Dipartimento di Politiche Pubbliche e Scelte Collettive – POLIS
Department of Public Policy and Public Choice – POLIS

Working paper n. 145

July 2009

Impact identification strategies for evaluating business incentive programs

Daniele Bondonio

UNIVERSITA' DEL PIEMONTE ORIENTALE "*Amedeo Avogadro*" ALESSANDRIA

Periodico mensile on-line "POLIS Working Papers" - Iscrizione n.591 del 12/05/2006 - Tribunale di Alessandria

Impact Identification Strategies for Evaluating Business Incentive Programs*

June 2009

Daniele Bondonio**

Università del Piemonte Orientale
(Ph.D Carnegie Mellon University)

Abstract

Although business incentive programs of different forms have been the bulk of local economic development policies in many industrialized countries for more than the last three decades, evaluating their impact on employment or local economic growth outcomes remains a challenging task due to the persisting lack of randomized experiments and the presence of many confounding factors which affect firms and economic growth outcomes. Moreover, much of the recent advancements in the statistical program evaluation methodology applicable to non-experimental settings do not make any direct reference to the specificities posed by business incentive policies. This paper aims at offering some clear guidance on how to choose the appropriate focus of the evaluation, the policy relevant evaluation parameters and empirical impact identification strategies when applying statistical methods attempting to estimate how much of the different outcomes between treatment and control groups are attributable to the program/s being evaluated. Each methodological option discussed in the paper is linked to the different features of commonly implemented US and EU policies and to whether or not the analysis focuses on outcomes recorded at a firm-level or at the level of the geographic areas in which the assisted firms are located.

JEL Classification: C40, C80, H81

Keywords: Impact evaluation; Business Incentives Policies; Comparison-group designs, Identification strategies

* The work that provides the basis for this paper was supported by funding under the PRIN-2005 grant from MIUR (Italian "Ministry of Education, University and Research"). The author is solely responsible for the accuracy of the statements and interpretations contained in the paper. Such interpretations do not necessarily reflect the views of the Italian Ministry of Education, University and Research. The author thanks Robert Greenbaum for useful comments.

** Corso Cavour 84, 15100 Alessandria, ITALY; daniele.bondonio@sp.unipmn.it (e-mail)

1 Introduction

Although business incentive programs of different forms have been the bulk of local economic development policies in many industrialized countries for more than the last three decades, evaluating their impact on employment or local economic growth outcomes remains a challenging task. Retrieving reliable program impact estimates can be more difficult for business incentives than for other public policies due to the persisting lack of randomized experiments, the presence of many confounding factors which affect firms and economic growth outcomes, and the often simultaneous presence of a significant number of many different competing programs in addition to the policy/ies being evaluated.

As also well argued in Bartik (2004), Bartik and Bingham (1997), statistical methods attempting to estimate how much of the different outcomes between treatment and control groups are attributable to the program/s are a crucial tool in evaluating business incentive policies in non-experimental settings. Sounds counterfactual statistical analyses on the proximate employment or local economic growth outcomes of the policies, moving away from the mere description of program activities, provide vital empirical evidence that are also a necessary base for possible subsequent survey and focus group analyses and/or regional econometric models aimed at estimating (when the importance of the program is appropriate) more distant fiscal and employment benefits in terms of long-run or province/regional/state-economy effects.

Unfortunately, most of the recently developed impact identification approaches developed by the statistical literature devoted to counterfactual comparison-group program evaluations [e.g.: propensity score matching with program heterogeneity and generalized propensity score for continuous treatment, Imbens (2000), Lechner (2001, 2002), Joffe and Rosembaum (1999), Lu et. Al. (2001), Hirano and Imbens (2004), and Imai and Van Dyk (2004); specification tests for non-experimental estimators based on partially fuzzy regression discontinuity designs, Battistin and Rettore (2008); conditional difference in difference with propensity score matching, Heckman, et. al.(1998); propensity score estimations as first stage processing for reducing model dependence in parametric estimators, Ho et. Al. (2007), estimations of the distribution of treatment effects, Abadie et. al (2002), Cheser (2003), Carneiro et. Al. (2003), Chernozhukov and Hansen (2005, 2006)] do not make any direct reference to the specific issues posed by business incentive policies.

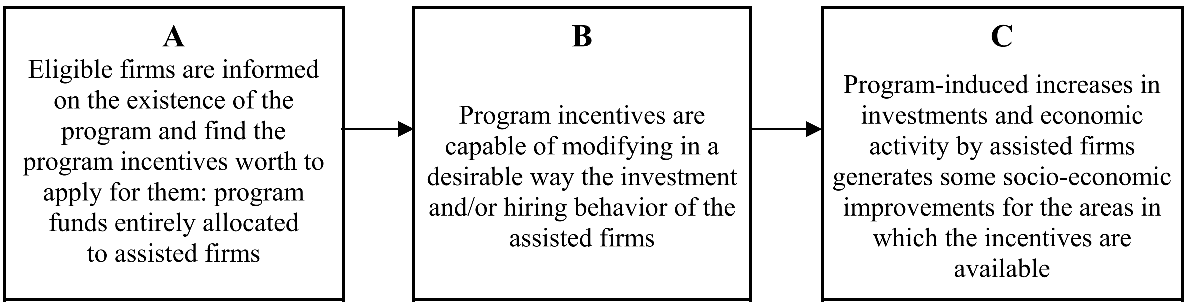
This paper review and discusses such recent statistical developments aiming at offering some clear guidance on how to choose the appropriate focus of the evaluation, the policy relevant evaluation parameters (average treatment effects versus distributions of the treatment effects) and the empirical impact identification methods for evaluating a variety of types of business incentive programs. Each methodological option discussed in the paper is linked to the different types of incentive policies commonly implemented in the US and EU countries and to whether or not the analysis focuses on outcomes recorded at a firm-level or at the level of the geographic areas in which the assisted firms are located. The paper extends and updates the work of Bartik (2004), Boarnet (2001), Bondonio (2000) and Bartik and Bigham (1997), which offered a first important contribution on statistical methods suited to evaluate local economic development and geographically targeted policies based on the provision of a variety of business incentives.

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 illustrates impact identification strategies in non-experimental settings. Section 5 discusses the issue of evaluations of single programs versus evaluations of multiple programs. Section 6 offers some concluding remarks.

1. 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):

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



In order for a program to succeed, eligible firms need to be well informed on the existence of the program and need to find the program incentives worthy 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 the program is managed so that all program funds are allocated to applicant firms, the actual impact of the program could still 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 between pre- and post-treatment times, needed to eliminate the selection bias due to correlation between unobserved fixed effects and the treatment, in the case of many business incentive policies, is very often to be performed as absolute changes rather percentage changes. This is because, for many business incentive policies 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.

Counterfactual statistical impact evaluation analyses 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 (e.g. Krupka and Noonan 2009, Bondonio and Greenbaum 2007, O'Keefe 2004, Greenbaum and Engberg 2004, Peters and Fisher 2002, Bondonio and Engberg 2000, Engberg and Greenbaum 1999, Boarnet and Bogart 1996, Papke 1994), the "Zones Franches Urbaine" of France (Rathelot and Sillard 2008), the proposed "Zone Franche Urbane" of Italy, and, by some degree, the incentives co-funded by the EU structural funds in "Objective 2 areas" (Bondonio and Greenbaum 2006). In such cases the economic weight of the program incentives is not disproportionably small compared to the size of the economy of the local target areas, and appropriate comparison group statistical models are capable of identifying the program impact on the target areas outcomes, controlling for the major confounding factors. Impact evaluations 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 (log) 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 or long-run 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 very little. As a result, any actual program impact (in the form of a positive impulse given to the province/region/state economy) becomes virtually undetectable from the changes to the outcome variable of the evaluation caused by many confounding factors (including, in many cases, the presence of other business incentive programs) of a much greater importance than the possible program-induced improvements in the economic activity of the assisted firms.

Using rigorous comparison-group statistical 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 considered unaffected by the program incentives and 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 only using regional macroeconomic simulation models (such as REMI - Regional Economic Models *inc.*, Fan, Treyz e Treyz 2000), as also suggested in Bartik (2004). Analyses with regional macroeconomic simulation models, however, should be performed only when the importance of the economic outputs of the assisted firms is not disproportionably smaller than the size of the local economy and 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 measures of program activity (such as 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.

2. Policy Relevant Impact Evaluation Parameters

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 $y^{(0)}$, $y^{(1)}$, $T \in \{0,1\}$.

3.1 Categorical Average Treatment Effects on the Treated (ATTs)

The policy-relevant parameters, which are often of most interest for evaluating business incentives programs, are ATTs (Average Treatment Effects on the Treated), estimated either for a single category of treatment on the entire population of treated units, or for a number of different categories of treatment on 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:

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

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

When, instead, the policy 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:

$$\tau(x, w) = E[\, \mathrm{y}_x^{(1)} - \mathrm{y}^{(0)} \mid T_x{=}1,\, W{=}w]. \tag{2}$$

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.

Estimating different impacts for different ranges of the economic value of the incentives is often 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 more suitable than continuous specifications. This is because, in many cases, the information leading to the operationalization of the incentive data are based on parameters such the Net Equivalent Subsidy (NES), which provides the net present value of the capital grant equivalent subsidy and represents the EU standard for measuring the public support offered by member-States national and regional governments, or, for tax reduction subsidies, on parameters such as the economic value of the incentives based on the difference between the internal rates of return of an investment made by a typical firm with and without the provision of the program incentives (with estimates of the monetary value of the incentives that are retrieved using an hypothetical firm approach, such as the Tax and Incentive Model –TAIM- developed by Peters and Fischer 2002). Computing NES or TAIM figures is very data demanding, often resulting in estimations of the economic values of the incentives that, because of the presence of potential significant noise, do signal actual differences in treatment intensity only across fairly wide apart values.

Categories of treatments based on different types of subsidies are policy relevant because, for example, in the EU and a number of US States, an increasingly strong debate is centered on the issue of how to allocate program funds between capital grants, below-market-rates and tax reduction subsidies, making of great interest to estimate categorical ATTs based on such policy distinctions[1].

3.2 Distributions of the Program Treatment Effects

A recent stream of the statistical literature on impact evaluation of social programs has focused on estimating the distribution of treatment effects represented, for example, by the proportion of treated units (considering either the entire population of the treated or each policy- relevant category of treated units) for whom the outcome with the treatment is greater or equal to the counterfactual outcome [i.e. $y^{(1)} - y^{(0)} \geq 0$ or $y_x^{(1)} - y^{(0)} \geq 0$]. Abadie et al. (2002), Cheser (2003), Carneiro, Hansen, Heckman (2003), Chernozhukov and Hansen (2005, 2006) and Firpo (2007)[2], among others, propose various identification strategies to estimate the distribution of treatment effects (often expressed as quantile treatment effects) and discuss the policy relevance of estimating impact results beyond mean effects on the treated in social and education programs.

In the case of business incentive policies, however, estimating the distribution of treatment effects (beyond identifying categorical ATTs) could be often of limited interest. This is because, for many business incentive programs (at least when estimating treatment impacts on economic or employment outcomes of assisted firms -outcomes of type B in Figure 1), socially desirable outcomes do not arise directly from the well-being of the treated entities receiving the program benefits (i.e. the assisted entrepreneurs or the stockholders of the assisted firms). Rather, socially desirable outcomes (in the form of economic or employment outcomes) are achieved indirectly when, for example, new jobs and/or investments attributable to the program incentives are generated in a declining local economy[3], or when they provide occupation to pockets of unemployment.

In such cases, similar socially desirable outcomes are often obtained whether or not the creation of new jobs or new investments is more concentrated among some of those assisted firms. Therefore, once policy-relevant categorical ATT parameters of the program

[1] Below-market-interest-rate loans, for example, are typically more economical than capital grants (with the same amount of public funds they allow the government to provide incentives to a much larger number of assisted firms than capital grants), while capital grants offer to assisted firms a financial advantage which is often superior to that of below-market interest rate loans (at least in times of not severe credit-crunch), embedding more potential to effectively modify the investment and hiring decisions of assisted firms. Tax reductions may not be very appealing to new firms and/or to research intensive ventures that do not expect to be profitable for their first periods of operations.

[2] See also Battistin and Fort (2008) for a concise review of such literature.

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

intervention are estimated (based on either different policy features or characteristics of the treated units), limited policy relevance is left for the distribution of treatment effects within each specific category of treatment may be of limited policy relevance.

4. Impact Identification Strategies and Applicable Statistical Methods

Table 1 summarizes the main identification strategies and statistical methods applicable in non-experimental comparison-group settings for estimating average treatment effects on the treated (ATTs) parameters in the context of the types of business incentive policies most commonly implemented in both the US and a number of EU countries. Compared to the impact identification strategies previously dicussed by Bartik (2004), Boarnet (2001), Bondonio (2000) and Bartik and Bigham (1997), the list of methods included in Table 1 adds a number of suitable approaches offered by the recent developments in the program evaluation statistical literature devoted to the estimation of ATTs parameters (namely: propensity score matching with program heterogeneity and generalized propensity score for continuous treatment; specification tests for non-experimental estimators based on partially fuzzy regression discontinuity design; conditional difference in difference with propensity score matching and propensity score estimations as first stage processing for reducing model dependence in parametric estimators). Regression discontinuity designs are also added among the methodological options of Table 1 as some relevant business incentive programs (especially in the EU) are based on rankings of applicant firms based on observable scores assigned to the proposed investment projects, with budget-induced cut-off points for program admissions.

Some well established statistical approaches used for causal inference estimations in empirical economic studies, such as instrumental variable estimators, shift share analyses and basic difference in difference models are instead excluded from the viable methodological options reported in Table 1. Instrumental variable estimators, together with “Inverse Mills Ratio” and “Heckman Selection” estimators (e.g. Heckman and Robb 1985, Robinson 1989, Angrist and Krueger 2001) require the availability of a subset of variables that affect the treatment assignment but do not have a direct effect on outcome of the evaluation. Such variables are extremely difficult to be found for most of the commonly implemented business incentive policies, since treated units (whether firms or geographic areas) are selected based on the same pre-existing characteristics that are among the major factors affecting the outcome of interest for the evaluation. The only exceptions are represented by the geographically-targeted programs in which the selection into the program can be explained by political variables measurable at a same geographic level of that of the target areas (e.g. Knight 2002, Wallace 2004).

Shift share analyses (SSA)s were used, up to the mid nineties, to analyze geographically targeted incentive policies such as the Enterprise Zone programs of some US states (e.g. Dowall 1996 and Rubin and Wilder 1989). In order to identify the treatment effects of business incentive policies, however, (as also noted in Boarnet 2001) SSA approaches assume that the counterfactual outcomes that would have been recorded in the treated areas in the absence of the program incentives is affected only by a linear economic growth (adjusted for industry mix) of exactly the same proportion as the one recorded in the larger geographic units surrounding the program target areas. In many instances such assumption is very hard t justify and, in light of the many recent methodological advancements offered by the statistical literature, impact evaluation studies of geographically targeted business incentive policies, in the last fifteen years, have progressively turned away from adopting SSA approaches.

Basic difference in difference (DD) estimators based on the mere availability of pre-post-intervention longitudinal outcome data (with no observable control variables) are capable of retrieving program impact estimates only under the assumption that every type of heterogeneity between treated- and non-treated-units do have a constant influence on the level of the outcome variable in any of the times considered in the analysis (DD estimators), or (for the case of DDD estimators) on the linear trends of the outcome variable (e.g. Moffit 1991). For most of the business incentive programs such strict functional form assumptions on the influence of unobserved heterogeneity on the outcome variable of the analysis may be hard to justify. This is because, in many cases, pre-intervention characteristics of firms or geographic areas (which may differ between treated and non-treated units) may generate multipliers effects with no constant influence on levels or on the linear trends of the outcome variable. As a result, for evaluating business incentive policies in the common cases when at least some data on observable pre-intervention characteristics are available, basic DD or DDD designs are often surpassed by the conditional difference in difference (CDD) approaches summarized in Table 1 that avoid relying on fixed-effects assumptions for all possible sources of heterogeneity between treated and non-treated units.

Table 1: ***Non-experimental impact identification strategies and statistical methods for business incentives programs in non-experimental settings***

Identification strategy	Description
Exploiting natural experiment conditions (NEC)	NEC 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 $Pr\{T=1\|B\}=0$
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 (2000), Lechner (2001, 2002) extensions of PSM to multiple treatment categories. Joffe and Rosembaum (1999) and Lu et al. (2001) matching estimator for programs with ordered doses of treatment. Hirano and Imbens (2004) or Imai and Van Dyk (2004), 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 estimators (3STG-CDD): PSM as a first-stage processing for reducing model dependence in parametric estimations	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

What follows is a concise review of each identification strategies and statistical methods reported in Table 1, discussing the extent to which each methodological option is a good fit for the different types of policies to be evaluated, the different choices of the outcome variable of the evaluation, the different scenarios of data availability and whether or not the analysis focuses on outcomes recorded at a firm-level or at the level of the geographic areas in which the assisted firms are located.

4.1 Exploiting Natural Experiment Conditions

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 which determine 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 their 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 (or part of) the 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.

4.2 Regression Discontinuity Designs (RDD)s and specification tests based on partially fuzzy RDDs

Sharp regression discontinuity designs (RDD) can be typically applied when the focus of the analysis is on evaluating proximate firm-level outcomes (i.e. outcomes of type B, Figure 1) and the programs being investigated offer the availability of data on rankings of applicants (an example of such programs is Italy's law 488/92 which has been evaluated with a sharp RDD approach by Bronzini De Blasio 2006 and Pellegrini Carlucci 2003). With sharp RDDs treatment impacts are estimated by comparing the outcomes from the applicant firms ranked just above and below the cut-off point $\bar{k}$ that determines the treatment status (e.g. Rubin 1977 and Trockim 1984). This is because in such neighborhood of $\bar{k}$ the treatment status can be though of being nearly randomly assigned (with firms in the treatment and comparison group having similar characteristics). As sharp 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 (which requires data availability on non-eligible units, eligible units that choose not to participate in the program, and program participants) 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 observable 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.

4.3 Conditional Difference in Difference with Propensity Score Matching

Conditional difference in difference designs, with propensity score matching *(CDD-PSM,* e.g. Heckman et al. 1998), are best applicable to programs for which a multitude of comparison units and some data on observable pre-intervention variables (which may be distributed differently between treated and non-treated units) are available for the analysis. With the CDD-PSM approach, the observable characteristics which my be different between units in the treatment and comparison group 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 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.

4.4 Propensity Score Matching with Program Heterogeneity

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 et. al. (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)[4]. In their pure forms, however, extended PSM, GPS and matching-with-ordered-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 et. al. 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).

4.5 Three Stages Conditional Difference in Difference

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 et.al. 2007). A suitable procedure of this sort could be the following three stages conditional difference in

[4] 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 et al. 2007).

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.

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

For business incentive policies, unlike the case of other public policies, multiple different programs are frequently available to the same eligible units (i.e. same types of eligible firms in a same geographic area). In many EU countries, in a same geographic area and for a same type of eligible firms, different EU-sponsored, National and Regional programs may coexist[5], and in some areas of the US, federal, state and local programs frequently overlap (Hultquist 2007). As a result, the different methodological options 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 in the literature 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 also from the different unobserved incentive programs for which they are eligible during the time span considered in the analysis:

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

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[6].

---

[5] In the Piemonte region of Italy, for example, in the last five years, the average number of such overlapping programs was in excess of twenty (Bondonio 2007).

[6] If $P(T_{x^*}=1 \mid T_x=1) \neq P(T_{x^*}=1 \mid T_x=0)$, 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: $\{ y^{(1)}_{x^*} = y^{(0)}_{x^*} \}_{x^* \in X^*}$.

Quite often such assumption is not much plausible[7] and findings 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 large external validity.

6. Concluding Remarks

This paper reviews and discusses the different options in choosing the appropriate outcome data, parameters of interest, impact identification strategies and statistical methods that are best suited to estimate, in non-experimental settings, how much of the different outcomes between treatment and comparison groups are attributable to business incentive policies.

Choosing the appropriate outcome variable/s for the analysis stems from clearly highlighting the chain of causal links from business incentives to desirable socio-economic outcomes, and from acknowledging the difference between enumerating program activities and estimating program effects. The choice between focusing on proximate firm-level outcomes rather than on community-level socio-economic outcomes is to be based on the types of programs analyzed and on the relative economic importance of their incentives compared to size of the local communities in which the assisted investment projects are located. Long-run macro-economic effects for an overall province/regional/state economy in which the program incentives are allocated should not be estimated with comparison-group statistical methods but rather with regional macroeconomic simulation models. Such province/regional/state-wide macroeconomic effects, however, should be estimated only when the importance of the economic outputs of the assisted firms is not disproportionably smaller than the size of the local economy and only after having rigorously estimated the program impact on proximate firm-level outcomes, using reliable comparison-group statistical methods.

When estimating the effects of business incentive policies with comparison-group statistical methods, the impact evaluation parameters of most interest are categorical average treatment effects on the treated based on either different ranges of economic values of the incentives or different types of benefits granted to the assisted firms. The recent developments in the program evaluation statistical literature offers a number of viable methods to estimate

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

such parameters. Choosing which methods are best suited for the analysis is to be based primarily on whether or not: the focus of the analysis is on proximate firm-level outcomes, rather than on community-level socio-economic outcomes; the program/s to be evaluated have many sources of treatment heterogeneity; the process to select assisted firms is based on rankings of applicants based on observable scores assigned to the proposed investment projects.

When evaluating business incentive policies is finally important to take into account that multiple different programs are often available to the same eligible firms in a same geographic area. As a result, in many instances, impact evaluations with statistical comparison-group designs can focus on investigating either the single effects of individual programs or the simultaneous effects of a set of different overlapping programs. Single-program evaluations are by far more frequent in the literature than simultaneous evaluations of multiple programs. It's important to notice, however, that single-program evaluations, in the presence of multiple overlapping incentive policies, can yield reliable empirical evidence only under the challenging assumption that treated- and non-treated- firms have the same conditional probability of receiving assistance also from the unobserved programs that operate in the same domain of the analysis.

References

Abadie A., Angrist J., Imbens G. (2002), Instrumental Variables estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings, Econometrica 70, 91-117.

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.

Angrist, J. D., Krueger A.B. (2001), Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments, *Journal of Economic Perspectives* 15(4): 69–85.

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.

Bartik T.J., Bingham R.D. (1997), Can economic development programs be evaluated?, in R.D. Bingham, R. Mier (eds.), Dilemmas fo Urban Economic Development, Thousand Oaks: Sage. Ch 10, pp. 246-277.

Battistin E. Fort. M. (2008), What's Missing from Policy Evaluation: Identification and Estimation of the Distribution of Treatment Effects, Proceedings from the XLIV National Conference of the Italian Statistical Society, Cosenza, Italy.

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.

Behrman J.R., Cheng Y., Todd P.E. (2004), Evaluating Preschool Programs when lengh of Exposure to the Program varies: A Nonparametric Approach, *Review of Economics and Statistics*, 86(1), 108-132.

Boarnet M.G. (2001), Enterprise Zones and Job Creation: Linking Evaluation to Practice, Economic Development Quarterly 15(3), 242-254.

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

Bondonio D. (2000), "Statistical methods to evaluate geographically-targeted economic development programs", *Statistica Applicata*, vol. 12, n.2, pp. 177-204.

Bondonio D. (2007), The Employment Impact of Business Incentive Policies: a Comparative Evaluation of Different Forms of Assistance, POLIS Working Paper 101/2007, Università del Piemonte Orientale.

Bondonio D., Engberg J. (2000), Enterprise Zones and Local Employment: Evidence from the States' Programs, Regional Science and Urban Economics, 30 (5), pp. 519-549.

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.

Carneiro P., Hansen K. T., Heckman J.J. (2003), Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice, IZA Working Papers n. 767.

Chernozhukov V., Hansen C. (2005), An IV Model of Quantile Treatment Effects, Econometrica 73, 245-261.

Chernozhukov V., Hansen C. (2006), Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, *Journal of Econometrics* 132, 491-525.

Cheser A. (2003), Identification in Nonseparable Models, *Econometrica* 71, 1405-1441.

Dowall, D.E., 1996, An evaluation of California's Enterprise Zone Programs, *Economic Development Quarterly* 10, 352-368.

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.

Firpo S. (2007), Efficient Semiparametric Estimation of Quantile Treatment Effects, *Econometrica* 73, 259-276.

Greenbaum, R., Engberg, J. (2004), The impact of state enterprise zones on urban manufacturing establishments, *Journal of Policy Analysis & Management* 23(2), 315-339.

Heckman J.J., Ichimura H., Todd P.E. (1998), Matching as an Econometric Evaluation Estimator, Review of Economic Studies 65(223), 261-294.

Heckman J., Robb R. (1985), Alternative methods for evaluating the impact of interventions: An overview, *Journal of Econometrics* 30:239-267.

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.

Hultquist A. (2007), An Evaluation and Comparison of Geographically Targeted Economic Development Programs in Ohio and Piedmont, Italy, Phd. Dissertation, Ohio State University.

Imai K., Van Dyk D.A. (2004), Causal inference with general treatment regimes: generalizing the propensity score,

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.

Knight B. (2002), Endogenous Federal Grants and Crowd-Out of State Government Spending: Theory and evidence from the Federal Highway Aid Program, American Economic Review, 92(1): 71-92.

Krupka D. J., Noonan. D. S. (2009) "Empowerment Zones, Neighborhood Change and Owner-Occupied Housing." Regional Science and Urban Economics 39(4): 386-396.

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.

Moffit R. (1991), Program evaluation with nonexperimental data, Evaluation Review 15, 3:291-314.

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.

Papke L. E. (1994), Tax policy and urban development: evidence from the Indiana enterprise zone program, *Journal of Public Economics* 54, 37-49.

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.

Peters A.H., S. Fisher (2002), State Enterprise Zone Programs. Have They Worked?, W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Rathelot R., Sillard P. (2008), Zones Franches Urbaines: quels effects sur l'emploi salarié et les créations d'établissements?, *Économie et Statistique* 415-416, 81-96.

Robinson C. (1989) The joint determination of union status and union wage effects: some test of alternative models. *Journal of Political Economy* 97: 639-667.

Rosembaum P., Rubin D. (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika,* 70, 41-55.

Rubin B.M., Wilder M.G. (1989), Urban Enterprise Zones: Employment Impacts and Fiscal Incentives, *Journal of the American Planning Association*, 55, 418-431.

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

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

Wallace M. (2004), Congressional Considerations and Urban Characteristics in the Selection of Empowerment Zones and Enterprise Communities, *Journal of Urban Affairs*, 26(5): 593-609.

Recent working papers

The complete list of working papers is can be found at http://polis.unipmn.it/pubbl

***Economics Series** ****Political Theory Series** **[ε] Al.Ex Series**

[T]Territories Series **[t]Transitions Series**

2009 n.145* Daniele Bondonio: *Impact identification strategies for evaluating business incentive programs*

2009 n.144* Barry Eichengreen, Rachita Gullapalli and Ugo Panizza: *Capital account liberalization, financial development and industry growth: a synthetic view*

2009 n.143* Emma Galli and Roberto Ricciuti: *Sulla* political economy *del deficit pubblico nell'Italia liberale*

2009 n.142* Matteo Migheli: *Religiosity and happiness: an ever-winning couple? An answer from India*

2009 n.141** Stefano Parodi: *I media dell'Alessandrino e l'Unione Europea*

2009 n.140* Matteo Migheli: *The two sides of a ghost: Twenty years without the wall*

2009 n.139[ε] Matteo Migheli and Francesco Scacciati: *How does labor supply react to different tax rates? A field enquiry*

2009 n.138[ε] Matteo Migheli and Guido Ortona: *Majority, proportionality, governability and factions*

2009 n.137** Noemi Podestà: *Strumenti di mediazione per la risoluzione di conflitti. L'esperienza dell'Osservatorio per il collegamento ferroviario Torino-Lione*

2009 n.136** Noemi Podestà and Alberto Chiari: *Esperimenti di democrazia deliberativa. Informazioni, preferenze e stili di conduzione in tre giurie di cittadini.*

2009 n.135** Andrea Lanza: *1848 comme reconfiguration des discours politiques.*

2009 n.134* Rongili Biswas, Nicolas Gravel and Rémy Oddou: *The segregative properties of endogenous jurisdictions formation with a welfarist central government*

2009 n.133[ε] Matteo Migheli: *Assessing trust through social capital? A possible experimental answer*

2009 n.132* Piero Cavaleri, Michael Keren, Giovanni B. Ramello and Vittorio Valli: *Publishing an E-journal on a shoe string: is it a sustainable project?*

2009 n.131* Alberto Cassone: *L'impatto economico e sociale dell'Universita' del Piemonte Orientale Amedeo Avogadro*

2009 n.130* Alberto Cassone and Pietro Zaccarella: *Il bilancio sociale delle universita'. Inventario dei problemi e analisi comparata delle esperienze italiane*

2009 n.129[ε] Matteo Migheli, Guido Ortona and Ferruccio Ponzano: *A preliminary simulative assessment of disproportionality indices*

2008 n.128* Fabio Privileggi: *On the transition dynamics in endogenous recombinant growth models*

2008 n.127* Roberto Zanola: *Who likes circus animals?*

2008 n.126* Michele Giuranno: *Regional income disparity and the size of the Public Sector*

2008 n.125* Giorgio Brosio and Roberto Zanola: *The welfare costs of national standards: a contribution to the debate on the outcomes of de/centralization*

2008 n.124[ε] Guido Ortona, Stefania Ottone, Ferruccio Ponzano and Francesco Scacciati: *Some differences in revealed behaviour under different inquiry methods*

2008 n.123* J. Stephen Ferris, Soo-Bin Park and Stanley L. Winer: *Studying the role of political competition in the evolution of government size over long horizons*

2008 n.122** Stefano Parodi: *Il funzionalismo di D. Mitrany: Dall'economia alla scienza politica*

2008 n.121** Joerg Luther: *L'antinegazionismo nell'esperienza giuridica tedesca e comparata*

2008 n.120* Roberto Zanola: *Consumer preferences for circus: a cluster approach*

2008 n.119* Roberto Ippoliti: *L'incentivazione economica nei problemi di agenzia: Il caso dell'Azienda Sanitaria Pubblica*

2008 n.118* Piermassimo Pavese and Roberto Zanola: *Autochthon vs. blended wines: Do objective and sensory characteristics matter?*

2008 n.117* Andrea Vindigni: *Uncertainty and the politics of employment protection*

2008 n.116* Carla Marchese: *The limits to growth then and now*

2008 n.115** Guido Ortona: *Perché in Italia le elezioni violano la legge di Duverger?*