

Alm, Bastian; Bade, Franz-Josef

Preprint

The impact of firm subsidies: Evaluating German regional policy

Suggested Citation: Alm, Bastian; Bade, Franz-Josef (2009) : The impact of firm subsidies: Evaluating German regional policy, Technische Universität Dortmund, Dortmund

This Version is available at:

<https://hdl.handle.net/10419/103402>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The impact of firm subsidies: Evaluating German regional policy

Bastian Alm and Franz-Josef Bade

October 2009

Abstract

Recently, the number of state interventions and of firm subsidies has largely increased. However, the possible outcome of such interventions is rather ambiguous. This paper investigates the impact of firm subsidies granted by German regional policy. The long experiences made with regional capital and employment assistances provide a good opportunity for evaluating the impact of firm subsidies in a systematic way. The first part of the paper begins with a brief introduction into the institutional structure of the German regional assistance scheme, its objectives and main instruments. It follows a discussion of earlier evaluation studies putting the emphasis on their methodological characteristics. In the second part, an alternative approach of impact analysis is developed concentrating on the economic performance of the individual firm assisted by investment grants. The empirical basis is given by linking data of the assisted firms with the social security data differentiated by firms. One main advantage of the uniform database is that it allows to measure and to compare the employment performance of assisted and non-assisted firms free of distortions caused by different data sources. In addition, the database provides information about the number and structure of firm employment for a longer period of time. Thus, micro-econometric methods of evaluation are applied such as matching procedures by pairing firms with and without financial subsidies and controlling them for other influences, simultaneously. First results for the period 1998-2008 will be presented.

JEL Classification C14, C21, R00, R11, R28, R58

Key Words Policy evaluation, Regional Policy, Capital Subsidies, Statistical Matching

Correspondence: Technische Universität Dortmund. Fakultät Raumplanung. Fachgebiet Volkswirtschaftslehre. August-Schmidt-Straße 6. 44227 Dortmund. Alemania. Telephone: +49 231-755-2262. Fax: +49 231-755-6439. E-Mail: bastian.alm@tu-dortmund.de, franz-josef.bade@tu-dortmund.de

1. INTRODUCTION

Aimed at giving structurally weak regions an opportunity to catch up with the overall economic process, the federal government and the federal states still invest considerable amounts of money in the field of German regional policy. Thereby, the Joint Task program of the Federation and the states "*Improvement of Regional Economic Structure*" (Joint Task program) still captures the lion's share of total funds, although the program funds have started to decline beginning in the nineteen nineties. Political decision makers justify the enduring commitment of the federation mainly with the explanation that the macroeconomic conditions and their continuous variation affect the regions in a different way, which requires for the reallocation of resources between regions in order to establish equal living conditions throughout the federal territory.

In view of the persistent divergence in the development of the different regions, as well as the scarcity of public funds, it is not surprising that the Joint Task program and specifically its main instrument, capital subsidies for firms, are constantly part of an intense and critical debate about the achievements in terms of the causal effect of German regional policy. The fundamentally different perceptions of the conception and, closely linked with this, the success of the Joint Task program became especially evident in the dispute about the future path of German regional policy. In this regard, at least part of the politicians regarded the prospective concentration of the investment aid on structurally strong regions and therefore a paradigm change as being a promising strategy.

There is numerous evidence of uncertainty concerning the consequences caused by the capital subsidies within the Joint Task program. For instance, the Court of Auditors (Bundesrechnungshof) stated in 2006 that the allocation of funds is controlled insufficiently and that the German Federal Parliament is incompletely informed about the impact of the program (Bundesrechnungshof: 2006). A so-called "oral question" (*Große Anfrage*) of the German Federal Par-

liament from the year 2006 is indicative for the accuracy of the latter diagnosis. Therein, the deputies demanded scientific evidence that could show an impact of the program on the creation and preservation of jobs (Deutscher Bundestag: 2006).

That the effects of capital subsidies on both the assisted firms and the assisted regions still seem to be a “black box” for the deputies is rather astonishing, as the necessity of an adequate evaluation of the Joint Task program was already evident when it was introduced in 1969. Already the first framework plan for the Joint Task program contained the passage (Deutscher Bundestag: 1971):

“The planning committee expects the assisted regions to be subject to evaluation.”

According to Article 7, paragraph 2 of the German Federal Budget Code (BHO), the Joint Task program is subject to investigation of its profitability. This is required for all political financial subsidies, which are defined by the fact that they have a direct or indirect effect on the revenue and expenditure of the federal budget. Evaluation of German regional policy is based on three different pillars: execution control, target control and impact analysis. Only the latter tries to detect the relationship between the program (treatment) and the program effects (impact). With regard to the evaluation of capital subsidies, generally two different levels of impact analyses can be distinguished. Microeconomic approaches focus on describing the performance of the subsidized firms, whereas macroeconomic approaches investigate the overall effects on the economy. Both types of impact analyses are characterized by advantages and drawbacks in estimating the effects of capital subsidies.

The remainder of the paper is organized as follows. Section 2 contains an overview of the concept and principal objectives of capital subsidies as a regional policy instrument. In Section 3, the specifics and restrictions of the available data for evaluation studies of the Joint Task program are described. The fourth section compares the different methodological approaches which can be

used for the estimation of the impacts of capital subsidies, focusing on the application of a matching estimator.

2. INSTITUTIONAL SETTING

The Basic Law of the Federal Republic of Germany (*Grundgesetz für die Bundesrepublik Deutschland*, GG) ascribes the basic competence in the field of regional policy affairs to the federal states and municipalities. As the general basis of this allocation of rights and duties serves the subsidiarity principle that stipulates a bottom-up structure of the state.

The cooperation between the federal government and the federal states within the Joint Task program is legitimized by article 91a GG and is elaborated in the law on the Joint Task program from October sixth in 1969. For a time frame of four years, the federation and the states agree on a framework plan that is to be adapted on a yearly basis and contains, inter alia, the regulations for assistance and the assisted areas map. According to article 30 GG, the sole responsibility for the implementation of the different measures lies with the states.

2.1 Capital subsidies within the Joint Task program

Since local firms are seen to be the drivers of the regional economic growth and development in German regional policy, capital subsidies to firms form the largest part of the overall Joint Task program budget. Between 1998 and 2008, the overall Joint Task program budget amounted to 21.6 bn €, with nearly 73 percent (15.7 bn €) allotted to capital subsidies. Beyond that, local economic infrastructure projects (5.9 bn €) and, starting in 1995, non-investive measures (72 m €) can also be granted.

By subsidizing investment in private business, the Joint Task program intends to modernize the capital stock, as well as to attract new economic activity into the assisted regions. This, in turn, should lead to an increase in regional

competitiveness, income and especially employment. For instance, in terms of the number of jobs promoted, there were 378,203 new jobs promoted over the period 1998-2008 according to the official statistics of the Joint Task program.

Besides, the number of safeguarded jobs amounted to 1,092,376 in the same period.

The appraisal of the need for each of the 270 German functionally defined labor market regions was conducted on the basis of the four indicators that are presented with their respective weights in figure 1.

Fig. 1: Indicators for identifying assisted areas within the Joint Task program (period 2000-2006)

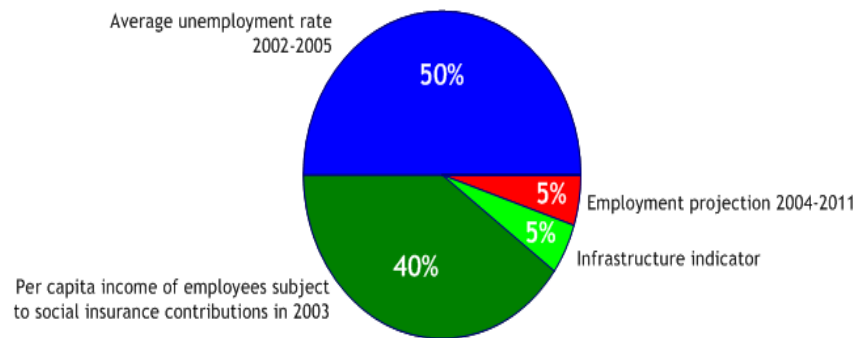
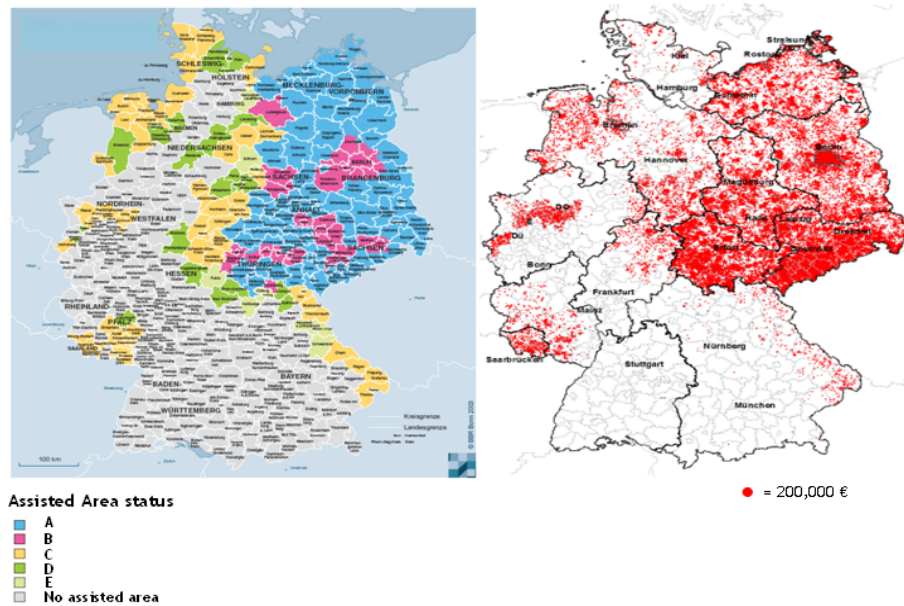


Figure 2 provides guidance on the assisted areas map 2000-2006 and the regional distribution of capital subsidies within the Joint Task program for the period covered in the ongoing evaluation. The respective status of the assisted areas ranged from A to E, with A denoting the highest assistance intensity and E the lowest. Apparently, the funding priority was unambiguously on eastern Germany. In western Germany, capital subsidies were especially provided in those regions that include primarily rural areas and old industrial regions undergoing structural changes, as well as the eastern Bavarian regions bordering with the Czech Republic.

Fig. 2: Assisted Areas map 2000-2006 and regional distribution of capital subsidies within Joint Task 1998-2008



Source: Ministry of Economics and Technology, own analysis

The principle eligibility for assistance depends not only on the type of investment and the belonging to an assisted area. An additional selection criterion is the assessment of the so-called “*primary effect*” (Tetsch et al.: 1996), which is only fulfilled if the majority of produced goods or services of the respective firm are suited for supra-regional sale. This is assumed for the manufacturing sector and 18 service sectors.

2.2 Principal objectives of capital subsidies within the Joint Task program

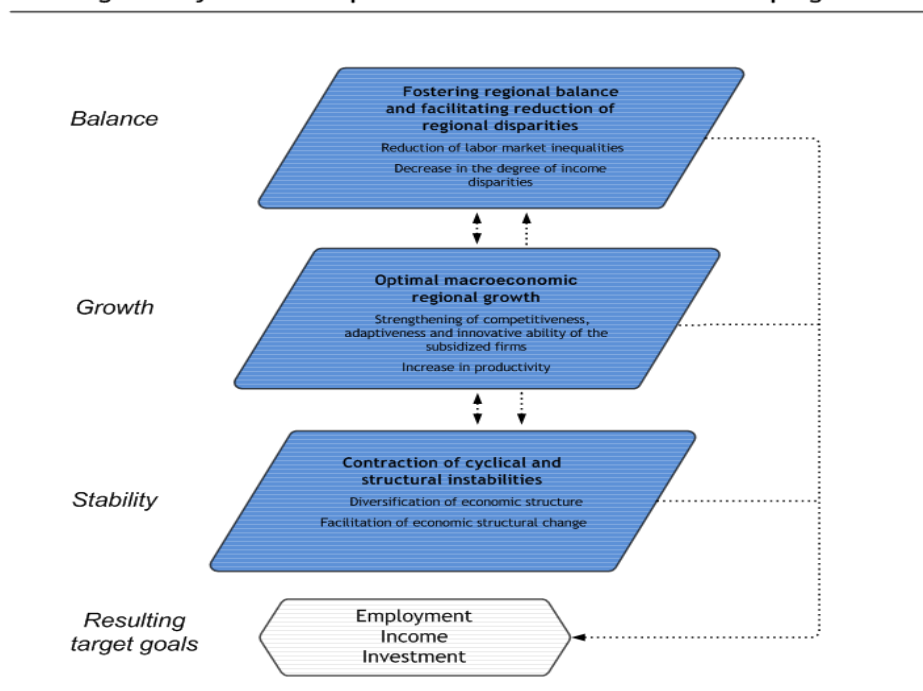
The primary emphasis of German regional policy is placed on enabling structurally weak regions to catch up with the overall economic development by compensating their locational disadvantages. The objective to foster the regional balance and to facilitate the reduction of regional disparities is derived from the constitutional claim of establishing equal living conditions in the entire federal territory (article 72 para. 2 and article 106 para. 3 of the GG). This general principle is in line with comprehensive regional planning at the federal level, which demands for balanced economic, infrastructural, social, ecological

and cultural conditions in the respective regions (sec. 2 para. 2 no.1 of Federal Regional Planning Act). From this (redistributive) perspective, the transfer mechanism of resources between the regions is regarded as being a necessary corrective of the market mechanism. Quoting the framework plan, especially the reduction of “*interregional differences in realized income and the availability of employment*” is a major objective (Deutscher Bundestag 2007).

But, the reallocation of resources between regions is not only justified with the balancing objective. Furthermore, granting of capital subsidies should lead optimal macroeconomic regional growth in the assisted areas. This is expected to happen through improved competitiveness, adaptiveness and innovative ability of the assisted firms, which in turn should increase their productivity.

Finally, the stability objective aims at keeping the consequences of cyclical and structural risks on the labor and income situation in the assisted areas at a low level. Thence, structure-conserving measures should be avoided as far as possible in favor of measures that facilitate the structural change. Figure 3 displays the objectives of capital subsidies and the resulting target goals within German regional policy.

Fig. 3: Objectives of capital subsidies within the Joint Task program



3. DATA

The data of the ongoing evaluation comes from two primary sources: The official statistics of the Joint Task program (Federal Office of Economics and Export Control, BAFA) and the statistics of employees subject to social insurance contributions (*Statistik der sozialversicherungspflichtig Beschäftigten*, Federal Employment Agency, BA).

3.1 Official statistics of the Joint Task program - BAFA data source

Subsequent to the selection of the funded projects, the federal states transmit the data from the approved grant applications to the BAFA that accounts for the acquisition of data. This data is the central element of the so-called *Bewilligungsstatistik*, which contains, beyond various identification attributes, projected figures concerning investment volume, approved funds and the number of promoted as well as safeguarded jobs for each investment project. The respective actual figures are collected after the completion of the investment project and gathered in the so-called *Verwendungsnachweisstatistik*.

According to the *Bewilligungsstatistik*, the Joint Task program allocated €15.7 bn € for capital subsidies to 26,762 firms that carried out 38,561 granted projects in the assisted areas during the 1998-2008 period.

3.2 Statistics of employees subject to social insurance contributions - BA data source

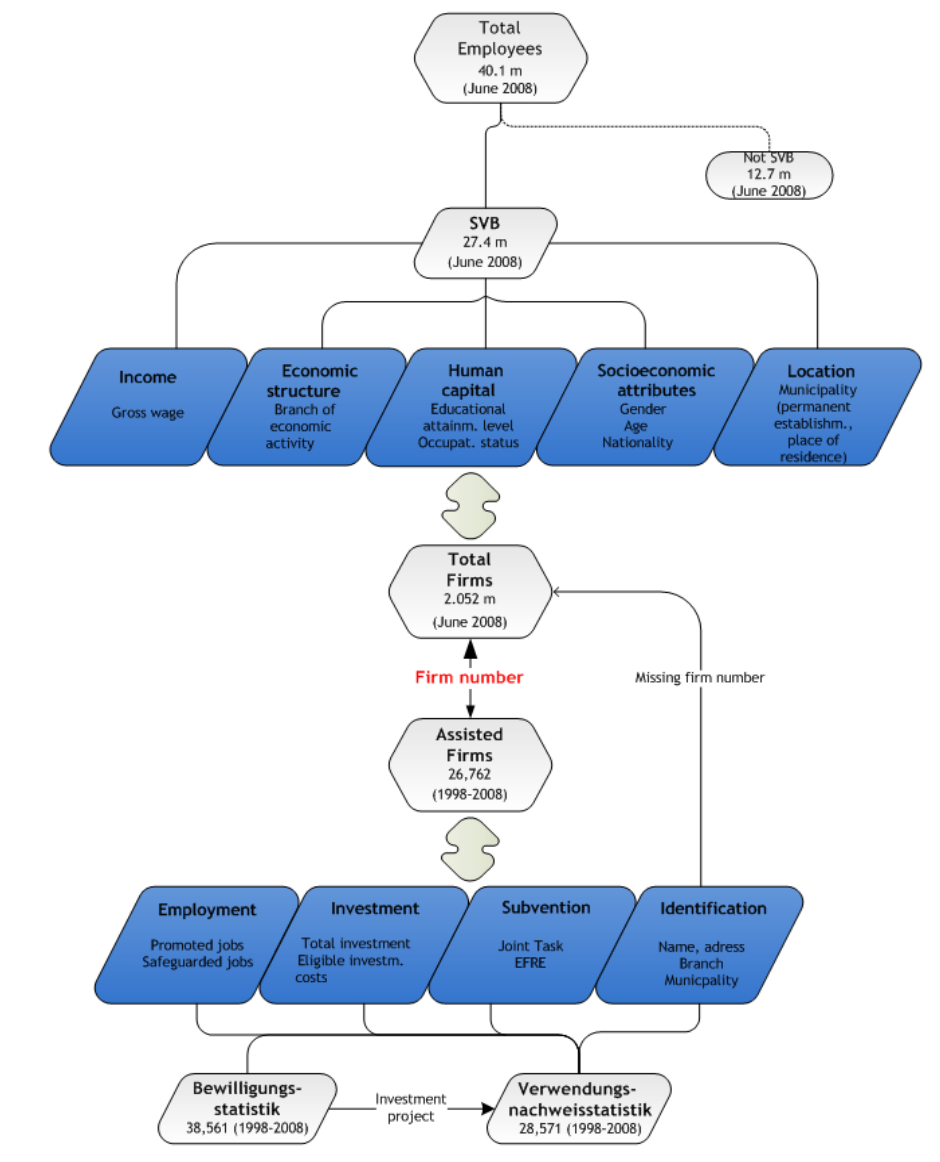
For various reasons, an adequate evaluation of capital subsidies within the Joint Task program cannot be performed solely on the basis of the official statistics of the Joint Task program. The most important drawback of the BATA data source is that both the *Bewilligungsstatistik* and the *Verwendungsnachweisstatistik* lack valid and appropriate indicators that could entail verifiable statements concerning the success of the program. Therefore, additional data is essential.

The statistics of employees subject to social insurance contributions of the Federal Employment Agency (BA) is based on reportable data by employers for their employees subject to social insurance contributions (SVB). However, this statistic does not show the total number of employees in Germany, as not all persons in employment are included in it: self-employed persons, officials and persons whose compensation does not exceed 400 € per month ("*minijob*"). In June 2008, the total number of employees exceeded the number of employees subject to social insurance contributions (SVB) by ca. 45 percent. A very valuable characteristic of the statistics of employees subject to social insurance contributions lies in the availability of quarterly retrospective structural data. Additionally, it provides information on gross wage and employment durations at annual intervals.

The empirical basis of the ongoing evaluation is given by the linkage of both data sources. For that purpose, the data of the statistics of employees subject to social insurance contributions has to be aggregated on the firm level instead of the employee level. The data thus obtained will be linked with the BAFA data source by using a component of both data sources: the firm number that is given to each firm with employees subject to social insurance contributions by the BA. In the case of a wrong or missing firm number in the BAFA data source, additional information can be used in order to identify the respective firm in the BA data source as, for instance, its name or address. After the linkage of these two data sources the sample therefore allows one to follow the development of each single identified subsidized firm and all non-subsidized firms until 1983 in western Germany and 1993 in eastern Germany.

Figure 4 illustrates the disposable parameters of both data sources.

Fig. 4: Data structure of ongoing Joint Task evaluation (1998-2008)



4. EVALUATION METHODS

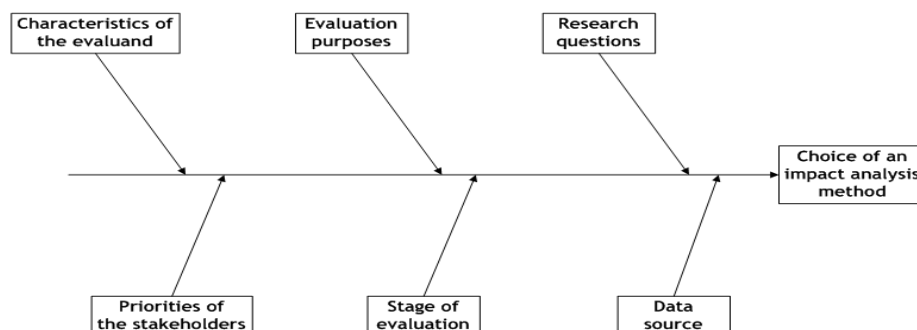
The main task of any scientific evaluation consists in the assessment of a certain policy, theme, program, project etc. (evaluand/treatment). For that purpose, the evaluator has to use an appropriate study design taking into account the specific circumstance of the case and whose results eventually reduce uncertainty about the impacts of the evaluand, as evaluation is not an end in itself. According to the Standards for Evaluation of the German Evaluation Society,

evaluations must be utile, feasible, proper and accurately at the same time, regardless to the specific evaluand (German Evaluation Society: 2003).¹

In German regional policy, evaluation of capital subsidies to industrial firms consists of three different types. The execution control (*Vollzugskontrolle*) examines whether the preconditions for allowing subventions were met and can be seen as a pure administrative control, while the target control (*Zielerreichungskontrolle*) investigates to what extent regional political objectives are accomplished at the time of evaluation. Only the impact analysis (*Wirkungsanalyse*) addresses if and to what extent capital subsidies are causative for the observed firm performance. Put in other words, exclusively the latter evaluation type is utile for measuring the causal effect of capital subsidies and therefore should be the core of any evaluation study of the Joint Task program.

In the course of designing an impact analysis method, the evaluator's role is to consider various factors, inasmuch as he or she aims at conducting a study that meets the four above-mentioned principles. In this context, attention has to be paid to not only to the characteristics of the specific evaluand and the priorities of the stakeholders. Moreover, the evaluation purposes and the stage of evaluation (ex-ante, mid-term, ex-post) have a bearing on the choice of the impact analysis method, as well as research questions and the specifics and restrictions of the available data.

Fig. 5: Determinants of choosing an impact evaluation method



¹ These principles are derived by the program evaluation standards by the Joint Committee on Standards for Educational Evaluation (1994).

Previous impact evaluation studies that were conducted starting in the late nineteen seventies differ particularly in their evaluation purposes, research questions, access to data and thus in their methodological background.² What they all have in common, however, is that they have to overcome the fundamental evaluation problem.

4.1 The fundamental evaluation problem

In a *randomized experiment*, the identification of the causal effect of an independent variable (treatment) on a dependent variable (outcome) is not challenging. After the random partition of subjects of a certain population into a treatment and a comparison group, the average treatment effect (α) can be calculated as follows:

$$\alpha = E[\bar{Y}^1 - \bar{Y}^0], \quad (1)$$

where $E[\bar{Y}^1 - \bar{Y}^0]$ stands for the expected difference in average outcomes between treated (\bar{Y}^1) and non-treated subjects (\bar{Y}^0) at any given time after the implementation of the treatment. As the subjects of both groups are drawn randomly from the same population, participation status is independent of all observed and unobserved baseline variables as a result of the law of large numbers.

Though in the case of the Joint Task program, the commissioning of capital subsidies does not rely on an experimental design for mainly two reasons. First, it seems doubtful that a random assignment to the program would be ethical: assigned firms would benefit from capital subsidies, regardless of their financial status, competitive position and even capital expenditure plan. This assuredly would give rise to a debate on whether the coercion to accept subsidies is tantamount with a high level of deadweight effects. Moreover, the according law as well as the framework plan for the Joint Task program define the prerequi-

² For a summary, see Lammers and Niebuhr (2004).

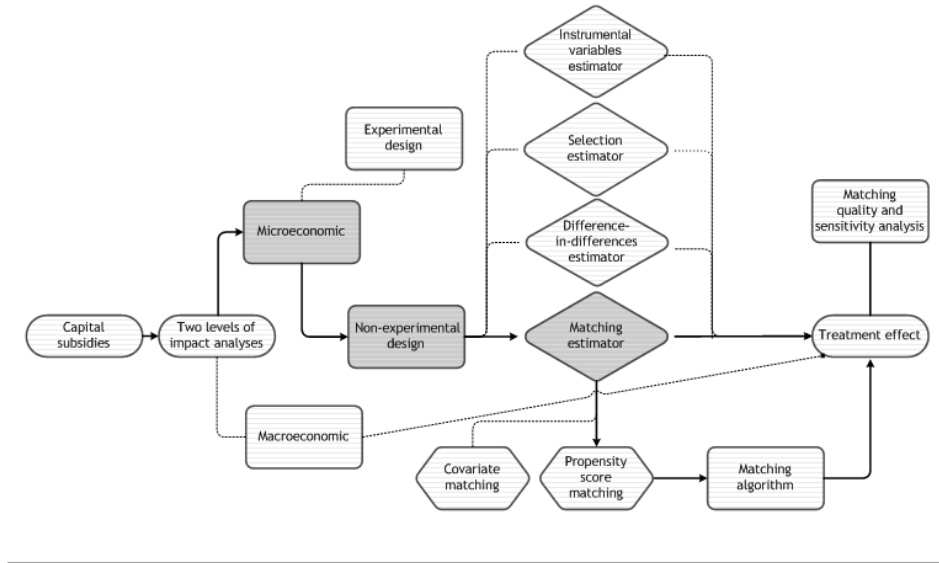
sites for the assignment and entitle eligible sectors. Since the program was already implemented in 1969, experimental design is not feasible for evaluation purposes of the Joint Task program.

As a consequence of *non-random* assignment, the two groups, on the one side the subsidized firms and on the other side the non-subsidized firms, may differ in their outcome variables even if the capital subsidies have no effect. Put in other words, differences in outcomes between the two groups cannot be attributed solely to the recipience of capital subsidies, since both groups may be incomparable prior to the treatment in their observable and unobservable variables (Rosenbaum: 1986). Therefore, the non-subsidized firms are no adequate predictor for the inherently non-observable counterfactual outcome of the subsidized firms in the event of their non-treatment. The fact that the counterfactual outcome is not observable by definition is often referred to as the fundamental evaluation problem. Given that, the calculation of the average treatment effect by means of equation (1) would yield biased results (Engel: 2001). In observational studies, different impact analyses methods are applied to steer around this fundamental evaluation problem. Thereby, *“research has moved a long way from relying on simple least squares methods for estimating average treatment effects”* (Imbens: 2004).

4.2 Impact analysis methods

The different impact analyses methods applicable to the assessment of the effect of capital subsidies fall into two different categories: macroeconomic and microeconomic methods. Figure 6 shows the broad range of impact analyses methods.

Fig. 6: Impact analysis methods of capital subsidies



4.2.1 Macroeconomic level

Macroeconomic approaches tend to simulate the impact of a specific treatment on the economy as a whole or specific sectors or regions of an economy. These approaches aim at taking both, substitution and external effects, into consideration.

To satisfy this objective, all interferences between the different components have to be brought into the model by means of a system of equations, guided by theoretical considerations. In the best knowledge of the author, the probably most advanced approach to estimate the impact of a public policy program on a macroeconomic level is the HERMIN modeling framework, which is based on the multi-sectoral HERMES model developed by the European Commission (D'Alcantara, Italianer: 1982). The HERMIN model that has been used frequently to evaluate the EU's Structural Funds Program, incorporates theoretical

foundations of the Keynesian small open economy model as well as elements of the neoclassical endogenous growth theory and typically consists of four sectors (agriculture, manufacturing, market services and non-market services). All in all, the model contains approximately 250 equations (Bradley et al.: 2006). Obviously, this kind of macroeconomic simulation models basically and essentially does not only require many assumptions, but make also high demands on the data source that is used.

The same applies to numerous impact analysis studies of regional investment incentives that have been conducted over the last few decades.³ Schalk and Untiedt (2000) estimated the effect of capital subsidies within the Joint Task program using a model with three equations: two factor demand functions for investment and labor and an output function. A key element of this model is the user cost of capital, which is reduced by capital subsidies and assumed to have a decisive influence on both, the factor demand and outcome decisions, and thus, substitution and outcome effects. The main drawback of this approach, comparable to the HERMIN model, is that its results are highly dependent on the derivation of equations convenient for estimation. Minor deviations from the approximations, that are necessary for the derivation of the model equations have a substantial influence on the results, and therefore could reduce empirical evidence of the causal effects of capital subsidies.⁴

Recently, Becker et al. (2008) applied a difference-in-differences regression-discontinuity design in order to evaluate the EU's Structural Funds Program. In the first step of their analysis, they exploited variation in GDP per capita at NUTS 3 level within NUTS 2 aggregates.⁵ Then, in order to estimate the effect of Objective 1 transfers, they compared the mean GDP per capita and employ-

³ See, for instance, Luger (1984), Faini and Schiantarelli (1985), Harris (1991), Daly et al. (1993), Schalk and Untiedt (2000) and Harris and Robinson (2005).

⁴ A common critique to this approach is further that spillover effects are not incorporated as input-output tables are not available for all regions in Germany, just as little as the effects on the factor prices, goods prices and substitution processes due to the missing regional price elasticity of demand data (Krieger-Boden, Lammers: 1996).

⁵ The acronym NUTS stands for Nomenclature of Territorial Units for Statistics. NUTS 2 is the regional level of aggregation for which most of the Structural Funds Program is assigned.

ment growth rates of regions with and without Objective 1 fund recipience. This comparison is not performed by contrasting the mean of the outcome variables of the assisted regions for one and the non-assisted regions for another. Rather, only those non-assisted regions are compared with the assisted regions that have approximately the same likelihood of belonging to the Objective 1 area.

In the case of Joint Task evaluation, this approach is inappropriate mainly because capital subsidies are provided in all labor market regions of eastern Germany. The problem of missing non-assisted comparison regions also arises in western Germany, given that the assisted areas differ not exclusively in their assignment status from the non-assisted areas.

4.2.2 Microeconomic level

Contrary to the mentioned macroeconomic approaches, microeconomic evaluation approaches can be applied to estimate the impact of capital subsidies on a set of outcome variables of interest on the firm level. Assume that the participation status of firm, d_i^* , can be written as:

$$d_i^* = Z_i\gamma + \vartheta_i, \quad (2)$$

where Z_i stands for various variables that affect the participation status with the respective coefficients γ , and ϑ_i is the error term. Further, let $d_i = 1$ if $d_i^* > 0$ and $d_i = 0$ if $d_i^* \leq 0$. Then, an outcome variable of interest for any firm under study (Y_i) is typically assumed to be a linear function of a vector of independent variables X_i with corresponding coefficients β and its participation in the Joint Task program:

$$\begin{aligned} Y_i &= X_i\beta + d_i\alpha + \varepsilon_i, \text{ in the case of treatment } (d_i = 1) \text{ or} \\ Y_i &= X_i\beta + \varepsilon_i, \text{ in the case of non-treatment } (d_i = 0), \end{aligned} \quad (3)$$

where α stands for the impact of the treatment on firm i and ε_i for the error term. If the error term ε is uncorrelated with the vector of independent vari-

ables X , the application of OLS regression techniques enables the measurement of the treatment effect on the outcome variable Y .

Due to the non-random granting of funds, there is no reason to believe that all firms have the same chance of receiving capital subsidies by the Joint Task program. Presumably, the participation status depends on firm characteristics that have a bearing on the outcome variable Y as well, so that a correlation between the participation status d and the error term ε is likely. Hence, an OLS regression of Y on X and d is not a valid instrument to measure the treatment effect of capital subsidies. Various methods can be applied to dissolve this problem that arises from selection bias: instrumental variables estimator, selection estimator, difference-in-differences estimator and matching estimator. While the two first mentioned are parametric methods, both the difference-in-differences estimator and the matching estimator are counted among the non-parametric approaches.

Instrumental variables estimator

Instrumental variables methodology was frequently used to evaluate treatment effects of social programs (Angrist, Imbens: 1995). The application of an instrumental variables method for estimating the effect of capital subsidies requires to identify at least one instrumental variable Z^* (instrument), that

- (i) determines participation status (d) of the firms and
- (ii) can be transformed into $g(Z^*)$ such that $g(Z^*)$ is uncorrelated with the error term (ε),
- (iii) is not completely (or almost) determined by independent variables X (Blundell, Costa Dias: 2000).

If all of these three conditions are satisfied, then the treatment effect can be estimated unbiasedly substituting d for $g(Z^*)$ and performing an OLS regression, as the instrument Z^* will affect the outcome variable Y only indirectly through the participation status (Caliendo, Hujer: 2005).

In the case of evaluating capital subsidies within the Joint Task program, it seems to be a sure bet that no instrument can be found that meets all mentioned conditions, especially that Z^* determines participation status without affecting the outcome variable directly. This condition is only fulfilled if the assignment of firms to the two treatment levels is random. Consequently, instrumental variables strategies are not suited to estimate the effects of the Joint Task program on the outcome variable Y .

Selection estimator

Another method that tries to address selection bias is the selection estimator based on the work of Heckman (1976, 1978, 1979), also referred to as the Heckit estimator. Compared to the instrumental variable approach, the assumptions of this method are even more demanding about the model structure:

- (i) at least one regressor that has a non-zero coefficient and that is independent of the error term \mathcal{G} is required in the decision rule (equation (2)) and
- (ii) the joint density of error terms ε and \mathcal{G} has to be either known or can be estimated (Blundell, Costa Dias: 2000).

The Heckit estimator involves first the calculation of the selection model by means of equation (2). Then, the part of the error term ε_i that is correlated with d_i (*Mill's Ratio*, $\hat{\lambda}$) has to be estimated for each firm, before $\hat{\lambda}$ is included as an additional regressor in the outcome equation, such that – under the common assumption that both error terms, \mathcal{G} and ε , follow the normal distribution and that $\sigma_{\mathcal{G}} = 1$ – we can rewrite equation (3):

$$Y_i = X_i \beta + d_i \alpha + \hat{\lambda} + \varepsilon_i. \quad (4)$$

$\hat{\lambda}$ takes the form $\rho \frac{\phi(Z_i \gamma)}{\Phi(Z_i \gamma)}$, if $d_i = 1$ and $-\rho \frac{\phi(Z_i \gamma)}{1 - \Phi(Z_i \gamma)}$, if $d_i = 0$, where ρ stands for the correlation coefficient of both error terms, ϕ for the density function- and Φ for the distribution function of the normal distribution.

Stierwald and Wiemers (2003) applied the Heckit estimator in an impact analysis of the Joint Task program in order to estimate the effect of capital subsidies on the development of firm investments in eastern Germany between 1999 and 2001. The results of the study indicated that, for an average subsidized firm, the investment level would have been significantly lower if the Joint Task program had not been in operation.

The fact that the Heckit estimator, unlike the instrumental variable method, does not rest on various exclusion restrictions entails that its robustness depends heavily on the validation of the assumptions about the functional form and the normal distribution of \mathcal{G} and ε . Imagine that \mathcal{G} is not normally distributed: then, the consistent estimation of γ in equation (2) is not possible. Furthermore, if the expectation of ε conditional on \mathcal{G} is not linear and/or \mathcal{G} is not normally distributed, $\hat{\lambda}$ specifies the relationship between $Z_i\gamma$ and d_i incorrectly, which will lead to biased results (Winship, Mare: 1992). Certainly, a major difficulty when using this selection model for evaluating treatment effects of capital subsidies is the estimation of the outcome equation (4). A correctly specified regression model should contain all relevant exogenous variables. Though, this does not seem to be applicable to the outcome variables of interest for the Joint Task program evaluation (e.g. employment and income), since their relevant exogenous variables are not known entirely and therefore not captured by the regression model.

Difference-in-differences estimator

The difference-in-differences estimator relies on the before-after estimator, which is the most intuitive method to estimate the impact of capital subsidies. The latter estimator (α^{BAE}) compares the mean outcome of the subsidized firms at a given point of time after receiving the subventions (\bar{Y}_t^1) with their mean outcome before the treatment (\bar{Y}_t^0):

$$\alpha^{BAE} = (\bar{Y}_t^1 - \bar{Y}_t^0)_1, \quad (5)$$

where the subscript 1 denotes conditioning on $d = 1$. This estimator assumes that the gain from the treatment can be written for each subsidized firm as $Y_t^1 - Y_t^0 = (Y_t^1 - Y_{t'}^0) + (Y_{t'}^0 - Y_t^0)$. Thereby, $(Y_{t'}^0 - Y_t^0)$, the difference of outcomes between the two points of time in the counterfactual situation of non-treatment, is referred to as the approximation error, which has to average out to zero in order to estimate equation (5) (Heckman, LaLonde, Smith: 1999). However, this assumption is usually violated –e.g., in the event of cyclical fluctuations.

The difference-in-differences method avoids this drawback by including the non-treated firms into the estimation of the treatment effects. The application of this approach requires

- (i) access to data in repeated cross-section or longitudinal format,
- (ii) that the mean difference in the non-treatment outcome measures is the same for the group of the subsidized and the group of the non-subsidized firms $(E(Y_t^0 - Y_{t'}^0 | d = 1) = E(Y_t^0 - Y_{t'}^0 | d = 0))$ and
- (iii) that these two groups do not differ systematically in their reactions on cyclical fluctuations.

If these conditions are fulfilled, then the impact of the Joint Task program can be estimated as follows:

$$\alpha^{DiD} = (\bar{Y}_t^1 - \bar{Y}_{t'}^0)_1 - (\bar{Y}_t^0 - \bar{Y}_{t'}^0)_0. \quad (6)$$

As the statistics of employees subject to social insurance contributions allows to access longitudinal data, condition (i) is met without reservation. Elsewise, assumption (ii) can be violated if the subsidized firms reduce their investment and/or employment prior to filing the grant application.⁶ This source of bias has become known in the literature as “*Ashenfelter’s dip*” (Ashenfelter: 1978). Condition (iii) is likely to be unreasonable as well, especially in the event of diverging pre-treatment characteristics among the two groups.

⁶ The regulations of the Joint Task program require that the amount of investment related to the year of the eligible investment project exceeds the average amortization of the last three years by at least 50 percent or that the firm has to increase the number of employees by not less than 15 percent.

Matching estimator

The key idea of this approach is to address the problem of pre-treatment differences in relevant characteristics among the two groups by finding at least one similar counterpart (*statistical twin*) for each subsidized firm. Thus, the pairing of treated and non-treated subjects rests on a set of attributes that are observable prior to treatment assignment.

Basic idea

Statistical matching estimators have their roots in the so-called “*potential outcome approach*”. The first attempts to accentuate this framework in conjunction with proving causal relationships date back to the middle of the eighteenth century. In 1748, David Hume presented his view of the concept of causation⁷ and constructed a kind of prototype for a counterfactual situation:

Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.

Writings of Fisher (1918), Kempthorne (1952), Cox (1958) and Cochran (1965) also laid the foundation for this framework. Neyman (1923) was the first to apply formal statistical models that focused on solving the problem of causation. In his study, Neyman draws inference on treatment effects based on a *potential yields*-approach within an agricultural field experiment. Another mode of inference was provided by Fisher’s proof of contradiction (1925).

Later, Rubin (e.g., 1974, 1976a, 1977, 1978) succeeded in transforming this approach into a formal framework that permits causal inference in observa-

⁷ For a review on the historical development within this area, see Bunge (1959).

tional studies. Accordingly, this approach is commonly known as “*Rubin’s causal model*” (RCM, Holland: 1986).⁸

For ease of understanding of the RCM, consider we have a sample of firms under study indexed by $i = 1, \dots, N$, where any firm can be potentially either granted at least once by the Joint Task program (treatment group, $d = 1$) or not (comparison group, $d = 0$). In this case, the average treatment effect on the treated (α_{TT}) is defined as the difference between the two mean outcomes of the treatment and comparison group:

$$\alpha_{TT} = E[Y^1 - Y^0 | d = 1] = E[Y^1 | d = 1] - E[Y^0 | d = 1]. \quad (7)$$

The average outcome of the treatment group, $E[Y^1 | d = 1]$, can be calculated (arithmetic or geometric mean) without any limitations. But as no treated firm can be observed in the other state (no treatment) simultaneously, the value of $E[Y^0 | d = 1]$ is hypothetical, and thence, unknown.

Vice versa, the average treatment effect on the non-treated firms (α_{TNT}), as the expected outcome difference for a firm randomly drawn from the comparison group, cannot be observed as the data for calculating $E[Y^1 | d = 0]$ is missing:

$$\alpha_{TNT} = E[Y^1 - Y^0 | d = 0] = E[Y^1 | d = 0] - E[Y^0 | d = 0], \quad (8)$$

Assumptions

As mentioned above, the observed outcomes of the non-treated firms do not serve as a proxy for $E[Y^0 | d = 1]$ in terms of non-random Joint Task program assignment, given that the decision on filing a subsidy application could be based on characteristics that are likely to affect the (potential) outcome. This means that the subsidized and non-subsidized firms would differ even in the event of absence of treatment by the Joint Task program, which can be noted as:

⁸ Analog frameworks in diverse fields of study were developed by Thurstone (1930), Haavelmo (1943), Roy (1951) and Quandt (1952).

$$E[Y^1|d = 1] - E[Y^0|d = 0] = \alpha_{TT} + E[Y^0|d = 1] - E[Y^0|d = 0] . \quad (9)$$

If the following assumptions hold, matching is a powerful instrument to set the difference of the last two terms of the right side of equation (9) to zero ($E[Y^0|d = 1] - E[Y^0|d = 0] = 0$) and therefore sort out this (selection) problem, so that the average treatment effect can be estimated.

(i) *Stable unit treatment value*

All matching estimators are designed to ascertain the causal effect of treatment on the treated subjects. As this estimation is normally employed for a *set* of subjects, the outcomes of the single subjects may depend, beyond their respective participation status, as well on the treatment assignment and the outcomes of other subjects.

The *stable subject treatment value assumption* (SUTVA) states that the outcome of each subject is independent from the treatment assignment of the other subjects. Furthermore, hidden versions of treatment are not supposed to exist (Rubin: 1980, 1986, 1990; Little and Rubin: 2000). I.e., this assumption requires that a variation in the treatment assignment of any firm has no effect on the other (treated and non-treated) firms. Since each firm has two potential outcomes under the two possible treatment assignments, the number of potential outcomes is limited to $N * 2$.

(ii) *Conditional independence*

This assumption posits that, conditional on the vector X of covariates x_i , outcomes are independent of participation status:

$$Y^1, Y^0 \perp d | X . \quad (10)$$

Thereby, the covariates x_i are exogenous variables that are measured already before treatment – i.e., they are not affected by the participation status.

Given this, systematic differences in post-treatment outcomes between treated and non-treated subjects can be ascribed to the treatment solely, as the treatment status is independent of potential outcomes, if the characteristics that determine pre-treatment outcome differences are exhaustively captured in X . This means that, after conditioning on X , treatment and comparison group are balanced with respect to the distribution of the covariates x_i . In order to justify this fundamental assumption, the researcher has to identify all variables that have a bearing on the participation status as well as on the outcomes.

The different versions of this assumption have been called *ignorable treatment assignment* (Rosenbaum and Rubin: 1983), *selection on observables* (Barnow, Cain, Goldenberger: 1980) or *conditional independence assumption* (CIA, Lechner: 1999, 2002).

In terms of estimating the average treatment effect on the treated ($\hat{\alpha}_{TT}$), the CIA can be weakened and one only has to assume that the non-participation outcome is independent of participation status after conditioning on X :

$$Y^0 \perp d | X . \quad (11)$$

(iii) *Common overlap*

The third assumption finally states that one observes subjects with the same conditioning covariates x_i . A match for all treated subjects ($d = 1$) can be found, if the likelihood of being treated is not only positive but also the same for subjects with the same values of X :

$$0 < P(d = 1 | X) < 1 . \quad (12)$$

If one is interested in estimating the average treatment effect on the treated ($\hat{\alpha}_{TT}$) instead of the average treatment effect ($\hat{\alpha}$), the condition reduces to:

$$P(d = 1 | X) < 1 . \quad (13)$$

When all assumptions are met, it can be hypothesized that the matched pairs of subsidized and non-subsidized firms do not differ in their likelihood of being treated by the Joint Task program. Then, if there are no interferences between firms (SUTVA) and all relevant differences between subsidized and non-subsidized firms are covered in their observable attributes (CIA), so that the probability of being treated is determined by the values of x_i for each firm (common support), the mean outcome of the non-subsidized firms can be used as a proxy for the potential outcome the subsidized firms would have had in the event of their non-treatment.

In the case of the Joint Task program evaluation, the most critical assumption of the RCM framework is the limitation of the number of potential outcomes of each subject to 2 (treatment/non-treatment) and their independence from the participation status of other subjects. Nevertheless, in terms of capital subsidies, it seems to be reasonable to assume that the extent of such spillover effects can be negligible, so that the data meets the requirement for SUTVA. Nonetheless, the potential problem of hidden biases due to unobserved covariates has to be addressed by using sensitivity analyses (Rosenbaum: 2002).

The role of the balancing scores and the propensity score

Typically, researchers that use matching estimators, face trade-off between two types of bias. Consider, for instance, a researcher who must decide either to maximize the *quantity* of matched pairs or to maximize the similarity between the single statistical twins as a measure for the matching *quality*. In the first case, numerous inexact matched pairs may be the result of the matching algorithm. Contrariwise, the attempt to maximize exact matches may be accompanied by the exclusion of matched pairs.

The mere maximization of matched pairs seems to be incompatible with the claim of evaluation. In view of matching quality, the most convincing method

to condition on the vector of covariates X is to exactly match on the respective covariates x_i (*covariate matching*). NB: In order to meet the CIA, X has to include numerous covariates x_i . The *dimensionality problem* arises in the case of a high dimensionality of X and/or continuous covariates x_i . The higher the dimensionality of X and/or the number of continuous covariates x_i , the higher the difficulty to obtain statistical twins with exactly the same value of the covariates x_i .

Since the data for the ongoing evaluation of the Joint Task program contains numerous covariates, some of them continuous, another method than covariate matching should be used. *Balancing scores*, introduced by Rosenbaum and Rubin (1983, 1985a, b), are adequate techniques to condition on X and reduce the dimensionality problem:

“A balancing score, $b(X)$, is a function of observed covariates x_i such that the conditional distribution of x_i given $b(X)$ is the same for treated ($d=1$) and control subjects ($d=0$)” (Rosenbaum and Rubin 1983).

Equation (14) expresses that, after conditioning on a balancing score $b(x)$, other conditioning on X is not needed to obtain knowledge about the participation status:

$$X \perp d | b(x). \quad (14)$$

Different types of balancing scores can be estimated, ranging from the covariate matching ($b(x) = x_i$) to the so-called “propensity score” ($e(X)$). Although covariate matching is the finest balancing score, as aforementioned, it is not appropriate to Joint Task program evaluation. We henceforce focus on the estimation of propensity scores as the coarsest, but simplest balancing score.

The propensity score $e(X)$ stands for the conditional probability (P) of assignment to the treatment ($d = 1$), given the vector of covariates X :

$$e(X) = P(d = 1|X). \quad (15)$$

Rosenbaum and Rubin (1983) showed that if SUTVA, CIA and the common support assumption hold, matching by means of the propensity score can be used to generate unbiased estimates of the average treatment effect or average treatment effect on the treated.

Like any other probability, a propensity score ranges from 0 to 1. In the case of a random mechanism for the assignment of subjects to treatment or control, each subject has an equal and non-zero chance of being selected to the treatment – e.g., if Joint Task program assigned firms to treatment by throwing a fair six-sides cube, where numbers 1 to 5 denote “no subsidy” and number 6 “subsidy”, the propensity score would be $e(X) = \frac{1}{6}$ for each firm.

In a non-randomized trial, the propensity score is, per definitionem, unknown, and has to be estimated by the data. The most commonly used method of propensity score estimation is a logistic regression model. Accordingly, the endogenous variable (treatment) is nominal scaled with two possible values (1, 0), while the exogenous variables can be categorical and/or metric.

A propensity score of, for instance, $e(X) = 0.5$ for two matched firms does not imply that both firms are equal in terms of the values of the single covariates x_i . Nevertheless, the mean values of metrical covariates will be the same for the treatment and the comparison group, as well as the proportion of ordinal and categorical covariates, such that both firms will have the same distribution of X .

The average treatment effect on the treated ($\hat{\alpha}_{TT}^{PS}$) then takes the form:

$$\hat{\alpha}_{TT}^{PS} = E\{E(Y^1|e(X), d = 1) - E(Y^0|e(X), d = 1)\}, \quad (16)$$

whereas the outer expectation is taken over the distribution of $e(X)|d = 1$.

Choice of a matching algorithm

After calculating the propensity score for each subsidized firm and all firms of the comparison group, one still has to apply a matching estimator in order to identify one or more adequate control units of each subsidized firm in terms of analogous characteristics. For $P = P(d = 1|X)$, a typical matching estimator can be written as:

$$\hat{\alpha}_M = \frac{1}{N^1} \sum_{i \in I^1 \cap S_p} \{Y_i^1 - \hat{E}(Y_i^0 | d_i = 1, P_i)\}, \quad (17)$$

where N^1 denotes the number of firms in the treatment group I^1 , I^0 the comparison group, S_p the region of common support and the estimator for the counterfactual mean takes the form $\hat{E}(Y_i^0 | d_i = 1, P_i) = \sum_{j \in I^0} w(i, j) Y_j^0$.

Expressed in words, equation (17) states that each match for a subsidized firm $i \in I^1 \cap S_p$ is constructed as a weighted average over the outcomes of the non-treated firms $\sum_{j \in I^0} w(i, j) Y_j^0$ (Smith and Todd: 2005). The single weights $w(i, j)$ depend on the distance between the propensity scores P_i and P_j .

Alternative matching estimators can be used to define a *neighborhood* $C(P_i)$ for each treated firm that comprises only those firms of the comparison group $j \in I^0$ that meet the condition $P_j \in C(P_i)$. Hence, each treated firm is juxtaposed with a group of non-treated subjects $A_i = \{j \in I^0 | P_j \in C(P_i)\}$. The various matching estimators differ in how the neighborhood is defined and in terms of the weights $w(i, j)$.

Nearest-neighbor matching

The nearest-neighbor matching estimator (NN) is intuitively easy to understand: For each treated firm i the most similar control unit out of the compari-

son group $j \in I^0$ is sought with respect to the propensity scores P_i and P_j , such that:

$$C(P_i) = \min_j \|P_i - P_j\|, j \in I^0. \quad (18)$$

Accordingly, $w(i, j)$ is 1 for each match j with the closest propensity score P_j in view of P_i and zero for all other subjects in the comparison group.

The so-called “*Mahalanobis metric matching*”⁹ (MM) is used in order to obtain the most similar control subject, if the balancing score contains other covariates apart from the propensity score:

$$MM_{ij} = \left\{ (X_i - X_j)' S^{-1} (X_i - X_j) \right\}^{\frac{1}{2}}, \quad (19)$$

where X_i and X_j denote the values of the multivariate vectors of the treated and non-treated firms and S^{-1} , the variance-covariance matrix of X_j .

If various control subjects $j \in I^0$ are used as matches for firm i , variance decreases as more information is available to construct the counterfactual situation. Otherwise, the number of (on average) bad matches and therefore bias tends to increase under this method.

Caliper matching

The NN can be modified in a way that the pairing of treatment and control subjects is not only based on a minimal distance between the propensity scores P_i and P_j , but also on a *caliper* ε that stands for a pre-specified tolerable distance between P_i and P_j :

$$\|P_i - P_j\| < \varepsilon, j \in I^0. \quad (20)$$

⁹ For further explanation, see Cochran and Rubin (1973), Rubin (1976b) and Carpenter (1977).

Caliper matching involves a trade-off between variance and bias. The latter decreases, if inadequate matches are excluded if the distance between the propensity scores P_i and P_j exceeds ε . Conversely, the variance of the matching estimator increases due to the reduced number of matches.

Dehejia and Wahba (1999, 2002) introduced a special form of caliper matching called “*radius matching*”. The advantage of this estimator lies in its reduced variance, since all control subjects within the tolerable maximum propensity score distance are used to create the counterfactual situation.

Stratification or interval matching

The stratification matching method consists of grouping subjects into various strata with the objective of estimating the treatment effect within each stratum.

The first study that used stratification was conducted by Cochran (1968), who created five equal strata for one covariate and showed that this subclassification at the quintiles removes 95 percent of the bias associated with that covariate. Later, Rosenbaum and Rubin (1983) showed that stratification on the propensity score also balances the covariates that are used to estimate the propensity scores.

Dehejia and Wahba (1999) suggested to partition all subjects into intervals such that within each stratum there is no significant difference between P_i and P_j .

Kernel and local linear matching

The matching estimators that have been described so far are characterized by the fact that they use one or a few subjects of the comparison group to construct matches. Heckman, Ichimura, Todd (1998) and Heckman, Ichimura, Smith and Todd (1998) introduced kernel-based matching estimators as a possibility to incorporate virtually all control firms to construct the outcome for each treated firm in the case of non-treatment. The weights that the firms of the comparison

group $j \in I^0$ receive by the kernel function $G(\cdot)$ depend on their distance to P_i .

In terms of equation (17), $w(i, j)$ equates $G\left(\frac{P_j - P_i}{h}\right) / \sum_{k \in I^0} G\left(\frac{P_k - P_i}{h}\right)$, where $G(\cdot)$ is a kernel function and h a bandwidth parameter. For the kernel-based matching estimator then follows:

$$\hat{\alpha}_{KM} = \frac{1}{N^1} \sum_{i \in I^1} \left\{ Y_i^1 - \left(\sum_{j \in I^0} w(i, j) Y_j^0 \right) \right\}. \quad (21)$$

The weights $w(i, j)$ and the neighborhood of each treated firm $C(P_i)$ depend on the specific kernel function and the bandwidth parameter h that are used.

In place of conventional kernel weights, local linear weights may be applied since the latter converge faster around boundary points and yield better results in the case of different data designs (Galdo, Smith, Black: 2007). Local linear weights are calculated as follows:

$$w(i, j) = \frac{G_{ij} \sum_{k \in I^0} G_{ik} (P_k - P_i)^2 - [G_{ij} (P_j - P_i)] \left[\sum_{k \in I^0} G_{ik} (P_k - P_i) \right]}{\sum_{j \in I^0} G_{ij} \sum_{k \in I^C} G_{ik} (P_k - P_i)^2 - \left(\sum_{k \in I^0} G_{ik} (P_k - P_i) \right)^2}, \quad (22)$$

with $G_{ij} = \left(\frac{P_i - P_j}{h} \right) * G$.

All matching algorithms allow that each matched firm j can be either replaced to the comparison group or not. If a matching algorithm *with* replacement is used, a control firm j can be selected more than once as a match. This will lead to an improvement of matching quality, especially in the case of considerable differences in the propensity scores between the two groups of subsidized and non-subsidized firms. In contrast, the number of different matched non-treated firms declines if replacement is allowed.

Valuation of the matching estimator

Generally, there is no infallible matching algorithm as each has its advantages and disadvantages, so that various approaches should be applied in order

to find the most appropriate one. Even if only comparable non-subsidized firms are detected in the statistics subject to social insurance contributions and matched with the subsidized firms it is important to remind, however, that interferences in terms of substitution processes between firms are not captured with this technique. Certainly, this is the main shortcoming of all microeconomic impact analyses methods.

After performing an convincing matching algorithm, the quality of the matching procedure has to be assessed by using indicators such as standardized bias, t – Test, pseudo – R^2 or stratification test (Caliendo, Kopeinig: 2005). All of these indicators are suited for testing if differences in covariate means still exist between both groups after conditioning on the propensity score. If the verification of matching quality leads to a satisfactory outcome, the researcher might proceed to the estimation of the treatment effect, before sensitivity analyses have to be carried out.

Smith and Todd (2005) pointed out, that a combination of the difference-in-differences method with the matching technique (DiD matching estimator) can yield more robust results of the treatment effect than cross-section matching estimators, as the DiD matching estimator is not affected by bias in consequence of time-invariant observables. Blundell and Costa Dias (2000) arrive at a similar conclusion: “[...] *an approach that combines propensity score matching with the difference-in-differences technique is quite robust. It allows matching on pre-program ‘shocks’ and, by collecting good local pre-program labor market history data, allows the comparison group to be ‘placed’ in the same labor market.*”

5. CONCLUSION

Impact evaluation studies of capital subsidies within regional policy focus on estimating the causal effect of this investment incentive by drawing causal interferences. As no experimental data exists in this area, researchers must resort

to the utilization of observational data from the official statistics of the Joint Task program, administrative statistics and/or surveys.

The impact of capital subsidies can be measured on two different levels, whereas both, microeconomic and macroeconomic approaches are characterized by strengths and weaknesses. While the former are suited for the analysis of the performance of the subsidized firms but fail to assess the aggregate impact of capital subsidies in the case of substitution and external effects, the latter focus on the estimation of the effects on the economy or a specific sector of the economy. But macroeconomic models are not only highly dependent on various assumptions and extremely data hungry. Another drawback is that they strand in describing individual firm performance, which is explicitly desired by the relevant stakeholders.

The non-parametric matching estimator in combination with the difference-in-differences technique is, under mentioned conditions, a powerful approach to contrast the performance of treated and adequate non-treated subjects in observational studies. It enables to answer the question what would have happened to the outcome of subsidized firms if they were not treated. For that purpose, subsidized and non-subsidized firms will be carefully paired controlling on a set of pre-treatment characteristics in the ongoing Joint Task evaluation. The advocated matching procedure in this paper is based on numerous covariates and abetted by the estimation of propensity scores. This approach has not been used in the field of German regional policy evaluation heretofore.

References

- Angrist, J. D., Imbens, G. W. (1995). Identification and estimation of local average treatment effects. National Bureau of Economic Research (NBER), Technical Working Papers 118.
- Ashenfelter, O. C. (1978). Estimating the effect of training programs on earnings. *Review of Economics and Statistics*, 60, 47-57.
- Barnow, B. S., Cain, G. G., Goldberger, A. S. (1981). Issues in the analysis of selectivity bias. *Evaluation Studies Review Annual*, 5, 43-59.
- Becker, S. O., Egger, P. H., von Ehrlich, M., Fenge, R. (2008). Going NUTS: The effect of EU Structural Funds on regional performance. CESifo, Working Paper 2495.
- Blundell, R., Costa Dias, M. (2000). Evaluation methods for non-experimental data. *Fiscal Studies*, 21, 427-468.
- Bradley, J., Mitze, T., Morgenroth, E., Untiedt, G. (2006). How can we know if EU cohesion policy is successful? Integrating micro and macro approaches to the evaluation of Structural Funds. Gesellschaft für Finanz- und Regionalanalysen (GEFRA), Working Paper 1/2006.
- Bundesrechnungshof (2006). Bemerkungen des Bundesrechnungshofes 2006 zur Haushalts- und Wirtschaftsführung des Bundes (einschließlich der Feststellungen zur Jahresrechnung 2005). Drucksache 16/3200, Berlin.
- Bunge, M. (1959). Causality. The place of the causal principle in modern science. Harvard University Press, Cambridge.
- Caliendo, M., Hujer, R. (2005). The microeconomic estimation of treatment effects – an overview. Institut zur Zukunft der Arbeit (IZA), Discussion Paper 1653.
- Caliendo, M., Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. Deutsches Institut für Wirtschaftsforschung (DIW), Discussion Paper 485.
- Carpenter, R. G. (1977). Matching when covariables are normally distributed. *Biometrika*, 64, 299-307.
- Cox, D. R. (1958). The planning of experiments. Wiley, New York.
- Cochran, W. G. (1965). Planning of observational studies of human populations. *Journal of the Royal Statistical Society, Series A*, 128, 234-265.
- Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 24, 295-313.

- Cochran, W. G., Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya, Series A*, 35, 417-446.
- D'Alcantara, G., Italianer, A. (1982). A European project for a multinational macrosectoral model. Document MS 11, DG XII, Commission of the European Communities, Brussels.
- Daly, M., Gorman, I., Lenjosek, G., MacNevin, A., Phiriyapreunt, W. (1993). The impact of regional investment incentives on employment and productivity. *Regional Science and Urban Economics*, 23, 559-575.
- Dehejia, R.H., Wahba, S. (1999). Causal effects in nonexperimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053-1062.
- Dehejia, R.H., Wahba, S. (2002). Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84, 151-161.
- Deutscher Bundestag (1971). Rahmenplan der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur" für den Zeitraum 1972 bis 1975. Drucksache 6/2451, Berlin.
- Deutscher Bundestag (2006). Große Anfrage zu Subventionen in der Bundesrepublik Deutschland. Drucksache 16/8441, Berlin.
- Deutscher Bundestag (2007). Sechsendreißigster Rahmenplan der Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur" für den Zeitraum 2007 bis 2010. Drucksache 16/5215, Berlin.
- Engel, D. (2001). Höheres Beschäftigungswachstum durch Venture Capital? Zentrum für Europäische Wirtschaftsforschung (ZEW), Discussion Paper 01-34.
- Faini, R., Schiantarelli, F. (1985). Oligopolistic models of investment and employment decisions in a regional context. Theory and empirical evidence from a putty-clay model. *European Economic Review*, 27, 221-242.
- Fisher, R. A. (1918). The causes of human variability. *Eugenics Review*, 10, 213-220.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver & Boyd, London.
- Galdo, J., Smith, J. A., Black, D. (2007). Bandwidth selection and the estimation of treatment effects with unbalanced data. Institut zur Zukunft der Arbeit (IZA), Discussion Paper 3095.
- German Evaluation Society (2003). Selected comments to the Standards for Evaluation of the German Evaluation Society – English edition. Wolfgang Bewyl (ed.), Alfter.
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1-12.

- Harris, R. (1991). The employment creation effects of factor subsidies: some estimates for Northern Ireland manufacturing industry. *Journal of Regional Science*, 31, 49-64.
- Harris, R., Robinson, C. (2005). Impact of regional selective assistance on sources of productivity growth: plant-level evidence from UK manufacturing, 1990-98. *Regional Studies*, 39, 751-765.
- 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. (1978). Dummy endogenous variables in a simultaneous equation systems. *Econometrica*, 46, 931-959.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153-161.
- Heckman, J. J., Ichimura, H., Todd, P. (1998). Matching as an econometric evaluation estimator: evidence from evaluation a job training programme. *Review of Economic Studies*, 64, 605-654.
- Heckman, J. J., Ichimura, H., Smith, J. A., Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66, 1017-1098.
- Heckman, J. J., LaLonde, R. J., Smith J. A. (1999). The economics and econometrics of active labor market programs. In: O. Ashenfelter and D. Card, *Handbook of Labor Economics*, IV, 1865-2073.
- Holland, P. (1986). Statistics and causal interference. *Journal of the American Statistical Association*, 81, 945-970.
- Hume, D. (1748). Enquiries concerning human understanding and concerning the principles of morals. In: L. A. Selby-Bigge (1975), University Press, Oxford.
- Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics*, 86, 4-29.
- Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards* (2nd ed.). Sage, Thousand Oaks, CA.
- Kempthorne, O. (1952). *The design and analysis of experiments*. Wiley, New York.
- Krieger-Boden, C., Lammers, K. (1996). Subventionsabbau in räumlicher Perspektive: Wirkungszusammenhänge und Schlussfolgerungen. Institut für Weltwirtschaft Kiel (IfW), Discussion Paper 280.
- Lammers, K., Niebuhr, A. (2004). Erfolgskontrolle in der deutschen Regionalpolitik: Überblick und Bewertung. *Hamburgisches Welt-Wirtschafts-Archiv (HWWA)*, HWWA-Report 214, Hamburg.

- Lechner, M. (1999). Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business and Economic Statistics*, 17, 74-90.
- Lechner, M. (2001). Identification of causal effects of multiple treatments under the conditional independence assumption. In: *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner and F. Pfeiffer, 43-58.
- Luger, M. I. (1984). Investment incentives and the demand for labor in U.S. regions. *Regional Science and Urban Economics*, 14, 481-503.
- Little, R. J., Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health*, 21, 121-145.
- Neyman, J. (1923). On the application of the probability theory to agricultural experiments. Essay on principles. Section 9. Translated and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science*, 5, 465-480.
- Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53, 873-880.
- Rosenbaum, P. R., Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rosenbaum, P. R., Rubin, D. B. (1985a). Constructing a comparison group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39, 33-38.
- Rosenbaum, P. R., Rubin, D. B. (1985b). The bias due to incomplete matching. *Biometrics*, 41, 103-116.
- Rosenbaum, P. R. (1986). Dropping out of high school in the United States: an observational study. *Journal of Educational Statistics*, 11, 207-224.
- Rosenbaum, P. R. (2002). *Observational studies*. Springer-Verlag, New York.
- Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3, 135-146.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1976a). Interference and missing data. *Biometrika*, 63, 581-592.
- Rubin, D. B. (1976b). Multivariate matching methods that are equal percent bias reducing: some examples. *Biometrics*, 32, 109-120.
- Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-21.
- Rubin, D. B. (1978). Bayesian interference for causal effects: the role of randomization. *The Annals of Statistics*, 7, 34-58.

- Rubin, D. B (1980). Discussion of "Randomization analysis of experimental data in Fisher randomization test" by Basu. *Journal of the American Statistical Association*, 75, 591-593.
- Rubin, D. B (1986). Which ifs have causal answers? Discussion of Holland's "Statistics and causal interference". *Journal of the American Statistical Association*, 81, 961-962.
- Rubin, D. B (1990). Formal modes of statistical interference for causal effects. *Journal of Statistical Planning and Inference*, 25, 279-292.
- Schalk, H. J., Untiedt, G. (2000). Regional investment incentives in Germany: impacts on factor demand and growth. *Annals of Regional Science*, 34, 173-195.
- Smith, J. A., Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125, 305-353.
- Stierwald, A., Wiemers, J. (2003). Auswirkungen der Gemeinschaftsaufgabe zur "Verbesserung der regionalen Wirtschaftsstruktur" auf die Investitionstätigkeit – Eine einzelwirtschaftliche Wirkungsanalyse für Betriebe des Verarbeitenden Gewerbes in Ostdeutschland. Institut für Wirtschaftsforschung Halle (IWH), Discussion Paper 185.
- Winship, C., Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, 18, 327-350.
- Thurstone, L. (1930). *The fundamentals of statistics*. The Macmillan Firm, New York.