

# Comparing continuous treatment matching methods in policy evaluation

Valentina Adorno, Cristina Bernini, and Guido Pellegrini

**Abstract** The paper evaluates the statistical properties of two different matching estimators in the case of continuous treatment, using a Montecarlo experiment. The traditional generalized propensity score matching estimator is compared with a new two steps matching estimator for the continuous treatment case, recently developed [1]. It compares treatment and control units similar in terms of their observable characteristics in both selection processes (the participation decision and the treatment level assignment), where the generalized propensity score matching estimator collapses the two processes into one single step matching. The results show that the two steps estimator has better finite sample properties if some institutional rules define the level of treatment with respect to the characteristics of treated units.

## 1 Introduction

The interest on the generalization of the programme evaluation framework from a binary treatment setting to a more general structure for the treatment has increased rapidly in the last years [4] and [6]. The policy mechanism can be away from an experimental data framework because of the presence of multiple non random selection processes, related not only to the participation decision but also to the treatment level assignment. In these cases, the selection bias problem cannot be tackled

---

Valentina Adorno

Department of Economics, University of Bologna, Piazza Scaravilli, 2 Bologna, e-mail: valentina.adorno@unibo.it

Cristina Bernini

Department of Statistics, University of Bologna, Via Belle Arti 41, Bologna e-mail: cristina.bernini@unibo.it

Guido Pellegrini

Department of Economic Theory and Quantitative Methods for Political Choices, University La Sapienza, Piazzale Aldo Moro 5, Roma e-mail: guido.pellegrini@uniroma1.it

using the estimation methods developed for the binary treatment case. The literature proposes few matching estimators for continuous treatment ([4],[2] and [5]). In all cases the analysis does not concern on the comparison between treated and non-treated units and on the selection process related to the treatment level assignment. We have recently developed a novel two steps matching approach to estimate the causal treatment effects as a function of the doses [1]. The novel estimator matches treatment and comparison units that are similar in terms of their observable characteristics in both the selection processes (the participation decision and the treatment level assignment). This is the main difference with respect to the generalized propensity score matching procedure that collapses the two processes into one single step matching.

The main empirical advantage of our method is its aptitude to incorporate in the matching procedure some recognized restrictions on the relation between the two selection processes, as it happens in a lot of empirical application where policy instruments have institutional restrictions. This is the case of public subsidies to private capital accumulation: in the European Union the (maximum) amount of subsidies is strictly linked to the firm dimension and its geographical localization. An important application of this institutional rules is the Law 488/1992 (L. 488) in Italy, the most important policy intervention to subsidize private capital accumulation in the poorest regions in the last decades. Moreover, subsidies by L.488 are allocated by mimicking an auction mechanism: the firm can choose the amount of subsidy, and lower the amount, higher the probability to receive it. This procedure generates heterogeneity in the amount of subsidy allocated to similar firms. Therefore, the L.488 case is an interesting experimental framework in order to test the statistical properties of continuous treatment matching estimators.

The aim of the paper is to explore the finite sample properties of a two steps matching estimator in the presence of a system of external constraints in the continuous treatment level assignment. The comparison is based on a Montecarlo experiments, mimicking the allocation mechanism of L.488, using different simulation settings.

## **2 Simulating a subsidies allocation mechanism: the case of L.488**

The L.488 allocates subsidies through a "rationing" system based on an auction mechanism which guarantees compatibility of subsidies demand and supply. This system is basically a regional tender for incentives where the subsidies allocation is based on general criteria expressing the policy preference. In each regional auction the investment projects are ranked by a score assigned on the base of five indicators (among them the share of the subsidy requested by the firm and the highest subsidy applicable, given the rules determined by the UE Commission). The rankings are drawn up through the decreasing order of the score awarded to each project and the subsidies are allocated to projects until the financial resources granted to each region are exhausted. The amount of financial resources to be allocated in every auction is

different across regions: for every auction exists a specific regional ranking and a regional allocation threshold. Then, if all the firm are considered, an overlapping area of firms with the same propensity to be subsidised for treated and control group is available and matching estimators can be correctly applied.

The institutional framework of L.488 is important for the choice about the level of the treatment. The maximum amount of subsidy (relative to the level of investment) allowable to a project depends on both the region where the investment is localised and the size of the firm. Therefore, it exists some institutional constrains related to the level of subsidies received by each firm, depending on some of its characteristics. This aspect can be fully exploited in the estimation of the treatment level decision. Moreover the amount of subsidy relative to the ceiling established by institutional rules is also a choice variable for the firm: lower the aid applied for by the firm, higher the probability to receive the subsidy. This is the key indicator that transforms the allocation procedure to an auction mechanism. On the other side, different amount of subsidies are allocated to similar firms, allowing for the implementation of a matching procedure on the subsidy level selection assignment.

The selection procedure of L. 488 can help in policy evaluation simulation. The procedure uses the indicators as selection variables: they explain the most part of the difference between subsidized and not subsidized firms. Therefore, the indicators can be very helpful in the construction of the counterfactual scenario. Moreover, different regional auctions (with different thresholds) can be easily replicated, generating a data set with treated and not treated firms having overlapping probability to be treated.

### 3 The matching methods in the continuous framework

In the continuous treatment framework  $Y(T)$  represents the set of potential outcomes, for each unit  $i$ , given a random sample indexed by  $i = 1 \dots N$  and  $T$  represents the continuous variable indicating the treatment level. The observed outcome  $Y$  can be written as  $y_i = d_i y_i(t_i) + (1 - d_i) y_i(0)$ .  $D$  is a dummy variable indicating the treatment status ( $d_i = 1$  if the individual has been treated);  $y_i(t_i)$  is the potential outcome at the observed level  $t_i$ . The participation decision assignment will determine the treatment status  $d_i$ , while the treatment level process will determine the dose  $t_i$ . Even if they can occur simultaneously, we suppose that can be logically separated in the analysis. Let assume assignment to treatment is made on the basis of:

$$t_i = \begin{cases} g(Z_i) + u_i & \text{if } d_i = 1 \\ 0 & \text{otherwise} \end{cases}, \quad d_i = \begin{cases} 1 & \text{if } I_i > 0 \\ 0 & \text{otherwise} \end{cases}, \quad I_i = h(w_i) + v_i \quad (1)$$

where  $W$ ,  $Z$  and  $U$ ,  $V$  represent a set of observable and unobservable variables respectively available at a fixed time, when the selection processes occur. This structure represents the basis of our approach: differently from the previous literature, it specifies separately the two selection processes. Adopting different specifications

for the two processes might be helpful for considering different influencing factors and for estimating the selection process in a more efficient way. The first step identifies the participation decision rule and units will be matched on the basis of similar value of the set of covariates, using the propensity score function  $p(w) = P(d_i|W_i)$ , instead of the full set of covariates  $W$  and  $Z$ . Among matched units in the first step, the next matching procedure will pair units with similar value in the covariates identifying the treatment level assignment process. Let  $\theta(Z) = E(T|Z, p(W))$  the parameter that uniquely represents the propensity function ([6]), then matching on it can be easily accomplished by matching on  $\theta$ . Instead, the one step matching procedure is based on the propensity score function  $p(w, z) = P(d_i|W_i, Z_i)$ , exploiting the full set of observable variables  $W$  and  $Z$ .

As regards the parameter of interest, a natural development in the continuous case of the traditional treatment effect estimation (the average treatment effect on the treated  $ATT = E[Y(1) - Y(0)|D = 1]$ ), is what we named the average treatment level effect on the treated  $ATLE = \alpha(T) = E[Y(T) - Y(0)|T = t]$  for a person randomly drawn from the subpopulation of the participants at the level  $t$ . Once estimated the ATLE for each observed treatment level, the next interesting object to estimate regards the specification of the relation between effects and levels:

$$\hat{\alpha} = f(t, \varepsilon) \quad (2)$$

in order to estimate the entire function of average treatment effects over all possible values of the treatment doses.

## 4 The Montecarlo experiment

There are very few papers that use Montecarlo simulations to evaluate matching estimators ([3], [7], [8]). All the studies investigate different aspects of matching estimators (finite sample properties, sensitiveness to propensity score specifications, etc.) but they restrict their attention to the binary treatment case and do not consider the small-sample property in the continuous treatment cases. This is the main contribution of our paper: our focus is on comparing one step and two steps matching estimators by Montecarlo experiments in the case of continuous treatment. The Montecarlo simulation mimics the L. 448 two steps allocation mechanism:

1. generating different data set by unit, including for each unit the value of 3 indicators ( $I_1, I_2, I_3$ ), i.e. covariates affecting the participation decision and the treatment level assignment, the outcome for treated ( $Y(1)$ ) and not treated units ( $Y(0)$ ). The indicators are generated as random numbers drawn from a standardized normal distribution ( $N(0, 1)$ ).
2. generating different thresholds for each dataset and creating the treated units and control groups. We generated  $\tau$  as a random number drawn from a uniform distribution (0.3; 0.8). The presence of different thresholds allows to define an overlapping area for the matching experiment, mimicking the L. 488 allocation setting.

3. calculating the amount of treatment ( $T$ ), as a function of two indicators (eq.4).  
The treatment level indicator  $T$  enters in the participation selection function  $D$ , as in the L. 488 selection mechanism.(eq.3).

The selection mechanism is defined by an index function  $I$ , depending on the two indicators  $I_1, I_2$  and the treatment level  $T$  (eq.3). In each dataset 10 thresholds  $\tau$  are generated. Assignment to treatment of each unit is made on the basis of the following rule:

$$D = \begin{cases} 1 & \text{if } \beta_1 I_1 + \beta_2 I_2 + \beta_3 T > \tau \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$T = \beta_4 I_2 + \beta_5 I_3 + \varepsilon_0 \quad (4)$$

$$Y(0) = \beta_6 I_1 + \beta_7 I_3 + \varepsilon_1 \quad \text{if } D = 0 \quad (5)$$

$$Y(1) = \begin{cases} \beta_6 I_1 + \beta_7 I_3 + \gamma_1 T + \varepsilon_1 & \text{in the linear case} \\ \beta_6 I_1 + \beta_7 I_3 + \gamma_2 T + \gamma_3 T^2 + \varepsilon_1 & \text{in the non linear case} \end{cases} \quad \text{if } D = 1 \quad (6)$$

The treatment level assignment depends also on  $I_2$  and on the other index  $I_3$ , creating a positive correlation with the selection mechanism. A standardized normally distributed error term  $\varepsilon_0$  is added (eq.4). The outcome variable  $Y(i)$  is observed after the treatment, for the control and treated group ( $i=0$  and  $i=1$ , respectively). In the untreated state,  $Y(0)$  depends on indicators  $I_1$  and  $I_3$ , that are also in the two selection processes, and on a standardized normally distributed error term  $\varepsilon_1$ . The outcome in the treated state  $Y(1)$  is generated by adding the effect of the treatment to the outcome of the untreated state. We are interested in capturing the difference in the treatment effect by the treatment level. Therefore a linear and a quadratic treatment effect are experimented. In the last case we imposed that the maximum of the curve lies inside the range of the generated treatment level. Then, the outcome of the treated units is defined as in eq. (6). We allow perfect correlation between the error terms in the outcome equations, but do not allow correlation between the error term in the selection equations and the error terms in the outcome equations, i.e.  $cov(\varepsilon_0, \varepsilon_1) = 0$ . So unconfoundedness is satisfied.  $\beta_i$  are fixed parameters in our Montecarlo experiment<sup>1</sup>. The parameters  $\gamma_i$  are the policy impact coefficients, explaining the relation between treatment level and outcome.

We are interested in comparing the properties of the two continuous matching estimators with respect to the traditional ATT and the parameters  $\gamma_i$ , estimated by an OLS regression of the ATLE on the treatment levels (eq. 2). We investigate this issue in different designs, changing the impact coefficients and the selection processes error variance. In the linear experiment we impose  $\gamma_1$  equal to 0.2 and 1.2. In the quadratic case we impose  $(\gamma_1; \gamma_2) = (6; -0.3)$ ,  $(9; -0.5)$ . The variance of  $\varepsilon_0$  assumes the values 0.5; 1 and 2 in both experiments. For each of these 12 different combinations, 100 dataset are simulated, each of 10,000 observations, coming from the simulation of 1,000 observations replicated for 10 different thresholds.

Among the matching algorithm proposed in literature we choose the stratification matching, properly adapted for the continuous case. In particular, for the two steps

<sup>1</sup> The set of parameters we adopted in our simulation is:  $\beta_1, \beta_2, \beta_3 = 0.33; \beta_4, \beta_5 = 0.4; \beta_6, \beta_7 = 0.1$ .

estimation, we first compute the stratification with respect to the propensity function that identifies the participation process ( $p(w) = P(d_i|I_1, I_2, T)$ ) and then with respect to the treatment level assignment  $T$ . We compute the ATLE (for each stratum of the treatment level) as the weighted average among the mean differences between the outcome of treated and not treated units, for each stratum of the propensity function  $p(w)$ . Indeed, for the one step case, we compute a stratification matching on the basis of a unique propensity function  $p(w) = P(d_i|I_1, I_2, T, I_3)$ . The ATLE are computed at the same stratum of the treatment level used in the two steps case.

## 5 Results and conclusions

Tables 1 and 2 report estimates of ATT and  $\gamma_i$  (Mean column), bias (the difference between the estimated and the true effect/parameter) and mean square error (MSE). The parameters  $\gamma_i$  are estimated by an OLS regression of the ATLE values on treatment levels. Fig. 1 plots this regression in one of the setting of the Montecarlo experiment. In the linear case both the 1-step and the 2-steps estimator show always

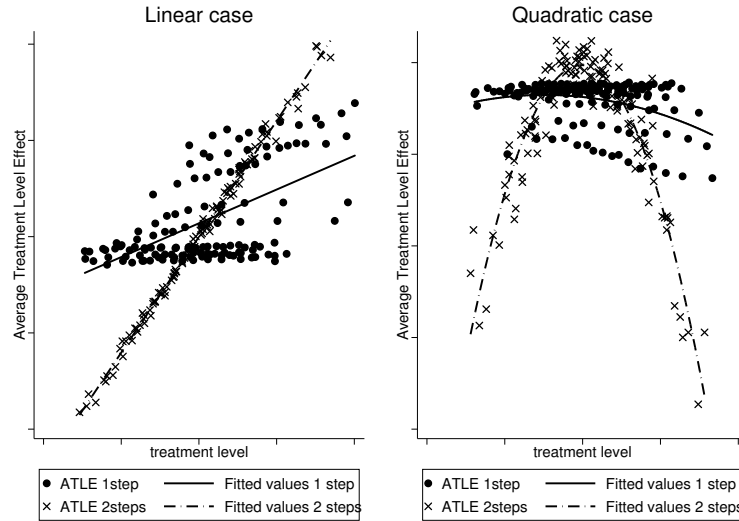
**Table 1** Sensitivity to the treatment level effect: linear case

$\gamma_i$	$\sigma^2$	2 STEPS CASE				1 STEP CASE			
		ATT		$\hat{\gamma}_i$		ATT		$\hat{\gamma}_i$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
0.2	0.5	2.040	0.068	0.199	0.050	2.052	0.080	0.017	0.026
	1	2.037	0.082	0.192	0.041	2.050	0.087	0.015	0.022
	2	2.039	0.088	0.203	0.050	2.042	0.099	0.013	0.026
1.2	0.5	12.265	0.452	1.197	0.045	12.299	0.458	0.072	0.056
	1	12.271	0.468	1.198	0.044	12.307	0.474	0.082	0.058
	2	12.233	0.507	1.204	0.047	12.281	0.502	0.079	0.065
$\gamma_i$	$\sigma^2$	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2	0.5	0.040	0.006	-0.001	0.002	0.052	0.009	-0.183	0.034
	1	0.037	0.008	-0.008	0.002	0.050	0.010	-0.185	0.035
	2	0.039	0.009	0.003	0.003	0.042	0.011	-0.187	0.036
1.2	0.5	0.265	0.275	-0.003	0.002	0.299	0.299	-1.128	1.275
	1	0.271	0.293	-0.002	0.002	0.307	0.319	-1.118	1.252
	2	0.233	0.311	0.004	0.002	0.281	0.331	-1.121	1.260

a small upward biased ATT (Tab. 1). In both cases higher the error variance, higher the MSE. However, the MSE is always lower in the 2-steps case than in the 1-step one. The bias on  $\gamma_i$  is substantially higher in the 1-step case, as well as the MSE. Fig. 1 shows that the downward biased coefficient generates a flatter regression line for the 1-step estimator: even if the estimated average ATT is close to the true ATT, the treatment effect is less affected by changes in the treatment level.

**Table 2** Sensitivity to the treatment level effect: non-linear case

			2 STEPS CASE						1 STEP CASE						
$\gamma_2$	$\gamma_3$	$\sigma^2$	ATT		$\hat{\gamma}_1$		$\hat{\gamma}_2$		ATT		$\hat{\gamma}_1$		$\hat{\gamma}_2$		
			Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	
6	-0.3	0.5	29.91	0.057	6.123	1.888	-0.305	0.092	29.90	0.064	0.057	0.224	-0.003	0.011	
			1	29.90	0.065	6.437	1.357	-0.322	0.067	29.89	0.079	0.064	0.262	-0.004	0.013
			2	29.88	0.074	5.833	1.617	-0.292	0.078	29.87	0.087	0.073	0.269	-0.004	0.013
9	-0.5	0.5	39.63	0.442	8.975	1.451	-0.499	0.071	39.60	0.446	-0.123	0.751	0.002	0.035	
			1	39.60	0.488	9.171	1.782	-0.509	0.087	39.57	0.497	-0.151	0.583	0.004	0.027
			2	39.57	0.448	9.033	1.605	-0.501	0.078	39.53	0.468	-0.120	0.805	0.002	0.038
$\gamma_2$	$\gamma_3$	$\sigma^2$	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	
6	-0.3	0.5	-0.090	0.011	0.123	3.581	-0.005	0.008	-0.103	0.015	-5.943	35.37	0.297	0.088	
			1	-0.098	0.014	0.437	2.033	-0.022	0.005	-0.105	0.017	-5.936	35.30	0.296	0.088
			2	-0.116	0.019	-0.167	2.644	0.008	0.006	-0.127	0.024	-5.927	35.20	0.296	0.088
9	-0.5	0.5	-0.367	0.330	-0.025	2.106	0.001	0.005	-0.402	0.361	-9.123	83.78	0.502	0.253	
			1	-0.398	0.396	0.171	3.204	-0.009	0.008	-0.433	0.434	-9.151	84.09	0.504	0.254
			2	-0.431	0.387	0.033	2.575	-0.001	0.006	-0.472	0.442	-9.120	83.82	0.502	0.253



**Fig. 1** OLS regression of ATLE on treatment level. Linear case ( $\gamma_1 = 1.2, \sigma^2 = 0.5$ ) and quadratic case ( $\gamma_2 = 6, \gamma_3 = -0.3, \sigma^2 = 0.5$ )

Similar results are derived in the non linear case. We register a small downward biased average ATT in both cases, with a slight higher MSE for the 1-step estimator. However, the coefficients  $\gamma_2$  and  $\gamma_3$  are poorly estimated by the 1-step matching procedure in every simulation. The quadratic curve is much more flatter than the 2-

steps case (Fig. 1), and it does not capture the strong heterogeneity of the treatment outcome with respect to different treatment levels.

To understand the finding, we note that the 2-steps procedure enhances the quality of the matching: if the treatment effect depends on the treatment level, comparing units with the same potential amount of treatment improves the accuracy of the ATLE estimation. The result is more evident in presence of some institutional rules that relate the amount of subsidies to the characteristics of the treated units. This is the case of our Montecarlo experiment, where the variables influencing the selection rule and the treatment level assignment are different. By adopting different specifications for the two processes we improve the selection process estimation, incorporating information on the institutional framework.

To conclude, the major finding from our simulations is that, even if the statistical performances of the two matching procedures are similar in the estimation of the ATT, they deeply differ on the estimation of the effect-treatment level relationship. In fact, the treatment impact coefficients are poorly estimated in the generalized (1-step) propensity score procedure, in particular in the non linear case. Our Monte Carlo results show an overall underestimation of the elasticity of the treatment effect to changes in the treatment level in the case of the 1-step estimator. The reason is that the 2-steps estimator sharpens the matching procedure, comparing units with the same potential amount of treatment. The finding can be empirically relevant if there are strict rules that relate the amount of treatment to the characteristics of treated units, as several economic policies. Therefore, both methods we used can have a wide application field. Nevertheless, the 2-step matching methods offers more and better information to the empirical instrument evaluation: the comparison between the treated and the non treated units is more homogeneous with respect to the treatment level, and a less biased measure of the impact of all different treatment levels to the treated units can be derived.

## References

1. Adorno V, Bernini C, Pellegrini G (2007) The impact of capital subsidies: new estimations under continuous treatment. *g. econ.* 66 1:67-92
2. Behrman J, Cheng Y, and Todd P (2004) Evaluating preschool programs when length of exposure to the program varies: a non parametric approach. *rev. econ. stat.* 86 1: 108-132
3. Frolich M (2004) Finite sample properties of propensity-score matching and weighting estimators. *rev. econ. stat.* 86: 77-90.
4. Hirano K, Imbens G (2004) The propensity score with continuous treatment In: *Missing data and bayesian methods in practise: contributions by Donald Rubin's Statistical family*, Wiley.
5. Hornik R, Rosenbaum P R, Lu B, Zanutto E (2001) Matching with doses in an observational study of a media campaign against drug abuse. *JASA appl. case stud.* 96 456: 1245-1253. .
6. Imai K, Van Dyk D (2004) Causal inference with general treatment regimes: Generalizing the propensity score. *JASA* 99 467: 854-866.
7. Zhao Z (2004) Using matching to estimate treatment effects: data requirements, matching metrics and Montecarlo evidence. *rev. econ. stat* 86 1: 91-107.
8. Zhao Z (2007) Sensitivity of propensity score methods to the specifications. *Forthcom. econ. lett.*