Lavoro: partecipazione, dinamica e valutazione di politiche

Misura, metodi, modelli

Progetto anno 2005

Evaluating the Effects of Business Incentives Policies: A Discussion Note on Identification Strategies

Daniele Bondonio

Dipartimento di Politiche Pubbliche e Scelte Collettive
Università del Piemonte Orientale

Working Paper n. 76, luglio 2008

La valutazione dell'impatto di interventi pubblici: metodi e studi di caso.

Cofinanziamento MIUR, anno 2005

Unità locali del progetto:

Dip. di Economia "S. Cognetti De Martiis", Univ. di Torino	(coord. B. Contini)
Dip. di Statistica "G.Parenti", Univ. Di Firenze	(coord. F. Mealli)
Dip. di Scienze Statistiche, Univ. di Padova	(coord. E. Rettore)
Dip. di Scienze Economiche, Univ.di Salerno	(coord. S. Destefanis)
Dip. di Politiche Pubbliche e Scelte Collettive, Univ. del Piemonte Orientale	(coord. D. Bondonio)

Dipartimento di Scienze Statistiche
via C. Battisti 241-243, 35121 Padova
www.valutazione2003.stat.unipd.it

Per la presente pubblicazione sono stati adempiuti gli obblighi previsti dalle norme per la consegna obbligatoria di esemplari degli stampati e delle pubblicazioni di cui alla legge del 2 febbraio 1939 n.374 e successive modificazioni, presso la Procura della Repubblica e la Prefettura di Padova.

1. Introduction[*]

In recent years the importance of business incentives policies as economic development tools has been increasing throughout the EU and the US. Despite the importance of such policies, the empirical evidence based on rigorous evaluations on their impact on desirable economic outcomes is still confined to a limited number of programs. Impact evaluations, produced with a rigorous counterfactual approach, do not constitute yet a widespread tool adopted by EU policy makers (and, with more exceptions, by US policy makers) in order to refine future business incentive interventions. Such lack of a greater diffusion of rigorous evaluation studies on business incentive policies may be due to the fact that assessing the causal link between the incentives and the outcome of the evaluation is a particularly difficult task. Many confounding factors (often including a number of competing public programs other than the policy being evaluated) affects the outcome of the evaluation at the same time as the program intervention. Moreover, extensive firm-level (or geographically aggregated) databases, often difficult to assemble, are needed by rigorous impact evaluations.

This paper aims at contributing to the evaluation literature on business incentive policies by offering a useful tool guide to navigate the many possible options in choosing the appropriate outcome data and impact identification strategy for the analysis. Each option presented in the paper is discussed based on the program characteristics and whether or not the analysis focus on socially desirable outcomes recorded at the level of assisted firms or at the level of the geographic areas in which the program incentives are in place.

The reminder of the paper is organized as follows. Section 2 is devoted to the choice of the outcome variable for the evaluation. Section 3 illustrates the policy relevant impact evaluation parameters. Section 4 discusses the feasibility of randomized experiments. Section 5 illustrates impact identification strategies in non-experimental settings. Section 6 offers some concluding remarks, discussing the issue of evaluations of single programs versus evaluations of multiple programs.

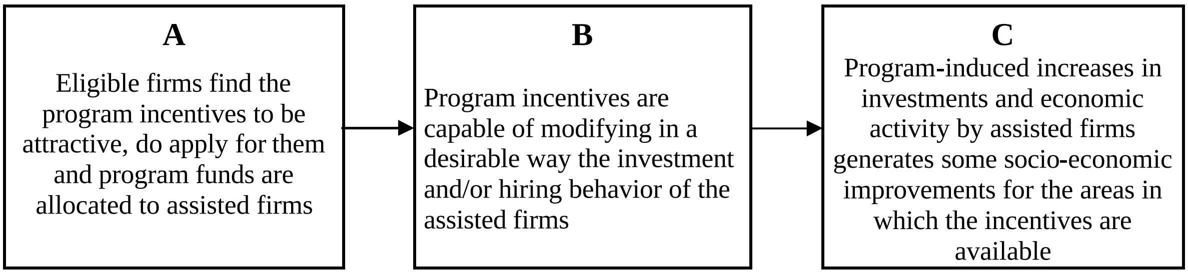
[*] The work that provides the basis for this paper was supported by funding under the PRIN-COFIN 2005 grant from MIUR (Italian “Ministry of Education, University and Research”). A shorter version of this paper is forthcoming in the Proceedings of the XLIV Conference of the Italian Society of Statistics.

ADDRESS FOR CORRESPONDENCE: daniele.bondonio@sp.unipmn.it

2. Choosing the outcome variable of the analysis

In general terms, business incentive programs can produce desirable socio-economic outcomes through the following chain of causal links:

Figure 1: *Causal links from business incentives to desirable socio-economic outcomes*



In order for a program to succeed, eligible firms need to find the program incentives to be attractive and they do have to apply for them. Measuring whether or not a program is capable of producing outcomes A, however, does not provide any kind of impact evaluation of the policy. This is because, even if all program funds are allocated to applicant firms, the actual impact of the program could be zero in the event that all assisted firms would have made the same investment, or hired the same number of workers, even in the absence of the incentives.

Focusing the analysis on outcomes A, therefore, is only aimed at assessing whether or not the management of the program was effective in designing desirable incentive packages, marketing the program among eligible firms and properly handling the program application process. Even if the program activity data are produced in the form of business outcomes, such as the number of jobs or the volume of investments generated by assisted firms, this type of analysis is not to be mistaken as an actual impact evaluation of the program, as it is still done in quite some number of reports commissioned by regional or state economic development agencies (both in the EU and in the US, as reported in Bartik 2004 and Bondonio and Greenbaum 2006) which erroneously assume that none of the business activity recorded in the assisted firms would have occurred in the absence of the program.

Proper impact evaluation analyses involve assessing whether or not the program incentives produce outcomes of type B or C. Assessing whether or not business incentive programs achieve outcomes of type B, very often, requires acquiring, for both assisted and non-assisted firms, longitudinal data recording firm-level employment, capital expenditures, or sales. Differentiation of the firm-level outcome data (y) between pre- and post-treatment times, needed to eliminate the selection bias due to correlation between unobserved fixed effects and the treatment, should be performed as absolute changes rather percentage changes. This is because the social benefit of each additional job/unit-of-investment/sales generated by the program incentives (compared to what would have happened in the absence of the program) is to be weighted equally whether or not the job/unit-of-investments/sales is generated in a small or large firm. In cases in which the data on the incentive payments are not very precise as the exact location in time, long differencing of y (between a pre- and a post-treatment time) is to be preferred to differencing with very detailed measures of time changes.

Impact evaluations focusing on the distant outcomes of type C should be performed mainly for programs targeting only specific geographic areas, such as, for example, the US state and federal Enterprise Zones, the "Zones Franches Urbaine" of France, the proposed "Zone Franche Urbane" of Italy, and, by some degree, the incentives co-funded by the EU structural funds in "Objective 2 areas" (see for example: Bondonio and Greenbaum 2007, O'Keefe 2004, Engberg and Greenbaum 1999, Boarnet and Bogart 1996). In such cases the economic weight of the program incentives is not disproportionally small compared to the size of the economy of the target areas, and appropriate evaluation models are capable of identifying the program impact on the target areas outcomes, controlling for the major confounding factors. Impact evaluation focused on outcomes of type C call for using geographically aggregated data on firm-outputs (such as employment, capital investments, sales), residents employment rate, per-capita income or indicators of improvements on the overall desirability of the target areas (such as housing values). In general terms, differentiation of the outcome variable $y_{i.t}$ (being i the geographic unit of the analysis) should take the form of percentage changes rather than absolute changes. This is because for outcomes of type C, the intensity of the social benefits of the program-induced absolute change in y depends on the pre-intervention size of the target areas communities.

In some cases, policy makers do also show interest in knowing program impacts on outcomes of type C even for incentives programs lacking specific geographic targeting. In principle, business incentives programs of all sorts are somehow capable of affecting distant outcomes, such as macro-economic indicators of the well-being of residents measured at the level of the entire provinces, regions, or states in which eligible firms are located. In the vast majority of cases, however, the economic importance of the group of assisted firms, compared to the size of the province/region/state economy in which they are located is almost negligible. As a result, any actual program impact (in the form of a positive impulse given to the province/region/state economy) become virtually undetectable from the changes to the outcome variable of the evaluation caused by many confounding factors (including, in many cases, other competing public programs) of a much greater importance than the possible program-induced improvements in the economic activity of the assisted firms.

Using rigorous impact evaluation designs to assess whether or not business incentives had long-lasting impacts on employment or economic activity outcomes of assisted firms is also often to be avoided. Assisted firms are economic units embedded in many ways in a network of economic transactions from ones to the others. In the medium/long-run, a possible positive program impulse produced on the assisted firms employment or economic activity is likely to have enough time to generate subsequent impacts also on non-assisted firms, those outcome data become endogenous to the treatment and cannot anymore be used to retrieve counterfactual estimates.

As a result, estimating the impact of business incentive policies in terms of long-run macro-economic or employment benefits for an overall province/regional/state economy, should be attempted exclusively when the importance of the economic outputs of the assisted firms is not disproportionally smaller than the size of the local economy, and only using regional macroeconomic simulation models (such as REMI - Regional Economic Models *inc.*, (Fan, Treyz e Treyz 2000). In such cases, analyses with regional macroeconomic simulation models, however, should be performed only after having rigorously estimated the program impact on outcomes of type B. Lacking reliable evidence on the program impacts on the proximate outcomes recorded at the level of the assisted firms, the evaluation outcomes produced by regional macroeconomic simulation models would be upward biased. This is because the set of multipliers used by such models would be applied directly to the entire volume of jobs or investments generated by the assisted firms, instead to only the number of additional jobs or new investments that the assisted firms would have not generated being absent the program incentives.

3. Policy relevant parameters: ATTs versus distributions of the treatment effects

Let's define for each unit of observation a set of potential outcomes, one denoted by $y^{(0)}$, indicating the outcome that would be observed if unit *i* received no treatment of any kind, and the other ones denoted by $\{ y_x^{(1)} \}_{x \in X}$, indicating the outcome of receiving a categorical treatment of type *x*, with $\{x = 1,2,\ldots,X\}$ being the different discrete treatment categories. $T_x \in \{0,1\}$ is a binary indicator for the treatment of category *x* received (with $T_x=0$ corresponding to no treatment, and $T_x=1$ corresponding to treatment). In case of a single treatment category, notations simplifies to $y^{(0)}$, $y^{(1)}$, $T \in \{0,1\}$.

The policy-relevant parameters, which are of most interest in the impact evaluation of business incentives programs, are ATTs (Average Treatment Effect on the Treated), estimated either for a single category of treatment and for the entire population of treated units, or for a number of different categories of treatment and for different subpopulations of treated units.

In cases when both the characteristics of the incentives and the pre-intervention observable covariates of the treated units are all fairly homogeneous, to obtain policy relevant empirical evidence is sufficient to estimate:

$$t = E[\,y^{(1)} - y^{(0)} \mid T=1], \qquad (1)$$

which represents the classic ATT parameter, for an homogenous binary treatment.

When, instead, the treatment has quite different economic values across the population of treated units, or when the treatment impact is expected to be different according to different pre-intervention observable characteristics (*W*) of the treated units, policy relevant empirical evidence is obtainable by estimating different ATTs for different subpopulations of the treated units and/or for different treatment categories:

$$t(x, w) = E[\,y_x^{(1)} - y^{(0)} \mid T_x=1, W=w]. \qquad (2)$$

In such cases, policy relevant empirical evidence is typically obtainable when the different treatment categories x are in the form of different ranges of economic values of the incentives, and/or in the form of different types of benefits granted to assisted firms, such as below-market-interest-rate loans versus capital grants or tax credits.

Estimating different impacts for different categories of the economic value of the incentives is of interest to policy makers because one of the most useful pieces of empirical evidence (in order to redefine future policy interventions) is the cost of the program per each additional unit of desirable outcome induced by the program. Discrete categories of the economic treatment intensities are often of more policy interest than continuous specifications. This is because, often, the information leading to the operationalization of data on the economic value of the incentives are based on Net Equivalent Subsidy (NES) figures. Computing such figures is very data demanding, and often results in computations of the economic values of the incentives that, because of the presence of significant noise, do signal actual differences in treatment intensity only across fairly wide apart NES values.

Estimating different impacts for categories of treatments based on whether or not the program incentives are under the form capital grants or below-market-interest-rate loans is of interest to policy makers for the following reasons. Below-market-interest-rate loans are more economical than capital grants, in the sense that with the same amount of public funds, the loans allow the government to provide incentives to a much larger number of assisted firms, generating a greater leverage than with capital grants. The latter, however, by giving firms a financial advantage largely superior to that of below-market interest rate loans, are less prone to the risk of dead-weight loss. In fact, they hold the potential of having a larger impact on the decisions that assisted firms make regarding their investment and employment levels, when compared to what it would have occurred without the benefits of the program.

While ATTs parameters do have a great policy relevance for business incentive programs, this may not be the case for estimates of the distribution of treatment effects (measured, for example, as the proportion of assisted firms for whom there is $y^{(1)}$- $y^{(0)}$ =0 or $y_x^{(1)}$ - $y^{(0)}$ =0). This is because (at least for estimating treatment impacts on economic or employment outcomes of assisted firms -outcomes of type B in Figure 1) socially desirable outcomes do not arise from the well-being of assisted entrepreneurs or of stockholders of assisted firms. Rather, socially desirable outcomes (in the form of economic or employment outcomes) are achieved, for example, when new jobs attributable to the program incentives are generated in a declining local economy[1]. In such cases, it could happen that similar socially desirable outcomes are obtained whether or not the creation of new jobs occurs evenly on the entire spectrum of assisted firms or it is concentrated only among some of the assisted firms. Therefore, once ATTs of the program are estimated, the distribution of treatment effects may play a smaller role in being of significant policy relevance.

4. Randomized experiments

Both in the EU and in the US, virtually no business incentives policy has been implemented with a randomized experiment scheme (the only exception being a small US entrepreneurship training program sponsored by the Department of Labor in the early nineties, Bartik 2004). Ethical and political difficulties in excluding some eligible firms or target areas from the incentives are likely to have prevented a more significant adoption of randomized experiments.

In the case of business incentive polices with no specific geographic target, however, as suggested in Bartik (2004), such difficulties could be eased if the experimentation would take the form of random selection of firms for targeted marketing of the program. If such randomly assigned marketing efforts are strong enough, the result should be some sharp difference in the usage of the program incentives between the treated firms (those receiving the marketing efforts) and the control-group firms (those not receiving the marketing efforts). The results would be a source of variation in program usage that does not directly affect the outcome variable of the evaluation.

[1] This is not necessarily the case when the analysis focus instead on estimating the impact of R&D incentives on measures of firms' innovation outcomes.

For incentives programs with specific geographic target, such as the proposed "Zone Franche Urbane" (ZFU) of Italy, ethical and political difficulties associated with experiments could be eased with a delay-of-treatment randomization scheme having the following features. First, a number of eligible target areas are designated based on the policy indicators of socio-economic distress. Second, data collection starts for all eligible target areas, gathering information (measured at the same precise geographic level of the target areas) that can be used as outcome variables of the evaluation (such as, for example, employment, per-capita income, poverty rate). Third, eligible areas are randomly assigned to a treatment and a control group. In the treatment group, the ZFU incentives are started immediately, while the control group (instead of having denied the ZFU incentives all together) experience a delay of two-three years in the beginning of the program. In such period of time, randomized experimental conditions are in place, while ethical and political difficulties are eased by the fact that treatment in the control group is delayed and not denied.

5. Impact identification strategies with non-experimental data

Table 1 summarizes some of the identification strategies that has been used (or that are usable) in rigorous counterfactual evaluation studies of US and EU incentives programs of significant importance. What follows is a concise review of each identification strategy, discussing the extent to which each strategy is a good fit for the different types of programs to be evaluated, the different choices of the outcome variables of the evaluation and the different scenarios of data availability.

Exploiting geographical natural experiment conditions *(GNEC)* has been used to evaluate state-wide incentives/tax programs in the US, focusing on data from communities crossed by state borders determining differences in the treatment status of firms (e.g. Holmes 1998). In general terms, such identification strategy is best applicable to evaluate the impact of incentive programs at the regional/state/province level, with eligible firms selling goods and/or services predominantly within the local markets in which they are located, and with focus on outcomes of type B (Figure 1). In such cases, threats to the validity of the analysis come, by the most part, from changes (exogenous to the program incentives) that may occur in the economy of the local communities in which assisted and non-assisted firms are located. In order to control for such confounding factors all other identification strategies have to rely either on conditional independence assumption (CIA, i.e. selection into treatment is based on observables characteristics of firms' local markets/communities) or on the hypothesis that all unobserved heterogeneity of firms' local markets/communities are fixed effects (in case of DD schemes applied to comparisons of firms outcomes), or at least fixed linear growth trends (in case of DDD schemes). Exploiting *GNEC* enables to identify program impacts without having to rely on such assumptions, at the risk, however, of producing results with weaker external validity, if *GNEC* can be found only for a small percentages of assisted firms, and neither the program incentives nor the assisted firms have fairly homogenous characteristics.

Table 1: *Non-experimental impact identification strategies for business incentives programs*

Identification strategy	Description
Exploiting geographic natural experiment conditions (GNEC)	GNEC can be exploited when it's possible to compare outcomes from units located within a same cohesive local community crossed by some administrative boundaries which generates two different areas A, B. In area A the program incentives become available $Pr\{T=1\|A\}>0$. No incentives are available in area B $Pr\{T=1\|B\}=0$
Sharp regression discontinuity designs (RDD)	Sharp RDD can be applied when applicant firms are ranked based on observable characteristics K and program incentives are awarded only to firms with K> $\bar{k}$. In such cases, for firms in a neighbourhood of $\bar{k}$, the treatment status is nearly randomly assigned, enabling treatment estimates to be based on $E\{y^{(1)}\|\bar{k}^{+}\}-E\{y^{(0)}\|\bar{k}^{-}\}$
Conditional difference in difference with binary propensity score statistical matching (CDD-PSM)	Observable pre-intervention differences between assisted and non-assisted units are controlled for by PSM. Fixed-effects unobserved characteristics are controlled for by a DD design on outcomes of matched units
Propensity score matching (PSM) with program heterogeneity: PSM with discrete treatment categories; Generalizes PS for continuous treatments	Imbens (1999), Lechner (2001, 2002) extensions of PSM to multiple treatment categories. Joffe and Rosembaum (1999) and Lu, Zanutto, Hornik and Rosembaum (2001) matching estimator for programs with ordered doses of treatment. Hirano and Imbens (2004) or Imai and Van extension of PSM to continuous treatments. Without implementing a DD design on outcomes of matched units (or without properly differencing the outcome variable), impact identification relies on pure selection on observables assumptions
Three stages conditional difference in difference (3STG-CDD)	I) based on each categorical binary variable $T_{x,w}$ ($x \in X$ incentives types, $w \in W$ firm characteristics), a set of PS vectors are estimated. II) for each treatment category, units outside the PS common support regions are eliminated III) a CDD parametric model (with categorical treatments and control variables with flexible functional forms) is estimated on units with common support

Sharp regression discontinuity designs (RDD) can be typically applied to programs without specific local geographic targeting and with the availability of data on rankings of applicants (for example, some RDD have been applied to the evaluation of Italy's law 488/92, in Bronzini De Blasio 2006 and Pellegrini Carlucci 2003). If the analysis is focused on the firms outcomes of type B (Figure 1), program impact identification is possible by comparing outcomes from applicant firms ranked in a neighborhood of a cut-off point $\bar{k}$

that determines the treatment status (this is because in such neighborhood of $\bar{k}$ treatment status can be though of being nearly randomly assigned). As RDD can identify mean impact estimates only for the assisted firms in the neighborhood of $\bar{k}$, results are typically of acceptable external validity (from a policy-relevance point of view) only if both incentive payments and the characteristics of the assisted firms are fairly homogeneous throughout the entire population of treated. As proposed in Battistin and Rettore (2008), "partially fuzzy" RDD set-ups could yield a specification test (in the neighborhood of $\bar{k}$) to assess the local properties of any non-experimental estimators usable to retrieve the treatment impacts on the whole population of treated. For business incentives policies, however, "partially fuzzy" RDD conditions may be quite a rare occurrence. This is because for programs with incentive payments based on a competitive auction process, data on the final rankings of applicant firms do typically exclude firms that drop-off from the auction. As a results virtually all firms above the cut-off threshold $\bar{k}$ do receive the program incentives, while all firms below $\bar{k}$ do not. Programs with no competitive auctions do not maintain lists of eligible firms. As a result, either available firm-level data are not sufficient to disentangle eligible non-treated firms from non-eligible firms, or the program eligibility rule (based for example on a simple binary coding of firms' sector classification) is such that a neighborhood of the eligibility threshold is hard to find, and eligible and non-eligible firms are likely to be exposed to quite different economic exogenous dynamics in times during the program implementation.

Conditional difference in difference designs, with propensity score matching *(CDD-PSM)*, are best applicable to either programs with specific local geographic target, in cases in which impact estimates are focused on outcomes measuring economic improvements of the target areas (O'keefe 2004), or programs focusing on assisted firms (such as manufacturers) that predominantly operate on national or international markets. In such cases, data on observable pre-intervention variables which may be distributed differently between treated and non-treated units are often available. Such confounding factors are controlled for by a PSM design, which (through it's well known balancing property, Rosembaum and Rubin 1983) surpasses the difficulties of choosing the proper functional forms of the observable control variables. Unobserved heterogeneity between treated and non-treated units is then controlled for by relying on DD (or DDD) schemes applied on the outcomes of the matched units. Such procedure, ensure that ATT parameters are identified relying on fixed-effects assumptions (or fixed linear growth rate assumptions in the case of DDD) only for unobserved heterogeneity, while observed heterogeneity is controlled for without such assumptions.

Imbens (2000), Lechner (2001, 2002) and Imai and Van Dyk (2004) extensions of PSM estimators to cases of multiple treatment categories are valuable alternatives to evaluate programs with treatment heterogeneity (related either to different ranges of economic values of the incentives, and/or to different types of benefits granted to assisted firms) and/or programs with heterogeneity of the treated units. In the case of programs with ordered doses of treatment, the matching estimator of Joffe and Rosembaum (1999) and Lu, Zanutto, Hornik and Rosembaum (2001), which entails a single scalar propensity score for all dose levels, is another possible option. Some type of Generalized propensity score estimator (GPS), finally, could also be applied for evaluating programs with continuous levels of the economic value of the incentives (Hirano and Imbens 2004 and Imai and Van Dyk 2004)[2]. In their pure forms, however, extended PSM, GPS and matching-with-ordered-

[2] To evaluate the impact of a business capital subsidy policy with continuous treatment (the Italian Law 488/92), a two steps matching estimators has also been proposed (Adorno, Bernini and Pellegrini 2007).

treatment-doses estimators strictly rely on selection on observables identification assumptions (either for both the selection into the program and the selection into the different treatment categories/doses/levels, or exclusively for the latter -as in one estimator proposed in Behrman, Cheng, Todd 2004). In order to control for fixed-effects- or fixed-linear-growth-rate- unobserved heterogeneity between treated and non-treated units (or among different treatment categories/doses/levels), therefore, it's advisable, also in these cases, to complement such estimators with a DD (or DDD) scheme applied on the outcomes of the matched units (or to transform the outcome variable into differences, in the case of the GPS estimator).

For evaluating multiple programs with many sources of treatment heterogeneity, variation in the economic level of the incentives and heterogeneity of treated units, a further possible option is to implement PS as a "nonparametric" first-stage processing for reducing model dependence in parametric estimators of treatment impact estimates (Ho, Imai, King and Stuart 2007). A suitable procedure of this sort could be the following three stages conditional difference in difference estimator *(3STG-CDD)*: I) based on each categorical binary variable $T_{x,w}$ ($x \in X$ incentives types, $w \in W$ firm characteristics), a set of PS vectors are estimated; II) for each treatment category, units outside the PS common support regions are eliminated; III) a CDD parametric model (with categorical treatments and control variables with flexible functional forms) is estimated on units with common support. Since such *3STG-CDD* procedure cannot exploit the PSM balancing property, extensive sensitivity analysis is to be performed to test how impact results may differ based on different functional forms of controls.

6. Single-program- versus multiple-programs- evaluations

This paper reviews and discusses the different options in choosing the appropriate outcome data, parameters of interest and impact identification strategies that are best suited for evaluating business incentives policies. Since, very often, multiple different programs are available to the same types of eligible firms in a same geographic area[3], all of the different issues discussed in the paper are applicable to either single-program evaluations or comparative joint evaluations of multiple programs.

Single-program evaluation studies (SPEs) are by far more frequent than evaluations of multiple programs (MPEs). SPEs require much less data collection efforts, and often, they can rely on simpler operationalizational rules for coding the treatment variables. In order to identify policy-relevant average treatment effects on the treated (or on subpopulations of the treated units), however, SPEs have to rely on the crucial assumption that treated- and non-treated- firms have the same conditional probability of receiving assistance from other different incentive programs (from which no data are available) during the time span considered in the analysis:

$$P(T_{x^*}=1 \mid W, T_x=1) = P(T_{x^*}=1 \mid W, T_x=0), \text{ for all } x^* \in X^* \text{ and } x \in X \qquad (3)$$

[3] In Italy, for example, in recent years, an average or twenty or more different incentives policies (from National, Regional and EU sponsored programs) have been available in the same geographic areas to the same types of eligible firms.

where $\{x = 1,2,\ldots,X\}$ represents the set of treatments being the focus of the impact evaluation analysis, and $\{x^* = 1,2,\ldots,X^*\}$ represents the set of treatments from the different unobserved incentive programs that may be available to treated and non-treated firms[4].

Quite often such assumption is not much plausible[5] and results from SPEs can suffer from attenuation bias (in the most frequent cases in which non-assisted firms are more likely to gain access to other forms of incentives than assisted firms). MPEs, instead, although requiring extensive data collection efforts, do not have to rely on such crucial assumption (as all of the sources of incentive payments are typically observed), and they are often capable of exploiting the across-programs heterogeneity of incentives and designation rules to provide findings with larger external validity.

[4] If $P(T_{x^*}=1 \mid T_x=1)$? $P(T_{x^*}=1 \mid T_x=0)$, identifying policy relevant average treatment effects on the treated with SPEs would require to assume that none of the incentive programs, other than the ones being evaluated, may affect the firm outcome considered in the analysis: $\{ y^{(1)}_{x^*} = y^{(0)}_{x^*} \}_{x^* \in X^*}$.

[5] With some few exceptions such as Italy's Law 488/92.

References

Adorno V., Bernini C., Pellegrini G. (2007), "The impact of capital subsidies: new estimations under continuos treatment", *Giornale degli Economisti ed Annali di Economia*, 66(1), 67-92.

Bartik T.J. (2004), "Evaluating the impacts of local economic development policies on local economic outcomes: What has been done and what is doable", in: *Evaluating local economic and employment development: How to assess what works among programmes and policies, OECD, Paris*, 113-141.

Battistin E., Rettore E. (2008), "Ineligibles and eligible non-participants as a double comparison group in regression discontinuity designs", *Journal of Econometrics*, 142(2), 715-730.

Boarnet M.G., Bogart W.T. (1996), "Enterprise zones and employment: Evidence from New Jersey", *Journal of Urban Economics*, 40, 98-215.

Bondonio D. Greenbaum R. (2006), "Do Business Investment Incentives Promote Employment in Declining Areas? Evidence from EU Objective 2 Regions", *European Urban and Regional Studies*, 13(3), 225-244.

Bondonio D. e Greenbaum R. (2007), "Do Local Tax Incentives Affect Economic Growth? What Mean Impacts Miss in the Analysis of Enterprise Zone Policies", *Regional Science and Urban Economics*, 37(1), 121-136.

Bronzini R., De Blasio G. (2006), "Evaluating the impact of investment incentives: The case of Italy's Law 488/1992", *Journal of Urban Economics*, 60, 327-349.

Engberg, J., Greenbaum R. (1999), "State enterprise zones and local housing markets", *Journal of Housing Research*,10, 163-187.

Fan W., Treyz F, Treyz G. (2000), "An Evolutionary New Economic Geography Model", *Journal of Regional Science*, 40(4), 671–695.

Hirano K., Imbens G. (2004), "The propensity score with continuous treatment", in: *Applied Bayesian modeling and causal inference from missing data perspective*, Gelman and Meng (eds.), Wiley, New York.

Ho D. E., Imai K., King G., Stuart E. A. (2007), Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference, *Policy Analysis*, 15, 199-236.

Holmes T. (1998), "The effects of state policy on the location of industry: Evidence from State borders", *Journal of Political Economy*, 106, 667-705.

Imai K., Van Dyk (2004), "Causal inference with general treatment regimes: generalizing the propensity score", *Journal of the American Statistical Association*, 99, 854-866.

Imbens G. (2000), "The role of Propensity score in estimating dose –response functions", *Biometrika* 87, 706-710.

Joffe M. M., Rosembaum P.R. (1999), "Propensity scores", *American Journal of Epidemiology*, 150, 327-333.

Lechner M. (2001), “Identification and estimation of causal effects of multiple treatments under the conditional independence assumption”, in: *Econometrics evaluations of active labor market policies in Europe*, Lechner and Pfeiffer (eds.), Physica/Springer, Heidelberg.

Lechner M. (2002), “Program heterogeneity and propensity score matching: an application to the evaluation of active labor market policies”, *The Review of Economics and Statistics*, 84(2), 205-220.

Lu B., Zanutto E., Hornik R. and Rosembaum P. R. (2001), “Matching with doses in an observational study of a media campaign against drug abuse”, *Journal of the America Statistical Association*, 96(456), 1245-1253.

O’Keefe S. (2004), “Job creation in California’s enterprise zones: a comparison using a propensity score matching model”, *Journal of Urban Economics*, 55, 131-150.

Pellegrini G., Carlucci C. (2003), “Gli effetti della legge 488/92: una valutazione dell'impatto occupazionale sulle imprese agevolate”, *Rivista Italiana degli Economisti*, 2, 267-286.

Rosembaum P., Rubin D. (1983), “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, 70, 41-55.

Evaluating the Effects of Business Incentives Policies: A Discussion Note on Identification Strategies

Summary

This paper aims at contributing to the impact evaluation literature on business incentive policies by offering a concise discussion of the many possible options in choosing the appropriate outcome data and impact identification strategy for the analysis. Each option presented in the paper is discussed based on the program characteristics and whether or not the analysis focus on socially desirable outcomes recorded at the level of assisted firms or at the level of the geographic areas in which the program incentives are in place. Also reviewed in the paper are the role of randomized experiments and the policy relevance of estimating average treatment effects on the treated versus distributions of the treatment effects.

Keywords: Impact evaluation; Business Incentives Policies; Selection bias, Identification strategies

JEL Classification: C40, C80, H81

Working Papers

76. D. Bondonio, Evalluating the Effects of Business Incentives Policies: A Discussion Note on Identification Strategies, Luglio 2008
75. B. Anastasia, D. Maurizio, Salari e stipendi in Veneto 2001-2002, Febbraio 2007
74. G. Tattara, M. Valentini, Labour market segmentation, flexibility and precariousness in the Italian North East, Aprile 2006
73. B. Anastasia, D. Maurizio, I lavoratori dipendenti in Veneto 1998-2003: profili e percorsi. Statistiche sistematiche da “Giove 2005”, Dicembre 2005
72. A. Fossaluzza, Evaluation of the Effects of Two Classes of Active Labour Market Policies for Welfare Recipients: a Danish Study Case, Dicembre 2005
71. P. Cipollone, A. Guelfi, Financial support to permanent jobs. The Italian case, Novembre 2005
70. F. Bassi, U. Trivellato, The latent class approach to estimating gross flows affected by correlated classification errors, with application to data from the French Labour Force Survey, Novembre 2005
69. M. Fort, Education and the timing of births: evidence from a natural experiment in Italy, Novembre 2005
68. D. Contini, N. Negri, Would declining exit rates from welfare provide evidence of welfare dependence in homogeneous environments?, Ottobre 2005
67. E. Battistin, E. Rettore, U. Trivellato, Choosing among alternative classification criteria to measure the labour force state, Ottobre 2005
66. G. Tattara, M. Valentini, Evaluating the Italian training on the job contract (CFL), Novembre 2005
65. A. Paggiaro, E. Rettore, U. Trivellato, The impact of the Italian “Mobility Lists” on employment chances: new evidence from linked administrative archives, Ottobre 2005
64. D. Bondonio, R.T. Greenbaum, Do Tax Incentives Affect Local Economic Growth? What Mean Impacts Miss in the Analysis of Enterprise Zone Policies, Ottobre 2005
63. D. Bondonio, The employment impact of business incentive programs in declining areas. Mean impacts versus impacts by degrees of economic distress, Ottobre 2005
62. U. Trivellato, F. Bassi, A. Discenza, A. Giraldo, Transizioni e mobilità nel mercato del lavoro italiano, 1979-2003, Giugno 2005
61. A. De Angelini, A. Giraldo, La mobilità dei lavoratori nel Veneto.Confronto fra misure su dati RTFL e su dati NETLABOR, Settembre 2003
60. F. Bassi, E. Salvan, Dinamiche di ricollocamento per lavoratori che perdono un'occupazione stabile, Settembre 2003
59. M. Romano, Abbinamento tra gli archivi INPS e Netlabor: il caso degli iscritti alle liste di mobilità, Settembre 2003
58. P. Baretta, U. Trivellato, La mobilità dei lavoratori da fonti amministrative e da surveys sulle famiglie: un'analisi comparata, Settembre 2003
57. D. Bondonio, R.T. Greebaum, A comparative evaluation of spatially targeted economic revitalization programs in the European Union and the United States, Settembre 2003
56. F. Berton, L. Pacelli, G. Segre, Tra lavoro dipendente e lavoro parasubordinato: chi sono, da dove vengono e dove vanno i lavoratori parasubordinati, Settembre 2003

55. G. Betti, V. Verma, M. Natilli, F. Ballini, Statistical imputation in conjunction with micro-simulation of income data, Settembre 2003

54 V. Verma, G. Betti, F. Ballini, M. Natilli, S. Galgani, Personal income in the gross and net forms: applications of the Siena micro-simulation model (SM2), Settembre 2003

53. R. Canu, G. Tattara, Quando le farfalle mettono le ali. Osservazioni sull'ingresso delle donne nel lavoro dipendente, Settembre 2003

52. D. Favaro, S. Magrini, Gender wage differentials among young workers: methodological aspects and empirical results, Settembre 2003

51. F. Devicienti, Downward nominal wage rigidity in Italy: evidence and consequences, Novembre 2002

50. V. Verma, G. Betti, Longitudinal measures of income poverty and life-style deprivation, Novembre 2002.

49. A. Borgarello, F. Devicienti, Trends in the Italian earnings distribution, 1985-1996, Novembre 2002

48. S. Bragato, F. Occari, M. Valentini, I problemi di contabilizzazione dei lavoratori extracomunitari. Una verifica nelle province di Treviso e Vicenza, Novembre 2002

47. B. Anastasia, D. Maurizio, Misure dell'occupazione temporanea: consistenza, dinamica e caratteristiche di uno stock eterogeneo, Novembre 2002

46. E. Rettore, U. Trivellato, A. Martini, La valutazione delle politiche del lavoro in presenza di selezione: migliorare la teoria, i metodi o i dati?, Novembre 2002

45. U. Trivellato, A. Giraldo, Assessing the 'choosiness' of the job seekers. An exploratory approach and evidence for Italy, Novembre 2002

44. E. Battistin, E. Rettore, Another look at the regression discontinuity design, Novembre 2002

43. B. Contini, F. Cornaglia, C. Malpede, E. Rettore, Measuring the impact of the Italian CFL programme on the job opportunities for the youths, Novembre 2002

42. F. Devicienti, Estimating poverty persistence in Britain, Novembre 2001

41. A. Giraldo, E. Rettore, U. Trivellato, Attrition bias in the bank of Italy's survey on household income and wealth, Novembre 2001

40. A. Giraldo, E. Rettore, U. Trivellato, The persistence of poverty: true state dependence or unobserved heterogeneity? Some evidence form the Italian survey on household income and wealth, Novembre 2001

39. G. Barbieri, P. Gennari, P. Sestito, Do public employment services help people in finding a job? An evaluation of the italian case, Novembre 2001

38. N. Torelli, A. Paggiaro, Estimating transition models with misclassification, Novembre 2001

37. B. Anastasia, M. Gambuzza, M. Rasera, Le sorti dei flussi: dimensioni della domanda di lavoro, modalità di ingresso e rischio disoccupazione dei lavoratori extracomunitari in Veneto, Novembre 2001

36. L. Pacelli, Fixed term contracts, social security rebates and labour demand in Italy, Novembre 2001

35. A. Borgarello, F. Deficienti, C. Villosio, Mobilità retributiva in Italia 1985-1996, Novembre 2001

34. B. Contini, C. Villosio, Job changes and wage dynamics, Novembre 2001

33. B. Contini, Earnings mobility and labor market segmentation in Europe and USA: preliminary explorations, Novembre 2001

32 A. Borgarello, F. Devicienti, Trend nella distribuzione dei salari italiani 1985-1996, Novembre 2001

31. F. Bassi, U. Trivellato, Gross flows from the French labour force survey: a reanalysis, Novembre 2001

30. F. Bassi, M. Gambuzza, M. Rasera, Imprese e contratti di assunzione: prime analisi da Netlabor, Novembre 2001

29. N. Rosati, Further results on inequality in Italy in the 1980s and the 1990s, Aprile 2001

28. A. Ichino, M. Polo, E. Rettore, Are judges biased by labor market conditions?, Novembre 2000

27. E. Battistin, E. Rettore, Testing for the presence of a programme effect in a regression discontinuity design with non compliance, Novembre 2000

26. E. Battistin, E. Rettore, U. Trivellato, Measuring participation at work in the presence of fallible indicators of labour force state, Giugno 2000

25. A. D'Agostino, G. Ghellini, L. Neri, Percorsi di ingresso dei giovani nel mercato del lavoro, Giugno 2000

24. F. Bassi, M. Gambuzza, M. Rasera, E. Rettore, L'ingresso dei giovani nel mercato del lavoro: prime esplorazioni dall'archivio Netlabor, Giugno 2000

23. A. Paggiaro, 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

22. A. D'Agostino, G. Ghellini, L. Neri, Modelli statistici per l'analisi dei comportamenti di transizione scuola lavoro, Marzo 2000

21. G. Betti, B. Cheli, A. Lemmi, Analisi delle dinamiche di povertà e disoccupazione su uno pseudo panel italiano, Marzo 2000

20. N. Rosati, Permanent and Temporary Inequality in Italy in the 1980s and 1990s, Marzo 2000

19. F. Bassi, M. Gambizza, M. Rasera, Struttura e qualità delle informazioni del sistema NETLABOR. Una verifica sui dati delle Scica delle province di Belluno e Treviso, Marzo 2000

18. B. Anastasia, M. Gambuzza, M. Rasera, La durata dei rapporti di lavoro: evidenze da alcuni mercati locali del lavoro veneti, Marzo 2000

17. G. Betti, B. Cheli, A. Lemmi, Occupazione e condizioni di vita su uno pseudo panel italiano: primi risultati, avanzamenti e proposte metodologiche, Ottobre 1999

16. A. D'Agostino, G. Ghellini, L. Neri, A Multiple Imputation Method for School to Work Panel Data, Ottobre 1999

15. A. Paggiaro, N. Torelli, Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro, Ottobre 1999

14. S. Campostrini, A. Giraldo, N. Parise, U. Trivellato, La misura della partecipazione al lavoro in Italia: presupposti e problemi metodologici di un approccio "time use", Ottobre 1999

13. T. Di Fonzo, P. Gennari, Le serie storiche delle forze di lavoro per il periodo 1984.1-92.3: prospettive e problemi di ricostruzione, Giugno 1999

12. A. Paggiaro, Un modello di mistura per l'analisi della disoccupazione di lunga durata, Maggio 1999

11. M. Lalla, F. Pattarin, Alcuni modelli per l'analisi delle durate complete e incomplete della disoccupazione: il caso Emilia Romagna, Maggio 1999

10. F. Bassi, M. Gambuzza, M. Rasera, Il sistema informatizzato NETLABOR. Caratteristiche di una nuova fonte sul mercato del lavoro, Maggio 1999

9. B. Contini, M. Filippi, L. Pacelli, C. Villosio, Working careers of skilled vs. unskilled workers, Gennaio 1999

8. B. Contini, L. Pacelli, C. Villosio, Short employment spell in Italy, Germany and Great Britain: testing the "Port-of-entry" hypothesis, Gennaio 1999

7. B. Contini, Wage structures in Europe and in the USA: are they rigid, are they flexible?, Gennaio 1999

6. M. Forni, S. Paba, Industrial districts, social environment and local growth. Evidence from Italy, Novembre 1998

5. Ginzburg, M. Scaltriti, G. Solinas, R. Zoboli, Un nuovo autunno caldo nel Mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali, Ottobre 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

3. U. Trivellato, Il monitoraggio della povertà e della sua dinamica: questioni di misura e evidenze empiriche, Settembre 1998

2. N. Rosati, E. Rettore, G. Masarotto, A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models, Agosto 1998

1. E. Battistin, A. Gavosto, E. Rettore, Why do subsidized firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy, Agosto 1998