(X)$
2. Matching on the propensity score

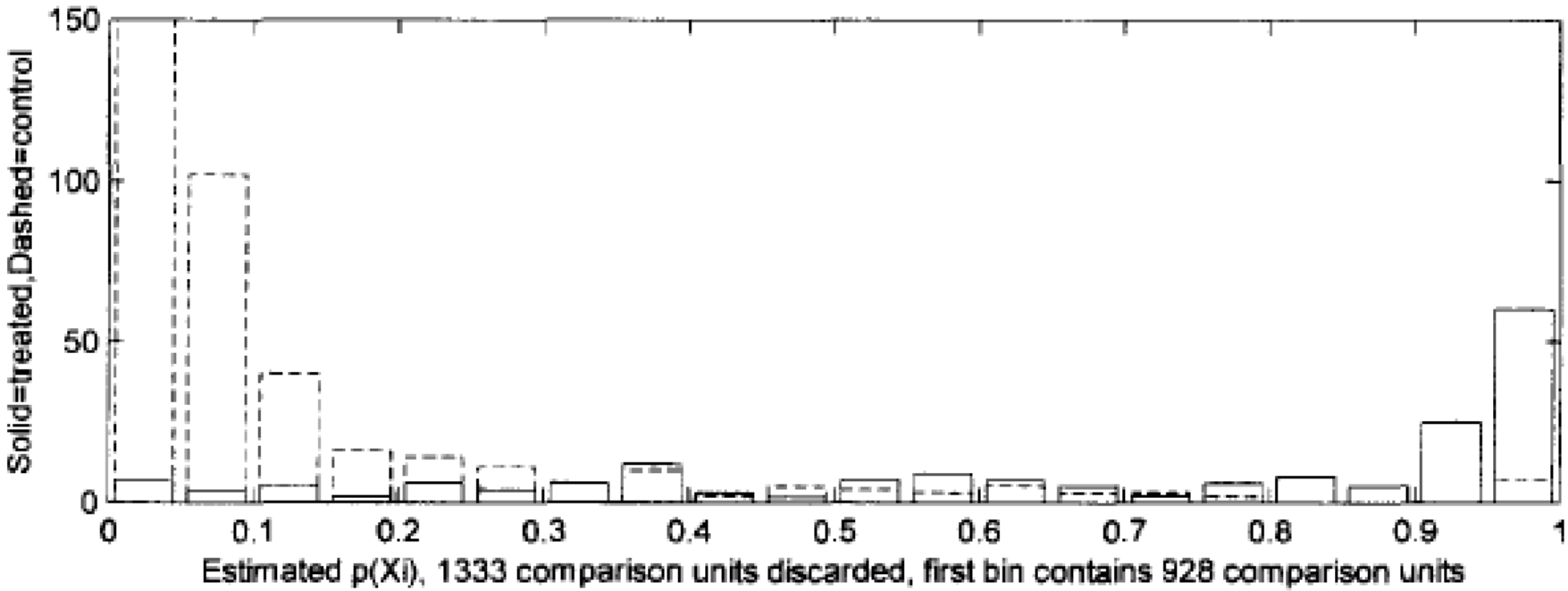
Why δ_{TOT}, not δ_{ATE}? NSW experiment gives δ_{TOT}; we can't use it to say much about the general population.

Estimation of the propensity scores

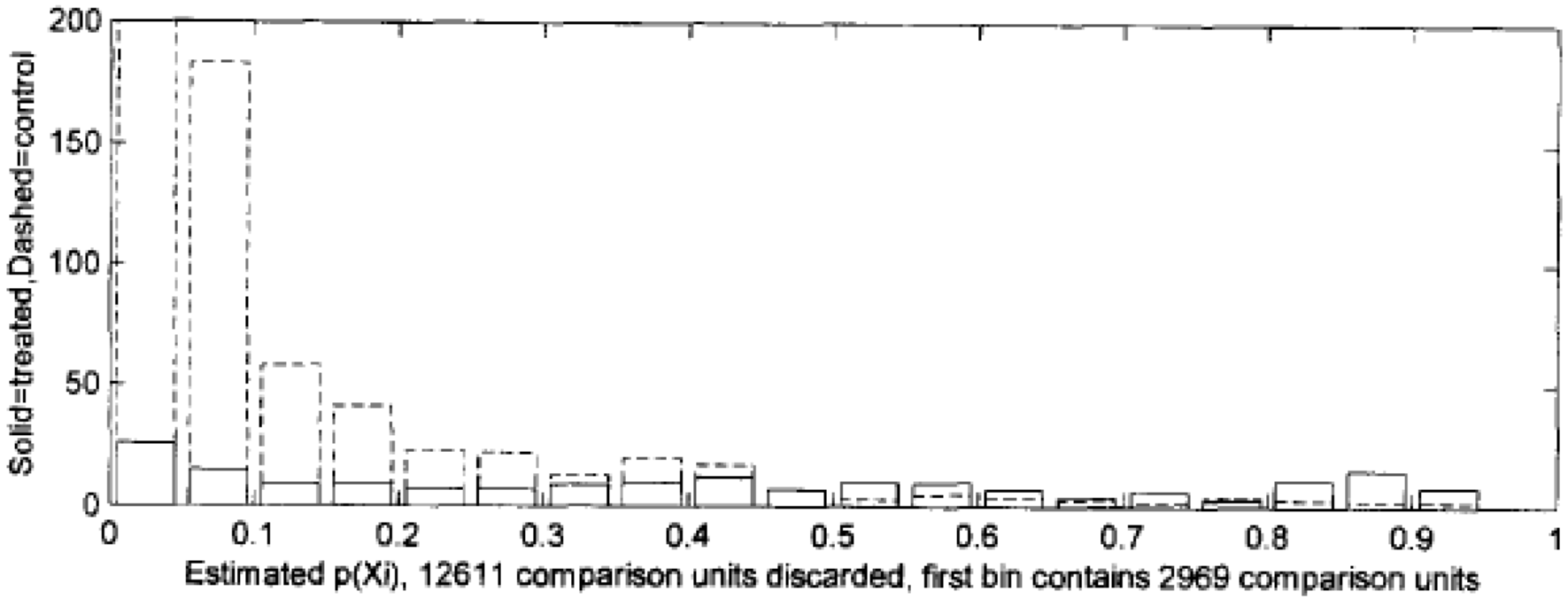
- Estimate a parsimonious logit for $P(D_i = 1|X_i)$
- Stratify the data by quintile blocks of $\hat{p}(X_i)$
- Compare $\bar{X}_C - \bar{X}_T$ in each block. Use a t-test or F-test of significant differences in means
  1. if X_i are balanced in each block STOP
  2. if not balanced, divide block in 2 parts and re-evaluate
  3. if X_i not balanced in all blocks re-specify the logit: add interaction terms and polynomials

See also: Caliendo, Marco and Sabine Kopeinig (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching" Journal of Economic Surveys, 22(1), 31-72.

PSID


150
100
50
0
Solid=treated,Dashed=control
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Estimated p(Xi), 1333 comparison units discarded, first bin contains 928 comparison units

CPS


Solid=treated,Dashed=control
200
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Estimated p(Xi), 12611 comparison units discarded, first bin contains 2969 comparison units

Propensity scores

- Limited overlap:
- Comparison units are dropped if $\hat{p}_j < min_{D_i=1}\hat{p}_i$, 12,611 dropped from CPS sample, 1333 from PSID sample
- Few observations in the control groups with high propensity scores.

Matching Methods

- Regression on the propensity score

$$\begin{aligned} Y_{it} &= \delta D_i + \gamma_1 \hat{p}_{it} + \gamma_2 \hat{p}_{it}^2 + \epsilon_{it} \quad \text{or} \\ Y_{it} &= \delta D_i + \gamma_1 \hat{p}_{it} + \gamma_2 \hat{p}_{it}^2 + \beta X_{it} + \epsilon_{it} \end{aligned}$$

- Stratification: Within block regression
- Matching on propensity score: find closest match based on the propensity score
  - with/without replacement
  - how many comparison units?
  - nearest neighbor matching
  - match within an interval with similar values of the propensity score

Table 3. Estimated Training Effects for the NSW Male Participants Using Comparison Groups From PSID and CPS

	NSW earnings less comparison group earnings		NSW treatment earnings less comparison group earnings, conditional on the estimated propensity score					
			Quadratic in score[b]	Stratifying on the score			Matching on the score	
	(1) Unadjusted	(2) Adjusted[a]	(3)	(4) Unadjusted	(5) Adjusted	(6) Observations[c]	(7) Unadjusted	(8) Adjusted[d]
NSW	1,794	1,672						
	(633)	(638)						
PSID-1[e]	−15,205	731	294	1,608	1,494	1,255	1,691	1,473
	(1,154)	(886)	(1,389)	(1,571)	(1,581)		(2,209)	(809)
PSID-2[f]	−3,647	683	496	2,220	2,235	389	1,455	1,480
	(959)	(1,028)	(1,193)	(1,768)	(1,793)		(2,303)	(808)
PSID-3[f]	1,069	825	647	2,321	1,870	247	2,120	1,549
	(899)	(1,104)	(1,383)	(1,994)	(2,002)		(2,335)	(826)
CPS-1[g]	−8,498	972	1,117	1,713	1,774	4,117	1,582	1,616
	(712)	(550)	(747)	(1,115)	(1,152)		(1,069)	(751)
CPS-2[g]	−3,822	790	505	1,543	1,622	1,493	1,788	1,563
	(670)	(658)	(847)	(1,461)	(1,346)		(1,205)	(753)
CPS-3[g]	−635	1,326	556	1,252	2,219	514	587	662
	(657)	(798)	(951)	(1,617)	(2,082)		(1,496)	(776)

Table 4. Sample Means of Characteristics for Matched Control Samples

Matched samples	*No. of observations*	*Age*	*Education*	*Black*	*Hispanic*	*No degree*	*Married*	*RE74 (U.S. $)*	*RE75 (U.S. $)*
NSW	185	25.81	10.35	.84	.06	.71	.19	2,096	1,532
MPSID-1	56	26.39	10.62	.86	.02	.55	.15	1,794	1,126
		[2.56]	[.63]	[.13]	[.06]	[.13]	[.12]	[1,406]	[1,146]
MPSID-2	49	25.32	11.10	.89	.02	.57	.19	1,599	2,225
		[2.63]	[.83]	[.14]	[.08]	[.16]	[.16]	[1,905]	[1,228]
MPSID-3	30	26.86	10.96	.91	.01	.52	.25	1,386	1,863
		[2.97]	[.84]	[.13]	[.08]	[.16]	[.16]	[1,680]	[1,494]
MCPS-1	119	26.91	10.52	.86	.04	.64	.19	2,110	1,396
		[1.25]	[.32]	[.06]	[.04]	[.07]	[.06]	[841]	[563]
MCPS-2	87	26.21	10.21	.85	.04	.68	.20	1,758	1,204
		[1.43]	[.37]	[.08]	[.05]	[.09]	.08	[896]	[661]
MCPS-3	63	25.94	10.69	.87	.06	.53	.13	2,709	1,587
		[1.68]	[.48]	[.09]	[.06]	[.10]	[.09]	[1,285]	[760]

Table 5. Sensitivity of Estimated Training Effects to Specification of the Propensity Score

	NSW earnings less comparison group earnings		*NSW treatment earnings less comparison group earnings, conditional on the estimated propensity score*					
			Quadratic in score[c]	*Stratifying on the score*			*Matching on the score*	
Comparison group	*(1) Unadjusted*	*(2) Adjusted*[a]	*(3)*	*(4) Unadjusted*	*(5) Adjusted*	*(6) Observations*[d]	*(7) Unadjusted*	*(8) Adjusted*[b]
NSW	1,794	1,672						
	(633)	(638)						
Dropping higher-order terms								
PSID-1:	−15,205	218	294	1,608	1,254	1,255	1,691	1,054
Specification 1	(1,154)	(866)	(1,389)	(1,571)	(1,616)		(2,209)	(831)
PSID-1:	−15,205	105	539	1,524	1,775	1,533	2,281	2,291
Specification 2	(1,154)	(863)	(1,344)	(1,527)	(1,538)		(1,732)	(796)
PSID-1:	−15,205	105	1,185	1,237	1,155	1,373	1,140	855
Specification 3	(1,154)	(863)	(1,233)	(1,144)	(1,280)		(1,720)	(906)
CPS-1:	−8,498	738	1,117	1,713	1,774	4,117	1,582	1,616
Specification 4	(712)	(547)	(747)	(1,115)	(1,152)		(1,069)	(751)
CPS-1:	−8,498	684	1,248	1,452	1,454	6,365	835	904
Specification 5	(712)	(546)	(731)	(632)	(2,713)		(1,007)	(769)
CPS-1:	−8,498	684	1,241	1,299	1,095	6,017	1,103	1,471
Specification 6	(712)	(546)	(671)	(547)	(925)		(877)	(787)
Dropping RE74								
PSID-1:	−15,205	−265	−697	−869	−1,023	1,284	1,727	1,340
Specification 7	(1,154)	(880)	(1,279)	(1,410)	(1,493)		(1,447)	(845)
PSID-2:	−3,647	297	521	405	304	356	530	276
Specification 8	(959)	(1,004)	(1,154)	(1,472)	(1,495)		(1,848)	(902)
PSID-3:	1,069	243	1,195	482	−53	248	87	11
Specification 8	(899)	(1,100)	(1,261)	(1,449)	(1,493)		(1,508)	(938)
CPS-1:	−8,498	525	1,181	1,234	1,347	4,558	1,402	861
Specification 9	(712)	(557)	(698)	(695)	(683)		(1,067)	(786)
CPS-2:	−3,822	371	482	1,473	1,588	1,222	1,941	1,668
Specification 9	(670)	(662)	(731)	(1,313)	(1,309)		(1,500)	(755)
CPS-3:	−635	844	722	1,348	1,262	504	1,097	1,120
Specification 9	(657)	(807)	(942)	(1,601)	(1,600)		(1,366)	(783)

NOTE: Specification 1: Same as Table 3, note (c). Specification 2: Specification 1 without higher powers. Specification 3: Specification 2 without higher-order terms. Specification 4: Same as Table 3, note (e). Specification 5: Specification 4 without higher powers. Specification 6: Specification 5 without higher-order terms. Specification 7: Same as Table 3, note (c), with RE74 removed. Specification 8: Same as Table 3, note (d), with RE74 removed. Specification 9: Same as Table 3, note (e), with RE74 removed.

[a] Least squares regression: RE78 on a constant, a treatment indicator, age, education, no degree, black, Hispanic, RE74, RE75.

[b] Weighted least squares: treatment observations weighted as 1, and control observations weighted by the number of times they are matched to a treatment observation [same covariates as (a)].

[c] Least squares regression of RE78 on a quadratic on the estimated propensity score and a treatment indicator, for observations used under stratification; see note (d).

[d] Number of observations refers to the actual number of comparison and treatment units used for (3)–(5); namely, all treatment units and those comparison units whose estimated propensity score is greater than the minimum, and less than the maximum, estimated propensity score for the treatment group.

Matching Results

- Baseline: experimental results
- Matching on propensity score works better than quadratic control
- Creation of ad hoc subsamples of the non-experimental comparision group neither necessary nor desirable (PSID-2, 3, CPS-2, 3).
  Don't stratify manually unless you know what you are doing.
- Sensitivity to specification of p-score — diagnose p-score issues by comparing treatment and control in terms of X within strata