Flexible, optimal matching for observational studies

Ben Hansen University of Michigan

15 June 2006

Optimal matching of two groups Comparing nuclear plants: an illustration

Generalizations of pair matching

The R implementation

・ロト・日本・日本・日本・日本・今々で

650

450

380

440

690

510

390

140

730

Existing site											
	date capacity										
Α	2.3	660									
В	3.0	660									
С	3.4	420									
D	3.4	130									
Е	3.9	650									
F	5.9	430									
G	5.1	420									

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

New site										
	date	capacity								
Н	3.6	290								
I.	2.3	660								
J	3.0	660								
K	2.9	110								
L	3.2	420								
Μ	3.4	60								
Ν	3.3	390								
0	3.6	160								
Р	3.8	390								
Q	3.4	130								
R	3.9	650								
S	3.9	450								
Т	3.4	380								
U	4.5	440								
V	4.2	690								
W	3.8	510								
Х	4.7	390								
Y	5.4	140								
Z	6.1	730								

Existing site				New	site
	date	capacity		date	capacity
_			H	3.6	290
Α	2.3	660 –		2.3	660
В	3.0	660	J	3.0	660
С	3.4	420	K	2.9	110
D	3.4	130	L	3.2	420
_			Μ	3.4	60
Е	3.9	650	Ν	3.3	390
F	5.9	430	0	3.6	160
G	5.1	420	Р	3.8	390
			Q	3.4	130

R

s

Т

U

V

W

Х

Υ

Ζ

3.9

3.9

3.4

4.5

4.2

3.8

4.7

5.4

6.1

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

	Existir	ng site		New	site
	date	capacity	- 	date	capacity
^			- Н	3.6	290
A	2.3	660 -		2.3	660
В	3.0	660 🚽	J	3.0	660
С	3.4	420	К	2.9	110
D	3.4	130	L	3.2	420
			M	3.4	60
Е	3.9	650	N	3.3	390
F	5.9	430	0	3.6	160
G	5.1	420	Р	3.8	390
	5		- Q	3.4	130
			R	3.9	650

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

		uale	capacity	
	Н	3.6	290	
_		2.3	660	
_	— J	3.0	660	
	K	2.9	110	
	<u> </u>	3.2	420	
	Μ	3.4	60	
	<u> </u>	3.3	390	
	0	3.6	160	
	Р	3.8	390	
	Q	3.4	130	
	R	3.9	650	
	S	3.9	450	
	Т	3.4	380	
	U	4.5	440	
	V	4.2	690	
	W	3.8	510	
	Х	4.7	390	
	Y	5.4	140	
	Z	6.1	730	

	Existir	ng site		New	site
	date	capacity	-	date	cap
Α	2.3	660 -	- H		
				2.3	
В	3.0	660 🚽	J	3.0	
С	3.4	420 🔪	ĸ	-	
D	3.4	130	// $-$ r	3.2	
			\backslash	1 3.4	
Е	3.9	650		3.3	
F	5.9	430) 3.6	
G	5.1	420	P	9 3.8	
<u> </u>	0.1	120	- \ c	3.4	

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

Existing site

	— I	2.3	660
	— J	3.0	660
	K	2.9	110
	<u> </u>	3.2	420
	Μ	3.4	60
	<u> </u>	3.3	390
$\overline{}$	0	3.6	160
	P	3.8	390
	Q	3.4	130
$\langle \cdot \rangle$	R	3.9	650
	S	3.9	450
	Ϋ́Τ	3.4	380
	U	4.5	440
	V	4.2	690
	W	3.8	510
	Х	4.7	390
	Y	5.4	140
	Z	6.1	730

New site date cap

capacity

capacity 290

・ロト・西ト・山下・山下・山下・

Existing site Now cito date capacity 660 Α 2.3 В 3.0 660 С 420 3.4 D 130 3.4 Е 650 3.9 F 430 5.9 G 5.1 420

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

	New site								
	date	capacity							
Н	3.6	290							
	2.3	660							
— J	3.0	660							
K	2.9	110							
<u> </u>	3.2	420							
— M	3.4	60							
∕ N	3.3	390							
0	3.6	160							
Р	3.8	390							
Q	3.4	130							
R	3.9	650							
S	3.9	450							
Ϋ́Τ	3.4	380							
U	4.5	440							
V	4.2	690							
W	3.8	510							
Х	4.7	390							
Y	5.4	140							
Z	6.1	730							

date capacity 2.3 Α 660 -В 3.0 660 👡 С 3.4 420 D 3.4 130 ~ Е 3.9 650 F 5.9 430 5.1 G 420

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

	н	3.6	290
	-1	2.3	660
	— J	3.0	660
\sim	К	2.9	110
\sim	~L	3.2	420
\mathcal{T}	— M	3.4	60
	<u> </u>	3.3	390
////	0	3.6	160
	P	3.8	390
	Q	3.4	130
	R	3.9	650
	S	3.9	450
oy a	Ϋ́	3.4	380
im.	U	4.5	440
	V	4.2	690
	W	3.8	510
n, in	Х	4.7	390
,	Y	5.4	140
et ca-	Z	6.1	730
MWe			

	Existing site				New	site
	date	capacity	-		date	capacity
Δ	2.3	660		Н	3.6	290
A				— I	2.3	660
В	3.0	660 👡		— J	3.0	660
С	3.4	420 🔪		Κ	2.9	110
D	3.4	130 –	V / -	∼L	3.2	420
				— M	3.4	60
E	3.9	650 🔍	\mathcal{N}	<u> </u>	3.3	390
F	5.9	430 🔪		0	3.6	160
G	5.1	420	$\forall \mid \not \mid \not \mid /$	P	3.8	390
	•		· // ///	∕Q	3.4	130
				R	3.9	650
				S	3.9	450
Example:	1:2	matching	by a \	`Τ	3.4	380
			· · · · · ·	<u> </u>		4.40

traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

		0.2	720	
_	— M	3.4	60	
	N	3.3	390	
	0	3.6	160	
	P	3.8	390	
	\ Q	3.4	130	
$\left[\right]$	R	3.9	650	
/	S	3.9	450	
	`⊤	3.4	380	
	∕ U	4.5	440	
	V	4.2	690	
/	W	3.8	510	
	∖x	4.7	390	
	Υ	5.4	140	
	Ζ	6.1	730	

◆□▶ ◆□▶ ◆ ≧▶ ◆ ≧▶ → ≧ → りへぐ

New and refurbished nuclear plants: discrepancies in capacity and year of construction

-																				
	Exist-		New sites																	
	ing	Н	Ι	J	Κ	L	Μ	Ν	0	Ρ	Q	R	S	Т	U	V	W	Х	Y	Ζ
	А	28	0	3	22	14	30	17	28	26	28	20	22	23	26	21	18	34	40	28
	В	24	3	0	22	10	27	14	26	24	24	16	19	20	23	18	16	31	37	25
	С	10	18	14	18	4	12	6	11	9	10	14	12	6	14	22	10	16	22	28
	D	7	28	24	8	14	2	10	6	12	0	24	22	4	24	32	20	18	16	38
	Е	17	20	16	32	18	26	20	18	12	24	0	2	20	6	8	4	14	20	14
	F	20	31	28	35	20	29	22	20	14	26	12	9	22	5	15	12	9	11	12
	G	14	32	29	30	18	24	17	16	10	22	12	10	17	6	16	14	4	8	17

	Existir	ng site		New	site
	date	capacity		date	capacity
_			H	3.6	290
А	2.3	660 -		2.3	660
В	3.0	660 🚽	J	3.0	660
С	3.4	420 🔪	К	2.9	11(
D	3.4	130 –	\mathbb{P} / \mathbb{P}	3.2	420
_			M	3.4	60
Е	3.9	650 🔍	N / N	3.3	390
F	5.9	430	0 ///	3.6	160
G	5.1	420	N N P	3.8	390
	5			3.4	130
				20	6E(

Example: 1:2 matching by a traditional, greedy algorithm.

"date" is date of construction, in years after 1965; "capacity" is net capacity of the power plant, in MWe above 400.

•	0.0	
∕ Q	3.4	130
∕ R	3.9	650
S	3.9	450
Т	3.4	380
∕ U	4.5	440
V	4.2	690
W	3.8	510
X	4.7	390
ŶΥ	5.4	140
Ζ	6.1	730

3.9

3.4

4.5

4.2

3.8

4.7

5.4

6.1

v

W

Y

١Z

450

380

440

690

510

390

140

730

・ 日本 4 日本 4 日本 4 日本 - うへの

	Existin	ng site		New	
	date	capacity		date	capacity
^			Н	3.6	290
Α	2.3	660		2.3	660
В	3.0	660 🚽	J	3.0	660
С	3.4	420 🥿	К	2.9	110
D	3.4	130 🚽		3.2	420
			M	3.4	60
Е	3.9	650 👡	N	3.3	390
F	5.9	430 💊	0	3.6	160
G	5.1	420	P	3.8	390
	-	-	Q A	3.4	130
			R	3.9	650

Optimal vs. Greedy matching

By evaluating potential matches all together rather than sequentially, optimal matching (blue lines) reduces the sum of distances from 82 to 63.

Introducing restrictions on who can be matched to whom: calipers

In the nuclear plants example, suppose we choose to insist upon a caliper of three years in the date of construction. This would forbid five potential matches, indicated below in red.

Exist-		New sites																	
ing	Н	I	J	Κ	L	Μ	Ν	0	Ρ	Q	R	S	Т	U	V	W	Х	Y	Ζ
A	28	0	3	22	14	30	17	28	26	28	20	22	23	26	21	18	34	40	28
В	24	3	0	22	10	27	14	26	24	24	16	19	20	23	18	16	31	37	25
С	10	18	14	18	4	12	6	11	9	10	14	12	6	14	22	10	16	22	28
D	7	28	24	8	14	2	10	6	12	0	24	22	4	24	32	20	18	16	38
E	17	20	16	32	18	26	20	18	12	24	0	2	20	6	8	4	14	20	14
F	20	31	28	35	20	29	22	20	14	26	12	9	22	5	15	12	9	11	12
G	14	32	29	30	18	24	17	16	10	22	12	10	17	6	16	14	4	8	17

Introducing restrictions on who can be matched to whom: calipers

With <code>optmatch</code>, matches are forbidden by placing ∞ 's in the distance matrix.

Exist-		New sites																	
ing	Н	Ι	J	Κ	L	Μ	Ν	0	Ρ	Q	R	S	Т	U	V	W	Х	Υ	Ζ
A	28	0	3	22	14	30	17	28	26	28	20	22	23	26	21	18	34	Inf	Inf
В	24	3	0	22	10	27	14	26	24	24	16	19	20	23	18	16	31	37	Inf
С	10	18	14	18	4	12	6	11	9	10	14	12	6	14	22	10	16	22	28
D	7	28	24	8	14	2	10	6	12	0	24	22	4	24	32	20	18	16	38
E	17	20	16	32	18	26	20	18	12	24	0	2	20	6	8	4	14	20	14
F	20	Inf	28	Inf	20	29	22	20	14	26	12	9	22	5	15	12	9	11	12
G	14	32	29	30	18	24	17	16	10	22	12	10	17	6	16	14	4	8	17

・ロト・日本・日本・日本・日本・日本

Outline

Optimal matching of two groups Comparing nuclear plants: an illustration

Generalizations of pair matching

The R implementation

・ロト・西ト・ヨト・ヨー うくぐ

Example # 2: Gender equity study for research scientists¹

Women and men scientists are to be matched on grant funding.

V	/omen	Men					
Subject	log ₁₀ (Grant)	Subject	log ₁₀ (Grant)				
A	5.7	V	5.5				
В	4.0	W	5.3				
С	3.4	Х	4.9				
D	3.1	Y	4.9				
		Z	3.9				

Full Matching² the Gender Equity Sample

V	/omen	Men					
Subject	log ₁₀ (Grant)	Subject	log ₁₀ (Grant)				
А	5.7	— V	5.5				
В	4.0	W	5.3				
С	3.4	<u> </u>	4.9				
D	3.1	Y	4.9				
		Z	3.9				

 Similar to matching with replacement, but creates disjoint matched sets — better for tests & CIs.

- In contrast to pair matching, it finds matches for everyone with a suitable counterpart.
- In contrast to multiple controls matching, it doesn't force poor matches.
- ▶ In optmatch, can be combined with structural restrictions.

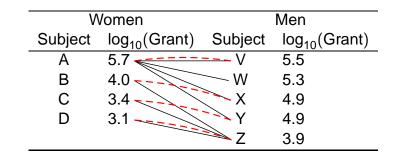
²(Rosenbaum, 1991; Hansen and Klopfer, 2005) א 🗆 א א 🖻 א א 🖹 א א 🖹 א א 🖹 א א צ

Full Matching² the Gender Equity Sample

V	/omen		Men
Subject	log ₁₀ (Grant)	Subject	log ₁₀ (Grant)
Α	5.7	<u> </u>	5.5
В	4.0	W	5.3
С	3.4	X	4.9
D	3.1	Y	4.9
		Z	3.9

- Similar to matching with replacement, but creates disjoint matched sets — better for tests & CIs.
- In contrast to pair matching, it finds matches for everyone with a suitable counterpart.
- In contrast to multiple controls matching, it doesn't force poor matches.
- ▶ In optmatch, can be combined with structural restrictions.

Full Matching² the Gender Equity Sample



- Similar to matching with replacement, but creates disjoint matched sets — better for tests & CIs.
- In contrast to pair matching, it finds matches for everyone with a suitable counterpart.
- In contrast to multiple controls matching, it doesn't force poor matches.
- ▶ In optmatch, can be combined with structural restrictions.

Full Matching² the Gender Equity Sample

V	Vomen		Men
Subject	log ₁₀ (Grant)	Subject	log ₁₀ (Grant)
Α	5.7	V	5.5
В	4.0	W	5.3
С	3.4	X	4.9
D	3.1	Y	4.9
		Z	3.9

- Similar to matching with replacement, but creates disjoint matched sets — better for tests & CIs.
- In contrast to pair matching, it finds matches for everyone with a suitable counterpart.
- In contrast to multiple controls matching, it doesn't force poor matches.
- ► In optmatch, can be combined with structural restrictions.

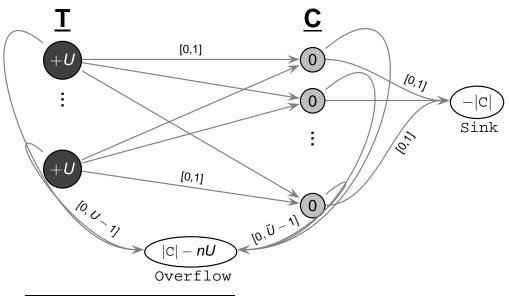
Outline

Optimal matching of two groups Comparing nuclear plants: an illustration

Generalizations of pair matching

The R implementation

Under the hood Full matching via network flows³



³(Hansen and Klopfer, 2005, Fig. 2). Time complexity of the algorithm is $O(n^3 \log(n \max(\text{dist})))$.

ロ・・日・・日・・日・ りゃぐ

Arguments to fullmatch()

- distance The argument demanding most attention from the user, b/c it defines "good" matches and because very large distance matrices can tax R's memory limits. A new helper function, makedist(), eases both of these efforts
- min.controls, max.controls In propensity matching, can be important for efficiency — see Hansen (2004), Augursky and Kluve (2004).
- omit.fraction for use in matched sampling (as opposed to matching all or most of a sample). Not needed for getting rid of subjects without suitable potential matches. If you're not specifically out to reduce the size of the control group, can be ignored.

Arguments to fullmatch()

- distance The argument demanding most attention from the user, b/c it defines "good" matches and because very large distance matrices can tax R's memory limits. A new helper function, makedist(), eases both of these efforts.
- min.controls, max.controls In propensity matching, can be important for efficiency — see Hansen (2004), Augursky and Kluve (2004).
- omit.fraction for use in matched sampling (as opposed to matching all or most of a sample). Not needed for getting rid of subjects without suitable potential matches. If you're not specifically out to reduce the size of the control group, can be ignored.

Arguments to fullmatch()

Summary

- distance The argument demanding most attention from the user, b/c it defines "good" matches and because very large distance matrices can tax R's memory limits. A new helper function, makedist(), eases both of these efforts.
- min.controls, max.controls In propensity matching, can be important for efficiency — see Hansen (2004), Augursky and Kluve (2004).
- omit.fraction for use in matched sampling (as opposed to matching all or most of a sample). Not needed for getting rid of subjects without suitable potential matches. If you're not specifically out to reduce the size of the control group, can be ignored.

With optmatch, R offers the most comprehensive optimal matching implementation for statistics.

- fullmatch() solves optimally such traditional problems as matched sampling, pair matching, and matching with k controls.
- fullmatch() can also solve matching problems flexibly, and far more effectively, by way of full matching, with or without structural restrictions (Hansen and Klopfer, 2005; Hansen, 2004).
- The effort required to code optimal and full matching algorithms seems to have dissuaded their widespread use. Now that I've made that effort, perhaps this situation can change! :)

▲□▶▲圖▶▲≣▶▲≣▶ ≣ の�?

Summary

- With optmatch, R offers the most comprehensive optimal matching implementation for statistics.
- fullmatch() solves optimally such traditional problems as matched sampling, pair matching, and matching with k controls.
- fullmatch() can also solve matching problems flexibly, and far more effectively, by way of full matching, with or without structural restrictions (Hansen and Klopfer, 2005; Hansen, 2004).
- The effort required to code optimal and full matching algorithms seems to have dissuaded their widespread use. Now that I've made that effort, perhaps this situation can change! :)

Summary

- With optmatch, R offers the most comprehensive optimal matching implementation for statistics.
- fullmatch() solves optimally such traditional problems as matched sampling, pair matching, and matching with k controls.
- fullmatch() can also solve matching problems flexibly, and far more effectively, by way of full matching, with or without structural restrictions (Hansen and Klopfer, 2005; Hansen, 2004).
- The effort required to code optimal and full matching algorithms seems to have dissuaded their widespread use. Now that I've made that effort, perhaps this situation can change! :)

◆□▶ ◆□▶ ◆ □ ▶ ◆ □ ▶ ● □ ● ○ ○ ○ ○

Summary

- With optmatch, R offers the most comprehensive optimal matching implementation for statistics.
- fullmatch() solves optimally such traditional problems as matched sampling, pair matching, and matching with k controls.
- fullmatch() can also solve matching problems flexibly, and far more effectively, by way of full matching, with or without structural restrictions (Hansen and Klopfer, 2005; Hansen, 2004).
- The effort required to code optimal and full matching algorithms seems to have dissuaded their widespread use. Now that I've made that effort, perhaps this situation can change! :)

Example with propensity scores and stratification prior to matching

```
>nuclear$pscore <- glm(pr~.-cost,</pre>
```

+ family=binomial(),data=nuclear)\$linear.predictors

◆□▶ ◆□▶ ◆□▶ ◆□▶ □ ● のくぐ

```
> pscorediffs <- function(trtvar,data) {
+ pscr <- data[names(trtvar), 'pscore']
+ abs(outer(pscr[trtvar],pscr[!trtvar], '-'))
</pre>
```

```
+ }
```

```
> psd2 <- makedist(pr~pt, nuclear, pscorediffs)</pre>
```

```
> fullmatch(psd2)
```

```
> fullmatch(psd2, min.controls=1, max.controls=3)
> fullmatch(psd2, min=1, max=c('0'=3, '1'=2))
```

Jake Bowers' and my RItools package provides diagnostics...

Example with propensity scores and stratification prior to matching

```
>nuclear$pscore <- glm(pr~.-cost,
+ family=binomial(),data=nuclear)$linear.predictors
> pscorediffs <- function(trtvar,data) {
+ pscr <- data[names(trtvar), 'pscore']
+ abs(outer(pscr[trtvar],pscr[!trtvar], '-'))
+ }
> psd2 <- makedist(pr~pt, nuclear, pscorediffs)
> fullmatch(psd2)
> fullmatch(psd2, min.controls=1, max.controls=3)
> fullmatch(psd2, min=1, max=c('0'=3, '1'=2))
Jake Bowers' and my RItools package provides
diagnostics...
```

Example with propensity scores and stratification prior to matching

```
>nuclear$pscore <- glm(pr~.-cost,
+ family=binomial(),data=nuclear)$linear.predictors
> pscorediffs <- function(trtvar,data) {
+ pscr <- data[names(trtvar), 'pscore']
+ abs(outer(pscr[trtvar],pscr[!trtvar], '-'))
+ }
> psd2 <- makedist(pr~pt, nuclear, pscorediffs)
> fullmatch(psd2)
```

```
> fullmatch(psd2, min.controls=1, max.controls=3
> fullmatch(psd2, min=1, max=c('0'=3, '1'=2))
```

Jake Bowers' and my RItools package provides diagnostics...

Example with propensity scores and stratification prior to matching

>nuclear\$pscore <- glm(pr~.-cost,</pre>

+ family=binomial(),data=nuclear)\$linear.predictors

```
> pscorediffs <- function(trtvar,data) {
+ pscr <- data[names(trtvar), 'pscore']
+ abs(outer(pscr[trtvar],pscr[!trtvar], '-'))
+ }</pre>
```

> psd2 <- makedist(pr~pt, nuclear, pscorediffs)</pre>

```
> fullmatch(psd2)
```

- > fullmatch(psd2, min.controls=1, max.controls=3)
- > fullmatch(psd2, min=1, max=c('0'=3, '1'=2))

Jake Bowers' and my RItools package provides diagnostics...

Modes of estimation for treatment effects

Preferred		Туре о	f outcome	
mode of infer-	Categorical		Continuous	
ence	-			
Randomization	Agresti	(2002),	Rosenbaum	(2002c),
	Categorical	Data	Observational	Studies;
	Analysis;	Rosenbaum	Rosenbaum (200)2b), "Cov-
	(2002a),	"Atributing	ariance adjustme	nt"
	effects to trea	atment "		
Conditional ^a	Agresti (2002	2); Cox and	ordinary OLS ^b is	s fine; see
	Snell (1989),	Analysis of	also Rubin (197	9), "Using
	binary data		multivariate match	ned"
Bayes, esp.	Agresti (2002	2)	Smith (1997),	"Matching
hierarchical			with multiple c	ontrols";
linear models			Raudenbush a	nd Bryk
С			(2002), Hierarch	-
			models	

^aUses a fixed effect for each matched set.

^bi.e., OLS with a fixed effect for each matched set plus treatment effect(s)

^cUses a random effect for each matched set.

Augursky, B. and Kluve, J. (2004), "Assessing the performance of matching algorithms when selection into treatment is strong," Tech. rep., RWI-Essen.

Cox, D. R. and Snell, E. J. (1989), Analysis of binary data, Chapman & Hall Ltd.

Hansen, B. B. (2004), "Full matching in an observational study of coaching for the SAT," *Journal of the American Statistical Association*, 99, 609–618.

— (2005), OPTMATCH, an add-on package for R.

- Hansen, B. B. and Klopfer, S. O. (2005), "Optimal full matching and related designs via network flows," Tech. Rep. 416, Statistics Department, University of Michigan.
- Raudenbush, S. W. and Bryk, A. S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications Inc.
- Rosenbaum, P. R. (1991), "A Characterization of Optimal Designs for Observational Studies," *Journal of the Royal Statistical Society*, 53, 597–610.
- (2002a), "Attributing effects to treatment in matched observational studies," *Journal* of the American Statistical Association, 97, 183–192.
- (2002b), "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, 17, 286–327.
- (2002c), Observational Studies, Springer-Verlag, 2nd ed.
- Rubin, D. B. (1979), "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, 318–328.
- Smith, H. (1997), "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies," *Sociological Methodology*, 27, 325–353.