“New Developments in Econometrics”

Lecture 2

Estimation of Average Treatment Effects

Under Unconfoundedness, Part II

Guido Imbens

Cemmap Lectures, UCL, June 2009

Outline

1. Assessing Unconfoundedness (not testable)

2. Overlap

3. Illustration based on Lalonde Data

4. Illustration based on Lottery Data

1.I Assessing Unconfoundedness: Multiple Control Groups

Suppose we have a three-valued indicator $T_i \in \{-0, 1, 1\}$ for the groups (e.g., ineligibles, eligible nonnonparticipants and participants), with the treatment indicator equal to $W_i = 1\{T_i = 1\}$, so that

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i \in \{-1, 0\} \\ Y_i(1) & \text{if } T_i = 1. \end{cases}$$

Suppose we extend the unconfoundedness assumption to independence of the potential outcomes and the three-valued group indicator given covariates,

$$Y_i(0), Y_i(1) \perp\!\!\!\perp T_i \,\Big|\, X_i$$

Now a testable implication is

$$Y_i(0) \perp\!\!\!\perp 1\{T_i = 0\} \;\Big|\; X_i, T_i \in \{-1, 0\},$$

and thus

$$Y_i \perp\!\!\!\perp 1\{T_i = 0\} \;\Big|\; X_i, T_i \in \{-1, 0\}.$$

An implication of this independence condition is being tested by the tests discussed above. Whether this test has much bearing on the unconfoundedness assumption, depends on whether the extension of the assumption is plausible given unconfoundedness itself.

1.II Assessing Unconfoundedness: Estimate Effects on Pseudo Outcomes

Partition the covariate vector into $X_i = (X_i^p, X_i^r)$, X_i^p scalar.

Unconfoundedness assumes

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp W_i \mid (X_i^p, X_i^r)$$

Suppose we are willing to assume X_i^r is sufficient:

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp W_i \mid X_i^r$$

and suppose X_i^p is a good proxy for $Y_i(0)$, then we can test

$$X_i^p \perp\!\!\!\perp W_i \mid X_i^r$$

Most useful implementations with X_i^p a lagged outcome.

Suppose the covariates consist of a number of lagged outcomes $Y_{i,-1},\ldots,Y_{i,-T}$ as well as time-invariant individual characteristics Z_i, so that $X_i=(X_i^p,X_i^r)$, with $X_i^p=Y_{i,-1}$ and $X_i^r=(Y_{i,-2},\ldots,Y_{i,-T},Z_i)$. Outcome is $Y_i=Y_{i,0}$.

Now consider the following two assumptions. The first is unconfoundedness given only $T-1$ lags of the outcome:

$$Y_{i,0}(1),Y_{i,0}(0)\ \perp\!\!\!\perp\ W_i\mid Y_{i,-1},\ldots,Y_{i,-(T-1)},Z_i,$$

Then, under stationarity it seems reasonable to expect Then it follows that

$$Y_{i,-1}\ \perp\!\!\!\perp\ W_i\mid\ Y_{i,-2},\ldots,Y_{i,-T},Z_i,$$

which is testable.

2.I Assessing Overlap

The first method to detect lack of overlap is to look at summary statistics for the covariates by treatment group.

Most important here is the normalized difference in covariates:

$$\text{nor}-\text{dif} = \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{S^2_{X,0} + S^2_{X,1}}}$$

$$\overline{X}_w = \frac{1}{N_w} \sum_{i:W_i=w} X_i \quad \text{and} \quad S^2_{X,w} = \frac{1}{N_w - 1} \sum_{i:W_i=w} \left(X_i - \overline{X}_w\right)^2$$

Note that we do not report the t-statistic for the difference,

$$t = \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{S^2_{X,0}/N_0 + S^2_{X,1}/N_1}}$$

The t-statistic partly reflects the sample size. Given the normalized difference, a larger t-statistic just indicates a larger sample size, and therefore in fact an easier problem in terms of finding credible estimators for average treatment effects.

In general a difference in average means bigger than 0.25 standard deviations is substantial. In that case one may want to be suspicious of simple methods like linear regression with a dummy for the treatment variable.

Recall that estimating the average effect essentially amounts to using the controls to estimate $\mu_0(x) = \mathbb{E}[Y_i|W_i = 0, X_i = x]$ and using this estimated regression function to predict the (missing) control outcomes for the treated units.

With a large difference between the two groups, linear regression is going to rely heavily on extrapolation, and thus will be sensitive to the exact functional form.

Assessing Overlap by Inspecting the Propensity Score Distribution

The second method for assessing overlap is more directly focused on the overlap assumption.

It involves inspecting the marginal distribution of the propensity score in both treatment groups.

Any difference in covariate distribution shows up in differences in the average propensity score between the two groups.

Moreoever, any area of non-overlap shows up in zero or one values for the propensity score.

2.II Selecting a Subsample with Overlap: Matching

Appropriate when the focus is on the average effect for treated, $\mathbb{E}[Y_i(1)-Y_i(0)|W_i = 1]$, and when there is a relatively large pool of potential controls.

Order treated units by estimated propensity score, highest first.

Match highest propensity score treated unit to closest control on estimated propensity score, without replacement.

Only to create balanced sample, not as final analysis.

2.III Selecting a Subsample with Overlap: Trimming

Define average effects for subsamples $\mathbb{A}$:

$$\tau(\mathbb{A}) = \sum_{i=1}^{N} 1\{X_i \in \mathbb{A}\} \cdot \tau(X_i) / \sum_{i=1}^{N} 1\{X_i \in \mathbb{A}\}.$$

The efficiency bound for $\tau(\mathbb{A})$, assuming homoskedasticity, is

$$\frac{\sigma^2}{q(\mathbb{A})} \cdot \mathbb{E}\left[\frac{1}{e(X)} + \frac{1}{1 - e(X)} \middle| X \in \mathbb{A} \right],$$

where $q(\mathbb{A}) = \Pr(X \in \mathbb{A})$.

They derive the characterization for the set $\mathbb{A}$ that minimizes the asymptotic variance .

The optimal set has the form

$$\mathbb{A}^* = \{x \in \mathbb{X} | \alpha \leq e(X) \leq 1 - \alpha\},$$

dropping observations with extreme values for the propensity score, with the cutoff value α determined by the equation

$$\frac{1}{\alpha \cdot (1-\alpha)} =$$

$$2 \cdot \mathbb{E}\left[\frac{1}{e(X) \cdot (1 - e(X))} \middle| \frac{1}{e(X) \cdot (1 - e(X))} \leq \frac{1}{\alpha \cdot (1-\alpha)} \right].$$

Note that this subsample is selected solely on the basis of the joint distribution of the treatment indicators and the covariates, and therefore does not introduce biases associated with selection based on the outcomes.

Calculations for Beta distributions for the propensity score suggest that $\alpha = 0.1$ approximates the optimal set well in practice.

3. Applic. to Lalonde Data (Dehejia-Wahba Sample)

Data on job training program, first used by Lalonde (1986), See also Heckman and Hotz (1989), Dehejia and Wahba (1999).

Small experimental evaluation, 185 trainees, 260 controls, group of very disadvantaged in labor market. Experimental estimate of average effect is 1.7 (in thousands of dollars, substantively important, statistically significant).

Large, non-experimental comparison group from CPS (15,992 observations). Very different in distribution of covariates.

How well do the non-experimental results replicate the experimental ones? Is non-experimental analysis credible? Would we have known whether it was credible without experiments results?

Summary Statistics for Lalonde Data

	CPS (N=15,992)		trainees (N=185)		
Covariate	mean	(s.d.)	mean	(s.d.)	nor-dif
black	0.07	0.26	0.84	0.36	**2.43**
hispanic	0.07	0.26	0.06	0.24	-0.05
age	33.2	11.0	25.8	7.2	-0.80
married	0.71	0.45	0.19	0.39	**-1.23**
nodegree	0.30	0.46	0.71	0.46	0.90
education	12.0	2.9	10.4	2.0	-0.68
earn '74	14.02	9.57	2.10	4.89	**-1.57**
unempl '74	0.12	0.32	0.71	0.46	**1.49**
earn '75	13.65	9.27	1.53	3.22	**-1.75**
unempl '75	0.11	0.31	0.60	0.49	**1.19**

With the CPS comparison group the differences between the averages are up to 2.4 standard deviations from zero, suggesting there will be serious issues in obtaining credible estimates of the average effect of the treatment.

Simple linear regression is unlikely to work

Next, create a matched sample to improve balance.

First estimate the propensity score. Select linear and second order terms through stepwise procedure (See Imbens, 2009).

Start with preselected covariates. Here: `earn '74`, `unempl '74`, `earn '75`, and `unempl '75`.

Add additional linear terms sequentially if likelihood ratio test statistic is larger than 1.

Add second order terms sequentially if likelihood ratio test statistic is larger than 2.7.

Estimated Propensity Score for the Lalonde Data

	Variable	est	s.e.	t-stat
preselect lin terms	earn '74	0.41	0.11	3.7
	unempl '74	0.42	0.41	1.0
	earn '75	-0.33	0.06	-5.5
	unempl '75	-2.44	0.77	-3.2
add lin terms	black	4.00	0.26	15.1
	married	-1.84	0.30	-6.1
	nodegree	1.60	0.22	7.2
	hispanic	1.61	0.41	4.0
	age	0.73	0.09	7.8
2nd order terms	age × age	-0.01	0.00	-7.5
	unempl '74 × unempl '75	3.41	0.85	4.0
	earn '74× age	-0.01	0.00	-3.3
	earn '75 × married	0.15	0.06	2.6
	unempl '74 × earn '75	0.22	0.08	2.6

Order treated observations on estimated propensity score.

Starting with the highest propensity score, match each treated observation to single closest control, without replacement. Match on the propensity score.

Leads to matched sample with 185 treated, 185 controls.

Summary Statistics for Matched Lalonde Data

Covariate	CPS controls (N=185) mean	(s.d.)	trainees (N=185) mean	(s.d.)	nor-dif
black	0.84	0.36	0.84	0.36	0.00
hispanic	0.06	0.24	0.06	0.24	0.00
age	27.14	9.99	25.82	7.16	-0.15
married	0.31	0.46	0.19	0.39	-0.28
nodegree	0.59	0.49	0.71	0.46	0.25
education	10.72	2.28	10.35	2.01	-0.18
earn '74	2.24	4.63	2.10	4.89	-0.03
unempl '74	0.70	0.46	0.71	0.46	0.02
earn '75	1.76	3.47	1.53	3.22	-0.07
unempl '75	0.59	0.49	0.60	0.49	0.02

In the matched sample the normalized differences are small, but not neglible.

Re-estimate propensity score using same algorithm for covariate selection.

Estimated Propensity Score for Matched Lalonde Data

	Variable	est	s.e.	t-stat
presel lin terms	`earn '74`	0.03	0.04	0.7
	`unempl '74`	-0.00	0.42	-0.0
	`earn '75`	-0.06	0.05	-1.2
	`unempl '75`	0.26	0.36	0.7
add lin terms	`married`	-0.52	0.55	-0.9
	`nodegree`	0.26	0.26	1.0
2 nd order terms	`unempl '75` × `married`	-1.24	0.55	-2.3
	`married` × `nodegree`	1.10	0.55	2.0

Two estimators

Use on matched sample.

I. Blocking on Propensity Score with Regression
Number of blocks and boundaries data driven. Use regression with all covariates within blocks.

II. Matching with Regression
Match on all covariates, with replacement. Use regression on matched sample, with all covariates.

Lalonde Data: Estimates of Average Treatment Effects

	Pre-sel Cov		Blocking		Matching	
Outcome	p-score	regression	est	s.e.	est	s.e.
Earnings 1978	none	all	1.56	(0.70)	2.06	(0.85)
	selected	all	1.93	(0.68)	2.06	(0.85)

Robust against choice of estimator and specification, and close to experimental estimate of 1.7.

Next, let us assess unconfoundedness in this sample using earnings in 1975 as the pseudo outcome.

Assessing Unconfoundedness for the Lalonde Data

	Pre-sel	Covs	Blocking		Matching	
Pseudo Outcome	p-score	regress	est	s.e.	est	s.e.
Earnings 1975	none	all	-1.22	(0.27)	-1.25	(0.30)
	selected	all	-1.22	(0.27)	-1.25	(0.30)

Results are mixed. Consistently find substantial "effects" on earnings in 1975, statistically and economically significant

The sensitivity is not surprising given substantial differences in covariate distributions.

So, even though point estimates for effect on '78 earnings are accurate, we would not have known that based on assesment of unconfoundedness assumption.

4. Applic. to Imbens-Rubin-Sacerdote Lottery Data

Effect of winning large prize in lottery on average labor market earnings in six years post winning. Controls are lottery players who won small prizes.

Covariates include six years pre-winning earnings.

Exact same estimators, same selection algorithm for propensity score specification.

Pre-selected covariates: `Tix Bought`, `Years of Ed`, `Working Then`, and `Earn Year -1`.

Summary Statistics Lottery Data

Covariate	Losers (N=259) mean	(s.d.)	Winners (N=237) mean	(s.d.)	nor-dif
Year Won	6.38	1.04	6.06	1.29	-0.27
# Tickets	2.19	1.77	4.57	3.28	0.90
Age	53.2	12.9	46.9	13.8	-0.47
Male	0.67	0.47	0.58	0.49	-0.19
Education	14.43	1.97	12.97	2.19	-0.70
Work Then	0.77	0.42	0.80	0.40	0.08
Earn Year -6	15.6	14.5	12.0	11.8	-0.27
⋮					
Earn Year -1	18.0	17.2	14.5	13.6	-0.23
Pos Earn Year -6	0.69	0.46	0.70	0.46	0.03
⋮					
Pos Earn Year -1	0.69	0.46	0.74	0.44	0.10

Estimated Parameters of Propensity Score for the Lottery Data

	Variable	est	s.e.	t-stat
presel lin terms	Tix Bought	0.56	0.38	1.5
	Years of Ed	0.87	0.62	1.4
	Working Then	1.71	0.55	3.1
	Earn Year -1	-0.37	0.09	-4.0
add lin terms	Age	-0.27	0.08	-3.4
	Year Won	-6.93	1.41	-4.9
	Pos Earn Year -5	0.83	0.36	2.3
	Male	-4.01	1.71	-2.3
2nd order terms	Year Won × Year Won	0.50	0.11	4.7
	Earn Year -1 × Male	0.06	0.02	2.7
	Tix Bought ×Tix Bought	-0.05	0.02	-2.6
	⋮ (six more interactions)			
	Year Won × Male	0.44	0.25	1.7

Subsample Sizes after Trimming ($\alpha = 0.0891$).

	low $e(x) < \alpha$	middle $\alpha \leq e(X) \leq 1 - \alpha$	high $1 - \alpha < e(X)$	All
Losers	82	172	5	259
Winners	4	151	82	237
All	86	323	87	496

Summary Statistics Trimmed Lottery Data

Covariate	Losers (N=172) mean	(s.d.)	Winners (N=151) mean	(s.d.)	nor-dif
Year Won	6.40	1.12	6.32	1.18	-0.06
# Tickets	2.40	1.88	3.67	2.95	0.51
Age	51.5	13.4	50.4	13.1	-0.08
Male	0.65	0.48	0.60	0.49	-0.11
Education	14.01	1.94	13.03	2.21	-0.47
Work Then	0.79	0.41	0.80	0.40	0.03
Earn Year -6	15.5	14.0	13.0	12.4	-0.19
⋮					
Earn Year -1	18.4	16.6	15.4	14.4	-0.19
Pos Earn Year -6	0.71	0.46	0.71	0.46	-0.00
⋮					
Pos Earn Year -1	0.72	0.45	0.71	0.46	-0.01

Estimated Propensity Score for the Trimmed Lottery Data

	Variable	est	s.e.	t-stat
presel lin terms	Tix Bought	-0.08	0.46	-0.2
	Years of Schooling	-0.45	0.08	-5.7
	Working Then	3.32	1.95	1.7
	Earnings Year -1	-0.02	0.01	-1.4
add lin terms	Age	-0.05	0.01	-3.7
	Pos Earn Year -5	1.27	0.42	3.0
	Year Won	-4.84	1.53	-3.2
	Earn Year -5	-0.04	0.02	-2.1
2nd order terms	Year Won × Year Won	0.37	0.12	3.2
	Tix Bought × Year Won	0.14	0.06	2.2
	Tix Bought ×Tix Bought	-0.04	0.02	-1.8
	Working Then × Year Won	-0.49	0.30	-1.6

Lottery Data: Estimates of Average Treatment Effects

			Blocking		Matching	
	Pre-sel Cov					
Outcome	p-score	regress	est	s.e.	est	s.e.
Ave Earn Year 1-6	none	all	-5.18	(1.11)	-4.54	(1.36)
	selected	all	-5.74	(1.14)	-4.54	(1.36)

Estimates robust, and credible.

Assessing Unconfoundedness for the Lottery Data: Estimates of Average Treatment Effects for Pseudo Outcomes

	Pre-sel	Cov	Blocking		Matching	
Outcome	p-score	regr	est	s.e.	est	s.e.
Earn Year -1	none	all	-0.20	(0.55)	-0.22	(0.95)
	selec	all	-0.53	(0.58)	-0.22	(0.95)
Ave Earn Year -1,-2	none	all	-1.29	(0.67)	-1.05	(0.98)
	selec	all	-1.16	(0.71)	-0.93	(0.94)

Estimates all close to zero substantively, statistically not significant at 5% level.

Conclusion

Overall results suggest estimates for effect of lottery are credible.

Important to assess and address lack of overlap.

In reasonably balanced samples choice of estimator is less important.

Combining regression & matching, or regression & propensity score blocking is preferred method for robustness properties.