

SIMULATION-BASED SENSITIVITY ANALYSIS FOR MATCHING ESTIMATORS

TOMMASO NANNICINI
UNIVERSIDAD CARLOS III DE MADRID

UK Stata Users Group Meeting
London, September 10, 2007

CONTENT

Presentation of a Stata program - `sensatt` - that implements the sensitivity analysis for propensity-score matching estimators proposed by:

- Ichino A., F. Mealli, and T. Nannicini (2007), “From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity?”, *Journal of Applied Econometrics*, forthcoming.

In this paper, we present an econometric tool that builds on Rosenbaum and Rubin (1983a) and Rosenbaum (1987a), and aims at assessing the robustness of average treatment effects estimated with matching methods.

MOTIVATION

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

In Stata, one can use the following routine programs:

- `nnmatch`
- `psmatch2`
- `attnd`, `attnw`, `attr`, `attk`

Matching estimators are now easy to use and perhaps too many users adopt them without checking **both** the conditions for their application and the sensitivity of the results to possible deviations from these conditions.

We propose a simulation-based sensitivity analysis for matching estimators, aimed at assessing their robustness to specific failures of the **Conditional Independence Assumption (CIA)** in the evaluation problem at hand.

INTUITION FOR THE SENSITIVITY ANALYSIS

Matching estimators rely crucially on the CIA to identify treatment effects. Suppose that this condition is not satisfied given observables, but would be satisfied if we could observe another 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 CIA.

The simulated values of the binary variable can be constructed to capture different hypotheses on the nature of potential confounding factors.

FRAMEWORK

Consider Rubin's (1974) potential-outcome framework for causal inference. Our goal is to estimate the Average Treatment effect on the Treated (ATT):

$$E(Y_1 - Y_0 | T = 1). \quad (1)$$

One possible identification strategy is to impose the **CIA**:

$$Y_0 \perp\!\!\!\perp T | W. \quad (2)$$

A further requirement for identification is the **overlap condition**:

$$Pr(T = 1 | W) < 1. \quad (3)$$

Under assumptions (2) and (3), within each cell defined by W , treatment assignment is random, and the outcome of control subjects can be used to estimate the counterfactual outcome of the treated in the case of no treatment.

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 the CIA, Y_0 is independent of T given the propensity score (Rosenbaum and Rubin, 1983b).

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

$$E(Y_1 - Y_0|T = 1) = E_{\{p(W)|T=1\}}[E(Y_1|p(W), T = 1) - E(Y_0|p(W), T = 0)] \quad (4)$$

In practice, $p(W)$ has to be estimated, and an algorithm has to be used in order to match treated and control units on the basis of their estimated score.

The program `sensatt` makes use of three algorithms: **nearest neighbor**, **radius**, **kernel**.

FIVE STEPS FOR A “CORRECT” MATCHING

First step. To use data where the treated and control units come from the same local area and are asked the same set of questions (HIT, 1997).

Second step. To discuss why the CIA should be verified in the specific context of the evaluation question at hand.

Third step. To test (indirectly) whether the available empirical evidence casts doubt on the plausibility of the CIA (Rosenbaum, 1987b; Imbens, 2004).

Fourth step. To inspect how the observations are distributed across the common support and how sensitive the estimates are with respect to the utilization of observations in the tails (Black and Smith, 2004).

Fifth step. To assess whether (and to what extent) the estimated average treatment effects are robust to possible deviations from the CIA.

SENSITIVITY ANALYSIS

Identification of the ATT relies crucially on the validity of the CIA.

The CIA is untestable, since the data are completely uninformative about the distribution of Y_0 for treated subjects, but its credibility can be supported/rejected by theoretical reasoning and additional evidence.

Moreover, one should try to assess whether (and to what extent) the estimated average treatment effects are robust to possible deviations from the CIA

The sensitivity analysis proposed by Ichino, Mealli, and Nannicini (2007) allows applied researchers who make use of matching estimators to tackle this task.

CENTRAL ASSUMPTION

Assume that the CIA does not hold given the set of observable covariates W , but it holds given W and an unobserved binary variable U :

$$Y_0 \perp\!\!\!\perp T \mid (W, U). \quad (5)$$

Using Rosenbaum's (1987b) terminology, we are moving from $(Y_0|W)$ -adjustable treatment assignment of condition 2 to $(Y_0|W, U)$ -adjustable treatment assignment of condition 5.

A similar assumption in:

- Rosenbaum and Rubin (1983b)
- Rosenbaum (1987a)
- Imbens (2003)
- Altonji et al. (2005)

CHARACTERIZING THE DISTRIBUTION OF U

For simplicity, consider binary potential outcomes: $Y_0, Y_1 \in \{0, 1\}$. The observed outcome is given by: $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$.

The distribution of the binary confounding factor U is fully characterized by the choice of four parameters:

$$p_{ij} \equiv Pr(U = 1|T = i, Y = j) = Pr(U = 1|T = i, Y = j, W) \quad (6)$$

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

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

$$p_{i.} \equiv Pr(U = 1|T = i) = \sum_{j=0}^1 p_{ij} \cdot Pr(Y = j|T = i). \quad (7)$$

Two simplifying assumptions are made: a) binary U , b) conditional independence of U with respect to W . Monte Carlo exercise shows they are innocuous.

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 matching estimate of the ATT.

For each set of values of the sensitivity parameters, we repeat the matching estimation many times (e.g., 1,000) in order to obtain an estimate of the ATT, which is an average of the ATTs over the distribution of U .

These ATT estimates are identified under our extended CIA, i.e., the assumption that treatment assignment is unconfounded conditioning on W and a confounder U that behaves according to the chosen configuration of the p_{ij} .

STANDARD ERRORS

A solution is to see the simulated confounder as a problem of missing data that can be solved by multiply imputing the missing values of U .

Let m be the number of imputations of the missing U , and let \hat{ATT}_i and se_i^2 be the point estimate and the estimated variance of the ATT estimator at the i -th imputed data set. The *within-imputation* and the *between-imputation* variances are defined as, respectively:

$$se_W^2 = \frac{1}{m} \sum_{i=1}^m se_i^2 \quad (8)$$

$$se_B^2 = \frac{1}{m-1} \sum_{i=1}^m (\hat{ATT}_i - \hat{ATT})^2. \quad (9)$$

The total variance associated with \hat{ATT} is given by:

$$T = se_W^2 + \left(1 + \frac{1}{m}\right) se_B^2. \quad (10)$$

MULTIVALUED OR CONTINUOUS OUTCOMES

With multivalued or continuous outcomes the same sensitivity analysis can be applied by defining:

$$p_{ij} \equiv Pr(U = 1 | T = i, I(Y > y^*) = j), \quad (11)$$

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

Of course, the ATT is estimated for the multivalued or continuous outcome Y .

The program `sensatt` allows for the utilization of four y^* : mean, median, 25th centile, 75th centile.

GUIDELINES FOR THE SIMULATIONS

Note that the confounder U would be “dangerous” if we had that:

$$Pr(Y_0 = 1|T, W, U) \neq Pr(Y_0 = 1|T, W), \quad (12)$$

$$Pr(T = 1|W, U) \neq Pr(T = 1|W). \quad (13)$$

These expressions, unlike the parameters p_{ij} , both include W and refer to the potential (not observed) outcome in case of no treatment.

However, Ichino, Mealli, and Nannicini (2007) show that:

$$p_{01} > p_{00} \Rightarrow Pr(Y_0 = 1|T = 0, U = 1, W) > Pr(Y_0 = 1|T = 0, U = 0, W),$$

$$p_{1.} > p_{0.} \Rightarrow Pr(T = 1|U = 1, W) > Pr(T = 1|U = 0, W).$$

Hence, by simply assuming that $d = p_{01} - p_{00} > 0$, one can simulate a confounder that has a positive effect on Y_0 (conditioning on W). And, by setting $s = p_{1.} - p_{0.} > 0$, one can simulate a confounder with a positive effect on T .

THE OUTCOME AND SELECTION EFFECTS

However, by setting the quantities d and s , we can control the **sign** but not the **magnitude** of the conditional association of U with Y_0 and T .

To sidestep this problem, the program `sensatt` estimates a logit model of $Pr(Y = 1|T = 0, U, W)$ at every iteration. The average odds ratio of U ($\equiv \Gamma$) is reported as the **outcome effect** of the simulated confounder.

Similarly, the logit model of $Pr(T = 1|U, W)$ is estimated at every iteration, and the average odds ratio of U ($\equiv \Lambda$) is reported as the **selection effect**.

By setting $d > 0$ and $s > 0$, both the outcome and selection effects must be positive (i.e., $\Gamma > 1$ and $\Lambda > 1$). By displaying the associated Γ and Λ , we can assess the magnitude of these effects, which end up characterizing U .

TWO SIMULATION EXERCISES

a) **“Calibrated” confounders.** Pick the parameters p_{ij} to make the distribution of U similar to the empirical distribution of important binary covariates.

This simulation exercise reveals the extent to which the baseline estimates are robust to deviations from the CIA induced by the impossibility of observing factors similar to the observed covariates.

b) **“Killer” confounders.** Search for a configuration of p_{ij} such that, if U were observed, the estimated ATT would be driven to zero, and then assess the plausibility of this particular configuration.

One can build a table of simulated ATTs such that d increases by 0.1 along each column, and s increases by 0.1 along each column, looking for those configurations that kill the ATT.

THE SYNTAX OF THE STATA PROGRAM

```
sensatt outcome treatment [varlist] [weight] [if exp] [in range] [ ,  
alg(att*) reps(#) p(varname) p11(#) p10(#) p01(#)  
p00(#) se(se_type) ycent(#) pscore(scorevar) logit index  
comsup bootstrap ]
```

The following remarks should be taken into account:

- The program makes use of the commands for the propensity-score matching estimation of average treatment effects written by Becker and Ichino (2002): `attnd`, `attnw`, `attk`, `attr`.
- The treatment must be binary.

OPTIONS

`alg(att*)` specifies the name of the command (i.e., of the matching algorithm) that is used in the ATT estimation. One of the following commands can be specified: `attnd`, `attnw`, `attk`, `attr`. The default is `attnd`.

`p(varname)` indicates the binary variable which is used to simulate the confounder. The parameters p_{ij} used to simulate U are set equal to the ones observed for *varname*.

`p11(#)`, `p10(#)`, `p01(#)` and `p00(#)` jointly specify the parameters p_{ij} used to simulate U in the data. Since they are probabilities, they must be between zero and one. For each parameter, the default is zero.

`reps(#)` specifies the number of iterations, i.e., how many times the simulation of U and the ATT estimation are replicated. The default is 1,000.

OPTIONS (cont.)

`se(se_type)` allows the user to decide which standard error should be displayed with the simulated ATT. Three `se_types` are possible: `set` uses the total variance in a multiple-imputation setting; `sew` uses the within-imputation variance; `seb` uses the between-imputation variance. The default is `set`.

`ycent(#)` is relevant only with continuous outcomes. It means that U is simulated on the basis of the binary transformation of the outcome: $I(Y > y^*)$, where y^* is the $\#$ th centile of the distribution of Y . Three centiles are allowed: 25, 50, 75. If `ycent(#)` is not specified by the user, y^* is the mean of Y .

EXAMPLES

For concrete examples of the utilization of the sensitivity analysis performed by the program `sensatt`, see:

- Ichino A., F. Mealli, and T. Nannicini (2007), “From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity?”, *Journal of Applied Econometrics*, forthcoming.
- Nannicini T. (2007), “Simulation-Based Sensitivity Analysis for Matching Estimators”, *The Stata Journal*, Issue 3, forthcoming.