Sensitivity of matching estimators to unconfoundedness.

An application to the effect of temporary work on future employment.

Andrea Ichino
EUROPEAN UNIVERSITY INSTITUTE

Fabrizia Mealli
UNIVERSITY OF FLORENCE

Tommaso Nannicini
EUROPEAN UNIVERSITY INSTITUTE

October 20, 2004

Motivation and content

Increasing recent use of matching estimators in evaluation studies for which a convincing source of exogenous random variation of assignment to treatment does not exist.

Matching estimators are now easy to use and perhaps too many users adopt them without really checking that the conditions for their application are satisfied.

We propose a sensitivity analysis for matching estimators aimed at assessing their robustness to failures of the unconfoundedness assumption in a specific evaluation problem.

We describe how this sensitivity analysis can be applied using data from two Italian regions collected to evaluate the effect of a temporary work experience on future employment probabilities.

Intuition for the sensitivity analysis

Suppose that unconfoundedness is not satisfied given observables but would be satisfied if we could observe an additional binary variable.

This binary variable can be simulated in the data and used as an additional matching factor in combination with the preferred matching estimator.

A comparison of the estimates obtained with and without matching on this simulated binary variable tells us to what extent the estimator is robust to this specific source of failure of the unconfoundedness assumption.

Moreover, the simulated values of the binary variable can be constructed to capture different hypotheses on the nature of potential counfounding factors.

This is convenient and instructive particularly when the simulated confounder is designed to mimic the distribution of important observed covariates.

Results of the application to Italian data

A TWA assignment increases the probability of finding a permanent job by 17 percentage points in Tuscany and by 8 percentage points in Sicily.

Note that the observed probabilities of finding a permanent job in the treated group are respectively 31% and 23% in the two regions.

The sensitivity analysis complements in an important way these results.

In Tuscany, but not in Sicily, the ATT is robust to deviations from unconfoundedness caused by binary confounders distributed similarly to gender, education, marital status and previous employment history.

Only when the unobservable confounding factor is calibrated so as to have a selection effect much larger than the one associated to observed covariates, the ATT for Tuscany is estimated to be close to zero.

A standard evaluation problem

Our goal is to estimate

$$E(Y_1 - Y_0|T = 1). \tag{1}$$

where:

$$V = f(Z, \epsilon_v) \quad T = I(V > 0) \tag{2}$$

$$Y_1 = g_1(X, \epsilon_y) \tag{3}$$

$$Y_0 = g_0(X, \epsilon_y). \tag{4}$$

Let $W = (X, Z)$. Unconfoundedness implies:

$$(Y_1, Y_0) \perp T|W \tag{5}$$

$$0 < Pr(T = 1|W) < 1. \tag{6}$$

Under unconfoundedness, the ATT is given by

$$E(Y_1 - Y_0|T = 1) = E(E(Y_1 - Y_0|T = 1, W)) = \tag{7}$$
$$= E(E(Y_1|T = 1, W) - E(Y_0|T = 0, W)|T = 1).$$

ATT estimation based on Propensity Score matching

The Propensity Score is the individual probability of receiving the treatment given the observed covariates: $p(W) = P(T = 1|W)$.

Under unconfoundedness, T is independent of W given $p(W)$, and Y_0 and Y_1 are independent of T given $p(W)$.

If $p(W)$ is known, the ATT can be estimated as follows:

$$\begin{aligned} \tau &\equiv E(Y_1 - Y_0|T = 1) = \\ &= E(E(Y_1 - Y_0|p(W), T = 1)) = \\ &= E(E(Y_1|p(W), T = 1) - E(Y_0|p(W), T = 0)|T = 1) \end{aligned} \tag{8}$$

where the outer expectation is over the distribution of $(p(W)|T = 1)$.

The nearest neighbor matching estimator

Denote by $C(h)$ the set of control units k matched to the treated unit h with an estimated value of the Propensity Score of p_h.

$$C(h) = \{k \mid k = \text{arg}\min_k \parallel p_h - p_k \parallel\}, \tag{9}$$

LetY_h^D and Y_k^C be the observed outcomes of the treated and control units, respectively. The Nearest Neighbor Propensity Score matching estimator is:

$$\begin{aligned}
\tau^M &= \frac{1}{N^D} \sum_{h \in T} \left[Y_h^D - \sum_{k \in C(h)} \omega_{hk} Y_k^C \right] \qquad (10) \\
&= \frac{1}{N^D} \left[\sum_{h \in D} Y_h^D - \sum_{h \in D} \sum_{k \in C(h)} \omega_{hk} Y_k^C \right] \\
&= \frac{1}{N^D} \sum_{h \in D} Y_h^D - \frac{1}{N^D} \sum_{k \in C} \omega_k Y_k^C
\end{aligned}$$

where the weights ω_k are defined by $\omega_k = \Sigma_h \omega_{hk}$. Standard errors can be derived both analytically and by bootstrapping.

Sensitivity analysis

Identification of the ATT relies crucially on the validity of the unconfoundedness assumption.

Parametric selection models are formally identified thanks to other types of non-necessarily preferable hypotheses.

If such hypotheses cannot be accepted, assuming unconfoundedness is an alternative option, but it requires to check the robustness of results to departures from this assumption.

This is the spirit of the sensitivity analysis proposed in our paper.

Central assumption of the sensitivity analysis

Assignment to treatment is not unconfounded given the set of observable variables W, i.e.,

$$Pr(T = 1|Y_0, Y_1, W) \neq Pr(T = 1|W) \tag{11}$$

but unconfoundedness holds given W and an unobserved binary covariate U, that is

$$Pr(T = 1|Y_0, Y_1, W, U) = Pr(T = 1|W, U). \tag{12}$$

A similar assumption in:

- Rosenbaum and Rubin (1983b)
- Imbens (2003)
- Rosenbaum (1987)

Parametrization ot the distribution of U

For simplicity, consider binary potential outcomes $Y_0, Y_1 \in \{0, 1\}$, and denote the observed outcome as $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$

The parameters

$$Pr(U = 1|T = i, Y = j, W) = Pr(U = 1|T = i, Y = j) \equiv p_{ij}, \qquad (13)$$

with $i, j = \{0, 1\}$ give the probability that $U = 1$ in each of the four groups defined by the treatment status and the outcome value.

Note that:

- U is considered as any other observed characteristics in W.
- The distribution of U given T and Y does not vary with W.
- But, in the population, U may be correlated with other variables.
- If U were fully correlated with one or more of the variables in W, it would not represent a problem.

Interpretation of the parametrization

Given p_{ij} and the observed probability $Pr(Y = i|T = j)$, we can compute:

$$p_{i.} \equiv Pr(U = 1|T = i) = \sum_{j=0}^{1} p_{ij} \cdot Pr(Y = j|T = i); \tag{14}$$

By setting the p_{ij} appropriately we can generate situations in which the fraction of subjects with $U = 1$ is greater

- among the treated ($p_{1.} > p_{0.}$)
- or among the controls ($p_{1.} < p_{0.}$).

Note that

- $p_{i1} > p_{i0} \Rightarrow Pr(Y = 1|T = i, U = 1) > Pr(Y = 1|T = i, U = 0)$
  within each $T = i$, the confounder U has a positive effect on the outcome.
- $p_{1.} > p_{0.} \Rightarrow Pr(T = 1|U = 1) > Pr(T = 1|U = 0)$
  U has a positive effect on the selection into treatment.

Simulation

Given arbitrary (but meaningful) values of the parameters p_{ij}, we attribute a value of U to each subject, according to her treatment status and outcome.

We then include U in the set of matching variables used to estimate the Propensity Score and to compute the Nearest Neighbor estimate of the ATT.

For each set of values of the sensitivity parameters, we repeat the matching estimation (e.g., 10,000 times), in order to obtain

- an estimate of the ATT, which is an average of the ATT's over the distribution of the unobserved U,
- a corresponding simulated standard error, representing the variability of the estimated ATT with respect to the distribution of U.

These ATT estimates are robust to failures of the unconfoundedness assumption implied by each specific configuration of the p_{ij}.

Advantages of this approach

The parameters p_{ij} and $p_{i.}$ can be chosen to make the distribution of the U similar to the empirical distribution of observable binary covariates.

We can search for the existence of a set of parameters p_{ij} and $p_{i.}$ such that if U were observed the estimated ATT would be zero, and then assess the plausibility of this configuration of parameters.

Like Rosenbaum (1987) and unlike Rosenbaum and Rubin (1983b) and Imbens (2003), we do not need to specify a parametric model for $T, Y(0), Y(1)$.

Unlike Rosenbaum (1987) we derive point estimates of the ATT under different possible scenarios of deviation from unconfoundedness.

Our simulation-based sensitivity analysis works irrespective of the algorithm used to match observations.

Generalization to multivalued or continuous outcomes

With multi-valued or continuous outcomes the same method can be applied by defining:

$$p_{ij} \equiv Pr(U = 1 | T = i, I(Y > y^*)), \tag{15}$$

where I is the indicator function and y^* is a chosen typical value, e.g. the median, of the distribution of Y.

Alternatively, one can specify the conditional distribution of U given T and Y.

A possible choice is, for example:

$$Pr(U = 1 | T = i, Y = y) = \frac{exp(\alpha_0 + \alpha_1 i + \beta_i y)}{1 + exp(\alpha_0 + \alpha_1 i + \beta_i y)}, \tag{16}$$

where α_0, α_1, β_0, and β_1 become the sensitivity parameters used to calibrate the distribution of U.

Relationship with other methods

Our method makes a point-identifying assumption and examines how results change as this assumption is weakened in specific ways.

A complement method is to identify "bounds" for the ATT and see if they can be narrowed with plausible assumptions.

The two approaches are closely linked.

Note that in the $ATT = E(Y_1|T = 1) - E(Y_0|T = 1)$ the second term cannot be observed, but bounds can be defined by substituting its smallest and biggest values:

$$ATT \in (Pr(Y_1 = 1|T = 1) - 1 \; ; \; Pr(Y_1 = 1|T = 1)). \tag{17}$$

Using our approach we can then see what are the assumptions on U which are needed for the ATT to reach the bounds.

The distribution of U and the ATT bounds

The lower bound is reached when:

- all the treated are skilled: i.e. $Pr(U = 1|T = 1) = 1$;
- all the skilled controls are employed: i.e, $Pr(Y_0 = 1|T = 0, U = 1) = 1$.

The upper bound is instead reached when:

- all the treated are skilled: i.e. $Pr(U = 1|T = 1) = 1$;
- all the skilled controls are unemployed: i.e, $Pr(Y_0 = 1|T = 0, U = 1) = 0$.

Both set of conditions, which can be translated in assumptions on the p_{ij}, appear implausible.

Different and more “plausible” assumptions on the p_{ij} can be used to define more “plausible” regions within the extreme bounds defined above.

An Application to Temporary Work Agencies in Italy

The growing share of temporary employment in many European countries raises concerns over the risk of an undesirable labor market "segmentation".

An intertemporal balance between flexibility and security may be possible if temporary jobs are an effective "springboard" toward permanent employment, as opposed to a "trap" of endless precariousness.

Theoretical arguments can be proposed in favour or against, but whether TWA employment is a springboard or a trap is ultimately an empirical question.

In Italy this question originated a very harsh debate, after the "Treu law" (n. 196/1997) which legalized TWA, but no serious evaluation study.

The data

In 2001, Sicily and Tuscany were among the remaining Italian regions with incomplete penetration of agencies. We selected the provinces:

- with Agency: Livorno, Pisa, Lucca, Catania, Palermo;
- without Agency: Grosseto, Massa, Messina, Trapani.

The presence of provinces with and without agency facilitates the use of the distance from an agency as a matching variable to proxy for local labor demand.

Treated Sample:

all residents in the 9 provinces who were on a TWA assignment through *Manpower* during the first semester of 2001;

Control Sample:

residents in the 9 provinces, aged 18-40, who belonged to the labor force but were not permanent employees as of January 1, 2001.

We have to take care of choice-based sampling and geographical stratification.

Data collection

Identical sets of questions for treated and controls:

1. demographic characteristics;
2. family background;
3. educational attainments;
4. work experience before 2001;
5. job characteristics during the first semester of 2001;
6. work experience from July 2001 to the end of 2002;
7. job characteristics at the end of 2002.

The final dataset contains 2030 individuals:

- 511 treated (temporary workers);
- 1519 controls (other "atypical" or unemployed workers).

Table 1: **Characteristics of the whole sample**

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Age	26.5	27.5	29.1	26.8	27.8	30.0
Male	0.56	0.41	0.29	0.67	0.57	0.29
Single	0.90	0.87	0.66	0.83	0.81	0.49
Children	0.09	0.16	0.45	0.20	0.23	0.86
Father school	9.3	9.2	8.6	8.7	9.2	7.6
Father blue	0.33	0.39	0.43	0.30	0.31	0.39
Father active	0.53	0.46	0.37	0.46	0.45	0.29
School	12.5	12.7	12.3	12.0	12.4	11.6
Grade	75.9	77.1	76.9	74.7	74.6	76.5
Training	0.32	0.30	0.28	0.42	0.42	0.34
Unemployment	0.38	0. 42	0.48	0.42	0.44	0.62
Employed 2000	0.35	0.36	0.42	0.34	0.35	0.30
Unemployed 2000	0.52	0.53	0.52	0.60	0.60	0.67
Out l.force 2000	0.13	0.10	0.05	0.06	0.05	0.03
Employed 2001	1.00	0.36	0.36	1.00	0.30	0.25
Unemployed 2001	0.00	0.64	0.64	0.00	0.70	0.75
Permanent 2002	0.31	0.16	0.17	0.23	0.14	0.13
Atypical 2002	0.42	0.36	0.31	0.39	0.17	0.18
Unemployed 2002	0.16	0.44	0.45	0.30	0.59	0.63
Out l.force 2002	0.11	0.04	0.07	0.07	0.09	0.07
N.individuals	281	135	628	230	128	891

Table 2: **Characteristics of the employed before the treatment**

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Permanent	0.16	0.22	0.26	0.14	0.16	0.36
Atypical	0.84	0.78	0.74	0.86	0.84	0.64
Blue-collar	0.62	0.59	0.39	0.44	0.24	0.22
White-collar	0.36	0.41	0.54	0.54	0.71	0.67
Self-empl.	0.02	0.00	0.07	0.01	0.04	0.10
Manufact.	0.53	0.41	0.23	0.39	0.20	0.15
Service	0.39	0.45	0.67	0.49	0.67	0.70
Other	0.08	0.14	0.11	0.11	0.13	0.15
Wage	5.2	5.6	6.8	5.6	7.6	7.0
Hours	38.0	36.3	33.3	34.5	32.1	31.1
N.individuals	98	49	266	79	45	267

Table 3: **Simulation-based sensitivity analysis for Tuscany**

	Fraction $U=1$ by treatment/outcome				Fraction $U=1$ by treatment		ATT	$\hat{s.e.}$	$s.e.$	Π
	p_{11}	p_{10}	p_{01}	p_{00}	$p_{1.}$	$p_{0.}$				
No confounder							0.17	-	0.06	-
U like:										
Male	0.54	0.56	0.32	0.28	0.56	0.29	0.14	0.05	0.06	0.70
Single	0.86	0.92	0.76	0.64	0.90	0.66	0.14	0.04	0.06	0.68
High school	0.75	0.74	0.69	0.71	0.74	0.71	0.16	0.03	0.06	0.92
College	0.14	0.12	0.13	0.19	0.12	0.18	0.15	0.03	0.06	0.89
Prev.employed	0.40	0.33	0.50	0.41	0.35	0.42	0.16	0.03	0.06	0.91
Father educ.	0.34	0.31	0.32	0.27	0.32	0.28	0.16	0.03	0.06	0.89
U kills ATT:										
Case I	0.94	0.12	0.83	0.19	0.38	0.30	0.07	0.06	0.06	0.28
Case II	0.86	0.46	0.76	0.32	0.59	0.39	0.08	0.05	0.06	0.27
Case III	0.90	0.50	0.40	0.10	0.63	0.15	-0.01	0.08	0.08	0.00
Other U:										
Case IV	0.90	0.20	0.20	0.20	0.42	0.20	0.15	0.04	0.06	0.80
Case V	0.30	0.20	0.90	0.80	0.23	0.82	0.20	0.06	0.08	0.85
Case VI	0.90	0.50	0.80	0.60	0.63	0.63	0.16	0.03	0.06	0.93

Table 4: **Simulation-based sensitivity analysis for Sicily**

	Fraction $U=1$ by treatment/outcome				Fraction $U=1$ by treatment		ATT	$\hat{s.e.}$	$s.e.$	Π
	p_{11}	p_{10}	p_{01}	p_{00}	$p_{1.}$	$p_{0.}$				
No confounder							0.08	-	0.05	-
U like:										
Male	0.87	0.61	0.45	0.27	0.67	0.29	0.01	0.04	0.06	0.13
Single	0.87	0.82	0.59	0.47	0.83	0.49	0.03	0.04	0.05	0.14
High school	0.74	0.72	0.76	0.63	0.72	0.64	0.06	0.03	0.05	0.22
College	0.11	0.06	0.16	0.12	0.07	0.13	0.06	0.03	0.05	0.26
Prev.employed	0.98	0.30	0.62	0.25	0.46	0.30	0.02	0.04	0.05	0.14
Father educ.	0.24	0.25	0.27	0.18	0.25	0.19	0.06	0.03	0.05	0.21
U kills ATT:										
Case I	0.91	0.06	0.16	0.12	0.26	0.13	0.04	0.03	0.05	0.16
Case II	0.87	0.82	0.59	0.17	0.83	0.22	-0.13	0.06	0.07	0.12
Case III	0.90	0.50	0.40	0.10	0.59	0.14	-0.08	0.06	0.07	0.12
Other U:										
Case IV	0.90	0.20	0.20	0.20	0.36	0.20	0.05	0.03	0.05	0.17
Case V	0.30	0.20	0.90	0.80	0.22	0.81	0.08	0.05	0.07	0.35
Case VI	0.90	0.55	0.80	0.60	0.63	0.63	0.07	0.03	0.05	0.30

Conclusions

We propose a sensitivity analysis for matching estimators to assess whether the unconfoundedness assumption is plausible in a specific evaluation problem.

Our approach is easy to implement and has several advantages.

It is, however, still in a development phase in terms of:

- presentation of results;
- characterization of its relationship with respect to other methods to assess the sensitivity of ATT estimates with respect to failures of unconfoundedness.

An application based on Italian data, shows that this sensitivity analysis is helpful in identifying the cases in which the estimated effect should be considered not robust to failures of unconfoundedness.