Discussion Papers on Entrepreneurship, Growth and Public Policy

03005

Quantitative, Non-Experimental Approaches to the Microeconomic Evaluation of Public Policy Measures - A Survey

by

Max Keilbach
Max Planck Institute of Economics

Number of Pages: 11

The *Papers on Entrepreneurship, Growth and Public Policy* are edited by the Group Entrepreneurship, Growth and Public Policy, MPI Jena.
For editorial correspondence,
please contact: egppapers@mpiew-jena.mpg.de

ISSN 1613-8333

Max Planck Institute of Economics
Group Entrepreneurship, Growth and Public Policy
Kahlaische Str. 10
07745 Jena, Germany
Fax: ++49-3641-686710

Quantitative, Non-Experimental Approaches to the Microeconomic Evaluation of Public Policy Measures – A Survey –

Max Keilbach*

Abstract

The objective of evaluating public policy measures is to asses its implications and thus to obtain a measure for whether the respective program has been successful. In this paper, we consider and classify microeconomic and microeconometric approaches to measuring this success. To do so, the evaluation problem is outlined and three estimation principles are presented. For each of these, underlying assumptions are identified and the consequences of their violation discussed.

Keywords: Evaluation Methods, Public Policy Measures, Microeconometrics
JEL-classification: H43, O22, O38, O57

Contents

*Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena (DE)

1 Introduction

The objective of evaluating public policy measures is to asses its implications and thus to obtain a measure for whether the respective program has been successful. What *success* actually means depends very much on the targets of the policy measure to be evaluated.

In this paper, we consider and classify microeconomic and microeconometric approaches to measuring the success of public policy measures. Thus experimental studies are not considered. Moreover we do not consider macro analysis and meso analysis is only considered implicitly, when presenting estimation principles. No indirect effects, i.e. *external effects* are considered. This might be a crucial omission when considering e.g. innovation programs, since some of those might be designed exactly to create external effects (compare Klette, Møen and Griliches (2000) for an attempt).

The next section characterizes the evaluation problem. Section 3 then presents three estimation principles, the assumptions behind and the consequences of violating these assumptions. Most attention is devoted to cross-section estimations since they are the most important ones in the empirical literature.

2 The Evaluation Problem

One of the basic tasks of evaluation is to measure the impact of a policy measure on its participants, that is to measure the effect of the *treatment on the treated*. Before considering different approaches to doing so, let us state this task formally. Let $\mathbf{y}_t$ be a $(N \times 1)$ vector of realizations of a *target variable* at time t. N is the number of participating and non-participating agents that enter an evaluation study. Which variable is to be covered by $\mathbf{y}_t$ will be determined by the targets of the program to be evaluated. In a setting of technology programs one could think of a binary variable, denoting e.g. whether a firm has introduced an innovation, but it can also be a metric variable, measuring e.g. the amount of savings or some measure of firms' R&D-performance.

Further, let $\mathbf{X}_t$ be a $(N \times k)$ matrix of variables that can explain $\mathbf{y}_t$. The choice of which variables to enter $\mathbf{X}_t$ is driven by an economic model that is set up by the evaluator, i.e.

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t \sim i.i.d.$ denotes a vector of unobservables with $\mathrm{E}(\mathbf{u}_t) = \mathbf{0}$. Suppose, that agents participate at the program at time $h \in (t, t+\tau)$. Thus the task of the evaluator is to describe the part of evolution of $\mathbf{y}$ that is due to the program being evaluated. In its most general form, the model after the introduction of the program is

$$\mathbf{y}_{t+\tau} = \mathbf{X}_{t+\tau} \boldsymbol{\beta}_{t+\tau} + \mathbf{u}_{t+\tau}.$$

However, to be able to make comparisons between participants and non-participants and thus to obtain meaningful results of the evaluation, we often have to introduce assumptions on which part of model (1) to be actually affected by the program. Hence, suppose a program to affect either $\mathbf{X}$, e.g. by changing the *factor endowment* of firms, or $\boldsymbol{\beta}$, e.g. by modifying the *behaviour* of firms. Throughout this overview we will assume that the influence of the program can be measured additively, i.e.

$$\mathbf{y}_{t+\tau} = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{d}\alpha + \mathbf{u}_{t+\tau}, \tag{2}$$

where $\mathbf{d}$ is a vector with some non-zero value for participants and 0 for non-participants. In its simplest case, $\mathbf{d}$ can be a dummy vector where "1" denotes participation[1]. In other cases $\mathbf{d}$ might consist of data on received payments etc.

To summarize different approaches to the evaluation of public policy measures we need to deal with participants and non-participants separately. Therefore let $\mathbf{y}_t^{(1)}$ be a partition of length n of $\mathbf{y}_t$ of those agents that participate in the program (hence the superscript (1)). Correspondingly, let $\mathbf{y}_t^{(0)}$ denote a vector of length $N - n$ with the state of a sample of non-participants at time t.

Then $\mathbf{y}_t^{(1)}$ denotes the state of the participants *before* treatment and $\mathbf{y}_{t+\tau}^{(1)}$ thereafter; corresponding notation applies for non participants (0). The objective of an evaluation study consists in identifying the effects of a public program such that they can be separated from that evolution of $\mathbf{y}_t$ that would have occurred without the existence of the program under scrutiny. To express this formally, denote this hypothetical state as $\mathbf{y}_{t+\tau}^{(c)}$. This vector is called the *counterfactual*. Using this variable, the evaluation problem can be expressed as measuring the effect of *treatment on the treated*, i.e.

$$\mathbf{y}_{t+\tau}^{(1)} - \mathbf{y}_{t+\tau}^{(c)} = \boldsymbol{\Delta}_{t+\tau}^{(1,c)} \tag{3}$$

or for the i'th individual (e.g. firm), i.e. for the i'th elements of the above vectors

$$y_{i,t+\tau}^{(1)} - y_{i,t+\tau}^{(c)} = \delta_{i,\tau}^{(1)},$$

Obviously it is impossible to know both vectors simultaneously. This phenomenon has been denoted the *fundamental evaluation problem* (e.g. Heckman *et al.*, 1999). This evaluation problem would be easily solved if participating agents do not differ systematically from non-participants, both at time t (hence before treatment), i.e.

$$\mathrm{E}\left(\mathbf{y}_t^{(1)}\right) = \mathrm{E}\left(\mathbf{y}_t^{(0)}\right) \tag{4}$$

However, this is generally not the case since the aim of a program is usually to support exactly those agents, whose target variable *does* differ systematically and to select them for program participation.[2]

[1]Note that this implies that the outcome of the program would be identical for all participants, i.e. a shift by α. This case is however very often rejected (e.g. Heckman, LaLonde and Smith, 1999, p. 1885).

[2]Other, less obvious cases are possible. Thus evaluation studies of *SEMATECH*, a research consortium in the semiconductor industry suffered from the fact, that this consortium comprised all major firms and 80% of the turnaround of this industry. E.g. Irwin and Klenow (1996).

Thus, the choice is not random and it is therefore impossible to estimate the treatment effect by means of a simple comparison between participants and non-participants. This effect has been called the *sample selection bias* (e.g. Heckman and Hotz, 1989).

Intuitively, a natural approach to obtaining an estimate for $\mathbf{\Delta}_{t,t+\tau}^{(1)}$ would be to simply ask participants to quantify their benefits (or losses) due to their participation at the program. Of course, this approach bears the risk of receiving systematically biased estimates[3] e.g. due to strategic answering. Therefore it seems preferable to apply microeconometric methods to obtain an estimate of the counterfactual. In the following section we present different approaches to do so and hence to deal with the evaluation problem.

3 Different Principles of Estimating the Counterfactual

In estimating the counterfactual, three principles have been suggested in the literature (e.g. Heckman *et al.*, 1999). These are

1. the before-after estimator,
2. the difference in difference estimator and
3. cross section models.

We present these principles and their econometric correspondence in turn, describing underlying assumptions and implications for disaggregate analysis.

3.1 The Before-After Estimator

3.1.1 The Basic Principle

Denote $\bar{\mathbf{y}}_t^{(\cdot)}$ the *mean* of vector $\mathbf{y}_t^{(\cdot)}$. Assume, that the average outcome of the "no-treatment state of participants after treatment" (i.e. the counterfactual) can be approximated by the pre-program state. That is,

$$\mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(c)}\right) = \mathrm{E}\left(\bar{\mathbf{y}}_t^{(1)}\right).$$

Then, a policy measure's *average effect of treatment of the treated (ATE)*, $\bar{\mathbf{\Delta}}^{(1)}$, can be consistently estimated by the *before-after estimator*:

$$\widehat{\bar{\mathbf{\Delta}}}_{t,t+\tau}^{(1)} = \bar{\mathbf{y}}_{t+\tau}^{(1)} - \bar{\mathbf{y}}_t^{(1)} \quad = \quad \underbrace{\left(\bar{\mathbf{y}}_{t+\tau}^{(1)} - \bar{\mathbf{y}}_{t+\tau}^{(c)}\right)}_{a} + \underbrace{\left(\bar{\mathbf{y}}_{t+\tau}^{(c)} - \bar{\mathbf{y}}_t^{(1)}\right)}_{b}, \tag{5}$$

where a gives the evaluation equation (3) and b gives the approximation error. Hence, this approach yields a consistent estimate of (3) if $\mathrm{E}(b) = 0$, i.e. if the approximation error

[3] See e.g. Oldsman (1996) who used estimated savings due to the participation at the ITES-program.

averages out and there is no systematic evolution in $\bar{\mathbf{y}}_t^{(1)}$. Then, the counterfactual can be approximated by the state of the target variable before treatment (at time t). The major advantage of this estimator is its weak demand for data. The estimator can be implemented on panel data or even repeated cross-section data on the participants alone, which is not the case for the following estimators. Its major drawback is that the assumption that the approximation error averages out is easily violated, namely if systematic, non-idiosyncratic shocks occur within period $[t, t+\tau]$. There are at least two reasons, why such shocks can be expected

Exogenous evolution of explaining variables. Suppose that some of the elements of $\mathbf{X}$ evolve systematically over time even without program participation of agent i (i.e. exogenously). The before-after estimator would account for this evolution as contribution of the program, estimations thus would be biased. This bias can be expected to be the larger, the longer the observed time interval τ.

Strategic behaviour by the participants. This phenomenon is often encountered in the evaluation of labor market programs. However it can be expected in all kind of public policy measures. Suppose, that public authorities announce a measure to support R&D measures, say in a certain technological area. Firms that are eligible for participation and have planned similar R&D activities can be supposed to postpone these activities and to take them up only after participation. This behaviour can be expressed as a modification of vector $\boldsymbol{\beta}$. This behaviour will influence $\mathbf{y}$ at least temporarily. When measures are postponed, this will usually result in a temporary drop of $\mathbf{y}$ and a subsequent recovering once the measure is introduces. This behaviour has entered the literature under the notion of *Ashenfelter's dip* (Ashenfelter (1978), see Figure 1 for an illustration). Once this is the case, the result of estimator (5) depends crucially of the evaluator's choice of t and τ. To my knowledge, there is to date no attempt to quantify this phenomenon within the evaluation of industrial programs.

3.1.2 Application to Microdata

Considering microdata is appropriate if we are interested in the *distribution* of the outcomes of the public program, rather that its average. Applying the before-after estimator to microdata yields

$$\widehat{\boldsymbol{\Delta}}_{t,t+\tau}^{(1)} = \mathbf{y}_{t+\tau}^{(1)} - \mathbf{y}_t^{(1)} \quad = \quad \underbrace{\left(\mathbf{y}_{t+\tau}^{(1)} - \mathbf{y}_{t+\tau}^{(c)}\right)}_{a} + \underbrace{\left(\mathbf{y}_{t+\tau}^{(c)} - \mathbf{y}_t^{(1)}\right)}_{b}, \tag{6}$$

with terms a and b corresponding to equation (5). Consider the case of autonomous evolution of $\mathbf{X}$, the case of evolving $\boldsymbol{\beta}$ can be developed correspondingly. With equation

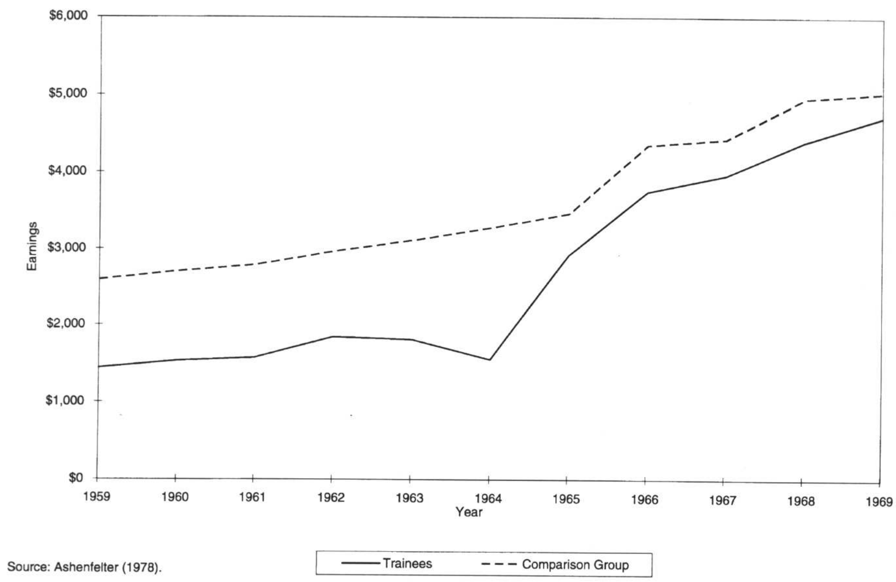

Figure 1: Mean annual earnings prior, during and subsequent to training for a 1964 training program and comparison group. An Illustration of "Ashenfelter's dip"

(1) the state of $\mathbf{y}^{(1)}$ before treatment can be explained as[4].

$$\mathbf{y}_t^{(1)} = \mathbf{X}_t^{(1)}\boldsymbol{\beta} + \mathbf{u}_t^{(1)}$$

and corresponding for $t + \tau$ and the counterfactual c. Suppose further that $\boldsymbol{\Delta}_\tau^{(1)}$, the outcome of participation at a program can be described by a model of type (2), i.e. by a simple shift in the intercept term of the regression. That is we can set

$$\mathbf{X}_{t+\tau}^{(1)}\boldsymbol{\beta} = \mathbf{X}_{t+\tau}^{(c)}\boldsymbol{\beta} + \mathbf{d}\alpha$$

hence

$$\mathbf{y}_{t+\tau}^{(1)} - \mathbf{y}_{t+\tau}^{(c)} = \underbrace{\mathbf{X}_{t+\tau}^{(1)}\boldsymbol{\beta} - \mathbf{X}_{t+\tau}^{(1)}\boldsymbol{\beta}}_{0} + \mathbf{d}\alpha + \boldsymbol{\varepsilon}_{t+\tau}$$

with $\boldsymbol{\varepsilon}_{t+\tau} = \mathbf{u}_{t+\tau}^{(1)} - \mathbf{u}_{t+1}^{(c)} \sim i.i.d.$ and $\mathrm{E}(\boldsymbol{\varepsilon}_{t+\tau}) = \mathbf{0}$. The bias of $\widehat{\boldsymbol{\Delta}}_{t,\tau}^{(1)}$, i.e. the expected value of term b in equation (6) can be expressed as

$$\mathrm{E}\left(\mathbf{y}_{t+\tau}^{(c)} - \mathbf{y}_t^{(1)}\right) = \mathrm{E}\left(\mathbf{X}_{t+\tau}^{(c)}\boldsymbol{\beta} - \mathbf{X}_t^{(1)}\boldsymbol{\beta}\right) = \mathrm{E}\left(\boldsymbol{\Delta}_{X^{(c)}}\boldsymbol{\beta}\right) \quad (7)$$

[4]This assumes that $E\left(\boldsymbol{\beta}^{(1)}\right) = E\left(\boldsymbol{\beta}^{(0)}\right) = E(\boldsymbol{\beta})$, i.e. that participants and non-participants do not differ significantly in their behaviour. This is far from being granted but can be tested with a simple test for structural change, e.g. Wald-test.

From this equation, we see, that the distortion is the bigger, the larger $\mathbf{\Delta}_{X^{(c)}}$, the *autonomous evolution* of $\mathbf{X}^{(1)}$. The principle of difference in difference estimation (section 3.3) corrects for this distortion.

3.2 Cross-Section Estimators

3.2.1 The Basic Principle

A second principle compares participants and non-participants at time $t + \tau$. The cross section estimators are based on the assumption that the target variable's average value does not differ significantly for non-participants and the participants' counterfactual value, i.e.

$$\mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(c)}\right) = \mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(0)}\right). \tag{8}$$

Then the average treatment effect $\bar{\mathbf{\Delta}}$ can be estimated as

$$\hat{\bar{\mathbf{\Delta}}}_{t+\tau}^{(1,0)} = \bar{\mathbf{y}}_{t+\tau}^{(1)} - \bar{\mathbf{y}}_{t+\tau}^{(0)}$$

Note that assumption (8) is stronger than assumption (11), since it does not correct for an initial state $\bar{\mathbf{y}}_t^{(\cdot)}$. Indeed, assumption (8) can rarely be met due to the selection bias (cf. equation 4 on page 2). Therefore, different modifications of this assumption have been suggested that lead to different approaches to eliminating this bias and thus lead to different instances of the cross-section estimator. These are

1. Matching methods

2. Microeconometric selection models

We present these approaches in turn.

3.2.2 Matching Methods

3.2.2.1 Direct Comparison of Participating and Non-Participating Agents – Exact Matching Suppose the following restriction of assumption (8):

$$\mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(c)} | \mathbf{X}_{t+\tau}^{(c)} = \mathbf{x}_{t+\tau}\right) = \mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(0)} | \mathbf{X}_{t+\tau}^{(0)} = \mathbf{x}_{t+\tau}\right). \tag{9}$$

That is the state of a non-participant and the counterfactual of a participant do not differ significantly, given that their respective realizations of the describing matrix $\mathbf{X}$ are identical. This assumption is called the *conditional independence assumption (CIA)* since is conditional on the realization of $\mathbf{X}$. Suppose further (as in section 3.1.2) that the outcome of a program can be described as a shift in the intercept of the regression

$$\mathbf{y}_{t+\tau}^{(1)} = \mathbf{X}_{t+\tau}^{(c)}\boldsymbol{\beta} + \mathbf{d}\boldsymbol{\alpha} + \mathbf{u}_{t+\tau}^{(1)}$$

Then a straightforward approach to estimate $\boldsymbol{\alpha}$ is to find a matrix $\mathbf{X}^{(0)}_{t+\tau}$ that *exactly matches* matrix $\mathbf{X}^{(c)}_{t+\tau}$, infer $\boldsymbol{\beta}$ from it and from corresponding depending variable $\mathbf{y}^{(0)}_{t+\tau}$ and finally deduce $\mathbf{d}\boldsymbol{\alpha}$ from $\mathbf{y}^{(1)}_{t+\tau} - \mathbf{X}^{(0)}_{t+\tau}\hat{\boldsymbol{\beta}}$. Hence this approach amounts to finding for each participant p a non-participant i whose realizations $\mathbf{x}^{(0)}_{i,t+\tau}$ are identical to those of the participant, i.e. to $\mathbf{x}^{(1)}_{p,t+\tau}$.

Therefore, this approach is sometimes called *exact matching approach*. Obviously, to find such a corresponding agent is a formidable task whose burden increases with the number of variables included in the explaining matrix $\mathbf{X}$. While the approach might still be feasible when $\mathbf{X}$ contains nominal or ordinal variables, it can be expected to be virtually impossible once metric variables are involved. Therefore, this approach does not seem to be useful when analyzing firm data.

3.2.2.2 Generalized Matching Methods Generalized Matching Methods (often simply called Matching Methods) can be interpreted as an extension of the comparison approach described above. Let $b : \mathbb{R}^k \mapsto \mathbb{R}^1$ be a homogeneous function (the *balancing score*, see Rosenbaum and Rubin, 1983). Based on it we can modify assumption (9) to

$$\mathrm{E}\left(\bar{\mathbf{y}}^{(c)}_{t+\tau} |\, b(\mathbf{X}^{(c)}_{t+\tau}) = b(\mathbf{x}_{t+\tau})\right) = \mathrm{E}\left(\bar{\mathbf{y}}^{(0)}_{t+\tau} |\, b(\mathbf{X}^{(0)}_{t+\tau}) = b(\mathbf{x}_{t+\tau})\right),$$

i.e. the multidimensional matching problem from section 3.2.2.1 is reduced to a one dimensional one. An intuitive and often used case of $b(\cdot)$ is the *propensity score* of agents that expresses the agents' conditional probability (conditional on $\mathbf{X}$) to participate at a public policy program. This probability can be estimated with a standard Probit or Logit-model. On the basis of this estimate a corresponding agent can be found through a nearest-neighbor Matching Method (Hagen and Steiner, 2000, after Heckman *et al.*, 1999, p. 1953):

1. Consider the set of participants $\{(1)\}$ and non-participants $\{(0)\}$
2. Choose a participating agent $i \in \{(1)\}$ and corresponding $b(\mathbf{x}_i)$. Eliminate i from $\{(1)\}$.
3. Find a non-participant $j \in \{(0)\}$ with *minimum distance* D to i such that
$$D_{ij} = \left(j|\, \underset{j\in\{(0)\}}{\mathrm{Min}} [b(\mathbf{x}_i) - b(\mathbf{x}_j)]\right).$$
4. Declare j being the agent matching i.
5. Delete j from $\{(0)\}$ and go back to the first step until $\{(1)\}$ is empty.

A number of other generalizations of the matching process have been suggested. Instead of referring to a function $b(\cdot)$ it is possible to define a metric

$$A_i = \left(j|\, \underset{j\in\{(0)\}}{\mathrm{Min}} \|\mathbf{X}_i - \mathbf{X}_j\|\right),$$

where $\|\cdot\|$ denotes some distance norm. Then agent $j \in \{(0)\}$ is given weight 1 if this condition is fulfilled. This case can be considered as a specialized case of *kernel matching* where each $i \in \{(1)\}$ is matched by a weighted sum of all $j \in \{(0)\}$ and the weights are constructed according to j's respective distance to the i under consideration.

3.2.3 Microeconometric Selection Models

Suppose the selection bias, i.e. the violation of assumption (8) can be explained by some variable or a set of variables. According to whether these variables are *observable* or *unobservable* we distinguish different approaches.

3.2.3.1 Modeling the Selection on the Basis of Observable Variables

3.2.3.1.1 Control Function Estimator A bias in the selection participants implies that vectors $\mathbf{d}$ and $\mathbf{u}$ in equation (2) are correlated. Suppose that an agent's decision to participate at a measure can be described as function of observable variables $\mathbf{Z}$. The consequences for this on assumption (8) are that

$$\mathrm{E}\left(\mathbf{y}_{t+\tau}^{(c)}|\mathbf{X}_{t+\tau}^{(c)}, \mathbf{d}\right) \neq \mathrm{E}\left(\mathbf{y}_{t+\tau}^{(0)}|\mathbf{X}_{t+\tau}^{(0)}\right).$$

but

$$\mathrm{E}\left(\mathbf{y}_{t+\tau}^{(c)}|\mathbf{X}_{t+\tau}^{(c)}, \mathbf{d}, \mathbf{Z}_{t+\tau}\right) = \mathrm{E}\left(\mathbf{y}_{t+\tau}^{(0)}|\mathbf{X}_{t+\tau}^{(0)}, \mathbf{Z}_{t+\tau}\right).$$

Assume, we can model the decision to participate with a latent model of the form

$$\mathbf{p} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{v} \tag{10}$$

where $p_i > 0$ if agent i participates, else $p_i < 0$. If this is the case, α the outcome of a measure might be estimated consistently by inclusion of $\mathbf{Z}$ as *control variables* in the regression. I.e. we have a model of the form

$$\mathbf{y}_{t+\tau} \quad = \quad \mathbf{X}_{t+\tau}\boldsymbol{\beta} + \mathbf{d}\alpha + \mathbf{Z}_{t+\tau}\boldsymbol{\gamma} + \mathbf{v}_{t+\tau}.$$

In practical implementations, this amounts to include all variables that influence an agent's participation decision as control variables in a reduced form estimator.

3.2.3.1.2 Instrumental Variable Estimator This estimator uses the matrix $\mathbf{Z}$ as instrument to regress on $\mathbf{d}$ and thus to eliminate the correlation between $\mathbf{d}$ and $\mathbf{u}$. That is, we use a model of the form (10) to be regressed directly on $\mathbf{d}$. An approach along these lines has been chosen by Arvanitis and Hollenstein (2001). The problem with this approach is that it is virtually impossible to identify variables $\mathbf{Z}$ that are uncorrelated with $\mathbf{u}$ but at the same time correlated with $\mathbf{d}$. Then estimates can be expected to be biased and this approach should therefore be used with caution.

3.2.3.2 Modeling the Selection on the Basis of Unobservable Variables An often encountered problem is that the selection bias occurs due to *unobservable* variables. Think e.g. of a firm's quality of management or the intensity of support by the public authorities during the program implementation. In that case, an inclusion of correction variables $\mathbf{Z}$ does not remove the correlation between $\mathbf{d}$ and $\mathbf{u}$. The remedy for this phenomenon differs according to the data availability.

3.2.3.2.1 Fixed- or Random Effects Estimator If we dispose of panel data, we can specify a latent model of the form

$$u_{i,t} = \phi_i + v_{i,t}$$

with $v \sim i.i.d.$ and mean 0. This effect vanishes when estimating in differences, i.e. α can be estimated consistently on the basis of the following model

$$(y_{i,t+\tau} - y_{i,t}) = (\mathbf{x}_{i,t+\tau} - \mathbf{x}_{i,t})\boldsymbol{\beta} + d_i\alpha_t + (v_{i,t+\tau} - v_{i,t}).$$

Correspondingly, α can be estimated consistently based on a *random effects model*, i.e. where the following specification is appropriate

$$u_{i,t} = \phi_i + t\theta_i + v_{i,t}.$$

Again, these individual effects vanish when building differences in the usual manner (Hsiao, 1986, or Heckman and Hotz, 1989).

3.2.3.2.2 The Heckman Selection Correction For cases, where only cross-section data are disposable, Heckman (1976) suggests an approach that interprets the selection bias as an omitted variables problem. In this case, Heckman suggests a two-equation approach to be built up of equations of type (2) and (10). Based on assumptions on the joint distribution of $\mathbf{u}$ and $\mathbf{v}$, α can be estimated in a simultaneous or in a sequential approach.

3.3 The Difference in Difference Estimator

3.3.1 The Basic Principle

If we have panel data or repeated cross-section data on participants and non-participants different approaches are possible to tackle the evaluation problem, i.e. the problem of sample selection bias. One is the *difference in difference (did)*-estimator. Suppose that the *autonomous evolution* of the target variable of participants can be approximated by the evolution of the target variable of non-participants, i.e.

$$\mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(c)} - \bar{\mathbf{y}}_t^{(1)}\right) = \mathrm{E}\left(\bar{\mathbf{y}}_{t+\tau}^{(0)} - \bar{\mathbf{y}}_t^{(0)}\right). \tag{11}$$

Then, the average treatment effect, $\bar{\mathbf{\Delta}}$ may be estimated consistently as

$$\widehat{\bar{\mathbf{\Delta}}}^{(1,0)}_{t,t+\tau} = \underbrace{\left(\bar{\mathbf{y}}^{(1)}_{t+\tau} - \bar{\mathbf{y}}^{(1)}_{t}\right)}_{c} - \underbrace{\left(\bar{\mathbf{y}}^{(0)}_{t+\tau} - \bar{\mathbf{y}}^{(0)}_{t}\right)}_{d}, \tag{12}$$

that is as the difference of two differences. Expanding (12) by (5) yields a consistent estimate of the evaluation problem (3) if assumption (11) is fulfilled.

3.3.2 Application to Microdata

Obviously, estimator (12) cannot be applied to individuals, since it is impossible to identify states c and d simultaneously for any participating agent. Therefore it is not straightforward to apply this estimation principle to microdata. Suppose however that we are able to construct a comparison group via a matching process such that $\mathbf{X}^{(0)}$ matches $\mathbf{X}^{(1)}$. This implies that both matrices are of the same dimension, i.e. $N - n = n$. Then we obtain from (12), extending with (6)

$$\widehat{\mathbf{\Delta}}^{(1,0)}_{t,t+\tau} = \underbrace{\left(\mathbf{y}^{(1)}_{t+\tau} - \mathbf{y}^{(c)}_{t+\tau}\right)}_{a} + \underbrace{\left(\mathbf{y}^{(c)}_{t+\tau} - \mathbf{y}^{(1)}_{t}\right)}_{b} - \underbrace{\left(\mathbf{y}^{(0)}_{t+\tau} - \mathbf{y}^{(0)}_{t}\right)}_{d}. \tag{13}$$

Inserting (7) we obtain the bias of (13)

$$\mathrm{E}(b+d) = \mathrm{E}\left[\left(\mathbf{\Delta}_{X^{(c)}} - \mathbf{\Delta}_{X^{(0)}}\right)\boldsymbol{\beta}\right],$$

where $\mathbf{\Delta}_{X^{(0)}}$ is the autonomous evolution of explaining variables for non-participants. This expression makes evident that the quality of the did-estimator depends on the quality of the matching process.

References

Arvanitis, S. and Hollenstein, H. (2001). The Determinants of Adoption of Advanced Manufacturing Technologies – An Empirical Analysis Based on Firm-level Data for Swiss Manufacturing. *Economics of Innovation and New Technology*, *19*(5), 377-414.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earning. *Review of Economics and Statistics*, *6*, 47-57.

David, P. A., Hall, B. H. and Toole, A. W. (2000). Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy*, *29*, 497-529.

Feller, I., Glasmeier, A. and Mark, M. (1996). Issues and perspectives on evaluating manufacturing modernization programs. *Research Policy*, *25*, 309-319.

Georghiou, L. and Roessner, D. (2000). Evaluating technolgy programs: tools and methods. *Research Policy*, *29*, 657-678.

Hagen, T. and Steiner, V. (2000). *Von der Finanzierung der Arbeitslosigkeit zur Förderung von Arbeit – Analysen und Handlungsempfehlungen zu Arbeitsmarktpolitik* (Vol. 51). Physica.

Hall, B. H. and Van Reenen, J. (2000). How effective are fiscal incentives for R&D? A review of the literature. *Research Policy*, *29*, 449-469.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement*, *5*, 475-492.

Heckman, J. J. and Hotz, J. (1989). Choosing Among Alternative Nonexperimental Methods for Estomating the Impact of Social Programs: The Case of Manpower Training. *Journal of the American Statistical Asociation*, *84*, 862-881.

Heckman, J. J., LaLonde, R. J. and Smith, J. A. (1999). The Economics And Econometrics of Active Labor Market Programs. In *Handbook of Labor Economics, Vol. 3A*. Elsevier.

Heckman, J. J. and Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions. An Overview. *Journal of Econometrics*, *30*, 239-267.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge.

Irwin, D. A. and Klenow, P. J. (1996). High-tech R&D subsidies. Estimating the effects of Sematech. *Journal of International Economics*, *40*, 323-344.

Klette, T. J., Møen, J. and Griliches, Z. (2000). Do subsidies to commercial R&D reduce market failures? Microeconometric evluation studies. *Research Policy*, *29*, 471-495.

Oldsman, E. (1996). Does manufacturing extension matter? Am evaluation of the Industrial Technology Extension Service in New York. *Research Policy*, *25*, 215-232.

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

Shapira, S., Youtie, J. and Roessner, J. D. (1996). Current practices in the evaluation of US industrial modernization programs. *Research Policy*, *25*, 185-214.