Occupazione e disoccupazione in Italia:
misura e analisi dei comportamenti

Progetto di ricerca cofinanziato dal Ministero per l'Università
e la Ricerca Scientifica e Tecnologica - Assegnazione: 1999
Coordinatore: Ugo Trivellato

Testing for the presence of a programme effect in a regression discontinuity design with non compliance

Erich Battistin, Enrico Rettore
Dip. di Scienze Statistiche, Univ. di Padova

Working Paper n. 27 novembre 2000

Unità locali del progetto:

Dip. di Economia "S. Cognetti De Martiis", Univ. di Torino (coord. Bruno Contini)
Dip. di Scienze Economiche, Univ. "Ca' Foscari" di Venezia (coord. Giuseppe Tattara)
Dip. di Metodi Quantitativi, Univ. di Siena (coord. Achille Lemmi)
Dip. di Scienze Statistiche, Univ. di Padova (coord. Ugo Trivellato)

Dip. di Scienze Statistiche
via S. Francesco 33, 35121 Padova

Testing for programme effects in a regression discontinuity design with imperfect compliance

Erich Battistin

Department of Statistics, University of Padova, Italy
Institute for Fiscal Studies, London, UK

Enrico Rettore

Department of Statistics, University of Padova, Italy

Summary. The administrators of an office automation training programme in Italy enrolled applicants on the basis of their score in an attitudinal test, with low scoring subjects mandated out of the programme. Some of the mandated out applicants resorted to an alternative programme. To identify the programme impact by comparing participants to non-participants one needs to properly account for both the selection on the score and the contamination of the comparison group by a number of non-complying subjects. The estimand resulting from using the mandated status as an Instrumental Variable (IV) for the actual status identifies the programme impact on complying subjects exhibiting a score at the attitudinal test in the neighborhood of the threshold for selection. Simple nonparametric IV estimators based on Robinson (1988) and on Hahn *et al.* (2001) reveal that the programme had no impact on the probability of being in work some months after its completion. Simulation results show that in spite of the small sample size the test for the no impact hypothesis has non-negligible power even at small departures from the null. As a side result the test based on Robinson (1988) turns out appreciably more powerful than the other one.

Keywords: Instrumental Variables; Local Average Treatment Effect; Rubin Causal Model

Acknowledgements: The paper benefited from helpful discussions with Adelchi Azzalini, Richard Blundell, Kei Hirano, Jack Porter, Wilbert van der Klaauw and comments by audiences at UCL/IFS and at the Royal Statistical Society Conference on the Evaluation of Economic and Social Policies, London, July 2000. Two anonymous referees provided helpful criticism and suggestions. We thank Federica Laudisa for the dataset. Part of this research was completed while the first author was visiting the Department of Economics at UCL, whose hospitality is gratefully acknowledged. Financial support from MURST to the project *Employment and Unemployment in Italy: measurement issues and behavioral analysis* is gratefully acknowledged. The usual disclaimer applies.

1. Introduction and problem setting

The municipality of Torino, a large city in the North West of Italy, run a menu of vocational training programmes targeted at specific sub-populations at risk of experiencing long spells of unemployment. Eligible subjects willing to participate chose one vocational training

Address for correspondence: Enrico Rettore, Dipartimento di Scienze Statistiche, Università di Padova, via Cesare Battisti 241/243, 35121 Padova, Italy. E-mail: enrico.rettore@stat.unipd.it

programme from the menu according to their preferences. This paper tests whether a specific programme - an office automation training programme - did increase participants' chance of getting a job.

The programme took place in the period October 1995 - June 1996 and was targeted at unemployed people, mostly holding at least a *diploma* (which means at least 13 years of formal education). Enrolled participants were offered an intensive training course lasting overall 600 hours to improve their skills in using different kinds of software (mostly word processors and accounting packages), including an apprenticeship period (100 hours) with an existing company.

Since the number of applicants was nearly twice as large the number of available slots, as a selection rule the administrators ranked applicants on the basis of the score in an attitudinal test and enrolled high scoring ones. A problem of *post assignment treatment choice* (PATC in the following) arose since, while all enrolled applicants did complete the training period, some selected out ones resorted to an alternative vocational training programme.

As a consequence, the pool of mandated non-participants turns out contaminated by a number of subjects self-selected in an alternative programme. Both the main and the alternative programme took place exactly over the same time span featuring the same eligibility criteria.

The general framework we refer to is the *Rubin Causal Model* (see Holland, 1986), which defines causal effects in terms of *potential outcomes.* To account for the way in which treatment and control groups was determined, we link the literature on programme evaluation when assignment to treatment is on the basis of observable covariates (see Cook and Campbell, 1979, and Rubin, 1977) to the literature on identification of causal parameters by means of *instrumental variables* (see Imbens and Angrist, 1994, and Angrist *et al.*, 1996; AIR in the following).

By analogy with Imbens and Angrist (1994), the resulting estimand can be interpreted as a causal effect for the subpopulation of *compliers* - that is subjects participating in the programme if and only if they are assigned to it - at the threshold value for selection at the attitudinal test. The resulting selection design - a discontinuity design with imperfect compliance - also straightforwardly fits the so called fuzzy *regression discontinuity design* (RDD in the following; see Trochim, 1984).

Statistical and econometric methods to evaluate results from assigning units to different groups only on the basis of a pre-programme measure have been widely used in the recent literature. Berk and De Leeuw (1999) extend the classical RDD to a generalized linear model with binary response variable and an arbitrary number of discrete or quantitative treatments to evaluate a prison classification system. Van der Klaauw (2000) exploits a RDD to evaluate the impact of financial aid offers on college enrollment. Angrist and Lavy (1999) study a regression discontinuity model arising from exogenously determined factors to obtain instrumental variable estimates of class size effect on scholastic achievement.

Hahn *et al.* (2001; HTVK in the following) formally derive theoretical links between regression discontinuity models and the literature about causality. They also provide a general discussion on conditions for identification of programme effects in a RDD and propose a nonparametric estimator deriving its asymptotic distribution.

Moving from their paper we point out that if rather than estimating the programme impact one is only willing to test the no impact hypothesis, an alternative nonparametric approach based on Robinson (1988) and Porter (1999) is feasible since some extra regularity conditions become available. According to some simulation results presented in our analysis the alternative testing procedure turns out more powerful.

The rest of this paper is organized as follows. Section 2 sets up the statistical framework needed to clarify which causal parameter can be identified from the available information. We start by discussing identification issues dealing with the selection on the score only, assuming perfect compliance with the mandated status; then, we account for the PATC problem and build on AIR to define programme effects. Section 3 reviews alternative estimation strategies for causal effects using regression discontinuity models. Section 4 shows the empirical results of testing the no impact hypothesis by means of the HTVK procedure as well as the alternative nonparametric procedure. Results from a small simulation exercise are presented in Section 5 to assess the power of the two tests. Section 6 concludes. Some more technical comments on our empirical findings are in Appendix A and Appendix B; the simulation experiment is described in Appendix C.

2. On the identifiability of programme impacts

Following the potential outcome approach to causal inference introduced by Rubin (1974), let (Y_i^T, Y_i^{NT}) be the potential outcomes the i-th subject would experience by taking and not taking part in the programme, respectively. The impact s/he would get by being exposed is $Y_i^T - Y_i^{NT}$, which is not observable since being exposed to (denied) the programme reveals Y_i^T (Y_i^{NT}) but conceals the other potential outcome.

Let S be the score in the attitudinal test programme administrators use to assign the subject status and let Z be the binary variable indexing the consequent assignment. At this step the selection process perfectly fits a RDD, since units are assigned to the treatment $(Z = 1)$ if and only if $S \geq \bar{s}$, where $\bar{s}$ is the threshold for selection.

Now, let compliance with assignment be imperfect so that some subjects can overturn the status mandated by programme administrators. Let D be the binary variable indexing the actual status, with $D = 1$ signaling that the subject took part in the programme. Presumably, the mandated status Z affects the actual status D: we stress such dependence writing D^Z. Then, (D^1, D^0) are the potential statuses resulting from being and not being assigned to the programme, respectively, and $D^1 - D^0$ is the impact of being assigned to the programme on the actual status.

The observed outcome for the i-th subject can be written as

$$Y_i = Y_i^{NT} + D_i^Z \beta_i, \tag{1}$$

where $\beta_i = Y_i^T - Y_i^{NT}$ is the gain that subject has by participating in the programme. In what follows we make the standard *Stable Unit Treatment Value Assumption* (SUTVA; Rubin, 1977) according to which the outcome experienced by subject i is not affected by assignment and receipt of treatment by other subjects.

As we explained in the Introduction, subjects featuring $(Z = 0, D = 1)$ take part in the *alternative* programme, not in the main one. This compels us to introduce a further potential outcome in the analysis, $\tilde{Y}^T$ say, representing the outcome a specific subject would obtain by resorting to the alternative programme.

To keep the analysis simple, in the following we maintain the assumption $\tilde{Y}^T = Y^T$. In the subsequent sections we shall point out that such an assumption might result in a biased estimate of the programme impact. We shall also provide an argument supporting the robustness of the actual results we get to violations of the assumption.

The next two sections show how regression discontinuity models allow representation of the two-step selection process so far discussed. In particular, Section 2.1 shows how - in a world of perfect compliance - a RDD can be thought as a quasi-experiment because of its strict analogy to a proper experimental design in a neighborhood of $\bar{s}$ (Cook and Campbell, 1979). Section 2.2 copes with further complications arising from imperfect compliance leading the discussion again onto a different class of discontinuity models.

2.1. The sharp regression discontinuity design

To begin with, let all subjects comply with the outcome of the assignment process, so that the actual status and the status determined on the basis of the attitudinal test coincide, $D^Z = Z$. It follows that the observed outcome in (1) can be written as $Y_i = Y_i^{NT} + Z_i\beta_i$.

Since $Z = \mathbb{1}(S \geq \bar{s})$, the *propensity score* $e(s)$ - that is the probability of selection into the treatment group given the score at the attitudinal test - is discontinuous at $\bar{s}$ stepping from zero to one as S crosses the threshold $\bar{s}$. Following Trochim (1984) we will refer to this situation as a *sharp* RDD, since the probability of being exposed is just a deterministic function of an observable pre-programme characteristic (this is an example of the so called *selection on observables*; see Rubin, 1977).

In this context, conditioning on S allows us to identify the average impact of the programme on subjects scoring $\bar{s}$ at the attitudinal test

$$\tau(\bar{s}) = E(\beta_i|S_i = \bar{s}),$$

the intuition being that for each unit in the neighborhood of $\bar{s}$ the following condition holds

$$(Y_i^T, Y_i^{NT}) \perp Z_i|S_i = \bar{s}, \tag{2}$$

that is the programme status Z and the potential outcomes are conditionally independent. Besides, in the neighborhood of $\bar{s}$ there are subjects both taking and not taking part in the programme implying that

$$0 < e(\bar{s}) < 1. \tag{3}$$

There is a slight abuse of language in (2) and (3) since we write $\bar{s}$ to mean its neighborhood. HTVK spell out the conditions needed for the identifiability of $\tau(\bar{s})$: loosely speaking, they require the continuity at $S = \bar{s}$ of the expected value of the random variables involved in the model conditional on S.

Conditions (2) and (3) together make up the *strong ignorability assumption* (see, for instance, Rosenbaum and Rubin, 1983), a major implication being that in the neighborhood of $\bar{s}$ we may think about the design as an almost experimental one.

2.2. The regression discontinuity design with imperfect compliance

When compliance is not perfect subjects can violate the assignment mandated by programme administrators in that selected out applicants might resort to an alternative programme while enrolled applicants might drop out of the main one.

In a randomized treatment assignment context AIR show that - provided the so called monotonicity assumption holds (see below) - the *average impact on compliers* τ_c is identifiable and can be written as

$$\frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i^Z|Z_i = 1) - E(D_i^Z|Z_i = 0)}. \tag{4}$$

This expression represents the ratio of the causal effect of the mandated status on the outcome of interest Y (that is the intention to treat effect, ITT in the following) to the causal effect of the mandated status on the actual status. They call (4) the *Local Average Treatment Effect* (LATE) of D^Z on Y and show that it is the parameter the instrumental variable estimator of the regression of Y on D^Z converges to using Z as the instrument. A discussion on the meaning and the policy relevance of this parameter originated from AIR (see the comments to AIR by Heckman, 1997, and Angrist and Imbens, 1999 and the reply by Heckman therein). Imbens and Rubin (1997) extend this result by showing that the marginal distributions of the potential outcomes (Y^T, Y^{NT}) for the subpopulation of compliers are identifiable, as well. Balke and Pearl (1997) derive the nonparametric tightest bound on the average treatment effect on the whole reference population.

To make the AIR analysis fits a RDD we follow their steps. Let us partition the reference population according to the value of potential outcomes (D_i^1, D_i^0) and label the resulting four groups as *compliers* ($D^1 = 1$ and $D^0 = 0$), *always takers* ($D^1 = 1$ and $D^0 = 1$), *never takers* ($D^1 = 0$ and $D^0 = 0$) and *defiers* ($D^1 = 0$ and $D^0 = 1$). Let the population frequencies of subjects belonging to the four groups at $\bar{s}$ be $\phi_c(\bar{s})$, $\phi_a(\bar{s})$, $\phi_n(\bar{s})$ and $\phi_d(\bar{s})$, respectively. For the regularity conditions needed to derive results reported in what follows the reader is referred to HVTK (1999) and Appendix A.

Because of the strong ignorability implied by design, the ITT effect on Y for units scoring $\bar{s}$ at the attitudinal test can be written as

$$E(D_i^1 \beta_i | S_i = \bar{s}) - E(D_i^0 \beta_i | S_i = \bar{s}),$$

which can be decomposed into the following expression

$$E(\beta_i | D_i^1 - D_i^0 = 1, S_i = \bar{s}) \phi_c(\bar{s}) + E(\beta_i | D_i^1 - D_i^0 = -1, S_i = \bar{s}) \phi_d(\bar{s}). \tag{5}$$

Following AIR, notice that the *monotonicity assumption*

$$\phi_d(\bar{s}) = \Pr(D_i^1 - D_i^0 = -1 | S_i = \bar{s}) = 0 \tag{6}$$

looks appealing: if subject i is denied the programme and nonetheless s/he gets it, namely $D_i^0 = 1$, *a fortiori* s/he gets the programme by being assigned to it, namely $D_i^1 = 1$, ruling out the existence of *defiers*. Whether or not such an assumption is reasonable it is problem specific; see the discussion in Heckman (1997) and Angrist and Imbens (1999).

Provided that the condition

$$\phi_c(\bar{s}) = \Pr(D_i^1 - D_i^0 = 1 | S_i = \bar{s}) > 0 \tag{7}$$

holds, the monotonicity restriction implies the identifiability of the average impact on compliers scoring $\bar{s}$ at the attitudinal test

$$\tau_c(\bar{s}) = E(\beta_i | D_i^1 - D_i^0 = 1, S_i = \bar{s}).$$

Note that (7) is testable since it is the impact of Z on D at $S = \bar{s}$; the sharp RDD straightforwardly applies.

Condition (6) restricts the pattern of the non-compliance behavior in the population and it is the counterpart of the monotonicity assumption by AIR in a neighborhood of the threshold $\bar{s}$; condition (7) states that the mandated status Z has a non-zero average impact on the actual status for units in the neighborhood of $\bar{s}$. These two conditions together imply that the inequality $D_i^1 \geq D_i^0$ holds at $\bar{s}$, with strict inequality for at least one unit; we will refer to these assumptions as *conditional monotonicity.*

Therefore, under SUTVA and conditional monotonicity the causal parameter $\tau_c(\bar{s})$ can be written as a local version of (4) with respect to S

$$\frac{E(Y_i|Z_i=1,S_i=\bar{s})-E(Y_i|Z_i=0,S_i=\bar{s})}{E(D_i^Z|Z_i=1,S_i=\bar{s})-E(D_i^Z|Z_i=0,S_i=\bar{s})}, \tag{8}$$

which only depends on easily estimated regression functions.

Notice that since the conditional monotonicity implies

$$E(D_i^Z|Z_i=1,S_i=\bar{s})>E(D_i^Z|Z_i=0,S_i=\bar{s}),$$

the propensity score $e(s)$ is still discontinuous at $\bar{s}$ even though - due to the presence of non-compliance - it is no longer equal to $\mathbb{1}(S \geq \bar{s})$. In this sense, a RDD in the presence of non-compliance belongs to the class of *fuzzy* rather than sharp designs (Trochim, 1984).

As for the possible bias resulting from the assumption $\tilde{Y}^T = Y^T$ we maintained throughout the previous analysis, consider the following argument. Subjects exhibiting $(Z = 0, D^0 = 1)$ (that is, exploiting AIR terminology, *always taker* units mandated out of the programme at selection) have been forced to resort to their second best choice - the alternative programme instead of the main one. Hence, one could argue that the outcome they actually experience, $\tilde{Y}^T$, is not better than the one they would have experienced, had they had access to their best choice, Y^T.

A direct implication of such a conjecture is that the intention to treat effect on Y (that is the mean causal effect of Z on Y) calculated from observed outcomes (at worse) overstates the true effect, since 'always takers' with $Z=0$ present $Y^T \geq \tilde{Y}^T$. This implies the ratio in (8) to be (at worse) upward biased for the true parameter $\tau_c(\bar{s})$.

3. Alternative estimation strategies

This section reviews two alternative estimation strategies for the parameter in (8) developed in the recent literature on programme evaluation. Broadly speaking, they both proceed directly estimating discontinuities in the regression functions of Y and D on S at $\bar{s}$, but differ in the way local information is exploited and in the set of regularity conditions required to achieve their asymptotic properties.

HTVK estimate $\tau_c(\bar{s})$ by applying a local linear regression procedure to each of the four one-sided expected values in (8). The size of the discontinuity jump at $\bar{s}$ for the regression function of X on S, where $X = \{Y, D\}$, is estimated solving the following (one-sided) optimization problems

$$\min_{x^+,b^+} \sum_{i=1}^n Z_i[X_i - x^+ - b^+(S_i-\bar{s})]^2 K(S_i-\bar{s}),$$

$$\min_{x^-,b^-} \sum_{i=1}^n (1-Z_i)[X_i - x^- - b^-(S_i-\bar{s})]^2 K(S_i-\bar{s}),$$

and taking the difference $(\hat{x}^+ - \hat{x}^-)$, where $K(.)$ are kernel weights depending on a smoothing parameter. Here $\hat{x}^+$ and $\hat{x}^-$ are the local polynomial regression estimators for

$$\lim_{s\downarrow\bar{s}} E(X|S=s) \qquad \lim_{s\uparrow\bar{s}} E(X|S=s).$$

It is worth noting that, in order to estimate the right-hand (left-hand) limit of the regression function at the discontinuity point, this procedure makes use of data only on the right (left) side of $\bar{s}$. In addition, using a local polynomial fitting technique ensures better properties of the nonparametric estimator in a neighborhood of the boundary (Fan and Gijbels, 1995). The parameter in (8) is then estimated taking the ratio of $(\hat{y}^+ - \hat{y}^-)$ to $(\hat{d}^+ - \hat{d}^-)$ which HTVK prove to be $n^{2/5}$-consistent and asymptotically normal.

As an alternative, let the regression function of X on S be

$$E(X|Z, S) = \delta Z + g(S), \tag{9}$$

where again $X = \{Y, D\}$. The discontinuity at $\bar{s}$ shifts the expected value for subjects presenting $Z = 1$ with respect to the remaining subjects over the whole support of S; therefore, the ITT effect on X corresponds to δ, the magnitude of the jump at $\bar{s}$. Intuitively, for identifiability to hold we need continuity of the unknown function $g(S)$ at $\bar{s}$ to ensure that the jump is entirely attributable to the effect of switching Z from zero to one.

Following Robinson (1988), the model in (9) is estimated in two steps taking at first the residuals ε_X and ε_Z from the nonparametric regression of X and Z on S, respectively, and then regressing ε_X on ε_Z to obtain $\hat{\delta}$ (see Appendix B for more details).

Porter (1999) shows that the resulting estimator is $\sqrt{nh}$-consistent and asymptotically normal for δ. The intuition for the less-than-$\sqrt{n}$-consistency is that unless one is willing to impose parametric restrictions on the regression curve away from the discontinuity point, only local data in a neighborhood of $\bar{s}$ contribute to estimate the jump. Asymptotically the neighborhood needs to shrink as with usual nonparametric estimation determining the rate of convergence. Exploiting the notation introduced above, it follows that

$$\delta = \lim_{s \downarrow \bar{s}} E(X|S = s) - \lim_{s \uparrow \bar{s}} E(X|S = s).$$

Hence, while HTVK (2001) estimate separately the right-hand and left-hand limits of the regression function taking the difference of the resulting estimates, Porter (1999) proceeds directly from the last expression estimating δ.

The ratio of estimated discontinuity jump sizes specifying both the regression function of Y and D on Z as in (9) is still asymptotically normally distributed and exhibits the same $\sqrt{nh}$ rate of convergence (see Porter, 1999).

It is worth making it explicit in what way the latter approach is different from the one proposed by HTVK. In essence, while the HTVK estimator requires the continuity of $g(S)$ at $\bar{s}$, the Robinson-Porter estimator also requires the continuity of some derivatives of $g(S)$ at $\bar{s}$ (see Porter, 1999). The payoff of this additional assumption will turn out clear-cut from our simulation study on the power of the two tests (see Section 5). The reader is referred to HTVK (2001) and Porter (1999) for a detailed discussion of regularity conditions needed to derive the asymptotic results mentioned so far. Small sample simulations we performed suggest that the asymptotic approximation to the behavior of both estimators works well (actually slightly better for the second one) even at the rather small sample size we work at. Results available on request.

On the other hand, such an additional assumption looks quite reasonable in our case, since rather than estimating $\tau_c(\bar{s})$ we are just testing whether it is zero. In the presence of a non-zero impact it makes sense to allow for a discontinuity of $g'(S)$ (and maybe even of higher order derivatives) at $\bar{s}$ to account for possible interaction between S and Z in which case the HTVK estimator should be preferred. On the other hand, under the hypothesis

Table 1. Data description

	Enrolled	Selected out	
	Main Programme	*Alternative Programme*	*Non-participants*
Sample size	100	20	92
Mean score at selection	69.99	41.79	41.26
At work	65%	30%	47%
Mean age	27.05	29.65	28.45
Male	79%	85%	89%
Compulsory education	2%	5%	11%
Diploma	87%	80%	81%
Degree	11%	15%	8%

'At work' is the percentage of units at work at the time of interview.

of no impact the interaction between S and Z - and the lack of smoothness in $g(S)$ thereof - is ruled out, which is exactly the (additional) requirement to apply the Robinson-Porter estimator.

4. Empirical results

The dataset we use was obtained by telephone interviewing 17 months after the programme completion (the response rate was 89%). The following information were collected: score at the attitudinal test (in the range 1 to 100), mandated programme status as determined on the basis of the attitudinal test, actual programme status and month by month labor force state in the period September 1996 - November 1997 together with the type of contract, i.e. whether the position is on a permanent basis or not (earnings are not collected). The composition of the actual data set with respect to the admittedly poor set of available characteristics is summarized in Table 1: 14 participants and 10 non-participants not in the labor force at the time of interview have been excluded from the analysis; 1 participant and 1 non-participant missing the score at the attitudinal test have been excluded, as well.

While we have some information on the main programme (see Section 1) we know very little about the alternative programme. From the interview we only know over which period it took place and the eligibility criteria.

The threshold for selection in the attitudinal test was $\bar{s} = 55$. The enrolment group turns out slightly younger and with an higher proportion of females. The mean and the median score at selection among selected out subjects are 41.35 and 45.38, respectively, with a standard deviation of 11.84; the corresponding values for the treatment group are 69.99, 68.29 and 9.23, respectively.

The outcome which we use to evaluate the programme impact is the labor force state at the interview time, namely 17 months after the programme completion. The percentage of subjects employed at that time is much higher in the treatment group (65% versus 44% in the 'selected out' group - see the third row in Table 1).

The programme impact on other outcomes such as whether or not the contract is permanent are also of primary interest. Notice however that the nature of the contract is observable only for subjects at work ($Y = 1$ in our notation) which raises a further non trivial selectivity problem (see Eberwein *et al.*, 1997). We will not deal with such a problem here.

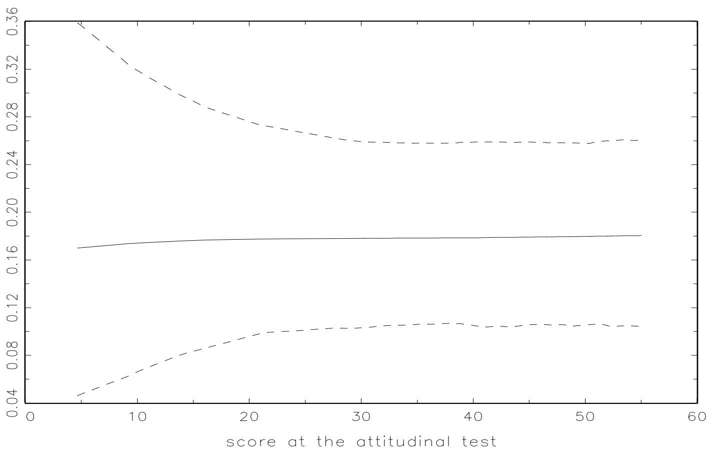

Fig. 1. Nonparametric estimate of $E(D^Z|Z=0,S)$

4.1. A preliminary result

As for the denominator in (8), we nonparametrically test for the dependence of the regression $E(D^Z|Z,S)$ on the score S. Since in our sample all subjects enrolled into the programme comply with the mandated status - namely, $Pr(D^Z=1|Z=1,S)=1$ - it only remains to test whether the equality

$$E(D^Z|Z=0,S)=E(D^Z|Z=0) \tag{10}$$

holds.

Figure 1 reports results from the nonparametric regression of D^Z on S conditional on $Z=0$ using a Nadaraya-Watson gaussian kernel estimator with constant bandwidth set at 13. The figure also reports the bootstrap 5% critical region resulting from the same smoothing procedure as applied to 1000 pseudo-samples simulated under the null hypothesis of independence.

We repeated the test over a wide range of values for the bandwidth involved in the kernel procedure coming out with no evidence of departure from (10) (the corresponding p-values are always greater than 0.4410 over the whole support of S). Accordingly, we estimate the denominator in (8) by the sample counterpart of $E(D^Z|Z=1)-E(D^Z|Z=0)$ which turns out to be as large as 0.82 with an associated standard error at 0.0363. Since under the monotonicity assumption this is an estimate of the size of the pool of compliers - see equations (5) and (6) -, the sub-population we are estimating the average programme impact on is fairly close to the whole population.

Therefore, in the next section we will estimate $\tau_c(\bar{s})$ as the ratio of the nonparametric estimates of the ITT effect on Y (exploiting the two methodologies described in Section 3) to 0.82, the estimated proportion of compliers in the population - which according to (10) - is constant over the whole support of S.

Table 2. Mean impact on compliers at selected values of the smoothing parameter

	Robinson-Porter		**HTVK**	
bandwidth	$\tau_c(\bar{s})$	*p-value*	$\tau_c(\bar{s})$	*p-value*
13	0.0996	0.3430	0.0875	0.3070
15	0.1194	0.3010	0.0757	0.3490
17	0.1377	0.2630	0.0691	0.3640

4.2. Testing the no impact hypothesis

In this section we test for the presence of any programme impact. Laudisa (1998) analyzes the same dataset considered in this paper finding no programme impact at all. Here we want to assess the robustness of such a result with respect to two critical choices she makes, namely using a linear probability model and ignoring the PATC problem by discarding non-compliers from the analysis. As for the first assumption, we do not impose the linear relationship between Y and S nor any other parametric specification. As for the second assumption, we explicitly account for the PATC problem exploiting the mandated status as an instrumental variable for the actual status. Battistin and Rettore (2000) gets the same qualitative result as Laudisa (1998) by specifying a fully parametric binary regression model to estimate each expected value in (8).

As for the monotonicity assumption required to secure the LATE interpretation of the IV estimand, we believe it is sensible in this context. It is also worth noting that since as a matter of fact $E(D^Z|Z = 1) = 1$, according to Angrist and Imbens (1991) the estimand in (8) is actually equal to $E(\beta_i|D_i^Z = 0, S_i = \bar{s})$, the mean impact of the programme on non-participants in the neighborhood of $\bar{s}$ whether or not the monotonicity assumption holds.

To begin with, let us move from the *naive* estimate one would obtain by contrasting the proportion of participants at work - including participants in the alternative programme - to the one of non-participants, ignoring the bias arising from both the selection on S and the PATC problem. The resulting figure is as large as 0.21 with an associated standard error of 0.0645. Accounting for the PATC problem (but not for the selection on S, yet) the instrumental variable estimate of τ_c as defined in (4) turns out to be 0.25 with an associated standard error of 0.0833. Apparently, it is a rather large and statistically significant programme impact.

We derive nonparametric estimates of $\tau_c(\bar{s})$ - at $h = 13, 15, 17$ (gaussian kernel, constant bandwidth) - for both the estimators presented in Section 3. To select the bandwidth constant needed for our nonparametric procedures we used a cross-validation method minimizing with respect to h the following 'leave-one-out' prediction criterion

$$\min_h \sum_{i=1}^{n} [Y_i - m^{(i)}(S_i)]^2,$$

where $m^{(i)}(s_i)$ is the estimate of the regression function of Y on S evaluated at S_i (exploiting the two alternative estimation strategies) dropping the i-th observation from the sample. A greed search method suggested $h = 15$ as the optimal value for both the estimators. We also keep $h = 13, 17$ to explore the sensitivity of our results in the neighborhood of the optimal value.

We also derive p-values for the no impact hypothesis against the alternative of positive impact simulating 1000 pseudo-samples under the null hypothesis and replicating the esti-

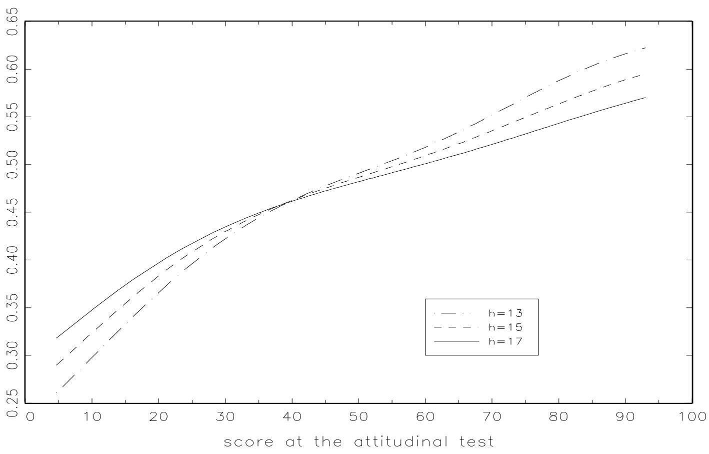

Fig. 2. Nonparametric estimate of $g(S)$ at selected value of the smoothing parameter

mation procedure on every pseudo-sample at each value of the bandwidth. Each p-value is evaluated as the proportion of pseudo-estimates exceeding the estimate obtained on the actual sample. The simulation procedure adopted is explained in Appendix C. Note that, due to the binary nature of the outcome, the disturbance in the regression of Y on S is heteroskedastic. We therefore performed a GLS-type estimation to run local linear (HTVK) and Robinson-Porter regressions, where the correction is obtained from a first round nonparametric regression of Y on S.

It is worth stressing once more that while under the alternative hypothesis the Robinson-Porter estimator might be inconsistent, under the null hypothesis it is a consistent estimator of the null impact (provided the required regularity conditions hold). Hence, it provides a legitimate basis for testing the no impact hypothesis.

Results reported in Table 2 are clear-cut: according to both the testing procedures, attending the programme does not enhance compliers' chance of being employed eighteen months after its completion. As for the estimated $\tau_c(\bar{s})$, the resulting order of magnitude is much smaller than the 0.25 estimate we got not accounting for the selection on S. Besides, estimates are remarkably robust with respect to the choice of the bandwidth value. Estimates of $g(S)$ at $h = 13, 15, 17$ reported in Figure 2 show that - as expected - the probability of being employed increases with the score S.

As we argued in Section 2.2, maintaining that the outcomes from participating in the main and in the alternative programme are equal, $\tilde{Y}^T = Y^T$, is likely to result in an overstatement of LATE. Such an argument straightforwardly implies that if the assumption was violated the results of our tests would hold *a fortiori*.

The policy relevance of our results is straightforward. From a policy point of view wondering whether any programme is worthwhile translates into the question whether its impact is large enough as compared to the cost of running it. This implies that the null

Table 3. Probability of rejecting the no impact hypothesis at selected values of LATE (the level of the test is 5%). Monte Carlo results for data generating process 1

	Robinson-Porter			**HTVK**		
$\tau_c(\bar{s})$	$h = 13$	$h = 15$	$h = 17$	$h = 13$	$h = 15$	$h = 17$
0.05	0.1110	0.1060	0.1050	0.0930	0.0760	0.0720
0.10	0.1880	0.1650	0.2310	0.1430	0.1030	0.1450
0.15	0.2980	0.3150	0.3780	0.1970	0.1840	0.2090
0.20	0.4410	0.4710	0.5670	0.2640	0.2640	0.3220
0.25	0.6260	0.6520	0.7500	0.4080	0.3690	0.4640
0.30	0.7790	0.8080	0.8700	0.5260	0.5050	0.5640
0.35	0.8760	0.9160	0.9460	0.6120	0.6250	0.7110
0.40	0.9510	0.9640	0.9870	0.7710	0.7730	0.8490

Table 4. Probability of rejecting the no impact hypothesis at selected values of LATE (the level of the test is 5%). Monte Carlo results for data generating process 2

	Robinson-Porter			**HTVK**		
$\tau_c(\bar{s})$	$h = 13$	$h = 15$	$h = 17$	$h = 13$	$h = 15$	$h = 17$
0.05	0.2160	0.3020	0.2970	0.0730	0.1320	0.1060
0.10	0.5270	0.6730	0.6790	0.2130	0.2730	0.2550
0.15	0.8030	0.8960	0.9050	0.4220	0.5130	0.5070
0.20	0.9520	0.9690	0.9760	0.6360	0.7290	0.7390
0.25	0.9800	0.9980	0.9990	0.8210	0.9080	0.9030
0.30	0.9970	0.9970	0.9980	0.9270	0.9520	0.9590
0.35	1.0000	1.0000	1.0000	0.9790	0.9870	0.9830
0.40	1.0000	1.0000	1.0000	0.9820	0.9910	0.9930

hypothesis should have been formalized as ‘impact not larger than x’, with x properly chosen, rather than ‘no impact’. Apparently, since the available evidence is not against the ‘no impact’ hypothesis, *a fortiori* it is not against the hypothesis ‘impact not larger than x’.

5. A simulation study on the power of the two tests

Given the small size of our sample one might legitimately wonder whether the considered nonparametric tests have any power at all in detecting positive programme effects, here we examine the issue in some detail.

To evaluate the power of the two tests at the size of our actual sample we run a Monte Carlo simulation drawing pseudo-samples from two sequences of data generating processes (DGP) whose elements move increasingly away from the no impact hypothesis.

The elements of the first (DGP1) and second (DGP2) sequence are represented in Figure 3 and Figure 4, respectively. All of them exhibit a positive step at $\bar{s}$; each step at the discontinuity point has exactly the same size in the two sequences. DGP1 differs from DGP2 in that its right-hand side first derivative at the discontinuity point is smaller than the left-hand side one while it is larger for DGP2. Under the no impact hypothesis DGP1 and DGP2 coincide.

Our experiment is therefore designed to assess the power of the two tests for the no impact hypothesis and to check whether the power is sensitive to the pattern of the first

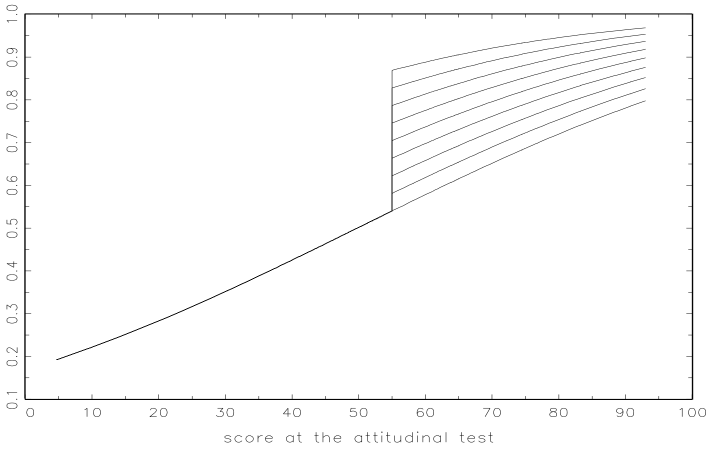

Fig. 3. Data generating process 1 for $E(Y|S,Z)$ under the alternative hypothesis

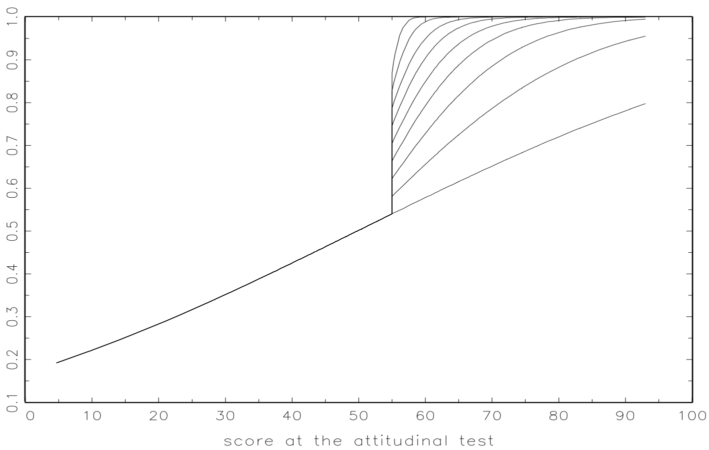

Fig. 4. Data generating process 2 for $E(Y|S,Z)$ under the alternative hypothesis

derivative at $\bar{s}$, whose critical role has been already discussed in Section 3. Both DGP's are specified building on parametric regressions exploiting the actual sample and are discussed in Appendix C.

Table 3 and Table 4 present the power function for the two tests of the no impact hypothesis against the alternative provided by DGP1 and DGP2, respectively, at different values of h. Critical (unilateral) regions are obtained from the empirical distributions of the Robinson-Porter and the HTVK procedures under the null hypothesis taking the proportion of estimates exceeding a certain critical threshold $\mathcal{C}$ (see Appendix C for more details). The size of the test is kept at 5%.

Overall, the probability of rejecting the null hypothesis is non-negligible even at very small values of the impact. This probability is quite a lot larger under DGP2 for both tests suggesting that the pattern of the first derivative at the discontinuity point plays a key role in determining the behavior of the two tests.

The Robinson-Porter test is definitely more powerful then the HTVK one, uniformly over h and over the considered range for $\tau_c(\bar{s})$. In particular, under DGP2 the probability of detecting a true positive effect is appreciably higher even at very small values of $\tau_c(\bar{s}) = 0.10$.

6. Conclusions

We summarize our results in few statements. Available evidence is definitely against the presence of a positive impact of the training programme on trial in this paper on the labor force status 17 months after the programme completion. We get such a conclusion testing the no impact hypothesis in a nonparametric setting exploiting results by Hahn *et al.* (2001) and by specifying a further test based on Robinson (1988) and Porter (1999).

The previous statement needs two qualifications. Firstly, to assess the power of the two tests we performed a small simulation exercise. The power turns out to be non-negligible even at small departures from the null hypothesis. Indeed, under one of the generating processes we considered it turns out appreciably high.

Secondly, strictly speaking the statement applies to compliers exhibiting a score in the attitudinal test in the neighborhood of the threshold for selection. On the one hand the presence of non-complying subjects does not seem to be a major problem here, since the proportion of non-compliers is not large (around 18%). The average impact on the whole population would be far away from the figure we found only if the average impact on non-complying subjects is very different from the one on compliers.

Besides, as a matter of fact in our case study all subjects assigned to the treatment comply with the mandated status implying that the average impact on compliers does coincide with the average impact on non-participants. This is an interesting policy parameter *per se* since it tells us that it does not make any sense trying to include non-participants (scoring close to the threshold for selection) in the programme.

On the other hand, what actually precludes generalizing the results to the whole population is the selection on the score at the attitudinal test with just one threshold value for selection. From the case study at hand we learn that the programme is not worthwhile for non-participants scoring $\bar{s}$ at the attitudinal test. From this result we cannot infer whether the programme benefits subjects at levels of the score far away from $\bar{s}$. The existence of multiple discontinuity points that densely cover the full support of S would lead us to identify the causal effect for compliers by averaging $\tau(\bar{s})$ over the support of S.

Lastly, the simulation exercise we performed - albeit limited in its extent since it is restricted to a given sample size as the one in our sample - provides clear-cut evidence that the Robinson (1988) - Porter (1999) test for the no impact hypothesis outperforms the one based on Hahn *et al.* (2001) in terms of power. As usual, there is a price to pay in terms of additional regularity condition requirements to achieve this higher power. Whether or not the additional requirements are reasonable is problem-specific but, broadly speaking, they seem very likely to fit most contexts.

7. Appendix A

Here is a brief outline based on Hahn *et al.* (1999) of the regularity conditions needed for identifiability of $\tau_c(\bar{s})$. Consider the behavior of the conditional means of Y and D^Z in a neighborhood of $\bar{s}$. Since $Z = \mathbb{1}(S \geq \bar{s})$, the difference

$$E(Y_i|Z_i = 1, S_i = \bar{s} + \varepsilon) - E(Y_i|Z_i = 0, S_i = \bar{s} - \varepsilon) \tag{11}$$

equals

$$E(Y_i|S_i = \bar{s} + \varepsilon) - E(Y_i|S_i = \bar{s} - \varepsilon),$$

where ε denotes an arbitrary small positive number.

Replacing Y_i by its definition in (1), we get the sum of the following two terms

$$\begin{gathered}\left[E(Y_i^{NT}|S_i = \bar{s} + \varepsilon) - E(Y_i^{NT}|S_i = \bar{s} - \varepsilon)\right], \\ \left[E(D_i^1\beta_i|S_i = \bar{s} + \varepsilon) - E(D_i^0\beta_i|S_i = \bar{s} - \varepsilon)\right].\end{gathered}$$

As $\varepsilon \to 0$ the first term in brackets vanishes if and only if

Condition 1: $E(Y^{NT}|S = s)$ is continuous at $S = \bar{s}$,

which is the key regularity assumption stated by HTVK (1999). As for the second term in brackets, by adding and subtracting $E(D_i^0\beta_i|S_i = \bar{s} + \varepsilon)$ it turns out equal to

$$E\left[(D_i^1 - D_i^0)\beta_i|S_i = \bar{s} + \varepsilon\right] + \left[E(D_i^0\beta_i|S_i = \bar{s} + \varepsilon) - E(D_i^0\beta_i|S_i = \bar{s} - \varepsilon)\right].$$

Provided that Condition 1 as well as the following

Condition 2: $E(D^1\beta|S = s)$ and $E(D^0\beta|S = s)$ are continuous at $S = \bar{s}$,

hold, the difference (11) tends to $E\left[(D_i^1 - D_i^0)\beta_i|S_i = \bar{s}\right]$ as $\varepsilon \to 0$, from which the decomposition (5) is straightforwardly derived.

By applying similar argument to

$$E(D_i^Z|Z_i = 1, S_i = \bar{s} + \varepsilon) - E(D_i^Z|Z_i = 0, S_i = \bar{s} - \varepsilon), \tag{12}$$

we get that the condition

Condition 3: $E(D^1|S = s)$ and $E(D^0|S = s)$ are continuous at $S = \bar{s}$

is sufficient for the convergence of (12) to the ITT of Z on D^Z at $\bar{s}$ as $\varepsilon \to 0$.

By means of the notion of continuity, Condition 1-3 formalizes the intuition that to identify the causal parameter in a RDD context subjects in the neighborhood of $\bar{s}$ have to be approximately alike one to each other (except for the mandated status).

8. Appendix B

Following the notation in Section 3, let the outcome Y be determined by the generating process

$$Y = \delta Z + g(S) + \varepsilon, \tag{13}$$

where $Z = \mathbb{1}(S \geq \bar{s})$. As an implication, the conditional mean value $E(Y|S)$ as a function of S is discontinuous at $S = \bar{s}$, the step being equal to δ.

Let

$$\bar{Y}(s) = \int E(Y|S=l)k(\frac{l-s}{h})dl \tag{14}$$

be the mean value of the smoothed regression of Y on S as a result of using a symmetric kernel function k depending on the constant bandwidth h. By applying the same smoother to the right hand side of (13), simple calculations yield

$$\bar{Y}(s) = \delta K(\frac{s-\bar{s}}{h}) + \bar{g}(s),$$

where $\bar{g}(s) = \int g(l)k(\frac{l-s}{h})dl$ and K is the cumulative distribution function associated to the kernel function.

By subtracting (14) from (13) we get

$$Y - \bar{Y}(s) = \delta \left[Z - K(\frac{s-\bar{s}}{h}) \right] + [g(S) - \bar{g}(s)] + \varepsilon.$$

Let $\varepsilon_Y(s) = Y - \bar{Y}(s)$, $\varepsilon_Z(s) = Z - K(\frac{s-\bar{s}}{h})$ and $\varepsilon_g(s) = g(S) - \bar{g}(s)$. The Robinson (1988) semiparametric estimator of δ is obtained by regressing (the sample analogue of) $\varepsilon_Y(s)$ on (the sample analogue of) $\varepsilon_Z(s)$.

The question is whether omitting the term $\varepsilon_g(s)$ (which we do not observe) from such regression results in any bias in the estimation of δ. Under the regularity conditions given by Porter (1999), asymptotically this omission does not bias the estimator as far as h goes to zero at an appropriate rate as the sample size grows to infinity.

Notice however that a stronger result might hold in a RDD context. Standard results from nonparametric estimation of functions (see for example Simonoff, 1996) yield

$$g(S) - \bar{g}(s) = -\frac{1}{2}g''(s)h^2 + o(h^2),$$

implying that the omitted regressor bias depends on the pattern of the second derivative of $g(s)$ in the neighborhood of $\bar{s}$. In particular, if $g(s)$ is approximately linear in the neighborhood of $\bar{s}$, the omitted term vanishes (up to the second order of the Taylor expansion) no matter for the bandwidth value and the Robinson estimator turns out unbiased even in finite samples. Notice that to preserve the finite sample unbiasedness of the Robinson-Porter estimator in the presence of a constant but non-zero second derivative in the neighborhood $S = \bar{s}$ it only takes to include an intercept in the regression of $\varepsilon_Y(s)$ on $\varepsilon_Z(s)$.

9. Appendix C

9.1. Evaluating the p-values for the no impact hypothesis

To evaluate the p-values for the no impact hypothesis against the alternative of a positive effect exploiting each of the two tests presented in Section 3 we perform the following simulation.

Table 5. Properties of the considered estimators under the null hypothesis

	Robinson-Porter			HTVK		
bandwidth	*Bias*	*Std.Dev.*	*MSE*	*Bias*	*Std.Dev.*	*MSE*
13	0.0581	0.1273	0.0196	0.0016	0.1749	0.0306
15	0.0608	0.1124	0.0163	-0.0006	0.1656	0.0274
17	0.0671	0.1056	0.0156	0.0011	0.1631	0.0266

(a) Select a value for the bandwidth parameter h (as discussed in Section 4.2) to obtain the Nadaraya-Watson estimate of $g(S)$ under the null hypothesis. Also obtain a Nadaraya-Watson estimate of the non-compliance generating process $Pr(D = 1|S, Y, Z = 0)$.
(b) Generate a pseudo-sample simulating a value Y for each unit in the actual sample from a Bernoulli random variable with probability $g(S)$. As for the simulation of the non-compliance behavior, conditional on the simulated value Y generate a value D for units exhibiting $Z = 0$ from a Bernoulli random variable with probability $Pr(D = 1|S, Y, Z = 0)$.
(c) Evaluate HTVK and Robinson-Porter estimates of (8) on the pseudo-sample.

Steps (b) and (c) are replicated 1000 times and the p-values are evaluated as the fraction of pseudo-estimates exceeding the actual one. Steps (a) to (c) are replicated at $h = 13, 15, 17$ leading to results reported in Table 2.

Table 5 reports estimated bias and mean squared error for both the estimators applying the procedure so far presented. HTVK performs definitely better in terms of bias while - as expected - the additional information exploited by Robinson-Porter leads to an appreciably lower variance. Overall, the comparison between the HTVK and the Robinson-Porter estimator in terms of MSE supports the latter one. Skewness, kurtosis and difference between mean and median over standard deviation for the empirical distributions under the null hypothesis do not vary a lot over the considered range for the smoothing parameter; their average values with respect to h are -0.0553, -0.0025 and 0.0020 for Robinson-Porter and -0.0832, -0.0164 and 0.0105 for HTVK, respectively.

9.2. Evaluating the power functions

The critical choice here concerns the path the true model takes departing from the no impact hypothesis towards positive values of the impact. To obtain a fair comparison between the power function of the two tests we move from the fact that HTVK's estimator is consistent whether or not $g(S)$ is differentiable at $\bar{s}$ whereas the alternative estimator needs $g(S)$ to be properly smooth in a neighborhood of the discontinuity point. We then specify models such that under the alternative hypothesis a step takes place at $\bar{s}$ both in the level and in the slope of $E(Y|S, Z)$.

(a) A Probit regression of Y on S is run on the actual sample yielding

$$E(Y|S) = \Phi(-0.9599 + 0.0193S).$$

(b) The variable Z is included in the previous model with a positive coefficient to get a positive step at $\bar{s}$, that is

$$E(Y|S, Z) = \Phi(-0.9599 + 0.0193S + \delta Z).$$

Because of the values for the intercept, the coefficient associated with S and the location of the discontinuity point this model also features a negative step for the slope at $\bar{s}$. We will refer to this model as DGP1.
As for the simulation of the non-compliance behavior, we simulate D conditional on Y and S from a Bernoulli random variable with probability

$$E(D|Y, S, Z = 0) = \Phi(-0.9644 + 0.0050S - 0.4178Y),$$

where the parameters have been estimated exploiting the actual sample. Notice that since in our case study non-compliance takes place only conditional on $Z = 0$, the probability of observing a non-complying subject is unaffected by the value we attach to δ. Hence, as we depart from the no impact hypothesis along the sequence of models indexed by the value of δ the denominator of the IV estimand in (8) remains constant. We chose the values for δ such that the resulting sequence of steps for $\tau_c(\bar{s})$ is as large as $0.05, 0.10, \ldots, 0.35, 0.40$.

(c) The 5% critical values $\mathcal{C}$ for both the HTVK and Robinson-Porter tests are obtained estimating (8) on 1000 pseudo-samples from the model in Step (a) and taking $\mathcal{C}$ such that the proportion of estimates larger than $\mathcal{C}$ is equal to 5%.

(d) Finally, the power of the two tests at δ is evaluated simulating 1000 pseudo-sample from the specified model, estimating (8) on each pseudo-sample and taking the proportion of estimates greater than $\mathcal{C}$.

(e) To further explore the performance of the two tests we consider an alternative model featuring a positive step for the slope at $\bar{s}$, instead of a negative one as in Step (b). This is accomplished by taking

$$E(Y|S, Z) = \Phi(-0.9599 + 0.0193S + \delta Z + \alpha Z[S - \bar{s}]),$$

with α set at zero when δ is zero and properly increasing with δ in order for the first derivative of $E(Y|S, Z)$ to feature a positive step at $\bar{s}$. We will refer to this model as DGP2. The values of δ are the same as those specified in Step (b); it follows that DGP1 and DGP2 features exactly the same step in the probability $E(Y|S, Z)$ at $\bar{s}$ but different values of their right-hand first derivatives at the discontinuity point. Step (d) above is then replicated simulating the pseudo-samples from the current model.

The resulting paths of departure from the no impact hypothesis under DGP1 and DGP2 as (δ, α) take values in the considered sets are represented in Figure 3 and Figure 4, respectively.

References

Angrist, J.D. and Lavy, V. (1999) Using Maimonides' Rule to Estimate the Effect of Class size on Scholastic Achievement. *The Quarterly Journal of Economics*, **114**, 533-575.

Angrist, J.D. and Imbens, G.W. (1991) Sources of Identifying Information in Evaluation Models. *NBER Technical Working Paper*, **117**.

Angrist, J.D. and Imbens, G.W. (1999) Comment on James J. Heckman 'Instrumental Variables. A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations'. *The Journal of Human Resources*, **XXXIV**, 823-827.

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996) Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, **91**, 443-455.

Balke, A. and Pearl, J. (1997) Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92**, 1171-1176.

Battistin, E. and Rettore, E. (2000) The Impact of a Vocational Training Program on the Probability to Get a Job. *Atti della XL Riunione Scientifica, Società Italiana di Statistica*, 533-536.

Berk, R.A. and De Leeuw, J. (1999) An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design. *Journal of the American Statistical Association*, **94**, 1045-1052.

Cook, T.D. and Campbell, D.T. (1979) *Quasi-Experimentation. Design and Analisys Issues for Field Settigs.* Houghton Mifflin Company, Boston.

Eberwein, C., Ham, J.C. and Lalonde, R.J. (1997) The impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data. *Review of Economic Studies*, **64**, 655-682.

Fan, J. and Gijbels, I. (1995) *Local Polynomial Modeling and Its Application - Theories and Methodologies.* New York: Chapman and Hall.

Hahn, J., Todd, P. and Van der Klaauw, W. (1999) Evaluating the Effect of an Antidiscrimination Law using a Regression-Discontinuity Design. *NBER Technical Working Paper*, **7131**.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001) Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, **69**, 201-209.

Heckman, J.J. (1997) Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources*, **XXXII**, 441-462.

Holland, P. (1986) Statistics and Causal Inference. *Journal of the American Statistical Association*, **81**, 945-970.

Imbens, G.W. and Angrist, J.D. (1994) Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**, 467-476.

Imbens, G.W. and Rubin, D.B. (1997) Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies*, **64**, 555-574.

Laudisa, F. (1998) *Come valutare l'efficacia dei corsi di formazione professionale.* Master Dissertation Thesis, COREP, Torino.

Porter, J.R. (1999) *A Note on the Estimation of Regression Discontinuities.* Unpublished manuscript, Harvard University.

Robinson, P.M. (1988) Root-n-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.

Rubin, D.B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psycology*, **66**, 688-701.

Rubin, D.B. (1977) Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, **2**, 4-58.

Simonoff, J.S. (1996) *Smoothing Methods in Statistics.* New York: Springer Verlag.

Trochim, W. (1984) *Research Design for Program Evaluation: the Regression-Discontinuity Approach.* Beverly Hills: Sage Publications.

Van der Klaauw, W. (2000) Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. Unpublished manuscript, University of North Carolina.

Working Papers già pubblicati

1. E. Battistin, A. Gavosto e E. Rettore, *Why do subsidized firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy*, Agosto 1998.
2. N. Rosati, E. Rettore e G. Masarotto, *A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models*, Agosto 1998.
3. U. Trivellato, *Il monitoraggio della povertà e della sua dinamica: questioni di misura e evidenze empiriche*, Settembre 1998.
4. F. Bassi*, Un modello per la stima di flussi nel mercato del lavoro affetti da errori di classificazione in rilevazioni retrospettive*, Ottobre 1998.
5. Ginzburg, M. Scaltriti, G. Solinas e R. Zoboli, *Un nuovo autunno caldo nel Mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali*, Ottobre 1998.
6. M. Forni e S. Paba, *Industrial districts, social environment and local growth. Evidence from Italy*, Novembre 1998.
7. B. Contini*, Wage structures in Europe and in the USA: are they rigid, are they flexible?*, Gennaio 1999.
8. B. Contini, L. Pacelli e C. Villosio, *Short employment spell in Italy, Germany and Great Britain: testing the "Port-of-entry" hypothesis*, Gennaio 1999
9. B. Contini, M. Filippi, L. Pacelli e C. Villosio*, Working careers of skilled vs. unskilled workers*, Gennaio 1999
10. F. Bassi, M. Gambuzza e M. Rasera, *Il sistema informatizzato NETLABOR. Caratteristiche di una nuova fonte sul mercato del lavoro*, Maggio 1999.
11. M. Lalla e F. Pattarin, *Alcuni modelli per l'analisi delle durate complete e incomplete della disoccupazione: il caso Emilia Romagna*, Maggio 1999.
12. A. Paggiaro*, Un modello di mistura per l'analisi della disoccupazione di lunga durata*, Maggio 1999.
13. T. Di Fonzo e P. Gennari*, Le serie storiche delle forze di lavoro per il periodo 1984.1-92.3: prospettive e problemi di ricostruzione*, Giugno 1999.
14. S. Campostrini, A. Giraldo, N. Parise e U. Trivellato*, La misura della partecipazione al lavoro in Italia: presupposti e problemi metodologici di un approccio "time use"*, Ottobre 1999.
15. A. Paggiaro e N. Torelli, *Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro*, Ottobre 1999.
16. A. D'Agostino, G. Ghellini e L. Neri*, A Multiple Imputation Method for School to Work Panel Data*, Ottobre 1999.
17. G. Betti, B. Cheli e A. Lemmi, *Occupazione e condizioni di vita su uno pseudo panel italiano: primi risultati, avanzamenti e proposte metodologiche*, Ottobre 1999.
18. B. Anastasia, M. Gambuzza e M. Rasera*, La durata dei rapporti di lavoro: evidenze da alcuni mercati locali del lavoro veneti*, Marzo 2000.
19. F. Bassi, M. Gambuzza e M. Rasera*, Struttura e qualità delle informazioni del sistema NETLABOR. Una verifica sui dati delle Scica delle province di Belluno e Treviso*, Marzo 2000.
20. N. Rosati, *Permanent and Temporary Inequality in Italy in the 1980s and 1990s*, Marzo 2000.
21. G. Betti, B. Cheli e A. Lemmi, *Analisi delle dinamiche di povertà e disoccupazione su uno pseudo panel italiano*, Marzo 2000.
22. A. D'Agostino, G. Ghellini e L. Neri, *Modelli statistici per l'analisi dei comportamenti di transizione scuola lavoro*, Marzo 2000.

23. A. Paggiaro e U. Trivellato, *Assessing the effects of the "Mobility List" programme in an Italian region: do (slightly) better data and more flexible models matter?,* Marzo 2000.
24. F. Bassi, M. Gambuzza, M. Rasera e E. Rettore, *L'ingresso dei giovani nel mercato del lavoro: prime esplorazioni dall'archivio Netlabor*, Giugno 2000.
25. A. D'Agostino, G. Ghellini e L. Neri, *Percorsi di ingresso dei giovani nel mercato del lavoro*, Giugno 2000.
26. E. Battistin, E. Rettore e U. Trivellato, *Measuring participation at work in the presence of fallible indicators of labour force state,* Giugno 2000.
27. E. Battistin e E. Rettore, *Testing for the presence of a programme effect in a regression discontinuity design with non compliance*, Novembre 2000.
