European Commission

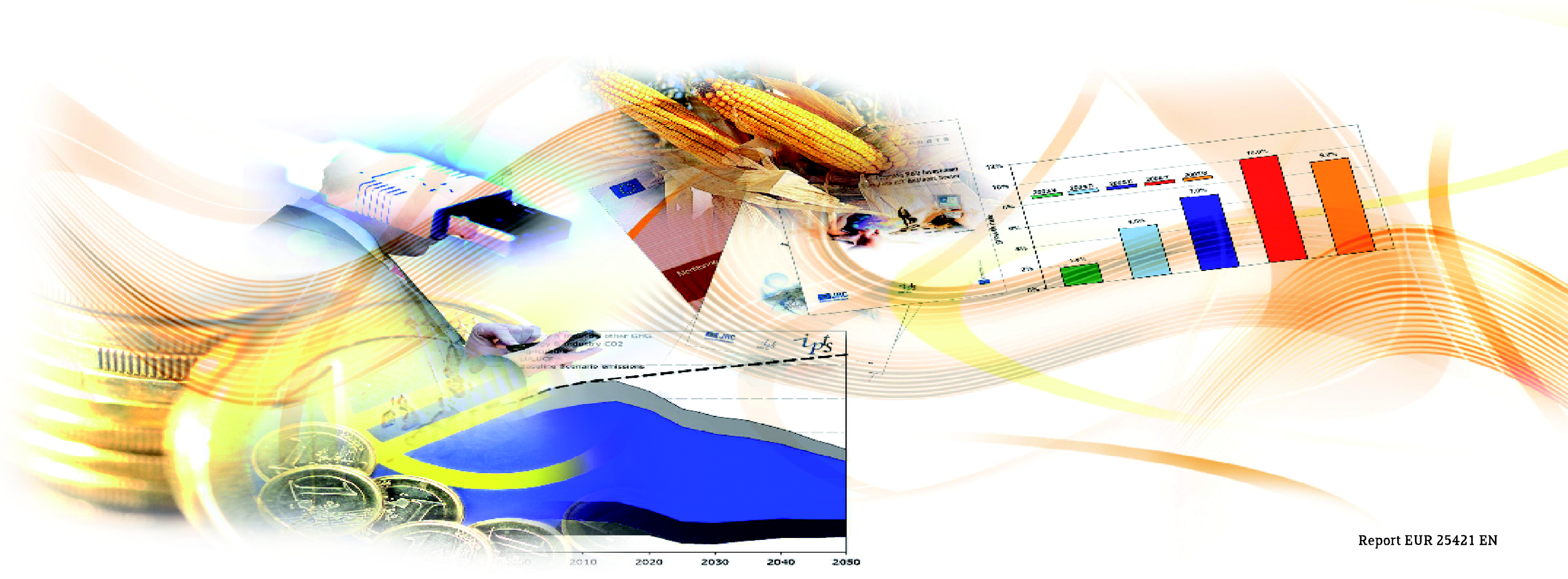
JRC SCIENTIFIC AND POLICY REPORTS

Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States

Volume 1: A micro-level approach

Author
Jerzy Michalek

2012



Report EUR 25421 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Prospective Technological Studies

Contact information
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-ipts-secretariat@ec.europa.eu
Tel.: +34 954488318
Fax: +34 954488300

http://ipts.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.

JRC 71977

EUR 25421 EN

ISBN 978-92-79-25681-3

ISSN 1831-9424

doi:10.2791/82592

Luxembourg: Publications Office of the European Union, 2012
Rome: Food an Agriculture Organization of the United Nations

Printed in Spain

Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States.

Volume 1: A micro-level approach

Author:
Jerzy Michalek

2012

EUR 25421 EN

Table of Contents

List of Figures

List of Tables

Summary

The main objective of this study is to show how various micro-economic *direct/indirect effects* (e.g. *deadweight loss, leverage effects*, etc.) and selected *general equilibrium effects* (e.g. *substitution and displacement effects)* of EU RD programmes can be calculated using recently developed advanced econometric semi-parametric evaluation methodologies. Answers to EU *Common Evaluation Questions* (CEQ) regarding the effects of an RD programme on programme beneficiaries at farm level (including deadweight loss and leverage effects) are provided by comparing changes in specific *result indicators* collected at a farm level (e.g. profits, employment, gross-value added, labour productivity, etc.) in the group of programme beneficiaries with an appropriately selected control group (counterfactual analysis - based on matching). Direct programme effects are calculated on the basis of Average Treatment on Treated (ATT) indicators (for programme beneficiaries), Average Treatment Effects on Non-Treated (ATNT) indicators (for programme non-beneficiaries) and Average Treatment Effects (for both groups) using a combination of propensity score matching (PSM) and difference in differences (DID) methods. A modified propensity score and difference in differences methodology (modified PSM-DID) is applied to derive various general equilibrium effects (e.g. substitution effects). The empirical analysis is focused on evaluation of effects of the SAPARD programme in Slovakia (years 2002-2005) and the Agrarinvestitionsförderungsprogramm (AFP) in Schleswig Holstein, Germany (2000-2006) using micro-economic data (balanced panels) of bookkeeping farms (including programme participants and non-participants) in respective countries.

1. Background

1.1. EU approach to the evaluation of the RD programmes

In recent years the evaluation of EU Member States' co-founded programmes was assigned particular importance. The administrative reform of the European Community (Agenda 2000) confirmed the significance of the monitoring and evaluation components, and extended periodic evaluation to all EU policies (Toulemonde et al., 2002). Meanwhile, evaluation has been recognized as a crucial component of policy development and became an integral part of EU programming at all levels, e.g. EU, national, and territorial, etc. (Vanhove, 1999; Ederveen, 2003; EC, 1999, 2002a, 2002b).

Evaluation of specific policy interventions can be undertaken for many reasons, for example: to assess a programme's impact, to improve programme management and administration (e.g. identify necessary improvements in the delivery of interventions) or to meet accountability requirements of funding institutions (Rossi, Freeman, 1993).

According to the EU definition, programme evaluation is a process that culminates in a judgement (assessment) of policy interventions according to their results, impacts and the needs they aim to satisfy[1]. In the case of structural and rural development (RD) programmes, EU regulations distinguish between ex-ante, mid-term, ex-post and ongoing evaluations. Ex-ante evaluations aim to optimize budgetary resources' allocation and improve the quality of programming by answering the question: "what impacts can be *expected* from a newly designed policy or programme?" Meanwhile the main purpose of mid-term and ex-post evaluations of EU programmes is to learn about:

- The programme's *effectiveness*, i.e. the degree to which a programme produced the desired outcome (an assessment of a programme's effectiveness implies a pre-definition of operationally defined objectives and criteria of its achievement), and

- Programme *efficiency*, i.e. the degree to which overall programme benefits relate to its costs.

In order to facilitate and improve the quality of evaluations, the EC issued several evaluation guidelines[2] laying out the principles and rules of the evaluation process. These, until now, serve as the main reference for evaluation of rural development programmes in all EU member states and EU accessing countries. The core element of the EC evaluation framework are **Common Evaluation Questions** (CEQ) (pre-defined by the EC) and programme-specific questions (to be defined by national programme authorities), both to be answered by external programme evaluators. Answering the EC common evaluation questions requires the use of the "intervention logic" concept pre-defined by the EC, i.e. differentiating between programme inputs, outputs, results, and

1 See: Evaluating EU activities – A practical guide for the Commission Services, DG Budget, July 2004.

2 Respective guidelines include: Evaluating EU activities – A practical guide for the Commission Services, DG Budget, July 2004; Evaluation of Rural Development Programmes 2000-2006 supported from the European Agricultural Guidance and Guarantee Fund – Guidelines; Guidelines for the Evaluation of Rural Development Programmes supported by SAPARD; Guidelines for the Mid-Term Evaluation of Rural Development Programmes funded by SAPARD; Handbook on Common Monitoring and Evaluation Framework for the programming period 2007-2013, Guidance document, September 2006.

impacts (by moving from a micro-level to the regional- or country levels)[3].

Standard evaluation questions focus for example on a direct effect of the RD programme on specific result indicators (e.g. farms income or employment) which requires a disentangling of programme effects from effects of other exogenously determined (programme independent) intervening factors. Furthermore, CEQ ask evaluators to quantify other programme effects. These include i) **deadweight loss** effects (i.e. to quantify changes observed in the situation of programme beneficiaries that would have occurred even without the programme); ii) **leverage** effects (i.e. the propensity of public intervention to induce private spending among direct beneficiaries); iii) **substitution** effects (i.e. effects obtained in favour of direct beneficiaries but at the expense of a person or organisation that does not qualify for the intervention (the latter are usually located in close neighbourhood of programme beneficiaries) (e.g. drop in profits of non-supported); and iv) **displacement** effect (i.e. effect obtained in an eligible area at the expense of another geographical area, e.g. shift of employment).

Although EC guidelines have been used as a main reference in all formal studies concerned with the mid-term and ex-post evaluation of EU RD programmes (programming period 2000-2006), some of the suggested methodologies appear as insufficiently rigorous to enable a correct answer to the CEQ. For example, the above guidelines, although fairly extensive and quantitatively oriented, allowed the usage of the so-called "naïve" evaluation techniques (e.g. before-after comparisons). As a consequence, in the huge majority of studies concerned with the quantitative assessment of socio-economic impacts of RD programmes in EU countries (programming period 2000-2006) "naïve" approaches were employed as a basic evaluation methodology[4]. While in some evaluation studies the authors attempted to build on counterfactuals, in most cases comparisons between supported and non-supported units were carried out without any consideration for *appropriate* matching. Usually, comparison groups were selected arbitrarily, leading to quantitative results that were statistically biased (i.e. selection bias). Moreover, in the majority of qualitative evaluations, knowledge about a specific programme's indirect effects, e.g. substitution, displacement, multiplier, etc. was "imputed" on the basis of anecdotal evidence or ad hoc surveys of a group of beneficiaries, opinions of administrative officials, etc. (CEAS, 2003; PCM, 2007; EENRD, 2010). As we show below, these techniques are in general unsuitable to address appropriately a number of issues generally considered crucial in any quantitative evaluation framework, i.e. the formulation of an unbiased baseline (construction of relevant control groups for estimation of counterfactual outcomes) or the estimation of the programme's general equilibrium effects (e.g. displacement or substitution effects).

Until recently, the major criticism of the existing EU common evaluation system and common indicators concerned: i) the relevance and appropriateness of particular indicators suggested by the EC; ii) the lack of a coherent evaluation framework linking inputs,

3 For assessment of programme results and impacts the evaluation guidelines recommend using various economic indicators (e.g. production, income, employment, etc.) whereby assessment of programme effects is to be carried out both at micro, regional and country levels. For example, in the case of the RD measure "Investments in agricultural holdings", methodological guidelines required, inter alia, answering specific measure-specific questions: A.I.1. To what extent have the supported investments contributed to the incomes improvement of beneficiary farmers? A.I.2. To what extent have the supported investments contributed to a better use of production factors on holdings?, etc. Programme evaluators are expected to provide empirical evidence that "due to participation in RD programme (…), e.g. gross value added (agriculture/non-agriculture) or employment (or gross number of jobs created) in supported enterprises increased by x%".

4 In approximately 75% of Mid-Term Evaluation (MTE) studies submitted to European Commission by the end of 2010 the impacts of EU RD programmes were assessed without any reference to a counterfactual situation (see: EC, European Commission, 2011)

outputs and outcomes; iii) gaps in data in the programmes' monitoring systems, and iv) the lack of prioritization between many indicators (e.g. Forstner and Plankl, 2004; CEAS 2003). While some of these problems have been addressed by the EC in the evaluation guidelines (EC, CMEF 2006 prepared for the programming period 2007-2013), in our view, the acceptance and overwhelming reliance on "naïve" evaluation techniques, which in extreme situations can bring about a considerable evaluation bias in the assessment of the real programme effects, remains especially problematic.

Clearly, far-reaching effects of inappropriate evaluation methodology could be the following:

- A lack of appropriate knowledge about the real impacts of the programme may result in the carrying out of policy interventions which, due to their low effectiveness/ efficiency, should have been discontinued or substantially re-designed.

- Poorly designed programmes may lead to an inefficient allocation of public and private resources, at the same time putting in jeopardy the achievement of policy objectives (e.g. poorly designed programmes may stimulate sectoral inefficiency, lead to a deterioration in competitiveness, and bring about progressing regional divergence). Lack of knowledge about the real programme impacts can reinforce these negative developments.

- Insufficient learning about programme effects can call into question not only the credibility of programme evaluations but also that of all institutions involved (conclusions of evaluation reports that used inappropriate and/or biased methods may be used selectively to support the interests of affected groups or may be contested if the evaluation does not conclude in favour of some interest groups).

Below we address the main methodological weaknesses of existing EU evaluation guidelines and suggest practical solutions enabling the provision of correct answers to EU Common Evaluation Questions. The analytical approach applied in this study draws on evaluation methodologies developed in: Rosenbaum and Rubin, 1983, 1985; Heckman, La Londe and Smith (1999); Heckman, et al (1998); Todd, P. (2008), and others. Recently developed advanced evaluation methodologies were successfully applied in a number of studies that focused on the measurement of effects of various structural, social and rural programmes in a number of countries, e.g. Dehejia and Wahba, 2002 (US); Newman et. al. 2002 (Bolivia); Venetokis, 2004 (Finland); Jalan and Ravallion, 2001 (Argentina); Lechner, 2002 (Switzerland); Larson, 2000 (Sweden); Pradhan and Rawlings, 2002 (Nicaragua), as well as in the studies focused on evaluations of social funds projects and other programmes aimed at eliminating poverty (Rawlings and Schady, 2002; Walle and Cratty, 2002; Bourguignon and Pereira da Silva, 2003; Ravallion, 2004). Yet, until recently their application to the evaluations of EU RD programmes was only sporadic (Schmitt et al. 2004; Pufahl and Weiss, 2007, Henning and Michalek, 2008).

2. Methodological Approach

The main principle of the analytical approach chosen to evaluate EU RD programmes is to infer about the economic return to resources employed in a RD programme by comparing this return to its opportunity costs and answering the question: what would have been earned in the next best alternative use[5].

2.1. Potential outcome model

A standard **potential outcome model** formalizes the problem of the inference about the impact of the participation in the given programme on the outcome of an individual unit (Roy, 1951; Rubin, 1974; Holland, 1986). The model adjusted to evaluations of RD programmes assumes that each unit/farm/region **i** potentially exposable to the RD programme/measure also fulfils all relevant programme participation criteria (e.g. programme general and specific eligibility criteria defined in a country's main programming document, i.e. Rural Development Plan). Observable variable **D** (a binary variable **0-1**) indicates whether an individual unit-**i** participated or did not participate in the RD programme. Furthermore, the simplified model assumes existence of a set of variables **X** representing pre-exposure attributes (covariates) for each individual unit **i**, of which some can be observable (**x**), and some other are not observable (**e**) as well as a set of variables **Y** which depend on **D**, representing the **potential response** of unit **i** to the RD programme $\mathbf{Y}_i\ (\mathbf{D}_i)$.

Obviously, **Y** may consist of outcome variables (e.g. **result indicators)** reflecting the effect of the programme at a micro-level: e.g. income, profits, employment, labour productivity, total factor productivity, etc.

In the case of EU RD programmes **Y** represents two variables standing for potential responses: $\mathbf{Y}_i\mathbf{(1)}$ in case of participation in the RD programme, and $\mathbf{Y}_i\mathbf{(0)}$ in case of non-participation in the same RD programme.

Using the potential outcome model, the effect of participation in an EU RD programme for an individual unit **i** (e.g. farm/region) can be written as:

$$\tau_i = Y_i\ (1) - Y_i\ (0) \qquad (1)$$

Where:

$\mathbf{Y}_i\ \mathbf{(1)}$ = potential outcome for unit **i** in case of participation in RD programme

$\mathbf{Y}_i\ \mathbf{(0)}$ = potential outcome for unit **i** in case of non participation in RD programme

τ_i = the effect of programme participation on unit **i,** relative to effect of non-participation on the basis of a response variable **Y**.

5 See Holland, 1986; Essama-Nssah, 2006,

While τ_i measures the effect of programme participation for **i**, only one of the potential outcomes, i.e. either $\mathbf{Y_i}$ **(1)** or $\mathbf{Y_i}$ **(0)** can be empirically observed at any given time for each individual unit **i**[6]. In other words, in standard (i.e. non-experimental) evaluation studies it is **impossible** to observe the value of the response variable (**Y**) for the same unit **i** under two mutually exclusive states of nature, i.e. participation in programme and non-participation (at the same time) (*The Fundamental Problem of Causal Inference (FPCI),* Holland, 1986).

While the FPCI makes *observing* causal effects impossible, this does **not** mean however that causal inference is impossible (see: Rubin, 1974; 1975). In fact, determining **unobservable** outcome in (eq.1) called *counterfactual outcome is possible and generally considered the core of each evaluation design*[7] (e.g. World Bank, 2002; Asian Development Bank, 2006).

The potential outcome model allows also for a more explicit consideration of time. In this case, for each programme eligible unit **i** there are two potential outcomes $(\mathbf{Y_{0it}, Y_{1it}})$ corresponding respectively to the non-participation **(0)** and participation **(1)** in an RD programme at a given time **t**. Given that, $\mathbf{D_i}$ =1 represents unit-**i's** participation in the RD programme, and $\mathbf{D_i}$ = 0 non-participation, the time-specific **potential outcome** on unit **i** can be described as:

$$Y_{it} = D_i Y_{1it} + (1-D_i) Y_{0it} \qquad (2)$$

The **potential outcome** equation in case of programme participation can also be expressed as:

$$Y_{1it} = \mu_1 (X_{it}) + U_{1it} \quad (3)$$

and the potential outcome in case of non-participation in RD programme as in (4):

$$Y_{0it} = \mu_0 (X_{it}) + U_{0it} \quad (4)$$

Where:

X_{it} is a vector of observed random variables **not affected** by treatment (programme participation), and

(U_{1it}, U_{0it}) are unobserved random variables which are distributed independently across units i's and satisfy conditions: $E(U_{1it})=0$ and $E(U_{0it}) = 0$

Given 3 and 4 and assuming that treatment (i.e. programme support) takes place in period **k** (t > k) the individual specific treatment effect, for any vector of covariates $\mathbf{X_i}$ can be described as (Blundell and Costa Dias, 2002):

$$\alpha_{it} (X_{it}) = [\mu_1 (X_{it}) - \mu_0 (X_{it})] + [U_{1it} - U_{0it}] \qquad (5)$$

where: $t > k$ and μ_0 and μ_1 are defined as in eq 3 and 4.

Typically we cannot expect that all **i**-units will be affected by the given RD programme in exactly the same way. Depending on an assumed individual programme response of each **i**-unit the explicit modelling and aggregation of programme effects can be carried out at various complexity levels.

6 Generally speaking, there are two major methods to determine the counterfactuals, i.e. experimental design and quasi-experimental design. In the experimental design that is generally viewed as the most robust evaluation approach (Burtless, 1995; Bryson, et. al. 2002) one would have to create a control group of units which are randomly denied access to a programme. In this random assignment a control group would comprise of firms/units/individuals with identical distribution of observable and unobservable characteristics to those in the supported group. In such an experiment the selection problem would be overcome because participation is randomly determined (Bryson, et. al, 2002). Yet, there is a vast amount of literature showing that social experiments (except of in sociology, psychology, etc.) are often too expensive and may require the unethical coercion of subjects unwilling to follow the experimental protocol (Winship and Morgan, 1999). As experimental designs (randomization) in the case of evaluation of RD programmes would be extremely cumbersome (for ethical and political reasons) a non-random method (quasi-experimental) will be used in this study. The basic idea behind quasi-experimental methods is that they generate comparison groups that are akin to the group of programme participants by using techniques described above.

7 Under this specification (eq.2) is equivalent to a switching regression model of Quandt (1972) or the Roy model of income distribution (Roy, 1951; Heckman and Honore, 1990) quoted in Aakvik, et al., 2000; Heckman and Vytlacil, 2005).

2.2. Homogenous Treatment Effects

Following Blundell and Costa Dias (2002) and Caliendo and Hujer (2005) a homogeneous treatment effect is the simplest case where the programme effect is assumed to be **constant** across individuals/units. Under the assumption that treatment takes place in a period **k,** the homogeneous (for all units i) treatment effect is defined as (6):

$$\alpha t = \alpha_{it} (X_{it}) = [\mu_1 (X_{it}) - \mu_0 (X_{it})] \text{ where: } t > k \quad (6)$$

where α_t is constant for any unit/individual **i**.

For the case of a homogeneous treatment effect μ_1 and μ_0 are two parallel curves only differing in level. Assuming homogenous treatment effects, the modelling of the aggregated programme impact can be carried out by means of an outcome equation (7) in which the participation specific error terms are not affected by the treatment status.

The corresponding outcome equation can be expressed as (Blundell and Costa Dias, 2002):

$$Y_{it} = \mu_0 (X_{it}) + \alpha_t D_{it} + U_i \qquad (7)$$

2.3. Heterogeneous Treatment Effects

In case of heterogeneous treatment effects it is assumed that treatment impact varies across individuals/units (a possible effect of an observable component or as a part of the unobservables).

In this case the outcome equation differs from eq 7 and can be rewritten (Blundell and Costa Dias, 2002; Caliendo and Hujer, 2005) as:

$$Y_{it} = D_i Y_{1it} + (1-D_i) Y_{0it} = \mu_0 (X_{it}) + \alpha_t (X_{it}) D_{it} + [U_{i0} + D_{it} (U_{1it} - U_{0it})] \qquad (8)$$

It is important to notice that the form of the **error term** differs across observations according to their treatment status. Contrary to the homogenous treatment effect this structure does **not** allow extrapolation to all population strata of units-i (e.g. to areas of the support of **X** that are not represented at least among the treated). Furthermore, if there is selection on unobservables, the OLS estimator after controlling for covariates **X** is inconsistent for α_t (X) (Blundell and Costa Dias, 2002).

As performance of farms supported by a RD programme cannot be *directly observed* in a "non-support" situation (a farm cannot simultaneously participate and not participate in the same programme) the economic performance of farms supported by the RD programme in a "non-support" situation (base-line) has to be simulated, using more advanced techniques.

Construction of an appropriate base-line should provide us with an answer to the question: "what would have been a given outcome for a farm supported by the RD programme if the programme had not been implemented?" By comparing performance outcomes of supported farms with a control group of farms in two data points; i.e. prior to support and after its conclusion, we can straightforwardly answer two questions: 1). What was the effect of exogenously determined factors[8] on the performance of farms supported by the programme?, and 2). What was the effect of the programme support?

8 All factors which influence performance of supported and non-supported regions and are not considered as RD programme related can be called exogenous.

3. Estimators commonly applied in evaluations of EU RD programmes

In the standard EU evaluation practice, where experimental studies of a random assignment to the group of programme participants and non-participants are not possible, evaluators of RD programmes usually apply four alternative naïve techniques to estimate the impact of the programme:

3.1. Naive "before-after" estimator for programme participants

Naive before-after estimator uses pre-programme data on **programme beneficiaries** to compute (counterfactual!) programme outcomes for programme participants defined in eq (1). Naive before-after estimator is defined in eq 9.

τ_i (naive "before-after") = $E_N\ [Y_{it=1}|D_i=1] - E_N\ [Y_{it=0}|D_i=1]$ (9)

where :

N is a sample size in observed survey of programme participants (i)

$E_N\ [Y_{it=1}|D_i=1]$ is the sample mean of the outcome for those observed as programme participants (i) **after** participation in programme (T=1)

$E_N\ [Y_{it=0}|D_i=1]$ is the sample mean of the outcome for those observed as programme participants (i) **before** their participation in the programme (T=0)

The problem with this approach is that information about $E_N\ [Y_{it=1}|D_i=1]$ and $E_N\ [Y_{it=0}|D_i=1]$ (usually obtained from quasi-scientific interviews carried out on sampled programme participants[9]) and related difference in the outcome indicators (e.g. profits, employment, etc) in time $T_0 \Rightarrow T_1$ is **arbitrarily** attributed to the effect of the RD programme.

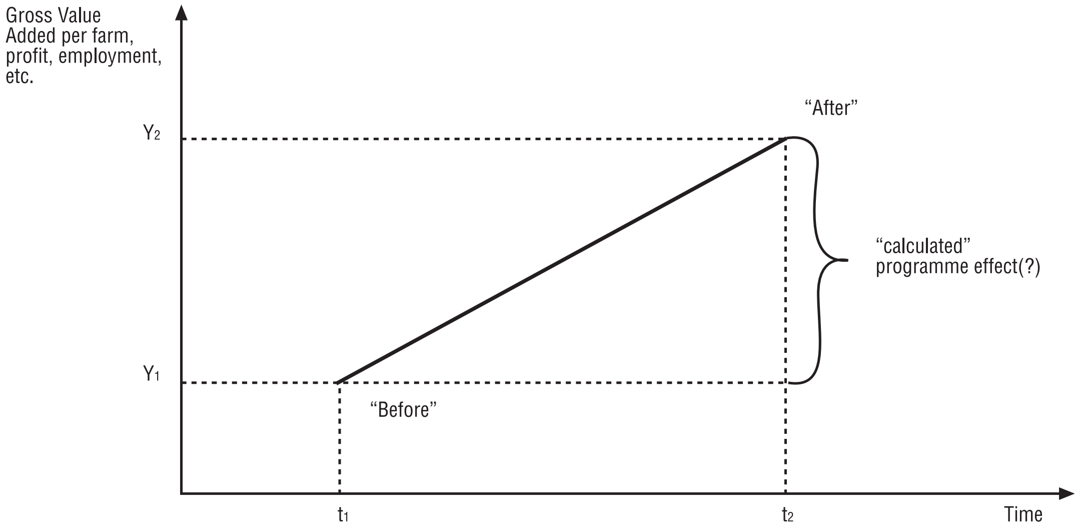
The implicit and rather unjustifiable assumptions of this evaluation technique are:

- In the absence of policy intervention (RD programme) the outcome indicator of programme participants would have been the same as before the programme.
- Changes in outcomes of programme participants are not affected by any other factor (e.g. macroeconomic, regional etc.) but are the effect of the RD programme only.

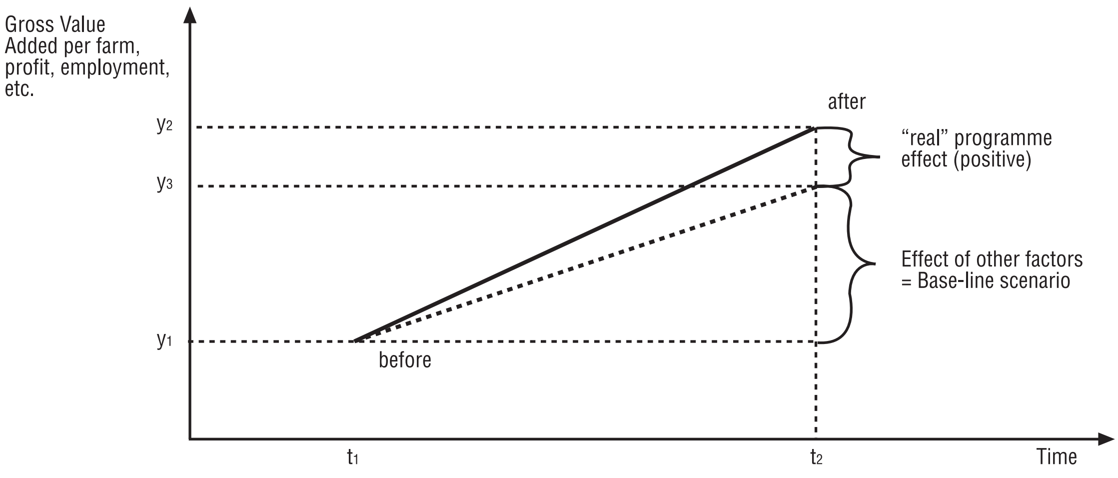
Although it is obvious that over years specific outcome indicators, e.g. gross income or profits do not remain unchanged, some evaluators assign the whole effect of observable change in an outcome indicator to the programme. By doing so the real impact of a given programme may be massively overstated (Graphs 1a -c).

9 In a huge majority of cases due to lack of data in monitoring systems phone interviews or the CATI (computed-assisted telephone interview) method (self-assessment) were used (PCM, 2007; CEAS, 2003).

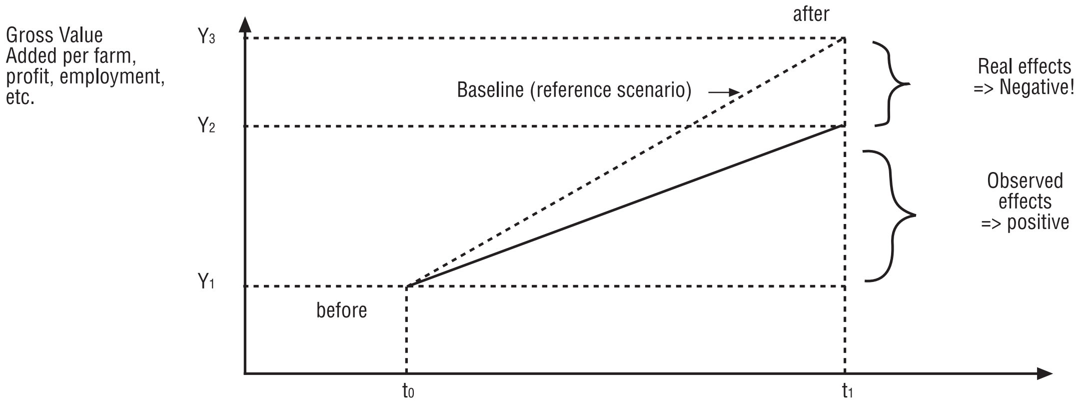
Graph 1a. Naive before-after estimation



Graph 1b. Significance of the relevant base-line ("no programme" scenario) for the same (!) farm/ enterprise (small positive real programme effect)



Graph 1c. Significance of the relevant base-line ("no programme" scenario) for the same (!) farm/ enterprise – negative real programme effect



3.2. Naive "participants vs. non-participants" estimator

Another technique commonly characterized as a naïve evaluation approach uses all non-participants as a control group.

τ_i (naïve "participants vs. non-participants") $= E_N\,[Y_{it=1}|D_i=1] - E_M\,[Y_{jt=1}|D_i=0]$ (10a)

where :

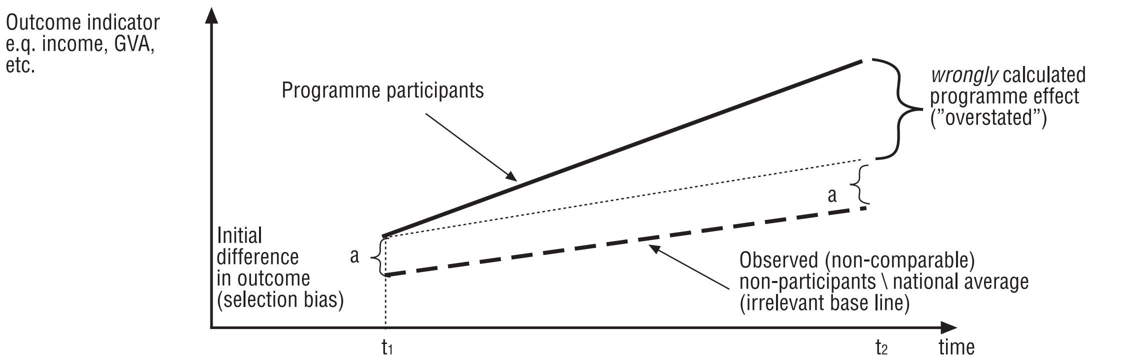
- N is a sample size in observed survey of programme participants (i);
- M is a sample size in observed survey of programme non-participants (j);
- EN $[Y_{it=1}|D_i=1]$ is the sample mean of the outcome for those observed as programme participants (i) after participation in programme (T=1);
- EM $[Y_{jt=1}|D_i=0]$ is the sample mean of the outcome for observed programme non-participants (j) in time T=1;

While monitoring systems of RD programmes usually do not contain any information on appropriate control groups of non-participants, the data about $E_M\,[Y_{jt=1}|D_i=0]$ is obtained on the basis of (rather ad-hoc) surveys carried out by programme evaluators on selected outcome indicators (e.g. profits, employment, etc.) for those who did not participate in RD programme (irrespectively on the level of similarity between these two groups) without any considerations regarding comparability between both groups (and eventual accounting for systematic differences). The approach relies on the assumption that in the absence of the programme the outcome indicator of programme participants would be the same as for programme non-participants. Yet, this would only be justifiable if the **systematic performance** of programme participants (measured by any arbitrary outcome indicator, e.g. income, profit or employment) was identical with the outcome performance of programme non-participants. Had this not been the case, the selection bias B(X) that results from using the outcomes of non-participants as proxy for the outcomes that programme participants would have experienced had they not participated can be very substantial and is equal to (Heckman, Ichimura, Smith and Todd, 1996):

$$B(X) = E\,(Y_0|X, D=1) - E\,(Y_0|\,X, D=0). \quad (10b)$$

Obviously programme effects shown in Graph 2 are **overstated** due to incorrectly calculated "base-line" (systematic performance of non-participants included in a control group differs from systematic performance of participants, even in the absence of a given programme).

Graph 2. Observable heterogeneity, e.g. participants are "better performing" than non-participants/or national average



3.3. Naive “participants vs. aggregated (i.e. participants and non-participants) sample average” estimator

Another naive estimator commonly applied in empirical evaluation studies of RD programmes uses a control group constructed as a population average (i.e. consisting of programme participants and non-participants).

τ_i (naive “participants vs. overall sample average”) = $E_N\ [Y_{it=1}|D_i=1] - E_{NM}\ [Y_{ijt=1}|D_i=0]$ (11)

where :

- N is a sample size in observed survey of programme participants (i);
- NM is a joint sample size in observed survey of programme participants (i) and non-participants (j);
- $E_N\ [Y_{it=1}|D_i=1]$ is the sample mean of the outcome for those observed as programme participants (i) after participation in programme (T=1);
- $E_{NM}\ [Y_{ijt=1}|D_i=0]$ is the sample mean of the outcome for observed joint sample of programme participants (i) and non-participants (j) in time (T=1);

In this evaluation “technique” necessary data on average outcome indicators in the group of “non-participants” is usually obtained from various national surveys. The approach relies on the similar assumption as in case of (3.2) that in the absence of the programme the outcome indicator of programme participants would be the same as the average of a joint group of programme participants and non-participants. This however would only be justifiable if systematic performance of the group of programme participants (measured by any arbitrary outcome indicator, e.g. income, profit or employment) was identical with the performance of the joint-group of programme participants and non-participants.

As shown in Lechner (2001), an effect based on comparisons of a treatment group to an aggregated comparison group of individuals has no meaningful casual interpretation and can lead to fairly misleading results.

3.4. Conventional “difference in differences” (DID) estimator (without appropriate matching between programme participants and the control group)

A conventional DID estimator can be expressed as in (12)

DID = $(E_N\ [Y_{it=1}|D_i=1] - EN\ [Y_{it=0}|D_i=1]) - (E_M\ [Y_{it=1}|D_i=0] - E_M\ [Y_{jt=0}|D_i=0])$ (12)

This estimator compares the before-after changes of programme participants (i) with the before-and-after changes of outcome indicators for *arbitrary selected* non-participants (j), whereby the estimation of the effect of the RD programme is usually obtained on the basis of panel data models involving group of programme participants and an arbitrary group of programme non-participants. The DID estimator is already more advanced compared with techniques described above as it additionally assumes that selection to a programme depends on both observables as well as **unobservables.** Although in this method any common trend in the outcomes of programme participants and non-participants (fixed selection bias) gets differenced out, the crucial assumption justifying this method is **that selection bias remains time invariant (so called fixed-effect).**

Although conventional DID, due to the lack of appropriate data has so far not been very popular in the evaluation of EU RD programmes, it can be easily shown that this estimator is problematic if *in the absence of policy intervention* the differences between performance of programme participants (i) and non-participants (j) do not remain constant over time. In this situation DID estimator will produce **biased estimates** of programme effects. Generally speaking, the available evidence suggests that conventional DID estimators, though supported by plausible stories about “fixed” differences in motivation, ability or performance, may be a poor choice in many evaluation contexts (Smith, 2000).

4. Basic methodological problems faced by RD programme evaluators

Given the above techniques and their methodological weaknesses it becomes obvious that basic problems faced by evaluators of RD programmes concern:

- *Elimination of a selection bias:* A selection bias in evaluating the impact of an RD programme occurs if the mean outcome of those units which participated in the RD programme differs from the mean outcome of non-supported units even in the absence of support. An important problem which usually arises while simply comparing average data for programme participants and non-participants is that many RD programmes/measures are not assigned randomly but: i) are designed to target specific beneficiaries with a certain performance characteristic (e.g. under performed producers/enterprises/areas, etc.), or ii) include various eligibility conditions (e.g. reimbursement of project costs *after* finalization of the project) which, in practice, can only be fulfilled by certain types of economic units, e.g. the best enterprises. In both cases, a supported group may easily outperform/under-perform specific control groups or national averages, making simple comparisons of both groups' performance statistically biased and unacceptable. Another type of distortion can be the so called "self-selection" bias[10]. To assess the programme's impact, one has to infer the counterfactual on what would have been in the absence of the programme (this calls for data on programme non-participants). But, even with good data on observable characteristics both for supported and non-supported units, a reliable comparison between those two groups is not easy. Ideally, control enterprises/producers should differ from the supported group only in so far as they do not receive any intervention. To be meaningful, a control group should therefore include only those enterprises which *match* in their observable characteristics with supported enterprises (*prior* to the programme). Moreover, the "similarity" of both groups should be statistically tested and all "undesired" differences explicitly accounted for by applying modern evaluation methodologies.

- *Disentangling an effect of the programme from other effects:* An assessment of a programme's impact requires a response to the question: What would have happened to supported enterprises *without* an RD programme? Clearly, a counterfactual performance of supported enterprises cannot be directly observed. For the same reason, in non-experimental studies a programme's impact (causal effects) should be assessed by making comparisons between supported enterprises with possibly identical ones which did not benefit from the programme.

A review of available mid-term and ex-post evaluation reports of EU RD programmes shows that answers to EC Common Evaluation Questions (CEQ) have mostly been provided by applying *qualitative* methods, i.e. interviews and surveys (sometimes complemented with ad hoc quantitative indicators).

10 Self-selection bias may appear if enterprises that anticipated participation in the RD programme already adjusted its own performance prior to the start of the programme, e.g. in order to comply with programme eligibility criteria. In such situation, even if the group of programme participants was very similar to a control group, making comparisons of both groups just "before" and "after" participation in the programme could lead a significant control bias. The important consequence for evaluation is that this type of bias should be eliminated first before the programme impact assessment is undertaken

In the majority of cases, "quantitative" effects of the programme have been assessed on the basis of interviews with programme beneficiaries, *without* formulation of a necessary *base-line*, i.e. without construction of an appropriate counterfactual situation[11]. In a few cases where comparisons between programme beneficiaries and non-beneficiaries were carried out this was done *without* any consideration for *appropriate* matching. Furthermore, quantitative knowledge about specific programme effects (i.e. substitution, displacement, etc.) was in most cases "imputed" on the basis of anecdotal evidence or ad hoc surveys of a group of beneficiaries, opinions of administrative officials, etc.[12]

The rigorous assessment of the impact of a policy intervention in the framework of rural development programmes proved to be difficult, in particular, because mainly crude evaluation techniques were used. Evidence shows that due to the application of inappropriate methodology and the lack of data, in the huge majority of cases important CEQs concerned with the evaluation of EU RD programmes were only partly answered or were not answered at all by evaluators (FAL, 2006).

Inappropriate methodology and problems with data resulted in the meagre quality of many evaluation reports. As Toulemonde, et al. stated: the "strength (of evaluation reports) has to be nuanced because often conclusions on impacts were purely descriptive and failed to provide a cause-and-effects analysis. This is why criterion – sound analysis was one of the most poorly rated"[13]. This development was also confirmed in other studies[14].

11 For example, assessment of the effect of an investment support under RD programmes (e.g. investments in agricultural holdings, renewal of villages, etc.) was in most of cases carried out by: i) Interviewing selected program beneficiaries on the impact of support, whereby the positive "impact" of supported investment on value added, competitiveness, etc. was "measured" and "evaluated" by referring to the number (%) of affirmative vs. negative responses obtained from interviewed beneficiaries (CEAS, 2003; Forstner and Plankl, 2004), ii) Deriving some quasi-quantitative information on the basis of interviews conducted among supported units (Tissen and Schrader, 1998); iii) Comparing various outcome indicators characterising supported enterprises (!) at the beginning of support and after it (e.g. before and after investment situation) (e.g. RDP 2004-2006 Slovakia; Mid-term evaluation of SAPARD in Slovakia, 2003); iv) Comparing some average outcome indicators between units which were supported by the program with those which were not. Yet, failure to control for differences in the pre-intervention characteristics of program participants and non-participants severely biased such comparisons.

12 See CEAS, 2003; Forstner and Plankl, 2004

13 See: Toulemonde et al. , 2002

14 For example, Forstner and Plankl, 2004 wrote: "Evaluations, as they were performed thus far, mainly use pragmatic approaches to keep up with given timetables… The trend is that profound scientific analysis is loosing ground in evaluation studies which are obligatory. Due to the public budget constraints this trend is continuing or even gaining momentum".

5. Advanced approach to the evaluation of RD programmes

5.1. Relevant policy indicators

Depending on concrete policy interest the EU common evaluation questions (CEQ) can be systematically answered by focusing on the impact of a given RD programme on various types of individuals/farms (groups) directly or indirectly affected by the RD programme. Answers to the CEQ (addressing a particular group of "gainers") may be provided using relevant policy indicators measuring the impact of the programme:

5.1.1. Average Treatment Effect (ATE)

The first indicator which can be applied to evaluate RD programmes is the (population) **average treatment effect (ATE)**. This indicator is simply the difference between the expected outcomes after participation in the RD programme and non-participation conditional on **X** (Heckman, 1996; Imbens, 2003; Imbens and Wooldridge, 2007).

$$\Delta^{ATE}(x) = E(\Delta | X = x), \text{ where: } \Delta = Y_1 - Y_0 \qquad (13)$$

ATE is the effect of assigning participation *randomly* to every unit **i** of type **X** (ignoring programme general equilibrium effects) and describes an expected gain from participating in the RD programme for a **randomly selected farm/individual from the joined sub-groups of programme participants and non-participants** in a given programme area. This policy indicator averages the effect of the programme over all units in the population, including both programme participants and non-participants.

Depending on the data set used for the calculation of this indicator, the **sample** average treatment effect **(SATE)** can be estimated by taking an average value of all (Y_{1it} - Y_{0it}) in a given sample k used for an analysis (Imbens and Wooldridge, 2007), i.e.

$$\mathbf{SATE = 1/k\ \Sigma}\ (Y_{1it} - Y_{0it}) \text{ for } i = 1 \ldots k \qquad (13a)$$

On the other hand, if the focus of policies is the estimation of average treatment effects for the **population at large** one can estimate the population average treatment effect **(PATE)** defined as:

$$\mathbf{PATE} = E(Y_1 - Y_0) \qquad (13b)$$

Whereby, the estimation of PATE requires some knowledge about distribution probability of individual units.

Although SATE is the best estimator for PATE one cannot estimate PATE without error because the potential outcomes for those population members *not included* in the sample are missing. According to Imbens and Wooldridge (2007) the implications of using SATE/PATE are as follows: i) one can estimate SATE at least as accurately as the PATE and typically more accurately, and ii) a good estimator for PATE is automatically a good estimator for SATE, and iii) as a given sample may not be representative for the population at large some caution is required if results of SATE are to be generalized.

Like every specific policy indicator, ATE also has some disadvantages. The first concerns the addressing of important policy aspects, i.e. clear targeting of intervention. Irrespective of whether the policy analysts use SATE or PATE to evaluate programme results, specific problems arise due to the fact that ATE **includes** the effect on units/farms/individuals for which the programme **was never intended/designed** (it may include the impact on units that may even be programme ineligible).

In non-experimental studies, provision of an empirical answer to the standard evaluation question always involves comparisons of programme participants with non-participants. Yet, a typical question which arises is: **which** units should be compared, i.e. which units best represent programme participants had they not participated?

While some evaluators try to estimate ATE using differences in means of $E(Y_1|D=1)$ and $E(Y_0|D=0)$ it can be shown that the **bias** resulting from this approach is equal to (eq 13c) (Heckmann and Lozano, 2003):

$$B(ATE) = E(Y_1|\ X, D=1) - E\ (Y_0|X, D=0) - [E\ (Y_1-Y_0)|X] \qquad (13c)$$

5.1.2. Average Treatment of Treated (ATT)

Given the deficiencies of ATE another evaluation indicator can be used, describing the average impact of programme participation on units/farms/individuals that participated in the programme, the so called: **the average treatment on the treated (ATT)** (see eq. 14):

$$\Delta^{ATT}(x) = E(\Delta|X=x, D=1) \qquad (14)$$

which is equivalent to:

$$E(Y_1-Y_0|D=1) = E\ (Y_1|\ D=1)–E\ (Y_0|D=1) \qquad (14a)$$

Where: $E\ (Y_0|D=1)$ is not directly observable (it describes the hypothetical outcome without a programme's support of those who participated in the programme)

In contrast to ATE, interpretation of ATT is much more policy relevant. While ATT focuses on the effect of the programme on programme participants, it also describes the gross gain accruing to the economy from the existence of the programme compared with an alternative of shutting it down (Heckman and Robb, 1985; Heckman, 1997; Smith, 2000; Smith and Todd, 2003). Combined with information on programme costs and general equilibrium effects the ATT indicator can therefore answer the policy question regarding the **net gain** to the economy[15].

Although ATT is generally applicable to provide answers to RD Common Evaluation Questions concerning the effect of the RD programme on units that participated in the programme, the **empirical estimation** of ATT is not straightforward. To illustrate the problem we consider both components of ATT (i.e. $E\ (Y_1|\ D=1)$ and $E\ (Y_0|D=1)$. It is obvious that $E\ (Y_1|\ D=1)$ can be easily identified from data on programme participants. In practical evaluations, the term $E\ (Y_1|\ D=1)$ describes specific outcomes (e.g. in form of result indicators), e.g. profits, employment, labour productivity or total productivity, etc. observable among programme beneficiaries after implementation of the given RD programme. On the other hand, the expected value of $(Y_0|D=1)$, i.e. the **counterfactual** mean in outcome (potential outcome in case of non-participation) of those who participated in the programme **cannot be directly observed**.

Given the above, one has to choose a proper *substitute* for unobservable $E(Y_0|\ D = 1)$ in order to estimate **ATT**.

So far, and only if the condition (14b) holds, one could use the non-participants directly as an adequate control group.

$$E\ (Y_0|\ D=1) = E\ (Y_0|D=0) \qquad (14b)$$

Yet, this condition is likely to hold only in randomized experiments (Caliendo and Hujer, 2005). In most of non-experimental studies estimation of ATT using the differences in outcome means of programme participants and non-participants **results in a selection bias (B)** defined as in eq (14c).

15 Depending on research interest it may be distinguished between the average treatment effect on treated sample (SATT) and the average treatment effect on treated for a population at large (PATT) in a manner similar to the one explained above for SATE and PATE (Imbens and Wooldridge, 2007).

$$B(X) = E(Y_0 | D=1) - E(Y_0 | D=0) \qquad (14c)$$

The selection bias arises because the means of $\mathbf{Y_0}$ for programme participants (D=1) and Y_0 for non-participants (D=0) may **differ systematically**, even in the absence of the programme.

ATT can also be defined conditional on **P(Z):**

$$\Delta^{ATT}(x) = E(\Delta | X=x, P(Z)=p, D=1) \qquad (14d)$$

Where: **P** is a probability distribution of observed covariances **Z**

As (14) and (14d) are equivalent, the latter formulation will be used in our study for the calculation of effects of an RD programme. Various methods which aim at the elimination of selection bias and estimation of it are described in Section 4.

5.1.3. Marginal Treatment Effect (MTE)

In some cases, policy makers are interested in indicators that show the impact of expansion of the programme to a marginal unit that is indifferent between participation and non-participation. The Marginal Treatment Effect **(MTE)** indicator (see: Björklund and Moffitt, 1987; Heckman, 1997; Heckman and Vytlacil 1999, 2000) is defined as follows:

$$\Delta^{MTE}(x, u) = E(\Delta | X = x, U_D = u) \qquad (15)$$

Δ^{MTE} (x, u) is the average effect of participation in the RD programme for those units **i** which are **on the margin of indifference** between participation in the programme (D=1) or non-participation (D = 0), where $\mathbf{u} = \mathbf{Z\beta_D}$. One can therefore interpret Δ^{MTE} (x, u) as the mean gain in terms of $\mathbf{Y_1 - Y_0}$ for units **i** with observed characteristics **X** which would be indifferent between participation in the RD programme and non-participation if they were exogenously assigned a value of **Z**, say **z**, such that $\mu_D(z) = u_D$ (Heckman, 2005).

For values of **u** close to **zero**, $\mathbf{\Delta^{MTE}}$ **(x, u)** is the average treatment effect for units **i** with **unobservable** characteristics that make them most likely to participate, and for values **u** close to **one** it is the average treatment effect for units **i** with **unobservable** characteristics that make them the least likely to participate.

Evaluation of the MTE parameter at low values of **u** averages the outcome gain for those with unobservables that make them least likely to participate. Evaluation of it at high values of **u** is the average gain for those individuals with unobservables that make them most likely to participate (Heckman, et. al, 2003).

While the estimate of $\mathbf{\Delta^{ATT}}$ provides an evaluator with some interesting information about a general impact of the RD programme, i.e. facilitates decision about abolition or retention of the RD measure, the $\mathbf{\Delta^{MTE}}$ is informative on the question of whether the units participating in the RD programme benefit from it in gross terms (net effects should also consider the costs of programme participation). The parameter $\mathbf{\Delta^{MTE}}$ estimates the gross gain from the marginal expansion of the programme.

The bias in MTE is the difference between average U_1 for programme participants and marginal U_1 minus the difference between average U0 for non-participants and marginal U0. Each of these terms is a bias which may be called **selection bias** (Heckmann and Navarro-Lozano, 2004).

5.1.4. Average Treatment on Non-Treated

Of considerable interest to evaluators of EU RD programmes can also be a measurement of the effect of a given RD programme on those who did **not** participate in it. The ATNT evaluation indicator is defined as

$$ATNT = E(Y_1 | D = 0) - E(Y_0 | D = 0). \qquad (16)$$

As E (Y_1|D=0) cannot be observed directly it must be calculated as counterfactual.

5.2. Construction of an appropriate baseline

Obviously, in the context of non-experimental studies the counterfactuals cannot be estimated directly, in a manner analogous to the one based on randomization. Given a possibility of a significant bias in results obtained from using crude evaluation techniques various other methods may be applied aiming at elimination/correction of this bias. The most prominent are: matching methods, the method of control functions and the method of instrumental variables (Heckman and Navarro-Lozano, 2004).

5.2.1. Matching methods

Matching methods seek to mimic conditions similar to experiments, in a way that the assessment of the impact of the RD programme can be based on comparison of outcomes for a group of programme participants (D=1) with those drawn from a comparison group of programme non-participants (Rosenbaum and Rubin, 1983; Smith and Todd, 2003; Heckman and Navarro-Lozano, 2004). In principle, matching can be viewed as a method of strategic sub-sampling from among programme participants and control cases whereby the selection of control cases for each programme participant is based on the **observable** characteristics of covariates $\mathbf{X_i}$.

Matching methods are based on the identifying assumption that conditional on some covariates **X**, the outcome **Y** is independent of **D**.

Application of matching to a consistent evaluation of programme effects makes the following two assumptions crucial (Rosenbaum and Rubin, 1983a)[16]:

Unconfoundedness assumption: $(Y_0, Y_1) \perp D \mid X$

Where: $\perp$ denotes independence

To yield consistent estimates of the programme impact, matching methods assume that conditional on observed covariates **X**, potential outcomes are independent of programme participation, or in other words, that (conditional on observed covariates **X**) the assignment (programme participation) probabilities do not depend on the potential outcomes. The unconfoundedness assumption is often controversial, as it assumes that beyond the observed covariates X there are no (unobserved) characteristics of the individual associated both with the potential outcomes and programme participation (Imbens and Wooldridge, 2007).

Overlap assumption: $0 < \Pr(D=1 \mid X) < 1$

The overlap assumption prevents **X** from being a perfect predictor in the sense that one can find for each programme participant a counterpart in the non-participant group and vice versa (Caliendo and Hujer, 2005). If there are regions where the support of **X** does not overlap for the participants and non-participants, matching has to be performed over the common support only (see below). A weaker version of overlap assumption implies the possible existence of a **non-participant similar** to **each participant**.

To avoid a lack of comparable units one can restrict matching and hence the estimation of the effect of programme participation to the region of common support, equivalent to the overlap condition. The latter not only rules out the phenomenon of perfect predictability of D given **X** but also ensures that units with the same **X** values have positive probability of being both participants and non-participants

16 Rosenbaum and Rubin (1983a) refer to the combination of the two assumptions (unconfoundedness and overlap) as "strongly ignorable treatment assignment"

(see: Caliendo and Kopeinig, 2005; Heckman, LaLonde and Smith, 1999). [17]

Matching assumes that there exists a set of **observable conditioning variables Z** (**which may be a subset** of **X**) for which the non-participation outcome **Y_0** is independent of participation status **D** conditional on **Z, or $Y_0 \perp D \mid Z$.** It is also assumes that **for all Z** there is a **positive probability** of either participating (D=1) or not participating (D=0) in a programme, which also implies that **a match can be found for all D=1 units** (Smith and Todd, 2003)[18]. Conditional on the observables **Z**, outcomes for the non-participants represent what the participants would have experienced had they not participated in the RD programme (under assumption that selection into the RD programme is based entirely on observable characteristics). For further explanation of consequences by choosing only a sub-set of conditional variables see: Chapter: 5.3 (below).

5.2.2. Application of the Propensity Score

Various empirical studies show that traditional matching may be difficult if the set of conditioning variables **Z** is large, due to the "curse of **dimensionality"** (problem of empty cells[19]) of the conditioning problem (Zhao, 2005; Todd, 2006; Black and Smith, 2004). As the number of observable characteristics in the group of programme participants increases linearly, the number of necessary observations in the control group increases exponentially. Moreover, matching on all the covariates using a distance measure, which effectively regards all interactions among the X covariates as equally important, does not work very well (Gu and Rosenbaum, 1993; Rubin and Thomas, 1996).

Rosenbaum and Rubin (1983) showed that the dimensionality of the conditioning problem can be dramatically reduced by implementing matching methods through the use of so-called **balancing scores b(Z)**, i.e. functions of the relevant observed covariates **Z** such that conditional distribution of **Z** given **b(Z)** is independent of the assignment into treatment. One possible balancing score is the **propensity score**, i.e. the probability of participating in a programme given observed characteristics **Z**.

For random variables **Y** and **Z** and for discrete variable **D,** Rosenbaum and Rubin (1983) defined the propensity score as the conditional probability of participating in a programme given pre-programme characteristics **Z**:

$$p(Z) \equiv \Pr(D=1|Z) = E(D|Z) \quad (17)$$

where **Z** is a multidimensional vector of pre-programme characteristics.

They showed that if the participation in a programme is random conditional on **Z**, it is also random conditional on **p(Z):**

$$E(D|Y, \Pr(D=1|Z)) = E(E(D|Y, Z)|Y, \Pr(D=1|Z)) \quad (18)$$

so that

$$E(D|Y,Z)=E(D|Z) \quad \text{implies} \quad E(D|Y, \Pr(D=1)|Z))=E(D|\Pr(D=1|Z)) \quad (19)$$

Where: Pr (D=1|Z) is a **propensity score**

The above equations imply that when **outcomes** are *independent of programme participation* **conditional on Z**, they are also

17 Following Heckman, Ichimura and Todd (1998), the importance of the overlap assumption can be illustrated on example of a situation where for some values of x we have either p(x) =0 or p(x)=1, i.e. in which one would find some units i with covariates implying that those units either always participate or never participate in the programme. If they always participated there would not have counterparts in the comparison group (non-participants). On the other hand, had they never participated, they would never have had counterparts in the group of programme participants.

18 It can be shown that assumption $Y_0 \perp D|Z$ is overly strong if parameter of interest is the mean impact of treatment on treated (TT) in which case conditional mean independence suffices: $E(Y_0|Z, D=1) = E(Y_0|Z, D=0) = E(Y0|Z)$, see: Smith and Todd, 2003.

19 For example with just 20 binary covariates there are 220 covariate patterns (1.04 mill possibilities).

independent of participation **conditional on the propensity score, Pr (D=1|Z)**. Thus, when matching on **Z** is valid, matching on the summary statistic **Pr(D = 1 | Z)** (the propensity score) is also valid (Todd, 2008). Conditional independence remains therefore valid if we use the propensity score **p(Z)** instead of covariates **Z** or **X**.

One of the important results of Rosenbaum and Rubin (1983) is the conclusion that there is nothing to be gained by matching (or stratifying) in a more refined way on the variables in **Z** than on the propensity score alone that is a function of the variables **Z. The propensity score** therefore contains all the information that is needed to create a balanced evaluation design (Winship and Morgan, 1999).

The major advantage of this result is that in empirical studies a conditional participation probability can be estimated using a parametric method, such as **probit** or **logit**, or **semi-parametrically** using a method that converges faster than the non-parametric rate. In such a situation the **dimensionality** of the matching problem can be **reduced** substantially by using a **one dimension** only, i.e. on the univariate propensity score.

An important feature of this method is that after the units are matched, the unmatched comparison units can easily be separated out and are not directly used in the estimation of programme effects.

The Propensity Score Matching (PSM) estimator for ATT can be written in general as:

$$\tau^{PSM} = E\ (P(Z)|D{=}1\ (E(Y1|D{=}1,\ P(Z)) - E(Y_0|D{=}0,\ P(Z)) \qquad (20)$$

which is simply the mean difference in outcomes over the common support, appropriately weighted by the propensity score distribution of programme participants (see: Caliendo and Kopeinig, 2005).

5.2.3. Propensity score matching algorithms

As the probability of observing two units with **exactly** the same value of the propensity score is in principle **zero** (since p(Z) is a continuous variable) the estimation of desirable programme effects requires the use of appropriate matching algorithms which define the measure of proximity in order to define programme non-participants who are acceptably close (e.g. in terms of the propensity score) to any given programme participant.

The most commonly used matching algorithms are: Nearest Neighbour Matching, Radius Matching, Stratification Matching and Kernel Matching (Cochran and Rubin, 1973; Dehejia and Wahba, 1999; Heckman, Ichimura and Todd. 1997, 1998; Heckman; Ichimura, Smith and Todd, 1998; Todd, 2006, 2008).

5.2.3.1. Nearest neighbour matching

In this matching method the non-participant with the value of Pj that is closest to participant's Pi is selected as the match:

$$C(P_i) = \min_j \ \| P_i - Pj \| \qquad (21)$$

Where: P is a propensity score

The most prominent variants of nearest matching are i) matching *with replacement*, i.e. farm/individual/unit which did not participate in the programme can be used more than once as a match; and ii) matching *without replacement* where respective programme non-participants can match only once. The biggest disadvantage of the nearest neighbour method is that it can result in bad matches if the closest neighbour (control unit) is situated far away (in terms of propensity score) from a supported unit.

5.2.3.2. Caliper matching

This method is to be considered as a variation of the nearest neighbour method. A match for a firm i is selected only if:

$$| P_i - P_j | < \varepsilon \qquad (22)$$

Where ε is pre-specified tolerance

By using caliper matching bad matches can be avoided by imposing a tolerance level on the maximum propensity score distance. The disadvantage of this method is the difficulty to know a priori what tolerance level is reasonable (Smith and Todd, 2005).

5.2.3.3. Kernel matching

Kernel matching is defined as:

$$W(i,j) = \frac{G\left(\frac{P_j - P_i}{\alpha_n}\right)}{\sum_{k \in I_0} G\left(\frac{P_k - P_i}{\alpha_n}\right)} \qquad (23)$$

Where:

W are weights for i and j

G is a kernel function

a_n stands for the bandwidth.

Various kernel functions can be used in applied work; the Gaussian, the Epanechnikow, biweight, triweight or the cosine functions. This non-parametric matching estimator (kernel) is especially interesting as it allows for a match of each programme participant with multiple units in a control group with weights which depend on the distance between the participant observation for which a counterfactual is being constructed and each comparison group observation. In this method weights are inversely proportional to the distance between the propensity scores of participants and controls within the common support level (the further away a comparison unit is from the participant unit, the lower the weight it receives in the computation of the counterfactual outcome). The main advantage of this method is that a lower variance is achieved because more information is used[20]. Another useful property of applying this method is a possibility of using standard bootstrap techniques for the estimation of standard errors for matching estimators that in generally should not be applied when using nearest neighbour matching (Abadie and Imbens, 2004; Todd, 2006).

5.2.3.4. Local linear weighting function

Local linear weighting function (Heckman, Ichimura and Todd, 1997; Smith and Todd, 2003)) can be defined as:

$$W(i,j) = \frac{G_{ij} \sum_{k \in I_0} G_{ik} (P_k - P_i)^2 - [G_{ij}(P_j - P_i)][\sum_{k \in I_0} G_{ik} (P_k - P_i)]}{\sum_{j \in I_0} G_{ij} \sum_{k \in I_0} G_{ij} (P_k - P_i)^2 - (\sum_{k \in I_0} G_{ik} (P_k - P_i))^2}$$

(24)

Where:

W = weights

The difference between kernel matching and local linear matching is that the latter includes in addition to the intercept a linear term in the propensity score of a treated individual. This is an advantage whenever comparison group observations are distributed asymmetrically around the treated observation, e.g. at boundary points, or when there are gaps in the propensity score distribution (Caliendo and Kopeinig, 2005).

5.2.4. Selection of appropriate matching algorithm

Obviously, the specification of a matching algorithm hinges on the two basics factors, i.e. definition of proximity (in the propensity score space) and determination of weights (weighting function) (Essama-Nssah, 2006). In some empirical studies 1-to-1 or 1-to-n nearest neighbour with calliper are used as a standard application. In others, the kernel matching is

20 A systematic analysis of the finite-sample properties of various propensity score matching and weighting estimators through Monte Carlo simulation can be found in: Frölich, 2004.

favoured. Empirical comparison of matching methods suggests that their performance can vary case-by-case thus no one method fits all circumstances and is therefore always preferable (Zhao, 2000; Caliendo and Kopeinig, 2005). Though asymptotically all PSM estimators should yield the same results (Smith, 2000), in small samples the choice of matching algorithm can be important (Heckman, Ichimura and Todd, 1997).

Among many methods allowing the assessment of the matching quality of the most popular approaches are: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); iii) joint significance and pseudo R^2 (Sianesi, 2004); or iv) stratification tests (Dehejia and Wahba 1999, 2002). If the quality indicators are not satisfactory, one reason might be misspecification of the propensity score model (Caliendo and Kopeinig, 2005) or failure of the CIA (Smith and Todd, 2005).

5.3. Selection of relevant conditioning variables

PSM is a suitable technique dealing with endogeneity problems in case a rich dataset is available and almost all important factors driving the potential bias can be observed. Yet, the success of matching estimator depends on the availability of observable data to construct the set **Z** such as appropriate conditions are satisfied (Heckman, Ichimura, Smith and Todd, 1998; Todd, 2006). If this not the case and if only a subset of Z, i.e. Z0 ε Z is observable, a bias may arise in matching.

As shown in (Todd, 2006) the propensity score matching estimator based on Z0 instead of Z converges to:

$$\alpha'_M = E_{P\,(Z0)|D=1}\,(E(Y_1|P\,(Z_0),\,D = 1) - E(Y_0|P\,(Z_0),\,D = 0)) \qquad (25a)$$

The bias for the parameter of interest, $E\,(Y_1 - Y_0|D = 1)$, is:

$$\text{Bias } M = E\,(Y_0|D = 1) - EP\,(Z0)|D{=}1\{E(Y_0|P\,(Z_0),\,D = 0)\}. \qquad (25b)$$

As (Todd, 2006) states "there is no way of *a priori* choosing the set of Z variables to satisfy the matching condition or of testing whether a particular set meets the requirements. In rare cases, where data are available on a randomized social experiment, it is sometimes possible to ascertain the bias". Heckman; Ichimura, Smith and Todd, 1998; Heckman, Ichimura and Todd 1997; Lechner 2001 argued that estimated bias depends on what variables are included in the propensity score, and showed that biases tend to be higher when the participation equation was estimated using a cruder set of conditioning variables. Heckman and Navarro-Lozano, 2004 offered examples where the application of the goodness-of-fit criteria advocated in the former literature resulted in a selection of conditioning sets that generated more bias compared with conditioning sets that were less successful in terms of the model selection criterion. By defining the concepts of *relevant information set* and *minimal relevant information set*, and distinguishing agent and analyst information sets Heckman and Navarro-Lozano, 2004 showed that when the analyst does not have access to the minimal relevant information, matching estimates of different treatment parameters are biased. Having more information, but not all of the minimal relevant information, can increase the bias over having less information. Yet, enlarging the analyst's information set with variables that do *not* belong into the relevant information set may either increase or decrease the bias from matching. While the econometric distinctions of exogeneity and endogeneity play crucial roles in applications of matching in the choice of appropriate conditioning sets, Heckman and Navarro-Lozano, 2004 argued in favour of the method of control functions that explicitly enables the modelling of omitted relevant conditioning variables.

The discussion in the literature as to the best method to help with the selection of relevant conditioning variables is still on-going. Some authors argue that the process of selection can be facilitated by using a method suggested by (Rosenbaum and Rubin, 1983). The method does not provide guidance in choosing which variables to include in Z, but can help to determine which interactions and higher order terms to include in the model for a given Z set (Todd, 2006).The proposed method helps to find the correct specification of a propensity score model by noting that after conditioning on P(D=1|Z) additional conditioning on Z should not provide new information about D. On the basis of this suggestion various specification tests on P(D=1|Z) after conditioning on P(Z) have been developed and implemented in the literature (Todd, 2006).

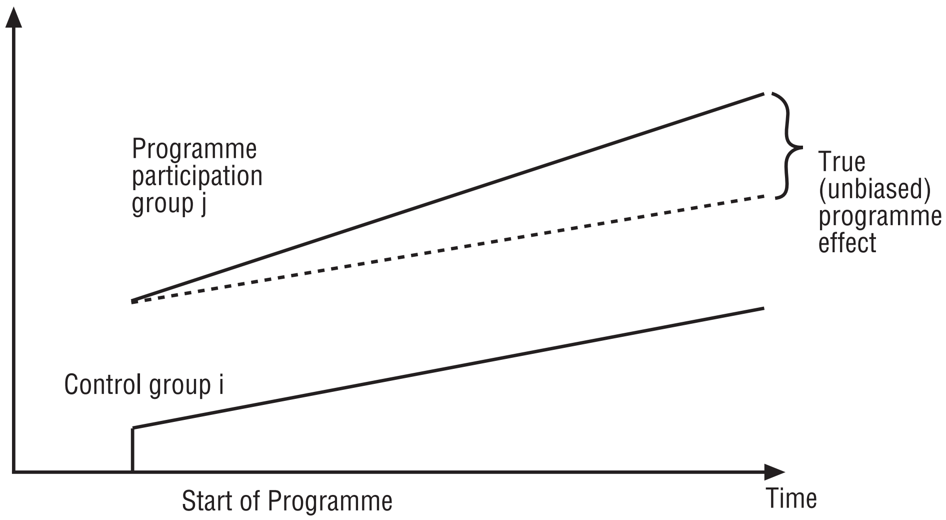
5.4. Unobserved heterogeneity

If there are unobserved variables which simultaneously affect assignment to treatment and outcome a hidden bias may arise. **Unobservable heterogeneity** can substantially affect estimated results of programme effects. This can be easily illustrated by showing the impact of unobserved heterogeneity on the estimated **ATE**.

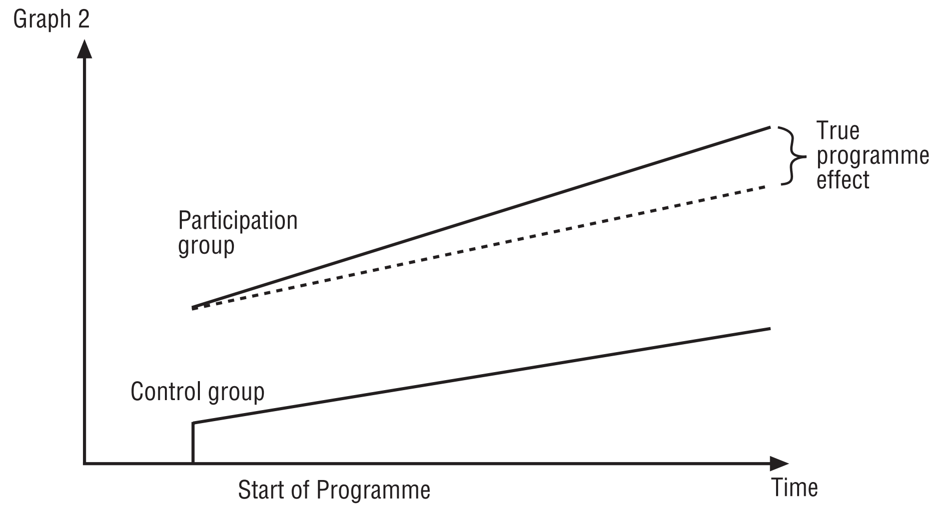
Assuming that the rate of unobservable heterogeneity does not change in time one can distinguish three major cases (Winship and Morgan, 1999):

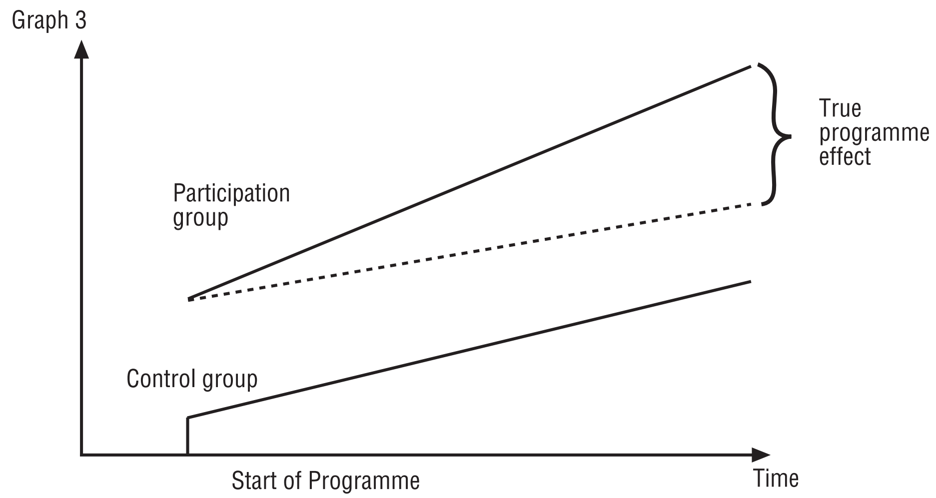
Case 1: Unobserved heterogeneity is **neutral** for both groups (i.e. it does not affect the growth rates in both groups differently). Unobserved differences between both groups, i.e. units which do not participate in the programme (e.g. units **i**) and units which participate (e.g. units **j**) are time invariant or fixed. In the absence of the programme, expected growth rates lines would be parallel to each other (Graph 1). The estimated ATE is unbiased.

Graph 1



Case 2. **Unobserved heterogeneity** in the group of programme participants **reduces** the estimated participation effects. *In the absence of the programme,* the growth rates of the outcomes would be different between two groups, e.g. growth rate in a group of programme participants would be higher than in the control group (Graph 2). This may occur for example, if the managerial skills of programme participants **unobserved** by the econometrician were higher than in the group of non-participants. The estimated ATE would be biased, without accounting for an unobservable heterogeneity.





Case 3. Unobserved heterogeneity effects in the group of programme participants **increase estimated programme effects**. *In the absence of the programme*, the growth rates of outcomes would be different between two groups, e.g. the growth rate in a group of programme participants would be lower than in the group of non-participants (Graph 3).

While the propensity score matching assumes **conditional independencies** (CIA) to exclude the problem of unobservable heterogeneity, in the evaluation literature arguments are provided that the **unconfoundedness** assumption holds even when two agents/units with the same values for observed characteristics differ in their treatment choices (participation or non-participation). The difference in their choices may be driven by differences in **unobserved characteristics** that are themselves unrelated to the outcomes of interest (Imbens, 2003). Yet, if there are unobserved variables that simultaneously affect assignment into the programme and the outcome variable, a **hidden bias** might arise to which matching estimators are not robust (Rosenbaum, 2002; Caliendo and Kopeinig, 2005; Becker and Caliendo, 2007).

A solution to the problem of a hidden bias can be a two-stage estimator (Heckman, 1976) that treats unobservable heterogeneity as a problem of an omitted variable, and solves this problem by including an estimate of the omitted variable as a regressor in the outcome equation along with the participation dummy and individual characteristics[21].

In our study the presence of hidden bias is formally tested using the approach described below in Chapter 5.6. <sensitivity analysis>.

5.5. Combined PSM and Difference-in-Differences estimator (conditional DID)

As shown above, conventional DID methods fail if the impact of unobservables is not time-invariant so that a group of programme participants and a control group are on different development trajectories. The probability of different development trajectories increases if already at the beginning of the programme the observed heterogeneity of both groups (and therefore the selection bias) is large. While propensity score matching can be applied to control for selection bias on **observables** at the beginning of the programme, a **combination** of PSM with DID methods (conditional DID estimator) allows for a better controlling of selection bias in **both** observables and unobservables. The combined PSM and DID method is a highly applicable estimator in case the outcome data on programme participants and non-participants is available both "**before**" and "**after**" periods (**t'** and **t**, respectively). The **PSM-DID** measures the impact of the RD programme by using the differences between comparable to each other programme participants **(D=1)** and non-participants (**D=0**) in the **before-after** situations. In this method observed changes over time for the **matched** (using PSM) programme non-participants are assumed to be appropriate counterfactual for programme participants.

The simplified notation for PSM-DID calculation can be described as follows:

$$\textbf{PSM-DID} = \{\Sigma\ (Y_{it} \mid (D=1) - Y_{it} \mid (D=0)) - \Sigma\ (Y_{it'} \mid (D=1) - Y_{it'} \mid (D=0))\}/n \qquad (26a)$$

Where:

$(Y_{it} \mid (D=1) - Y_{it} \mid (D=0))$ is the difference in mean outcomes between the **i** participants and the i *matched* comparison units *after* the access to the RD programme and

$(Y_{it'} \mid (D=1) - Y_{it'} \mid (D=0))$ is the difference in mean outcomes between the i participants and i *matched* comparison units at date 0 (*prior* to the RD programme).

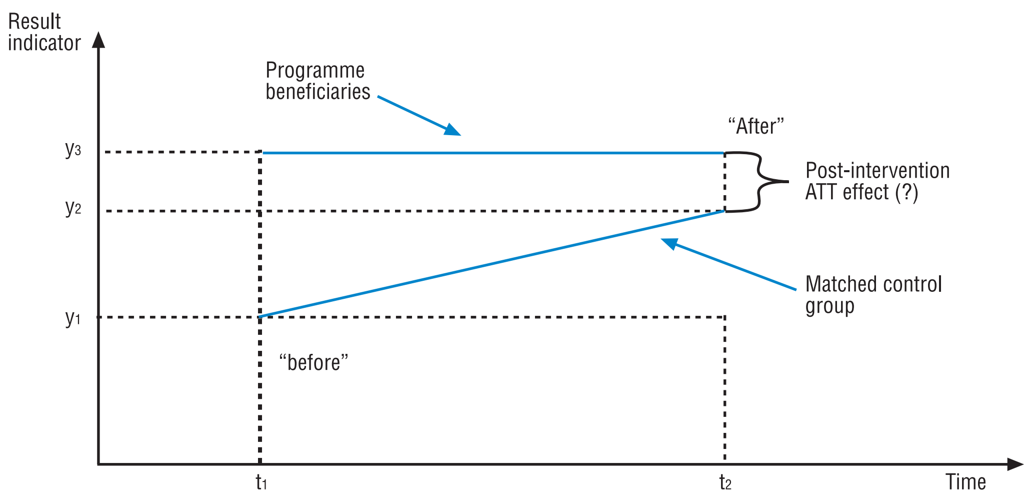
A decisive advantage of the **PSM-DID** estimator (conditional DID estimator), compared to a conventional DID estimator, is that by applying this methodology, initial conditions regarding observable heterogeneity of both groups (programme participants and non-participants) that could influence subsequent changes over time are largely eliminated[22]. Similarly, an application of a conditional DID estimator (PSM-DID) to the measurement of the effects of a given RD programme may greatly improve research findings compared with a situation where a standard PSM (e.g. for estimation of ATT) that uses post-intervention data only is applied.

21 A recent microeconomic evaluation literature focussed on constructing and estimating of models allowing for **heterogeneity** in response to programme participation among **otherwise observationally identical** units. Important outcome of these studies is the development of a new class of econometric estimators which allow for the possibility of selection to treatment (e.g. decision to participate in a programme) that is based on **unobserved components of heterogeneous responses** to treatment. (Heckman and Vytlacil, 2005).

22 Similar methodology was used by Ravallion, 2004;

The following example illustrates a potential qualitative difference in results (i.e. a conditional DID estimator (ATT-DID) vs. a standard ATT), see: Graph 4.

Graph 4. Comparison of ATT with ATT-DID estimator



In the above example (Graph 4) the use of a standard ATT estimator (based on post-intervention data only) calculated as a difference between mean values of a result indicator in the group of programme beneficiaries and matched control group (Y3-Y2) would have led policy makers to conclude incorrectly that the effect of a given RD programme was positive (the calculated post intervention ATT is higher than zero). Yet, had a ATT-DID estimator been applied, the effects of a given RD programme would have to be judged negatively, i.e. a mean value of a result indicator in the group programme beneficiaries remained unchanged (Y_1) while in the matched control group <i.e. without the programme> it increased in the examined period <before and after programme> from Y_1 to Y_2 (the calculated ATT-DID estimator is negative, i.e. $(Y_3\text{-}Y_2) - (Y_3\text{-}Y_1) < 0$).

5.6. Sensitivity analysis

5.6.1. Rosenbaum bounding approach

Since estimation of the magnitude of selection bias with non-experimental data is impossible one possibility to address the issue of unobservables is the **bounding approach** proposed by Rosenbaum, 2002. The approach allows determining how much hidden bias would need to be present to render plausible the null hypothesis of no effect (Rosenbaum, 2002) or in another words how strongly an unmeasured variable must influence the selection process in order to undermine the implications of matching analysis (Caliendo and Kopeinig, 2005).

As stated in (Becker and Caliendo, 2007) the bounding approach does not test the unconfoundedness assumption itself, because this would amount to testing that there are no (unobserved) variables that influence the selection into the programme, but instead this approach provides an evidence on the degree to which any significance results hinge on this untestable assumption.

While an extensive discussion of this approach was provided in Rosenbaum, 2002 an outline of this approach can also be found in (Caliendo and Kopeinig, 2005; Becker and Caliendo, 2007).

By assuming that the probability of participation can be expressed as:

$$P_i = P(x_i, u_i) = P(D_i=1|x_i, u_i) = F(\beta x_i + \lambda u_i) \quad (27)$$

Where:

D_i = equals 1 if an unit i participates in programme

x_i = are the observed characteristics for unit i

u_i = the unobserved variable

λ = is the effect of ui on the participation decision

If the study is free of hidden bias, $\boldsymbol{\lambda}$ will be zero and participation probability will be determined entirely by effects of $\mathbf{x_i}$. However, in the presence of hidden bias two matched units (with the same observed covariates x) will have different chances of programme participation. While the odds that both units **i** and **j** will participate are given by $P_i/(1-P_i)$ and $P_j/(1-Pj)$ the odds ratio is equal to [exp $(\beta x_i + \lambda u_i)$]/ [exp $(\beta x_j + \lambda u_j)$] which in case of identical observed covariates (implied by matching) reduces (the vector x cancels out) to exp $\{\lambda\ (u_i-u_j)\}$. Rosenbaum, 2002 showed that this implies the following bounds on the odds ratio that either of the two matched units will participate:

$$\frac{1}{e^{\lambda}} \le \frac{P_i(1-P_j)}{P_j(1-P_i)} \le e^{\lambda} \qquad (28)$$

If the odds ratio differs, i.e. departs from a value of 1 this can only be due to hidden bias. In this sense **eλ** is a measure of the degree of departure from a study that is free of hidden bias (Rosenbaum, 2002; Caliendo and Kopeinig, 2005; Becker and Caliendo, 2007). Sensitivity analysis means therefore examining the bounds on the odds ratio for programme participation that lie between 1/ **eλ** and **eλ.**[23]

Sensitivity analysis as above is applied in our study using formal (Mantel and Haenszel, 1959) tests statistics suggested by (Aakvik, 2001) and described in (Becker and Caliendo, 2007). Applications of sensitivity analysis can also be found in (Aakvik, 2001; DiPrete and Gangl, 2004; Caliendo, Hujer and Thomsen, 2005; Watson, 2005).

5.6.2. Other sensitivity checks

Sensitivity checks are carried out to test the stability of obtained results. Sensitivity checks embraced the response of estimated effects from programme participation to small changes, e.g. in the specification of the propensity score, number of selected companies, changes in covariates, changes in parameters of balancing properties, etc.

Given a standardized set of variables describing characteristics of agricultural enterprises (e.g. FADN data) an important sensitivity test was to find out what is a minimal/optimal set of conditional variables to be included in estimation of propensity scores.

23 With increasing $\mathbf{e}^{\lambda}$ the bounds move apart reflecting uncertainty in test statistics in the presence of unobserved hidden bias.

6. Estimation of other programme effects

6.1. Micro-economic effects

6.1.1. Deadweight loss effects

In some cases, RD programme support, may be mistargeted. **Deadweight loss effect** occurs if a participant of a RD programme (unit-**j**) would undertake a similar investment also *without* a RD programme support (i.e. RD support would not change investment behaviour of targeted enterprise). Deadweight loss effects can be measured using the following result indicators:

- investment value (in a given sector/branch/activity) per farm and year, or (as a second best);
- value of inventories (excluding private buildings) per farm and year;

Depending on data availability estimation of a possible deadweight loss can be carried out using two different approaches (Michalek, 2007):

Approach 1:

- Identification of RD-programme supported units **j** carrying out investments under specific RD measures (e.g. modernisation and restructuring of agricultural enterprises);
- Identification of a control group **k** (programme non-participants) matching with units **j** (similar distribution of all relevant covariates) in the period **t'** (i.e. before **j**'s access to the programme);
- Identification of units **m** in the control group **k** (where **m** is a sub-vector of **k**) which undertook the same of type of investment as **j** (in the period **between t'** and **t**);
- Calculation of **ATT** using data from both groups (i.e. **j** and **m**) on the basis of **DID,** using one of the matching techniques (i.e. kernel method);

It is expected that in case of deadweight loss the calculated DID-ATT between above groups (**j** and **m**) will be close to zero.

Approach 2:

- Identification of RD programme supported units **j** carrying out investments under specific RD measures (e.g. modernisation and restructuring of agricultural enterprises);
- Identification of a control group **k** (programme non-participants who were in the following period **willing to invest**[24]) matching with units **j** (similar distribution of all relevant covariates, including a covariate showing the **current level of investment**) in the period **t'** (i.e. before **j**'s access to the programme);
- Using a variable "value of inventories" or "value of investment" as a result indicator;
- Calculation of **DID-ATT** using the above result indicator and data from both groups (i.e. **j** and **k**);

It is expected that in case of deadweight loss the change in result indicators in the group of farms supported from the RD programme (between years **t** and **t'**) and the control group (between years **t** and **t'**) will be almost the same,

24 Some farms (programme non-participants) may not be willing to invest due to e.g. lack of farm successor (the latter is usually an unobservable, i.e. cannot be derived from available FADN data).

i.e. calculated DID-ATT between the above groups (**j** and **k**) will be close to zero.

6.1.2. Leverage effects

Leverage effect can be considered as important micro-economic consequence of RD support. It occurs if public funding (e.g. in form of RD programme) **induces private spending among the programme beneficiaries**[25]. Leverage effects can be measured using the following result indicators:

- Money transfers from farm to farm household for living
- Money transfers from farm to farm household for building of private assets
- Money transfers from farm to farm household (total)

Calculation of the leverage effect can be carried out by taking the following steps (Michalek, 2007):

- selection of individual units **j** supported by a RD programme
- identification of a comparison/control group **k** matching with units **j** (identical distribution of covariates) in the period **t'** (i.e. prior to **j**'s access to the programme)
- selection of outcome variables (result indicators) as **proxies for private spending**, e.g. money transfers from farm to farm households; level of private and farm consumption; other expenditures, except for those which were directly supported by RD projects
- calculation of **ATT** for given outcome variables between both groups (i.e. **j** and **k**) on the basis of **DID,** using one of the matching techniques (i.e. kernel method).

It is expected that in case of a significant leverage effect the calculated DID-ATT will be positive and significant.

6.2. Macro-economic/general equilibrium effects

General equilibrium effects occur when a programme affects persons/enterprises other than its participants (Smith, 2000). The most important possible impacts are the **substitution effect** and the **displacement effect** (Calmfors, 1994). Both effects play usually a more important role in the evaluation of large programmes than in the evaluation of small programmes. Yet, they cause problems for programme evaluators because most of the partial evaluation methods either miss these effects entirely or produce biased results in their respect. Due to a possibly negative/positive impact on programme **non-participants** the evaluation **of a given programme becomes more complex.** Specifically, standard propensity score matching methods assume that outcomes for non-participants in the control group are not affected by the programme (no general equilibrium effects). If general equilibrium effects had occurred during the implementation of a given RD programme (i.e. between years **t** and **t'**), i.e. if a given RD programme had a substantial impact (positive or negative) on farms which did not participate in this programme, partial equilibrium evaluation techniques such as standard PSM would produce biased estimates of programme effects. To overcome these problems we propose to use a modified approach (i.e. a two stage approach using a combination of a modified and a standard propensity score matching) which aims to achieve unbiased results (see below).

25 Obviously, additional private spending should exclude specific RD **supported** activities carried out by programme beneficiaries, e.g. supported investments. Had this been not a case, the scope of leverage effect would be proportional to the co-financed part of the RD project.

6.2.1. Substitution effect

Substitution effect belongs to **macro-economic** effects. It is normally defined as the effect obtained in favour of direct programme beneficiaries but **at the expense** of persons/farms/unit that do not qualify or participate in a given intervention. It occurs if, due to support provided from RD programme to units **j**, **available resources shift (e.g. due to increase of input prices (costs) or decrease of output prices (profits)**, **at the detriment of non-supported or non-eligible units i (the latter are usually located in close neighbourhood of j).**

For example**,** persons employed in units **i** (programme non-participants) become employees of programme assisted units **j;** input/factor prices **w** faced *prior to RD programme* by units **i** *and* **j** increase *after supporting of units* **j;** or producer prices **p** available *prior to RD programme* for units **i** *and* **j** decrease *after* support provided to units **j**. Substitution effect (in contrast to displacement effect) is expected to occur only in a *direct neighbourhood* of units **j** and may influence all major outcome indicators, e.g. profits, the level of employment, etc.

In some cases, substitution effects may have an especially strong impact on units **i** characterised by a **dissimilar distribution of covariates** compared to **j,** e.g. programme support of *large* companies in rural areas may bring about a clear deterioration of situation of *neighbouring small* companies. In this case, programme participants may differ markedly from non-participants (*supported large* company is surrounded by *small* ones **only) bringing about additional problems of using PSM methods** (e.g. due to insufficient common support regions). Should the distribution of propensity score in group of supported farms (large) be very different from non-supported farms (small) in this situation an application of **standard** matching techniques to remove a potential selection bias may not be justified and respective adjusted methodologies have to be applied.

Substitution effects can be measured using the following result indicators:

- profit per farm/year;
- gross value added per farm/year;

In order to estimate substitution effects of RD programme, depending on data availability and object of interest, we propose the following two approaches (Michalek, 2007):

Approach 1 (e.g. estimation of substitution effects on small units located in a close neighbourhood of supported (large) units):

- identification of units **j** supported by RD programme (using pre-defined characteristics, e.g. large companies);
- identification of *non-supported* units **m** located in a **close spatial neighbourhood** of **j** (closeness of the neighbourhood will be identified according to a pre-defined **radius r**);
- identification of *non-supported* units **k** matched to units **j** (using pre-defined characteristics, e.g. large companies) in other locations;
- identification of non-supported units **n,** matched to units **m** located in a close spatial neighbourhood of units **k** (radius **r);**
- redefining units **m** as "quasi-supported" , i.e. D=1;
- calculation of ATT between units **m** and units **n** *before* and *after* supporting period, given that **m** units are considered as "quasi-supported";
- calculation of **substitution** effects by subtracting a difference between ATT computed for small units/farms **m** (i.e. **quasi-**supported) and small units/farms **n** (non-

supported) in period t and t′ (*before* and *after* supporting period) using eq. 24;

Note, that units **m** and units **n** are expected to be *differently* affected by RD programmes which originally was provided only to units **j**. Units **m,** located in the close neighbourhood of **supported** units **j,** are expected to be under **indirect impact** of RD support provided for units **j** (quasi-supported), whether units **n** (located in the close neighbourhood of **non-supported** units **k**) are **not** expected to be affected by a RD support.

$$\mathbf{SUB} = \{\Sigma\,(Y_{mt} \mid (D\sim 1) - Y_{nt} \mid (D=0)) - \Sigma\,(Y_{mt'} \mid (D\sim 1) - Y_{nt'} \mid (D=0))\}/n \qquad (29)$$

Approach 2: Estimation of substitution effects on *similar* non-supported farms located in regions with a high programme intensity/exposure

- identification of non-supported units/farms **m** located in a close spatial neighbourhood of units **j** supported by the programme (closeness can be identified according to the programme intensity of a given region);
- identification of non-supported units **n**, matched to units **m** located in regions characterised by a low programme intensity (no programme effects on non-supported farms in regions characterised by a low programme intensity is assumed);
- redefining units **m** as “quasi-supported” , i.e. D~1;
- calculation of ATT between units **m** and units **n** before and after supporting period, given that **m** units are considered as “quasi-supported”;
- calculation of substitution effects by subtracting a difference between ATT computed for small units/farms **m** (i.e. quasi-supported) and small units/farms **n** (non-supported) in period **t** and **t′** (before and after supporting period) using eq. 24;

Should substitution effects be substantial, i.e. should all enterprises which did not participate in a given RD programme were found to be affected by this programme (positively or negatively), they would have to be eliminated from further PSM analysis. In this case, “true” ATTs have to be re-estimated *without farms (potential controls) considered to be a subject to substitution effects*.

6.2.2. Displacement effect

Displacement effect is the programme effect that occurs in a programme area at expense of another area. It takes place if farms **i** located in one geographical area (**a$_i$**), which is **not** a subject to RD support, becomes adversely affected by a support provided to farms **j** located in another geographically area (**aj**). For example, due to RD support to units **j** jobs are created in units **j** (located in programme assisted area **aj**) at the detriment of jobs *lost* in units **i** located outside of the area concerned.

One of the differences between substitution and displacement effects is the location of adversely affected units **i**, compared to units **j.** In case of substitution one expects possible effects to occur in the same region (programme area), whereas in case of displacement effect resources are expected to be shifted from non-supported programme areas to supported ones.

Displacement effects can be measured using the following results indicators:

- employment per farm (in case of a direct move of employees from non-supported to supported regions);
- profits per farm (in case of a move of distributors/providers of agricultural inputs from non-supported to supported regions);

Calculation of **displacement effects** can be carried out by taking the following steps (Michalek, 2007):

- identification of units **j** supported by the RD programme in a specific rural area **aj** (to be defined);
- identification of non-supported units **k,** located in the supported area **aj,** which match with units **j;**
- identification of *non-supported* areas (**ai**) (closeness of the areas **ai** from areas **aj** can be identified according to a pre-defined **radius r**);
- identification of non-supported units **m,** located in the *non-supported* area **ai,** which match to units **j;**
- calculation of DID-ATT between units **j** and units **k** as well as between units **j** and units **m (e.g. the only difference should be location of units k and m)** *before* the support and *after* support, taking into consideration selected specific outcome variable, e.g. level of employment;

The lack of displacement effects would result in similar differences in DID-ATT between units **j** and **k** compared with **j** and **m** (i.e. location of units **k** and **m** would be considered as irrelevant). Generally speaking, and assuming no other general equilibrium effects (e.g. substitution effects), the greater the difference in DID-ATT between both groups (**j-m**) and (**j-k**) after the programme, the higher the probability that the better performance of units **j** and **k** located in area **aj** occurred to the detriment of units **m** located in non-supported areas **ai**. For example, assuming that there is no simultaneous shift of employment (e.g. due to substitution effect) from farms **k** to farms **j** (both located in the supported area **aj**) it can be expected that considerable displacement effects would bring about a deterioration of DID-ATT between **j** and **k** compared with **j** and **m** (this applies both to employment as well as profits as result indicator). Should however substitution effects also take place at the same time (e.g. a simultaneous shift of employment from farms **k** to farms **j** (both located in the supported area **aj**) it would be difficult to separate one general equilibrium effect from the other.

7. Other evaluation problems

7.1. Slowly unfolding effects

In case of slowly unfolding programme effects the measurement of programme results may take place more than once (e.g. the measurement of selected result indicators can be carried out two or three times in subsequent time periods (years) after programme support was received). The major technical constraint is the data availability (many farms/units may drop from the panel data base in subsequent years). An important methodological restriction is the absence of any other major factor *additionally* influencing *either supported* farms or *non-supported* farms *only* (e.g. another programme targeting previously supported or non-supported farms) that started between the finalization of the original RD programme and the time the results indicators are measured.

8. Overview of the procedure applied for an empirical estimation of micro-economic and general equilibrium effects of an RD programme

The above evaluation methodology was applied to the assessment of RD programmes implemented both in new Member States (SAPARD programme in Slovakia in years 2002-2004) as well as old Member States (AFP Programme in Germany: Schleswig-Holstein in years 2000-2006). In both cases (SAPARD and the AFP programme) the assessments of programme effects concerned a similar measure, i.e. the investment in agricultural farms.

Due to data constraints, the assessment of programme effects of the SAPARD programme in Slovakia was limited to direct programme results (at a micro-level) only. In case of the AFP programme (Schleswig-Holstein in Germany) richness of data allowed for an estimation of both direct and indirect programme effects (including, deadweight loss, substitution and displacement effects).

The main analytical steps carried out to estimate direct and indirect programme effects followed the methodology described in Chapters 5 and 6. In both cases (Slovakia and Germany) the binary propensity score matching was the crucial methodology applied to evaluate direct and indirect programme results.

The above evaluation methodology was applied in the following steps:

Firstly, **Calculation of individual propensity scores**. The propensity scores for each observation in the supported and the non-supported sample of producers/enterprises were econometrically estimated using the predicted values from a standard logit-model. The estimated logit model of programme participation is a function of all the variables in the data describing farm/unit characteristics and economic performance that are likely to determine both participation and programme outcomes. Propensity scores are predicted values of the probability of participation obtained from the logit regression calculated individually for every sampled supported and non-supported unit.

Secondly, **Exclusion of non-similar enterprises from the control group**. Some of the supported and non-supported units were excluded from further comparisons because their propensity scores were outside the range calculated for supported units (outside of the common support region). Matched pairs of producers/enterprises/regions etc. were constructed on the basis of how close the estimated scores were across the two samples (programme participants vs. controls). Out of several alternative matching algorithms enabling calculation of the average outcome indicator of the matched supported and non-supported groups, ranging from "nearest neighbour" to kernel functions (Gaussian or Epanechnikov), the "best" matching algorithm (given a data base) was selected on the basis of three important criteria: minimization of the standardized bias (after matching), satisfaction of balancing property tests, and satisfaction of the pseudo R^2 test.

Thirdly, **Calculation of relevant outcome indicators**. The mean values of the outcome indicator for comparable supported and control units were computed using the matching algorithm selected above (e.g. Kernel method).

Fourthly, **Calculation of the most important policy parameters**. All the most important policy parameters, i.e. Average Treatment Effect on treated (ATT), average treatment effect (ATE) and average treatment effects on non-treated (ATNT) were calculated in this step.

Fifthly, **Estimation of programme effects.** Programme effects were computed on the basis of the estimated differences between respective policy parameters (ATT, ATE, ATNT) prior and after finalization of the programme (conditional DID method).

Sixthly, **Performing sensitivity analysis.** The sensitivity analysis was carried out in order to find out how much hidden bias would need to be present to render plausible the null hypothesis of no programme effect.

9. Data

Slovakia:

The dataset comprised FADN farm data collected for 232 Slovak large agricultural companies supported and non-supported through the SAPARD programme in the years 2002-2005 (balanced panel data).

Schleswig- Holstein, Germany

The main data source used for the assessment of the effects of the AFP programme in Schleswig-Holstein was farm bookkeeping data comprised of approximately 10 500 farms for the year 2000/2001 and 3 900 farms for the year 2007/2008).

II. Empirical Analysis

10. Estimated effects of the SAPARD programme in Slovakia

10.1. Scope and distribution of SAPARD funds under Measure 1: Investments in agricultural enterprises.

The SAPARD support provided under Measure 1 primarily targeted the following agricultural sectors: a) beef sector, b) pork sector, c) sheep sector, d) poultry sector, e) fruits and vegetables sector.

The main objectives of this measure were to:

- Assure compliance with EU animal welfare, hygiene and environmental requirements;
- Increase the labour productivity and improve working conditions;
- Increase quality of agricultural production;
- Increase competitiveness of products and producers;
- Improve storage and post-harvest infrastructure;
- Maintain and use the natural potential of the country and solve employment problems in marginal regions;

Programme support under Measure 1 had the form of a capital grant covering up to 50% of costs of eligible investments in the above sectors. The structure of allocated financial resources from the SAPARD programme to individual programme measures (1-9) shows that the Measure 1 (Investment in agricultural enterprises) was the most important single programme activity. Indeed, between the years 2002-2004 as much as 27.5 Mill EUR or 28% of the total available resources under SAPARD programme (97.3 Mill EUR) were allocated to Measure 1. After 2004, i.e. after Slovakia's EU accession, the amount of total funds (i.e. SAPARD + RDP) allocated to Measure 1 increased to 32.6 Mill EUR i.e. by additional 19%. Out of 450 project proposals submitted under this measure 343 projects (SAPARD and RDP) were contracted and concluded. The major share of available funds under Measure 1 was spent on the support of investments in the cattle sector (34% of funds and 149 projects), followed by the fruit sector (23% of funds and 67 projects), poultry sector (20% of funds and 57 projects), pork sector (18% of funds and 55 projects) and sheep sector (5% of funds and 15 projects). The major beneficiaries of programme support under this measure (receiving approximately 67% of funds available under this measure) were **large agricultural companies** located in relatively well-developed regions of West Slovakia (Nitra, Trnava and Bratislava).

At the beginning of the SAPARD programme there was a rather slow uptake of funds and a low level of participation of primary agricultural producers (Measure 1). In the case of large agricultural companies, this was mainly due to their difficulty in meeting the originally strict formal economic eligibility criteria, and their problems in securing external co-financing of their investment projects (50% or more) through commercial banking systems (many large agricultural companies in Slovakia were highly indebted at this time!). In the case of small individual firms, agricultural producers had problems with the interpretation of programme guidelines and were facing huge administrative costs for project preparations. Given the above situations, during the implementation of the SAPARD programme in Slovakia numerous changes to the programme were undertaken by the Programme Managing

Authority (i.e. via amendments to the Rural Development Plan) with the aim of facilitating the spending of available SAPARD resources. Many of these changes were initiated by lobbyists of large agricultural/food processing enterprises, which at that time, due to the overall difficulties in the agricultural sector but also due to management inefficiencies, were *economically too weak to qualify as eligible* enterprises under the original programme conditions. Although many studies pointed out that a drastic weakening of programme eligibility criteria may have diluted the potential impact of the programme, this trend was irreversible[26]. In fact, the distribution of funds under Measure 1 shows clearly that at that time an important goal of the Managing Authorities of the SAPARD programme was to strengthen the competitiveness and financial condition of large agricultural enterprises (in their majority former cooperatives or state farms) enabling a relatively smooth transition for them from a risky pre-accession period to a more stable EU membership (post-accession) stage[27].

10.2. Selection of companies

Deficiencies in the SAPARD monitoring system (especially problems with a data base describing economic performance of small individual farms), and a relatively abundant amount of data on large agricultural companies (part of Slovak FADN) resulted in the focus of the assessment of the programme on (large-sized) agricultural enterprises (average size approx 1500 ha). In order to ensure maximum comparability between SAPARD supported and non-supported agricultural companies the eligibility issues were explicitly accounted for[28].

The following steps were carried out:

- SAPARD beneficiaries were identified and selected from the existing FADN data bases. Data for each SAPARD beneficiary was collected in the years prior to their participation in SAPARD and in 2005 (after SAPARD).

- SAPARD general and specific eligibility criteria (e.g. pre-defined farm performance coefficients and farm profitability ratios; various minimum/maximum production-, age-, etc. thresholds; etc.) that were valid in individual years were translated into respective quantitative coefficients and applied to all non-SAPARD units included in FADN data bases.

- Agricultural companies, which did not receive a support from the SAPARD programme and which satisfied the above participation criteria in years 2002-2005 were selected as *eligible* non-participants.

- Respective balanced panels (i.e. embracing SAPARD beneficiaries and all non-SAPARD units that met SAPARD eligibility criteria in specific years) were constructed for the years 2002-2005, i.e. observations on the same units in period 2002-2005.

26 Proposed changes were accepted by the EC on the presumption that otherwise the programme funds would not be spent at all.

27 After EU accession, a stabilization of the situation for large agricultural companies was ensured by taking advantage of available direct payments.

28 Generally speaking, an individual agricultural company not participating in the SAPARD programme may have chosen not to do so, or may have been ineligible (eligibility criteria were set in the programming document "Rural Development Plan"). Ideally, supported and non-supported companies should only differ in their decision to participate. Yet, if a company is programme ineligible it means that its support (via a given programme) was not policy intended because some critical company background characteristics (e.g. prior economic performance, current capacities, etc.) significantly differed from targeted ones. By including ineligible programme non-beneficiaries (which markedly differ in their background characteristics from eligible firms) into the analysis of programme effectiveness the similarity (balancing property) between programme beneficiaries and the control group would be violated.

On the basis of the available Slovak FADN data base, 232 agricultural companies were selected for further analysis (balanced panel data), which was performed for the years 2003 (before SAPARD) and 2005 (after SAPARD)[29]. Of the selected 232 agricultural enterprises 51 agricultural farms were SAPARD participants and 181 farms SAPARD non-participants (yet, SAPARD eligible).

10.3. Differences between the groups of programme participants and non-participants

A brief analysis of some key characteristics of the selected groups of farms (SAPARD participants D=1 vs. non-participants D=0) shows that these two groups (both SAPARD eligible) differed considerably (Table 1).

Table 1. Slovakia: Major characteristics of agricultural companies supported (D=1) and non-supported (D=0) from the SAPARD programme (year 2003)

Participation	Own-land in ha	Agric. Land used in ha	Employment (persons)	Value of assets (buildings) in SKK (1000)	Value of assets (machinery and others) in SKK (1000)	Value of asset (livestock) in SKK (1000)	Profit in SKK (1000)	Profit per ha in SKK (1000)
D=0 (181)	870	1439	64	34747	8604	3492	-3338	-2.318
D=1 (51)	1507	1930	77	46154	14939	4709	-880	-0.456

Agricultural companies which received support from the SAPARD programme were generally much larger (ha), they employed more people and were more profitable (i.e. less unprofitable) compared with those agricultural companies which were non-supported.

10.4. Estimation of a logit function

10.4.1. Selection of covariates

Given the individual characteristics of agricultural companies (programme participants vs. programme non-participants), propensity scores (i.e. the conditional probability of a farm's participation in the SAPARD programme) were estimated for all selected enterprises using a logit function.

Since the matching strategy builds on the Conditional Independence Assumption (CIA) requiring that outcome variables must be independent of support conditional on the propensity score, the selection of variables into the logit function (also) has to meet these requirements. Generally, covariates entering the logit function are expected to determine both programme participation and outcomes (the latter are typically measured in terms of relevant result indicators at micro-level). Given that only variables that are unaffected by programme participation (including anticipation of participation) can be included into the model, the selection of covariates from a given set of available characteristics (FADN data) can in principle be carried out using two methods:

a. relying on experts choice (based on economic theory and some empirical evidence)

b. relying on statistical significance (by iteratively adding new variable to specification)

29 All selected beneficiaries received support from SAPARD in year 2004. Unfortunately, inclusion of the following years (2006 and 2007) was not possible due to dropping of many former agricultural companies from the data panel.

Table 2. Slovakia. List of selected variables

List if variables	
v2b364y03	Profit per company
vd37a23	Number of pigs for fattening
vh 72	Liabilities
vk61	Initial stock wheat
vf331	Costs of interest paid
vd37a20	Stock of other sheep
v1b541	Assets total non-current receivables
vc39	Employment (manual workers)
vk69	Initial stock beans, peas, etc.
vk85	Production of oat
vk665	Initial stock of grass and hay in haylage
vf323	Overhead costs (water)
vf335	Costs of interest and fees (total)
v1b52	Net value of current assets
v2b340	Interest income
v2b31	Revenue from sale of merchandise
vf311	Costs of own feedstuff for pigs
vk865	Production of grass and hay
vd37a4	Stock of heifers and bulls for fattening (6–12 months)
vd37a6	Stock of bulls for fattening (1-2 years)
v3b66	Land area hired from others (grass land and pastures)
vf35	Costs of cars
vc32	Employment (directors, chairmen, representatives, etc.)
vk813	Production of industrial potatoes for starch
v1b586	Total liabilities (external sources)
v2b325	Other operating income
vb2a3111	Costs of consulting and services
vk873	Production of grapes for wine
vc37	Employment (tractor drivers and mechanics)
v2b324	Value adjustment against operating expenses

In our study these two approaches were combined. This was done in three major steps:

Firstly, on the basis of expert knowledge the long-list of the most crucial variables determining both participation and outcomes was constructed, using the entire Slovak FADN data set. This activity resulted in a pre-selection of ca. 400 out of approximately 7400 potential variables and categories.

Secondly, the most important statistically significant variables (out of the pre-selected 400) which simultaneously satisfied balancing property tests were selected as relevant covariates in the estimated logit function. This was done by applying an iterative procedure (by iteratively adding new variables to the logit specification) whereby respective balancing property tests were carried out given imposed common support conditions.

Thirdly, step two was supplemented by an obligatory selection of those variables thought as critical for comparability of economic performance across agricultural companies (i.e. profit per company).[30]

By using this method 30 variables were selected that appeared statistically as the most significant and simultaneously satisfied the balancing property tests. The list of selected variables is given in Table 2.

10.4.2. Estimation results

The results of logit estimation (SAPARD Measure-1) are shown in Tab 3.

Table 3. Slovakia: Results of estimation of logit function

particip	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
v2b364y03	.0025436	.0008284	3.07	0.002	.0009199	.0041673
vd37a23y03	.0073994	.0027669	2.67	0.007	.0019763	.0128224
vh72y03	.0024133	.0007507	3.21	0.001	.0009421	.0038846
vk61y03	-.0179229	.0061878	-2.90	0.004	-.0300509	-.005795
vf331y03	-.0121055	.0038116	-3.18	0.001	-.0195761	-.004635
vd37a20y03	-.0289665	.0103319	-2.80	0.005	-.0492167	-.0087162
v1b541y03	-.0066384	.0024064	-2.76	0.006	-.0113549	-.0019219
vc39y03	-.6308535	.1907164	-3.31	0.001	-1.004651	-.2570562
vk69y03	.1495759	.0472736	3.16	0.002	.0569214	.2422304
vk85y03	-.069111	.0241572	-2.86	0.004	-.1164583	-.0217636
vk665y03	.0048039	.0015541	3.09	0.002	.0017579	.0078499
vf323y03	-.0386929	.0130836	-2.96	0.003	-.0643363	-.0130496
vf335y03	.0003901	.0001232	3.17	0.002	.0001485	.0006316
v1b52y03	-.1539366	.068769	-2.24	0.025	-.2887213	-.0191519
v2b340y03	.0343233	.0116338	2.95	0.003	.0115216	.0571251
v2b31y03	-.0016953	.0005189	-3.27	0.001	-.0027123	-.0006783
vf311y03	.001743	.0006322	2.76	0.006	.0005038	.0029822
vk865y03	.0029226	.000914	3.20	0.001	.0011311	.0047141
vd37a4y03	-.2366995	.0723047	-3.27	0.001	-.3784141	-.0949848
vd37a6y03	-.0293615	.0096498	-3.04	0.002	-.0482749	-.0104482
v3b66y03	-.0073581	.0025647	-2.87	0.004	-.0123848	-.0023315
vf35y03	-.0188592	.0061192	-3.08	0.002	-.0308527	-.0068657
vc32y03	-7.00138	2.474549	-2.83	0.005	-11.85141	-2.151352
vk813y03	.0100983	.0040284	2.51	0.012	.0022029	.0179938
v1b586y03	-.0002031	.0000718	-2.83	0.005	-.0003439	-.0000623
v2b325y03	.0004984	.0001745	2.86	0.004	.0001564	.0008405
vb2a3111y03	.0030658	.001674	1.83	0.067	-.0002152	.0063468
vk873y03	-.0281685	.013879	-2.03	0.042	-.0553709	-.0009661
vc37y03	.6128139	.1969328	3.11	0.002	.2268328	.998795
v2b324y03	-.0139305	.005634	-2.47	0.013	-.0249729	-.0028881
_cons	-7.079863	2.091144	-3.39	0.001	-11.17843	-2.981296
Logistic regression	Number of obs	LR chi2(30)	Prob > chi2	Log likelihood	Pseudo R2	
	= 232	= 194.98	= 0.0000	= -24.703442	= 0.7978	

30 Selection of this variable to the logit function was done after having verified that balancing property tests were satisfactory.

In the next step, the results of a logit function estimation were used to derive for all agricultural companies their individual probability (propensity scores) of receiving support from the SAPARD programme (Measure 1).

10.4.3. Selection of a matching algorithm

In order to ensure comparability, the estimated propensity scores of agricultural companies - programme participants (SAPARD, Measure 1) and their controls should be similar. As the probability of observing two units with exactly the same value of the propensity score is in principle zero (since p(Z) is a continuous variable), the estimation of programme results (ATT) requires the use of appropriate matching algorithms. The latter set up the measure of proximity thus enabling the definition of programme non-participants who are acceptably close (e.g. in terms of the propensity score) to any given programme participant. To avoid a lack of comparable units we restricted matching and hence estimation of the effect of programme participation to the **region of common support**.

The most commonly used matching algorithms involving the propensity score are: Nearest Neighbour Matching, Radius Matching, Stratification Matching and Kernel Matching (Cochran and Rubin, 1973; Dehejia and Wahba, 1999; Heckman, Ichimura and Todd. 1997, 1998; Heckman; Ichimura, Smith and Todd, 1998). While asymptotically all PSM matching techniques should yield the same results, the choice of the matching method (or applied matching parameters, e.g. number of nearest neighbours, radius magnitude, kernel type, etc.) can make a difference in small samples (Smith, 2000)[31]. As the quality of a given matching technique depends heavily on the data set, the selection of a relevant matching technique in our study was carried out using three independent criteria: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); and iii) joint significance and pseudo R^2 (Sianesi, 2004).

We found that the best results were achieved by using an iterative procedure (e.g. linear search) aiming at the minimization of the calculated standardized bias[32] (after matching) and applying min{min} as the main selection criterion. In all considered cases (various matching algorithms)[33] an optimal solution could easily be found due to local/global convexity of the objective function with respect to functional parameters characterizing each matching algorithm (e.g. radius magnitude in radius matching; or number of nearest neighbours in nearest neighbour matching). An overview of the results obtained using different matching algorithms is provided in Table 4.

Although in our example all matching algorithms lead to qualitatively similar results, i.e. irrespective of the matching algorithms used the estimated programme effects (obtained by using conditional DID, i.e. combining PSM and a traditional DID method) were found to be negative, the best results (the lowest bias) were obtained by employing kernel Epanechnikov matching (bandwidth 0.06).

The application of the above procedure and the imposition of common support restrictions resulted in dropping from further analysis 37 programme supported agricultural enterprises thus selecting 14 *comparable* (out of total 51) programme participants and 181 programme non-participants as relevant counterparts (Table 5).

31 Description of trade-offs linked to each of matching algorithms can be found in (Caliendo and Kopeinig, 2005).

32 The standardized bias is the difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin, 1985).

33 This does not apply to local linear weighting function matching which first smoothes the outcome and then performs nearest neighbour matching. In this case more controls are used to calculate the counterfactual outcome than the nearest neighbor only (Leuven and Sianesi, 2007).

Table 4. Slovak agricultural companies: Selection of matching algorithm Profit/ha. Difference in Differences (DID) of Average Treatment Effects on Treated (ATT): Leuven – Sianesi Method (2007)

Matching algorithms	ATT		DID (2003-2005)	Estimated standardized bias (after matching)	% bias reduction
	2003	2005			
Nearest neighbours N (1)	0.119	-0.001	-0.12	9.6	51%
Nearest neighbours N (5)	-0.314	-0.486	-0.172	10.7	45%
Radius caliper (max distance 0.01)	0.558	-0.463	-1.021	19.75	0%
Kernel gaussian	-0.139	-0.333	-0.194	9.01	54%
Kernel biweight	-0.009	-0.338	-0.329	7.51	61%
Kernel epanechnikov bandwidth 0.01	0.579	-0.446	-1.025	19.61	0%
Kernel epanechnikov bandwidth 0.06	0.0165	-0.345	-0.3615	7.45	62%
Kernel epanechnikov bandwidth 0.1	-0.185	-0.316	-0.131	9.06	54%
Kernel epanechnikov bandwidth 0.2	-0.125	-0.376	-0.251	9.14	53%

Table 5. Selection of Slovak agricultural companies: Results of applying Kernel method (Epanechnikov bw. 0.06) and imposing common support constraints

Treatment assignment	Common support		Total
	Off support	On support	
Untreated	0	181	181
Treated	37	14	51
Total	37	195	232

The balancing property tests (t-test)[34] show that the applied matching procedure (i.e. minimization of the standardized selection bias using kernel epanechnikov matching (bw. 0.06) considerably improved the comparability of both groups of agricultural companies, making a counterfactual analysis more realistic. Indeed, previously existing (i.e. prior to the SAPARD programme) significant differences (measured in terms of the t-test) in variables between the group of agricultural companies supported from the SAPARD programme (D=1) and the group of non-supported farms (D=0) dropped after matching (differences became no more statistically significant). This applies to all important variables determining both programme participation and outcomes, e.g. profit per company (prior to the SAPARD programme), liabilities, value of current assets, etc. Results of performed balancing property tests are shown in Table 6.

34 According to some authors, conventional t-tests are fallacious, see: Imai K, G. King and E. A. Stuart, 2006.

Table 6. Slovakia: Variables' balancing property test between selected programme supported and non-supported agricultural companies (common support imposed; matching algorithm: kernel epanechnikov bw 0.06)

		Mean		% Reduction		t-test	
Variable	**Sample**	**Treated**	**Control**	**%bias**	**bias**	**t**	**p> t**
v2b364y03	Unmatched	-880.9	-3338.6	45.4		2.66	0.008
	Matched	-1705	-1264.7	-8.1	82.1	-0.33	0.747
vd37a23y03	Unmatched	915.37	437.93	50.6		3.92	0.000
	Matched	419.07	415.75	0.4	99.3	0.02	0.982
vh72y03	Unmatched	16566	837.97	23.8		2.27	0.024
	Matched	968.21	1405.6	-0.7	97.2	-0.57	0.575
vk61y03	Unmatched	376.65	17192	-10.6		-0.53	0.594
	Matched	301.44	1670.1	-0.9	91.9	-0.07	0.941
vf331y03	Unmatched	632.78	450.09	15.0		1.15	0.253
	Matched	200.36	261.16	-5.0	66.7	-0.48	0.636
vd37a20y03	Unmatched	47.348	97.28	-31.4		-1.84	0.067
	Matched	88.409	84.811	2.3	92.8	0.06	0.955
v1b541y03	Unmatched	97.431	475.04	-27.4		-1.46	0.146
	Matched	-6.3571	73.567	-5.8	78.8	-0.51	0.614
vc39y03	Unmatched	3.0784	4.9227	-21.0		-1.10	0.273
	Matched	1.5714	2.4112	-9.6	54.5	-0.51	0.612
vk69y03	Unmatched	16.363	8.2929	19.1		1.18	0.240
	Matched	7.5964	2.0118	13.2	30.8	0.63	0.532
vk85y03	Unmatched	49.065	46.904	2.0		0.13	0.897
	Matched	32.857	25.141	7.3	-257.0	0.36	0.724
vk665y03	Unmatched	116.43	50.691	15.1		1.24	0.215
	Matched	66.091	109.32	-9.9	34.2	-0.38	0.707
vf323y03	Unmatched	253.1	219.91	8.0		0.49	0.622
	Matched	131.07	158.77	-6.7	16.5	-0.36	0.721
vf335y03	Unmatched	78538	52723	46.7		3.32	0.001
	Matched	54635	47188	13.5	71.1	0.54	0.595
v1b52y03	Unmatched	.54902	16.901	-16.2		-0.82	0.415
	Matched	.71429	1.4052	-0.7	95.8	-0.06	0.953
v2b340y03	Unmatched	168.78	53.249	41.3		3.41	0.001
	Matched	59.786	33.8	9.3	77.5	0.87	0.390
v2b31y03	Unmatched	1845.7	1954.5	-1.8		-0.10	0.920
	Matched	1008.9	611.53	6.5	-265.3	0.32	0.749
vf311y03	Unmatched	3967.3	2227.8	36.6		2.48	0.014
	Matched	2203.4	1393.9	17.0	53.5	0.89	0.384
vk865y03	Unmatched	1522.7	1249.7	12.1		0.84	0.401
	Matched	1183.4	1022.9	7.1	41.2	0.24	0.809
vd37a4y03	Unmatched	7.3876	8.8968	-6.3		-0.36	0.716
	Matched	2.4307	1.1453	5.4	14.8	0.42	0.675
vd37a6y03	Unmatched	62.847	57.025	5.1		0.31	0.757
	Matched	56.791	35.997	18.2	-257.2	0.76	0.454
v3b66y03	Unmatched	619.8	733	-11.7		-0.79	0.432
	Matched	493.18	474.28	1.9	83.3	0.09	0.929
vf35y03	Unmatched	174.25	114.41	17.4		1.27	0.205
	Matched	191.43	101.49	26.2	-50.3	0.59	0.563
vc32y03	Unmatched	.52941	.54144	-2.4		-0.15	0.880
	Matched	.57143	.50006	14.2	-493.5	0.36	0.723
vk813y03	Unmatched	115.66	85.037	9.0		0.63	0.531
	Matched	59.897	43.774	4.8	47.0	0.24	0.810
v1b586y03	Unmatched	28991	21514	24.9		1.86	0.064
	Matched	21105	19280	6.1	75.6	0.29	0.771

v2b325y03	Unmatched	9643.1	8049.8	22.4		1.52	0.130
	Matched	6506.1	7037.4	-7.5	66.7	-0.36	0.723
vb2a3111y03	Unmatched	410.9	352.56	10.2		0.64	0.520
	Matched	393.43	353.86	6.9	32.2	0.21	0.832
vk873y03	Unmatched	17.239	46.913	-20.0		-1.03	0.304
	Matched	36.708	47.799	-7.5	62.6	-0.32	0.752
vc37y03	Unmatched	18.431	14.337	29.2		1.92	0.057
	Matched	13.643	13.581	0.4	98.5	0.02	0.987
v2b324y03	Unmatched	145.22	162.58	-2.6		-0.17	0.867
	Matched	17.643	13.153	0.7	74.1	0.08	0.935

10.5. Result indicators

Generally speaking, the assessment of the micro-economic effects of a given RD programme can be carried out using various farm-specific economic coefficients as result indicators.

In our study, we selected seven relevant result indicators available from a standard FADN system:

- Profit per company;
- Profit per ha;
- Profit per person employed;
- Gross value added per company;
- Employment per company;
- Labour productivity (Gross value added per employed);
- Land productivity (Gross value added per ha);

In order to measure the effect of the SAPARD programme on agricultural companies which received support from the SAPARD programme (Measure 1) the ATT (Average Treatment on the Treated) coefficients were estimated for each result indicator separately at two data points: before the programme (year 2003) and after the programme (2005)[35]. The outcomes of ATT estimations for individual result indicators are shown in Tab 7 (below)[36].

10.6. Assessment of programme results

The assessment of results of SAPARD (Measure 1) on profit, employment, gross value added, etc. of agricultural companies that were supported by the programme was carried out by applying the conditional DID method (i.e. combination of a binary PSM method and DID technique) to ATT parameters calculated for respective result indicators (a-g) before and after the programme (years 2003-2005).

10.6.1. Estimation of SAPARD's impact using a traditional approach

Evaluation studies which employed naïve or traditional techniques for an estimation of

35 Unfortunately, the estimation of ATT in consecutive years (e.g. 2006, 2007) was not possible due to a growing fluctuation in the data base (dropping of many agricultural companies from the balanced panel).

36 Estimated values of ATT in years 2001, 2003 and 2005 were used for calculation of SAPARD results separately for each outcome indicator. The ATT-DID estimator measures the impact of the RD programme by comparing the differences between programme participants and non-participants before (i.e. years 2001 and 2003) and after (i.e. 2005) situations. Specifically, the difference "one" was the difference in mean outcomes between the programme beneficiaries and the matched controls after implementation of the RD programme (T_1), the difference "two" was the difference in mean outcomes between beneficiaries and matched controls at date T_0 (prior to the RD programme) and the difference "three" was the difference between difference "one" and difference "two".

programme effects found a very high impact of SAPARD on the performance of agricultural enterprises[37]. In a traditional (naive) approach to evaluation, the effects of a given programme are often calculated using data on supported companies before and after the programme. Application of this approach in the context of the SAPARD programme would indicate that the effect of SAPARD was very positive (e.g. an increase in profits per company from -800 thousand SKK in 2003 to 1589 thousand in 2005, i.e. a gain of 2496 thousand SKK; an increase in profits per person employed from -11.3 thousand SKK to 23.6 thousand SKK, i.e. a gain of 34.6 thousand SKK; or an increase in profits per ha +1.336 thousand SKK).[38] Yet, since these approaches completely ignore possible effects of other confounding factors (exogenous to the SAPARD programme) they are certainly not reliable.

Indeed, already a simple comparison with non-participants (1-0) or a country average (1-Ø) as control groups would in the case of a result indicator: "profits per company" or "profit per person" lead to completely different results, i.e. would indicate only a slightly positive or almost negligible effect of the SAPARD programme (3.5-4 thousand SKK per person or -30 thousand SKK per agricultural company). Yet, due to a significant selection bias involved even these calculations would be problematic.

37 Ex-post evaluation of the SAPARD programme in the Slovak Republic. Final Report. P.C.M. Group. December 2007.

38 Results of other evaluation studies that used traditional approach were even more peculiar. For example, the answer of "traditional" evaluators on one of many CEQ, e.g. "To what extent have supported investments contributed to improvement of the income of beneficiary farmers"? was as follows: "The average salary in the agricultural sector increased from 10958 SKK in 2003 to 13 340 SKK in 2006, i.e. nominally by 30.9%. ... It is logical to assume that the growth of income of beneficiaries was at least the same, more than likely even higher". Following, "the impact of the implementation of Measure 1 Investments in agricultural enterprises was … excellent" (PCM, 2007).

10.6.2. PSM-DID approach

The PSM-DID approach largely eliminates selection bias, thus making comparisons between supported and control groups more reliable. The PSM-DID results (presented in Table 7) show clearly that the impact of the SAPARD programme (Measure 1) on total profits per company and value added of supported agricultural companies was very different from those estimated by the ex-post SAPARD evaluators using naive approach (PCM, 2007). Indeed, our estimates show that profits per company in the matched **non-supported** group of similar agricultural companies increased from -1264 in 2003 to 815 thousand SKK in 2005 (that is by +2079 thousand SKK), i.e. they grew faster than in the matched **supported group** (+1836 thousand SKK). Subsequently, the effect of SAPARD (measured in terms of this result indicator) was found to be **either negative** (-243 thousand SKK, profit per company or -346 SKK per ha) or **close to zero** (profit per person employed).

Similar effects were found by applying other result indicators: i.e. gross value added per company, gross value added per employed person and gross value added per ha. Indeed, while gross value added per company, GVA per person employed and GVA per ha in the matched *non-supported* agricultural companies *increased* between 2003 and 2005 in *supported* companies they *either decreased* (e.g. GVA per company, labour productivity) or *increased at a lower rate* compared with matched non-beneficiaries (e.g. land productivity). As a consequence the estimated effect of SAPARD on the above result indicators in the examined period was found to be either almost zero (GVA per employed) or negative (GVA per company and GVA per ha).

Concerning the impact of SAPARD (Measure 1) on farm employment (see Table 9), we found that, contrary to some expectations, the total employment in the group of supported agricultural companies remained at the same level over the period of 2003-2005, i.e. 53 persons per company (employment in analysed

Table 7. Slovakia: Effect of SAPARD (Measure 1) on supported agricultural companies using profit as result indicator (PSM-DID method)

	Profit/company Tsd. SKK			Profit/person employed			Profit/ha		
	2003	2005	DID (2005-003)	2003	2005	DID (2003-2005)	2003	2005	DID (2003-2005)
Participants (1)	-880	1589	+2496	-11	23.6	34.6	-0.456	0.91	1.366
Non-participants (0)	-3338	-839	+2499	-42	-11	31	-2.32	-0.51	1.81
Country Average Ø	-2798	-305	+2493	-35	-3.9	31.1	-1.91	-0.19	1.72
Difference (1-0)	2458	2428	-30	31	35	4	1.86	1.42	-0.44
Difference (1- Ø)	1918	1894	-24	24	27.5	3.5	1.454	1.1	-0.354
Matched participants (1)	-1705	131	+ 1836	-30.8	7.8	38.6	-1.112	0.21	1.322
Matched control group (0)	-1264	815	+ 2079	-19.3	19	38.3	-1.128	0.54	1.668
ATT	-440	-683	-243	-11.5	-11.2	0.3	0.016	-0.33	0.346

Table 8. Slovakia: Effect of SAPARD (Measure 1) on supported agricultural companies using GVA as result indicator (PSM-DID method)

	GVA/company			Labour productivity (GVA/employed)			Land productivity (GVA/ha)		
	2003	2005	DID (2005-003)	2003	2005	DID (2003-2005)	2003	2005	DID (2003-2005)
Participants (1)	17727	18478	751	222	216	-6	10.6	11.5	0.9
Non-participants (0)	9950	9680	-270	130	150	20	7.2	6.6	-0.6
Country Average Ø	11660	11614	-46	151	164	13	7.9	7.7	0.2
Difference (1-0)	7777	8798	1021	92	66	-26	3.4	4.9	1.5
Difference (1- Ø)	6067	6864	797	71	52	-19	2.7	3.8	1.1
Matched participants (1)	11082	9610	-1472	206	147	-59	7.31	7.41	0.1
Matched control group (0)	9367	9701	334	164	168	4	6.85	7.13	0.28
ATT	1715	-90	-1805	41.4	-21.3	-62.7	0.46	0.28	0.18

companies did not decrease), whereas in comparable non-supported companies it dropped slightly (from 59 to 56 per company). As a consequence, the estimated effect of SAPARD on employment was found to be slightly positive.

Table 9. *Slovakia: Effect of SAPARD (Measure 1) on supported agricultural companies using employment as result indicator (PSM-DID method)*

Calculation basis	Employment total (per company)		
	2003	2005	D I D (2005-2003)
Unmatched P=1	85	82	-3
Unmatched P=0	68	57	-11
Average Ø	84	62	-22
Difference (1-0)	17	25	8
Difference (1- Ø)	1	20	19
Matched M= 1	53	53	0
Matched M= 0	59	56	-3
ATT	-5.53	-3.32	2.21

Sensitivity analysis

A sensitivity analysis was carried out using the Rosenbaum bounding approach methodology described in Chapter 5.6. The results show that estimated effects of the SAPARD programme are rather sensitive concerning the presence of unobservables (hidden bias). For example, in the case of the estimated effect of the SAPARD programme on labour productivity the sensitivity analysis shows that an increase of the odds ratio due to a hidden bias from 1 to 1.05 (by 5%) would make the obtained results statistically insignificant. The relatively high sensitivity of obtained results may originate from a small number of observations available in an existing data base. Still, even in the presence of high sensitivity, the results obtained using the PSM-DID method are valid.

Table 10. Rosenbaum bounds for labour productivity (2003) (N = 14 matched pairs)

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	.099061	.099061	34.8851	34.8851	-11.4305	91.8209
1.05	.113737	.085702	31.997	37.9813	-13.1665	92.5384
1.1	.128973	.074159	31.0288	38.8525	-13.9887	101.065
1.15	.144672	.064185	29.4552	39.555	-19.4491	101.802
1.2	.160743	.055566	29.1089	40.2927	-21.9205	105.29
1.25	.177104	.048115	28.5916	42.2066	-22.9366	113.308
1.3	.193678	.041673	28.2868	43.7802	-23.6743	114.352
1.35	.210399	0.36102	27.4646	46.8034	-25.408	115.78
1.4	.227204	.31283	26.1201	47.3787	-26.1457	115.926
1.45	.244039	.027114	18.1015	47.6255	-26.1457	115.926
1.5	.260856	.023506	16.3578	47.6255	-29.9391	117.5
1.55	.277611	.020382	14.614	48.1164	-32.2005	123.75
1.6	.294268	0.17677	13.8763	50.2252	-32.2005	123.75
1.65	.310794	.015335	12.963	50.5136	-33.4266	125.323
1.7	.327159	.013305	10.4916	51.6039	-34.1643	133.147
1.75	.34334	.011547	10.4916	51.6039	-34.1643	133.147
1.8	.359315	.010023	10.0749	51.7989	-34.6719	133.992

1.85	.375067	.008701	9.25277	52.6966	-34.6719	133.992
1.9	.390579	.007555	9.25277	52.6966	-36.9142	134.814
1.95	.405839	.006562	7.60353	54.2703	-36.9142	134.814
2	.420837	.0057	7.60353	59.6225	-37.6519	137.702
2.05	.435564	.004952	6.78136	59.6225	-37.6519	137.702
2.1	.450013	.004303	5.35012	62.0939	-38.3895	153.33
2.15	.46418	.003739	5.35012	62.0939	-38.3895	153.33
2.2	.478061	.00325	2.47301	66.1422	-42.6905	177.435
2.25	.491653	.002825	2.473	66.1422	-42.6905	177.435
2.3	.504954	.002456	-.415091	70.9084	-42.6905	177.435
2.35	.517965	.002136	-.415093	70.9084	-46.178	179.009
2.4	.530685	.001858	-1.01453	71.7306	-46.178	179.009
2.45	.543116	.001616	-1.01454	71.7306	-46.178	179.009
2.5	.55526	.001406	-1.1259	72.1063	-46.9157	186.833
2.55	.567118	.001223	-1.23726	72.4821	-46.9157	186.833
2.6	.578694	.001064	-1.23726	72.4821	-46.9157	186.833
2.65	.589992	.000926	-1.75221	73.3042	-55.4419	240.519
2.7	.601013	.000806	-1.75222	73.3042	-55.4419	240.519
2.75	.611763	.000701	-2.82742	73.9615	-55.4419	240.519
2.8	.622246	.000611	-3.90263	74.6187	-55.4419	240.519
2.85	.632465	.000532	-3.90263	74.6187	-99	99
2.9	.642426	.000463	-4.64031	76.1923	-99	99
2.95	.652132	.000403	-4.64031	76.1923	-99	99
3	.66159	.000351	-4.64031	76.1923	-99	99

** gamma - log odds of differential assignment due to unobserved factors*
sig+ - upper bound significance level
sig- - lower bound significance level
t-hat+ - upper bound Hodges-Lehmann point estimate
t-hat- - lower bound Hodges-Lehmann point estimate
CI+ - upper bound confidence interval (a= .95)
CI- - lower bound confidence interval (a= .95).

10.6.3. Possible explanations of obtained results

The possible explanations for the PSM-DID results are as follows:

- The period covered in the analysis of programme impact could be too short to allow for a full unfolding of effects of an investment process supported under the SAPARD programme. Yet, the deterioration of a stability of the balanced panel (the drop of many enterprises from the data base after inclusion of a new year) prevented estimation of programme effects for subsequent years (after 2006).

- Effects of the SAPARD programme on the big agricultural companies in Slovakia were indeed less encouraging than expected, mainly due to a progressive admission of less economically viable agricultural companies to the programme (during implementation of the SAPARD programme the official eligibility criteria (participation criteria) were adjusted several times to enable larger but less efficient agricultural companies to benefit from available EU subsidies, see: 10.1).

- Clearly the PSM-DID results show that traditional estimates of programme effects can be highly misleading, and that the application of advanced evaluation methodologies leads usually to much more reliable results. Yet, if the available data base is weak (e.g. a low number of observations and a high instability of balanced panel data) even a more sophisticated approach cannot provide all answers to relevant evaluation questions.

The main problems that occurred in applying a PSM-DID method to a rather weak data on agricultural enterprises in Slovakia appear to be as follows:

Firstly, in case of a low number of observations, setting common support conditions reduces the estimation of programme effects to a relatively small number of agricultural companies that received support from the SAPARD (included in our data base). Yet, a relatively narrow common support region can create some problems if obtained results are to be extrapolated to the whole population of supported enterprises,

Secondly, analysis of slowly unfolding impacts, i.e. estimation of ATT in consecutive years after investment (e.g. 2006, 2007) was not possible due to progressing fluctuations in the available data base (a dropping of many agricultural companies from the balanced panel).

Thirdly, a weak data base (small number of observations in panel FADN data base) also hindered the estimation of programme impacts involving general equilibrium effects (e.g. substitution or displacement effects).

All of the methodological problems above could be addressed in a satisfactory manner in our next example: the AFP programme in Schleswig-Holstein, Germany.

11. Estimation of programme direct and indirect effects in the example of the AFP programme in Schleswig-Holstein (Germany)

11.1. Description of the programme

The main objective of the Agrarinvestitionsförderungsprogramm (AFP) implemented in the region of Schleswig-Holstein (Germany) during the years 2000-2006 was to improve the structure and competitiveness of the agricultural sector through financial support provided for the modernisation of agricultural enterprises. The main mechanism of the AFP programme was the subsidy to a commercial interest rate paid by eligible agricultural enterprises for a loan on investment activities (total investment volume was allowed to vary between 175 000 EUR and 500 000 EUR) carried out mainly in the milk and beef, pork, and agro-tourism sectors. The subsidy to a commercial interest rate (approximately 13% of eligible investment volume) was provided to eligible individual farms for the period of 10 to 20 years on an average amount of 23 000-30 000 EUR/farm. During the years 2000-2006 total subsidies provided under AFP programme reached approximately 29.7 Mill EUR. During the period of 2000-2006 1513 farms received support from the AFP programme (net investment volume of 250 Mill EUR). The biggest part of the programme budget (approximately 80%) was provided for farm inventory (buildings) investment support, mainly in the milk and beef sectors. The rest was split up for investment support (including purchases of machinery or investments in alternative sources of energy) among the pork sector, the agro-tourism sector and the horticulture sector. Specific eligibility criteria, such as investment volume higher than 175 000 EUR, eligible personal income up to 90 000 EUR per person or 120 000 EUR per couple, excluded the smallest and the biggest agricultural farms from this programme.

11.2. Data

The main data source used for the assessment of the effects of the AFP programme in Schleswig-Holstein was farm bookkeeping data comprised of approximately 10 500 farms for the year 2000/2001 and 3 900 farms for the year 2007/2008). Furthermore, for specific comparisons approximately 400 datasets from "Testbetriebe" (part of FADN data set) were used.

Since the main focus of the AFP programme's support was the milk and beef sector 1333 bookkeeping farms specializing in milk/beef production were selected from the available data set and included in a panel for further analysis. The balanced panel (years 2001-2007) consisted of 101 milk/beef farms supported by the AFP programme and 1232 non-supported farms.

11.3. Major methodological assumptions of PSM-DID and the steps to be carried out

The main objective of our analysis was to estimate the micro-economic results of the AFP programme implemented during the years 2000-2006 on milk/beef farms located in Schleswig-Holstein, and especially to estimate selected direct and indirect effects of the programme (e.g. deadweight loss, substitution and leverage effects). As the main methodological approach a combination of PSM-DID methods was chosen. While the PSM-DID method is particularly useful for estimation of effects of a given RD programme at farm level, the applicability of a standard PSM method (based on estimation of the logit function) necessitates an assumption regarding the absence of general equilibrium effects. In other words, standard PSM estimates are only valid under an assumption of *no indirect effects* of a given RD programme on programme *non-beneficiaries*. In practice, an absence of indirect effects on non-beneficiaries has to be verified first in order to validate obtained results.

Given the above, the empirical assessment of the effect of the AFP programme at a micro-level involved the following stages:

Stage 1: Preliminary estimation of direct programme effects occurring at the level of direct programme beneficiaries (direct effect of the programme on Gross Value Added, profits or employment at a micro-level)

Stage 2: Preliminary estimation of specific indirect effects (e.g. deadweight loss and leverage effects) at the level of direct programme beneficiaries

Stage 3: Estimation of general equilibrium effects (e.g. substitution effects and replacement effects) at the level of programme non-beneficiaries

Stage 4: Re-estimation of Stage 1 and Stage 2 in case of a presence of general equilibrium effects (e.g. substitution effects) by dropping from further analysis all "programme affected non-beneficiaries"

11.4. Implementation of Stages 1-4

11.4.1. Preliminary estimation of direct programme effects occurring at the level of direct programme beneficiaries (Stage 1)

Here the following steps were carried out:

1. Using information about general- and measure-specific conditions for programme participation, potential programme eligible farms/enterprises (Measure 1: "Modernisation and Restructuring of Agricultural Enterprises") were identified and selected from the available data base (e.g. bookkeeping or FADN data)

2. The above group of farms was divided into beneficiaries vs. non-programme beneficiaries. A balanced panel for both sub-groups (direct programme beneficiaries vs. non-beneficiaries) was constructed for years 2000 (i.e. prior to the implementation of the programme) and 2007 i.e. (after the programme)

3. On the basis of expert knowledge the most important variables determining both economic outcomes, as well as the decision of farms specialized in milk/beef production to participate in the AFP programme, were selected from the list of variables/coefficients available in bookkeeping (or FADN) data set. Selected variables were included in the list of covariates in the estimated logit function.

4. Given information on GVA per enterprise, profits and other important farm characteristics (e.g. land area, employment, value of assets, etc) prior to the programme (T=0) a PSM matching method was applied in order to construct appropriate controls.

5. The selection of a relevant matching technique was carried out using three independent criteria: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); and iii) joint significance and pseudo R^2 (Sianesi, 2004), see: methodology described in the section: 10.4.

6. The "similarity" of both groups prior to their participation in the programme was verified statistically (e.g. by performing balancing property tests involving selected covariates)

7. Specific policy indicators, e.g. Average Treatment Effects on Treated (ATT) were estimated before the programme (T=0)

and after the programme (T=1), using GVA per enterprise, profit per employee, etc. as relevant result indicators

8. Conditional DID method (combination of ATT and standard DID) was applied to calculate the net effects of the RD programme (at micro-level)

9. Sensitivity analysis of obtained results was performed using Rosenbaum bounds.

11.4.1.1. Results of Stage 1 (preliminary results):

The preliminary results of the application of the PSM method (DID-ATT) to the evaluation of the RD Agrarinvestitionförderungsprogramm (AFP) in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants) specialized in milk production (panel for years 2001-2007) using profits as outcome indicator are presented in Table 11.

Table 11. Estimated effect of AFP programme on milk farms (Schleswig-Holstein, Germany) – preliminary results using PSM-DID methodology

Calculation basis	Profit per farm in EUR		
	2001 Prior to participation	2007 After implementation	D I D (2007-2001)
Unmatched AFP participants P=1 (101)	54,629	116,777	62,148
Unmatched Non-participants P=0 (1,232)	43,007	83,718	40,711
Ø (1,333)	43,888	86,222	42,334
Difference (1-0)	11,621	33,059	21,438
Difference (1- Ø)	10,741	30,555	19,814
Matched AFP participants M= 1 (101)	54,629	116,777	62,148
Matched Non-participants M= 0 (1,067)	55,266	106,752	51,486
ATT	-637	10,024	10,661

The application of the PSM (ATT-DID) method (given constraints on the common support region) resulted in the dropping of 165 (non-comparable) non-beneficiaries from further analysis. Preliminary results obtained on the basis of PSM-DID methodology showed positive effects of the AFP programme on farm profit (+10,661 EUR). While these effects were found to be much smaller compared with effects obtained using traditional methods[39] they are valid only in the absence of significant general equilibrium effects (e.g. substitution effects). The respective verification was carried out at Stage 3.

11.4.2. Estimation of specific indirect effects (e.g. deadweight loss and leverage effects) at the level of direct programme beneficiaries (Stage 2)

11.4.2.1. Estimation of a deadweight loss (preliminary results)

A deadweight loss effect occurs if a participant of a RD programme would also have undertaken a similar investment without the RD programme

39 Calculated programme effects using a comparison of programme beneficiaries with all (unmatched) non-beneficiaries would be +21 438 EUR, or +19 814 EUR if programme beneficiaries were compared to a country's average (see: Table 11).

Table 12. *Estimated deadweight loss effect of AFP programme on milk farms (Schleswig-Holstein, Germany)*

Calculation basis	Value of inventories in EUR		
	2001	2007	DID (2001-2007)
Participants (P=1) (83)	80,058	153,545	73,487
Non-participants (P=0) (293)	57,379	108,539	51,160
Matched participants (M=1) (83)	80,058	153,545	73,487 (+92%)
Matched non-participants (M=0) (263)	70,181	130,733	60,552 (+86%)
Deadweight loss (M)			93% = (86/92)

support (i.e. RD support would not change investment behaviour of a targeted enterprise). Deadweight loss effects can be measured by comparing performance of programme beneficiaries with respective controls and applying a relevant result indicator (e.g. investment value per farm/enterprise and year; or value of inventories per farm/enterprise and year) for calculations of the ATT prior to and after the programme.

Estimation of deadweight loss at the level of direct programme beneficiaries was carried out in the following steps:

- Identification of units/farms supported from the AFP programme carrying out investments under a specific RD measure (e.g. Measure 1: Modernisation and Restructuring of Agricultural Enterprises);
- Identification in the control group (i.e. similar programme non-participants) of a sub-vector of those farms which undertook similar investments as programme beneficiaries (in period between T=0 and T=1);
- Calculation of ATT using data from both groups and applying a selected result indicator (e.g. investment value per farms) before and after the programme;
- Applying DID on the estimated ATT;

While it is expected that in case of a deadweight loss the calculated DID-ATT between the above groups will be close to zero, the estimated percentage of deadweight loss (between 0% and 100%) should be used to correct the estimates of direct programme effects.

The above methodology was applied to the estimation of the deadweight loss effects of the AFP programme (Measure: Investments in milk and beef sectors) using data on 376 selected bookkeeping farms (83 AFP participants and 293 non-participants) specialized in milk production (panel for years 2001-2007) that undertook similar investments in the examined period. Our results show that, even without any support from the AFP programme, the value of inventories in the matched (control) group of non-beneficiaries (263 farms) increased in the examined period by 86% compared with the base period (prior to the programme) see: Table 12. While at the same time the value of inventories in the group of programme beneficiaries (83 farms) increased by 92% the estimated deadweight loss effects were as high as 93% (ratio of 86/92). This means that a huge portion of supported investment (i.e. 93%) would have taken place **even without** the AFP programme, probably due to very favourable changes in economic conditions for dairy farmers (i.e. significant increase of price for milk).

11.4.2.2. Estimation of leverage effects (preliminary results)

The leverage effect can be considered an important micro-economic consequence of RD support. It occurs if public funding (e.g. in form of a RD programme) induces private spending among the programme beneficiaries.

Calculation of the leverage effect was carried out by taking the following steps:

- selection of individual units j supported by a RD programme;
- identification of a comparison/control group matching with units j (identical distribution of covariates) in the period T=0 (i.e. prior to j's access to the programme) using PSM method;
- selection of relevant result indicators as proxies for private spending, e.g. money transfers from farm to farm households; level of private and farm consumption, etc.;
- calculation of ATT for selected result indicators between both groups (i.e. j and m);
- Applying DID on the estimated ATT;

It is expected that in case of a significant leverage effects the calculated DID-ATT will be positive and significant.

The application of the above methodology for the estimation of the leverage effects in the AFP programme in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants)

Table 13a. Estimation of the leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Money transfer from farm to farm household for living

Calculation basis	Variable: Money transfer from farm to farm households for living		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	30,072	43,810	13,738
Unmatched P=0 (1,232)	24,512	32,336	7,824
Ø (1,333)	24,933	33,206	8,273
Difference (1-0)	5,560	11,473	5,913
Difference (1- Ø)	5,139	10,604	5,465
Matched M= 1 (101)	30,072	43,810	13,738
Matched M= 0 (1,067)	27,647	36,732	9,085
ATT	2,424	7,077	4,653

Table 13b. Estimation of the leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Money transfer from farm for building of private assets

Calculation basis	Money transfers from farm for building of private assets		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	18,447	48,302	29,855
Unmatched P=0 (1,232)	11,632	31,926	20,294
Ø (1,333)	12,148	33,167	21,019
Difference (1-0)	6,814	16,376	9,562
Difference (1- Ø)	6,299	15,135	8,836
Matched M= 1 (101)	18,447	48,302	29,855
Matched M= 0 (1,067)	17,504	44,181	26,677
ATT	942	4,120	3,178
ATNT	1,865	1,781	-84
ATE	1,785	1,983	198

Table 13c. *Estimation of the leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Total money transfer from farm to farm household*

Calculation basis	Total money transfers from farm to farm household		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	75,415	137,886	62,471
Unmatched P=0 (1,232)	61,393	99,493	38,100
Ø (1,333)	62,455	102,402	39,947
Difference (1-0)	14,022	38,392	24,370
Difference (1- Ø)	12,960	35,484	22,524
Matched M= 1 (101)	75,415	137,886	62,471
Matched M= 0 (1,067)	76,181	124,100	47,919
ATT	-765	13,785	14,550
ATNT	-3,016	8,460	11,476
ATE	-2,821	8,920	11,741

Table 14. *Effects in AFP programme (Schleswig-Holstein). Result indicator: Total money transfers to farm ("Einlagen insgesamt")*

Calculation basis	Total money transfers to farm		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	25,604	46,426	21,362
Unmatched P=0 (1,232)	22,812	36,069	13,257
Ø (1,333)	23,024	36,853	13,829
Difference (1-0)	2,791	10,357	7,566
Difference (1- Ø)	2,580	9,573	6,993
Matched M= 1 (101)	25,604	46,426	21,362
Matched M= 0 (1,067)	26,823	46,036	17,413
ATT	-1,218	389	1,607
ATNT	-1,352	821	2,173
ATE	-1,341	783	2,124

specialized in milk production (panel for years 2001-2007) indicates considerable leverage effects. Indeed, investment support through the AFP programme brought about significant additional transfers of funds from farms to households (e.g. additional money transfers from farm to farm households for living on average EUR +4,653 per farm, see: Tab 13a; additional money transfers from farm to households for building of private assets on average EUR +3,178 per farm, see Tab 13b; additional total money transfers from farms to farm households on average EUR +14,550 per farm, see Tab 13c). The above figures show that the propensity to consume among farms that received support from the AFP programme was much higher compared with similar coefficient calculated for programme non-beneficiaries (i.e. the programme leverage effect was substantial).

While a high proportion of additional transfers from farms to farm households among programme beneficiaries could originate from higher farm profits, it appears that investment support induced farms' private spending much more strongly than the building of deposits ("Einlagen"). Indeed, total transfers from farms to farm households increased in the examined period on average by EUR 14,550 while in the same period total money transfers to farms (farm deposits) grew by only EUR 1,607 (see Table 14).

11.4.3. Estimation of specific indirect programme effects on non supported farms/enterprises (Stage 3)

General equilibrium (GE) effects occur when a given programme affects (positively or negatively) farms/enterprises other than direct programme participants. Important GE effects are substitution effect and displacement effect. The major methodological problems linked to the estimation of GE effects are discussed in Chapter 6.2.1.

11.4.3.1. Estimation of programme substitution effects

The substitution effect belongs to the indirect general equilibrium or macro-economic effects of a given programme. It is normally defined as the effect occurring in favour of direct programme beneficiaries but at the expense of persons/farms/units that do not participate in a given intervention. For example, due to a given RD programme input/factor prices in an affected region may increase; or regional produce prices may decrease compared with other regions (e.g. where the programme was not implemented or implementation intensity was low) which may finally affect profits/employment/gross value added etc. of farms which were not direct programme beneficiaries. The substitution effect (in contrast to the displacement effect) occurs primarily in a direct neighbourhood of units supported by a given programme. It can be expected that this effect will have an impact on all major programme result indicators, e.g. GVA per enterprise.

Learning about substitution effects of RD programmes is important for particular reasons:

- It facilitates the assessment of the net effectiveness of a given RD programme;
- In case of using PSM methodology, it provides additional information on the validity of preliminary results calculated at the level of direct programme beneficiaries (see: Chapter 11.2).

Generally, substitution effects can be measured using similar techniques as in the case with direct programme effects (i.e. by applying PSM-DID methodology). Yet, the basic difference in comparison with standard PSM is the necessity to redefine the "treatment" by using one of two alternative approaches defined in Chapter 6.2.1.

In case of the AFP programme (Schleswig-Holstein) Approach 2 (see: 6.2.1.) was chosen as it allows the estimation of the indirect impact of the programme on other (similar) farms located in a close neighbourhood of programme beneficiary farms. As the intensity of the AFP programme was the highest in two neighbouring sub-regions of Schleswig-Holstein (i.e. Nordfriesland (NF) and Schleswig-Flensburg (S-F)) it was assumed that in these two regions the probability of positive/negative **indirect** impact of the programme on programme non-beneficiaries was also the highest. The basic idea behind this approach was therefore to compare performance (e.g. profits, GVA, employment, etc.) of programme **non-beneficiaries** in regions where intensity of a given programme exposure was **high** (high probability of positive/negative effects from a given programme; P=1) with the performance of **similar** programme **non-beneficiaries** in other regions characterised by a **low** programme intensity (P=0). A high difference in the estimated ATT-DID between both groups should indicate the existence of substitution effects. No difference in calculated ATT-DID for non-participants in both regions would indicate the absence of substitution effects.

The approach was implemented in the following steps:

- Disregarding all programme participants;
- Performing PSM analysis by computing ATT for "seemingly affected" (non-participants) in the high intensity regions NF and S-F (P=1) versus non-affected (non-participants) in other regions (P=0), whereby the economic performance of non-participants in NF and

S-F regions can be described as a result of a "non-intended selection to programme" implemented at a given area;

- Carrying out all other steps as in standard PSM analysis (e.g. selection of matching method, testing similarity between matched and controls, sensitivity analysis, etc.);

- Calculation of ATTs and DID-ATT, whereby (i.e. depending on obtained results);

  – if the estimated DID-ATT is low or zero, this implies no significant general equilibrium effects (e.g. substitution effects). It also means that preliminary results of a standard PSM method are valid, or

  – if the estimated DID-ATT is high, this suggests a presence of significant general equilibrium effects (substitution effects) in regions where the programme intensity was the highest (NF and S-F). This means also that preliminary results of PSM applied under the Stage 1 are biased (especially for the year 2007!). Should this happen "affected non-participants" in respective regions (NF and S-F) would have to be dropped from further analysis and the DID-ATT should be re-estimated again**.**

The above methodology was applied to the estimation of the substitution effects in the AFP programme in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors). An analysis of substitution effects was carried out on the basis of bookkeeping data collected for 1,231 programme **non-beneficiaries** specialized in milk production (balanced panel for years 2001-2007), of which: 526 were located in regions with the highest exposure to the AFP programme (NF and S-F;) and 705 were located in other ("non-affected") regions.

Our results show that profits per farm among programme **non-beneficiaries** located in regions with zero or low intensity of AFP programme increased much stronger (EUR +41,371) in the years 2001-2007 compared with profits per farm in the group of farms (non-beneficiaries) located in the regions where the intensity of the AFP programme was the highest (EUR +37,824, see: Table 15). The estimated substitution effects lead therefore to a deterioration in the economic situation of farms which did not receive programme support (programme non-beneficiaries), i.e. through a reduction of profit by EUR -3,546 per farm on average.

Similar negative substitution effects of the AFP programme affecting non-programme participants located in regions with the highest

Table 15. Estimated substitution effects of AFP programme on milk farms (Schleswig-Holstein, Germany)

Calculation basis	Profit per farm		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (526)	46,349	84,703	38,354
Unmatched P=0 (705)	40,531	83,034	42,503
Ø (1,231)	43,017	83,747	37,398
Difference (1-0)	5,817	1,669	-4,148
Difference (1- Ø)	3,332	956	-2,376
Matched M =1 (517)	45,933	83,757	37,824
Matched M= 0 (677)	48,559	89,930	41,371
ATT	-2,626	-6,172	-3,546
ATNT	4,337	-2,414	-6,751
ATE	1,322	-4,041	-5,363

Table 16. Difference in the leasing price for agricultural land (in EUR per ha) paid by non-beneficiaries of the AFP programme in Schleswig-Holstein (2001-2007)

Regions NF and S-F (high intensity of AFP) (non beneficiaries, N = 517)		Other regions (low intensity of AFP) (non beneficiaries, N = 677)	
2001	2007	2001	2007
12001	11998	12461	11543

programme intensity were found in the cases of the following result indicators: economic corrected profit, milk production, corrected profit per person fully employed (AK), corrected profit per family labour, standard profit per fully employed, and standard profit per family labour.

The negative substitution effects could have occurred due to many factors. One possible explanation is that agricultural farms that were directly supported by the AFP programme considerably increased their demand for specific inputs, e.g. land (pastures or arable land) thus leading to an increase of input (e.g. land) prices. Indeed, while the leasing price for agricultural land remained at the same level in the regions where support from the AFP programme was very intensive it dropped by 7.3% in those regions where the programme was not implemented or the intensity of AFP implementation was low (see: Table 16)[40].

11.4.4. Re-estimation of Stage 1 (due to a presence of significant substitution effects) (Stage 4)

Considerable programme substitution effects imply the presence of a bias in the estimation of programme effects on direct programme beneficiaries (a control group is affected by a given programme). In order to eliminate this bias all programme non-beneficiaries located in regions with the highest programme intensity, i.e. regions NF and S-F (i.e. all programme affected non-beneficiaries) were dropped from further analysis and the results of Stage 1 were re-estimated without these farms.

A new assessment of the effect of the AFP programme on programme beneficiaries (re-estimation of results from Stage 1) was carried out on the basis of remaining 807 observations on bookkeeping farms in Schleswig Holstein (2001-2007) specialized in milk production (all farms that were not supported by the AFP programme but which were located in regions: NF and S-F were dropped from further analysis). The major steps of the further analysis were consistent with those in Stage 1, and included:

- Re-estimation of a logit function using the same covariates as in Stage 1 yet, based on a different number of observations (807 instead of 1333)
- Calculation of individual propensity score for programme beneficiaries and non-beneficiaries
- Imposing restrictions on the common support region (as both ATT, ATNT and ATE indicators were to be computed, comparable units had to be found in both groups)
- Selection of a relevant matching technique. This was carried out using three independent criteria: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); and iii) joint significance and pseudo R^2 (Sianesi, 2004) and applying methodology described in Section 10.4.
- Statistical verification of the “similarity” of both groups prior to their participation in the programme (e.g. by performing balancing property tests on the most important farm characteristics)

40 Obviously, the AFP programme support lead to an increase of economic capacities of these farms that could later afford to pay a higher leasing price for land.

- Calculation of specific policy indicators, e.g. Average Treatment Effects on Treated (ATT) estimated before the programme (T=0) and after the programme (T=1), using GVA per enterprise, profit per employed, etc. as relevant result indicators

- Application of conditional DID method (combination of ATT and standard DID) for calculation of the first component, i.e. the net effect of the RD programme on GVA generated by programme beneficiaries (at micro-level)

- Performing sensitivity analysis of obtained results using Rosenbaum bounds.

11.4.4.1. Re-estimation of a logit function

After cleaning the data base (by dropping from the set of potential controls those agricultural farms which were affected by the AFP programme) the logit function was re-estimated using 807 observations on bookkeeping farms (Schleswig-Holstein) specialized in milk production, of which 101 were programme beneficiaries and 706 programme non-beneficiaries. The list of variables (38) that determine both programme participation and outcomes and were included as relevant covariates is provided in Table 17 (below). Among the variables used to match programme beneficiaries with programme non-beneficiaries an important one was the covariate showing the former level of support obtained from the RD programme previously implemented in Schleswig-Holstein (vsupp). Inclusion of this variable allowed us to increase comparability and to overcome a problem mentioned in many evaluation studies concerning non-existence of non-supported farms (from current and previous RD programmes) in a specific programme area.

Table 17. Schleswig-Holstein: List of variables selected as covariates to estimation of logit function (excluding programme non-beneficiaries in regions with the highest programme exposure)

List of variables	
v1025i2	Value of fixed assets – buildings
v1030i2	Operating facilities (value)
v1031i2	Machinery (value)
v1091i2	Cattle (value)
v1110i2	Inventory stock
v1449i2	Capital stock (value)
v2129i5	Revenues beef/cattle/milk sales
v2705i5	Purchased concentrated feed for cattle
v2799i5	Labour costs (total)
v4116i2	Milk yield (per cow)
v5111i2	Fem. Calves > 0.5 year
v5112i2	Fem. Calves > 0.5 and < 1 year
v5113i2	Fem. Cattle > 1 and < 2 years
v5114i2	Breeding Heifer
v5115i2	Heifer
v5116i2	Milk cows
v5117i2	Suckler cows
v5118i2	Slaughter cows
v5120i2	Male calves > 0.5
v5121i2	Male cattle > 0.5 and < 1 year

v5122i2	Male cattle > 1 and < 1.5 years
v5123i2	Male cattle > 1.5 and < 2 years
v5124i2	Male cattle > 2 years
v5125i2	Breeding bulls
v6104i7	Pasture area
v6119i7	Agricultural area (total)
v7098i3	Non-family labour
v7099i3	Labour total
vmilkprod	Milk production
v8026i2	Excess milk quota
v9001	Equity capital formation
v9003	v9003
v9005	Labour productivity (cattle/beef/milk per total labour)
v9006	Labour productivity (milk per total labour)
profit01	profit01
v9004	Adjusted equity capital formation
profit_co~01	Profit per farm (adjusted)
v8213i2	Earnings from non-self-employment
v2381i5	Interest subsidy to investment
vsupp	Obtained level of support from previous programmes

The results of logit estimation are shown in Table 18.

Table 18. Schleswig-Holstein. Results of estimation of a logit function

Logistic regression **Log likelihood = 243.64496**				**Number of obs = 807** **LR chi2 (40) = 121.30** **Prob > chi2 = 0.0000** **Pseudo R2 = 0.1993**		
Particip	**Coef.**	**Std. Err.**	**z**	**P> Z**	**[95% Conf. Interval]**	
v1025i2_01	2.02e-06	2.35e-06	0.86	0.390	-2.59e-06	6.63e-06
v1030i2_01	-4.51e-06	7.51e-06	-0.60	0.548	-.0000192	.0000102
v1031i2_01	-.0000268	7.17e-06	-3.74	0.000	-.0000408	-.0000127
v1091i2_01	1.97e-06	.0000146	0.13	0.893	-.0000267	.0000306
v1110i2_01	.0000383	.0000487	0.79	0.432	-.0000572	.0001338
v1449i2_01	-2.54e-07	3.65e-07	-0.69	0.488	-9.70e-07	4.63e-07
v2129i5_01	6.66e-06	9.42e-06	0.71	0.480	-.0000118	.0000251
v2705i5_01	.0000454	.0000106	4.28	0.000	.0000246	.0000662
v2799i5_01	.0001077	.0004719	0.23	0.819	-.0008171	.0010326
v4116i2_01	-.0000613	.0002764	-0.22	0.825	-.000603	.0004805
v5111i2_01	.0186913	.0178942	1.04	0.296	-.0163807	.0537632
v5112i2_01	.0118835	.0167657	0.71	0.478	-.0209766	.0447436
v5113i2_01	-.0121226	.0153492	-0.79	0.430	-0422064	.0179613
v5114i2_01	-.0060769	.0137317	-0.44	0.658	-.0329905	.0208366

v5115i2_01	-.0134439	.0618279	-0.22	0.828	-.1346243	.1077365
v5116i2_01	-.0613138	.0338315	-1.81	0.070	-.1276224	.0049947
v5117i2_01	-.016113	.0720671	-0.22	0.823	-.1573618	.1251358
v5118i2_01	-.0048062	.0287148	-0.17	0.867	-.610862	.0514739
v5120i2_01	.0121035	.0156262	0.77	0.439	-.0185234	.0427303
v5121i2_01	.0165394	.0131412	1.26	0.208	-.0092169	.0422956
v5122i2_01	.014429	.013428	1.07	0.283	-.0118895	.0407475
v5123i2_01	.0051632	.0197474	0.26	0.794	-.0335411	.0438675
v5124i2_01	-.285279	.3196748	-0.89	0.372	-.9118302	.3412722
v5125i2_01	.1216614	.1539543	0.79	0.429	-.1800836	.4234063
v6104i7_01	.0072186	.0068231	1.06	0.290	-.0061544	.0205916
v6119i7_01	.0050058	.0079983	0.63	0.531	-.0106706	.0206822
v7098i3_01	-.581429	.4297761	-1.35	0.176	-1.423775	.2609166
v7099i3_01	.3884432	.3904466	0.99	0.320	-.376818	1.153704
vmilkprod_01	7.79e-06	5.58e-06	1.40	0.163	-3.15e-06	.0000187
v8026i2_01	1.93e-06	3.32e-06	0.58	0.562	-4.59e-06	8.44e-06
v9001_01	8.19e-07	1.47e-06	0.56	0.577	-2.06e-06	3.70e-06
v9003_01	-.0001288	.0004732	-0.27	0.786	-.0010563	.0007987
v9005_01	-3.84e-06	.0000143	-0.27	0.787	-.0000318	.0000241
v9006_01	.0005672	.0006534	0.87	0.385	-.0007134	.0018478
profit01	-4.90e-06	8.59e-06	-0.57	0.568	-.0000217	.0000119
v9004_01	2.55e-07	2.98e-06	0.09	0.932	-5.58e-06	6.09e-06
profit_co~01	1.37e-06	5.39e-06	0.25	0.800	-9.20e-06	.0000119
v8212i2_01	-.0005951	.0013484	-0.44	0.659	-.0032378	.0020476
v8213i2_01	.0000249	.000037	0.67	0.500	-.0000476	.0000975
vsupp_01	-1.32e-06	.0000126	-0.10	0.917	-.0000261	.0000234
_cons	-3.443257	2.004407	-1.72	0.086	-7.371823	.4853098

In the next step the results of a logit function estimation were used to derive for all agricultural farms specialized in milk production their individual probability (propensity scores) of participation in the AFP programme (Measure 1: Modernization of agricultural farms).

11.4.4.2. Selection of a matching algorithm

As the quality of a given matching algorithm depends strongly on a data set, the selection of a relevant matching technique was carried out using three independent criteria: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); and iii) joint significance and pseudo R^2 (Sianesi, 2004).

Similarly to the cases of other assessments of programme impact we found that the best results were achieved by using an iterative procedure (e.g. linear search) aimed to minimise the calculated standardized bias[41] (after matching) and applying min{min} as the main selection criterion. In all

41 The standardized bias is the difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin, 1985).

Table 19. Selection of a relevant matching algorithm

Matching method	Matching parameters	Estimated standardized bias (after matching)
Nearest neighbours	N (8)	4.30
	N (9)	3.90
	N (10)	4.02
Caliper	(0.08)	3.76
	(0.07)	Selected (min) => 3.70
	(0.06)	3.95
Kernel normal	bw (0.03)	4.22
	bw (0.04)	3.99
	bw (0.05)	4.13
Kernel biweight		4.65
Kernel epanechnikov	bw (0.10)	3.92
	bw (0.09)	3.76
	bw (0.08)	3.89

Table 20. Schleswig-Holstein: Overview of the matched sample of agricultural farms

Treatment	Common support		Total
	Off support	On support	
Untreated	44	662	706
Treated	2	99	101
Total	46	761	807

considered cases (various matching algorithms)[42] an optimal solution could easily be found due to local/global convexity of the objective function with respect to function parameters under each matching algorithm (e.g. radius magnitude in radius matching; or number of nearest neighbours in nearest neighbour matching). An overview of results obtained using different matching algorithms for the case of re-estimation of effects of the AFP programme in Schleswig-Holstein is provided in Table 19.

The lowest estimated standardized bias (after matching) was found in the case of caliper matching (0.07). This matching algorithm was therefore used in the further work for assessment of the effect of the AFP programme on direct programme beneficiaries[43].

The application of the above procedure and the imposition of common support restrictions resulted in the dropping of 46 farms (2 programme supported and 44 non-programme supported) from further analysis, thus selecting 761 *comparable* farms of which: 99 were programme participants and 662 were programme non-participants (Table 20).

11.4.4.3. Verification of the balancing property of matched variables

One of the important criteria applied for the assessment of the matching's quality can be the

42 This does not apply to local linear weighting function matching which first smoothes the outcome and then performs nearest neighbour matching. In this case more controls are used to calculate the counterfactual outcome than the nearest neighbor only (Leuven and Sianesi, 2007).

43 The caliper matching algorithm (0.07) was also found to perform satisfactory concerning other important Selection criteria, i.e. balancing property and pseudo R^2 tests (see below).

Table 21. Schleswig-Holstein. Balancing property tests

Variable-Name	variable	Sample	Treated	Control	%bias	|bias|
Long-term assets – buildings	v1025i2_01	Unmatched	78645	64423	26.4	
		Matched	77665	77949	-0.5	98.0
Operating facilities (value)	v1030i2_01	Unmatched	17355	16524	4.4	
		Matched	17400	17474	-0.4	91.1
Machinery (value)	v1031i2_01	Unmatched	28285	32066	-16.3	
		Matched	28410	28297	0.5	97.0
Cattle (value)	v1091i2_01	Unmatched	1.1e+05	93309	43.7	
		Matched	1.1e+05	1.1e+05	4.8	89.0
Inventory stock	v1110i2_01	Unmatched	174.12	93.661	4.3	
		Matched	177.64	115.81	3.3	23.2
Capital stock (value)	v1449i2_01	Unmatched	6.8e+05	6.6e+05	5.9	
		Matched	6.8e+05	6.7e+05	2.8	52.3
Revenues beef/cattle/milk sales	v2129i5_01	Unmatched	2.3e+05	1.7e+05	63.7	
		Matched	2.2e+05	2.2e+05	6.3	90.1
Purchased concentrated feed for cattle	v2705i5_01	Unmatched	-29362	-26278	-16.0	
		Matched	-29955	-30484	2.7	82.9
Labour costs (total)	v2799i5_01	Unmatched	-6808.1	-5562.6	-14.9	
		Matched	-6815.2	-6229.6	-7.0	53.0
Milk yield (per cow)	v4116i2_01	Unmatched	7351.9	6572	64.0	
		Matched	7340.2	7283.7	4.6	92.8
Fem. Calves > 0.5 year	v5111i2_01	Unmatched	17.089	13.544	35.7	
		Matched	16.929	16.114	8.2	77.0
Fem. Calves > 0.5 and < 1 year	v5112i2_01	Unmatched	21.911	19.007	25.4	
		Matched	21.788	21.116	5.9	76.9
Fem. Cattle > 1 and < 2 years	v5113i2_01	Unmatched	35.119	30.305	32.9	
		Matched	35.03	33.67	9.3	71.7
Breeding Heifer	v5114i2_01	Unmatched	19.218	19.221	-0.0	
		Matched	19.222	19.545	-2.6	-10189.4
Heifer	v5115i2_01	Unmatched	.18812	.30028	-6.4	
		Matched	.19192	.15312	2.2	65.4
Milk cows	v5116i2_01	Unmatched	71.861	61.584	38.6	
		Matched	71.404	70.437	3.6	90.6
Suckler cows	v5117i2_01	Unmatched	.13861	.25212	-6.8	
		Matched	.14141	.12746	0.8	87.7
Slaughter cows	v5118i2 _01	Unmatched	2.4158	1.5312	20.9	
		Matched	2.4646	2.2616	4.8	77.0
Male calves > 0.5	v5120i2 _01	Unmatched	14.762	10.374	41.7	
		Matched	14.525	14.631	-1.0	97.6
Male cattle > 0.5 and < 1 year	v5121i2 _01	Unmatched	19.465	13.006	44.7	
		Matched	19.364	20.036	-4.7	89.6
Male cattle > 1 and < 1.5 years	v5122i2_01	Unmatched	16.04	9.7578	43.3	
		Matched	15.818	15.918	-0.7	98.4
Male cattle > 1.5 and < 2 years	v5123i2_01	Unmatched	4.6337	2.6785	26.3	
		Matched	4.5556	4.4296	1.7	93.6

Male cattle > 2 years	v5124i2 _01	Unmatched	.05941	.2762	-15.4	
		Matched	.0404	.04363	-0.2	98.5
Breeding bulls	v5125i2 _01	Unmatched	.63366	.61331	2.4	
		Matched	.60606	.60544	0.1	96.9
Pasture area (ha)	v6104i7 _01	Unmatched	48.231	39.04	36.1	
		Matched	47.908	45.685	8.7	75.8
Agricultural area (total) (ha)	v6119i7 _01	Unmatched	94.335	83.954	26.9	
		Matched	93.834	92.596	3.2	88.1
Non-family labour (AK)	v7098i3_01	Unmatched	.17337	.18493	-2.5	
		Matched	.17586	.14761	6.2	-144.3
Labour total (AK)	v7099i3_01	Unmatched	1.7463	1.7426	0.5	
		Matched	1.7523	1.7325	2.7	-429.2
Milk production	vmilkprod_01	Unmatched	5.3e+05	4.1e+05	59.0	
		Matched	5.3e+05	5.1e+05	5.9	90.1
Excess milk quota	v8026i2_01	Unmatched	22801	15735	20.8	
		Matched	23064	20533	7.4	64.2
Equity capital formation	v9001 _01	Unmatched	1.6e+05	1.3e+05	23.5	
		Matched	1.6e+05	1.5e+05	5.4	77.1
v9003	v9003_01	Unmatched	-5374.4	-4303	-13.2	
		Matched	-5387.1	-4827.3	-6.9	47.8
Labour productivity (cattle/beef / milk per total labour)	v9005_01	Unmatched	1.4e+05	1.1e+05	69.6	
		Matched	1.4e+05	1.4e+05	0.5	99.2
Labour productivity (milk per total labour)	v9006 _01	Unmatched	3303	2487.6	64.8	
		Matched	3266.7	3255.9	0.9	98.7
profit01	profit01	Unmatched	54629	40518	48.8	
		Matched	54634	52293	8.1	83.4
Adjusted equity capital formation	v9004 _01	Unmatched	4818	2168.3	5.6	
		Matched	4847.6	6284	-3.0	45.8
Profit per farm (adjusted)	profit_co~01	Unmatched	35728	23889	35.3	
		Matched	35855	34159	5.1	85.7
Earnings from self-employment	v8212i2	Unmatched	9.8107	93.767	-10.2	
		Matched	10.009	11.991	-0.2	97.6
Earnings from non-self-employment	v8213i2	Unmatched	466.01	534.24	-2.3	
		Matched	475	389.37	2.9	-25.5
vsupp_01	vsupp	Unmatched	9340	8685.3	5.8	
		Matched	9206.3	8954.3	2.2	61.5

comparison of mean values of relevant covariates in both groups of farms (programme beneficiaries vs. controls) before and after matching (using the selected matching algorithm). It is expected that application of the selected matching algorithm (here: caliper matching 0.07) will lead to a considerable reduction in original differences in mean values of each individual variable included as a covariate in the logit function, between supported and non-supported groups of farms.

The comparison of mean values for all variables included as covariates in the estimated logit function in both groups of farms before and after matching is presented in Table 21. The results show that for almost all variables (except for the variables: number of breeding heifers, non-family labour and earnings from non-self employment) the selected matching procedure resulted in a significant reduction of differences in variables' means among both groups of farms, i.e. beneficiaries vs. controls thus making both groups of farms much more comparable. Furthermore, after the implementation of the above matching procedure the estimated standardized selection bias could be reduced from 25.6 (before matching) to 3.70 (after matching), i.e. it dropped by 86%. At the same time pseudo R^2 decreased as expected, i.e. dropped from 0.201 to 0.119 respectively, i.e. by 41%.

11.4.4.4. Results indicators

The assessment of the effect of the AFP programme (Schleswig-Holstein) on:

- direct programme beneficiaries (by means of ATT indicator);
- programme non-beneficiaries (potential impact by means of ATNT indicator);
- randomly selected unit from the sample of programme beneficiaries and non-beneficiaries (potential impact by means of ATE indicator);

was carried out using the following result indicators:

- Profit per farm;
- Corrected profit per farm[44];
- Addition to economic assets (capital formation)[45];
- Milk production (total per farm);
- Labour productivity (value of milk and beef production per fully employed persons (AK));
- Transfers from farm to household for living (for assessment of programme leverage effects);
- Transfers from farms to household for building of private assets (for assessment of programme leverage effects);
- Transfers from farm to household (total) (for assessment of programme leverage effects);
- Corrected profit (adjusted for taxes and other payments pre-paid)[46];
- Farm total employment (family labour + hired labour) in fully employed units (AK);
- Corrected profit per family labour[47];
- Corrected profit per fully employed person[48] (family labour + hired labour);
- Standard profit per family labour;
- Standard profit per fully employed person (family labour + hired labour);
- Extended profit per farm (profit + paid salaries/wages);

44 Corrected profit per farm = profit - (v2460i5 + v2461i5 + v2462i5 + v2463i5 + v2489i5 + v2492i5 + v2493i5 + v2494i5 + v2495i5 + v2496i5) + v2870i5 + v2871i5 + v2872i5 + v2873i5 + v2887i5 + v2888i5 + v2889i5 + v2890i5 + v2891i5 + v2894i5 + v2895i5 (i.e. current profits corrected for revenues and expenses linked to other periods: "Gewinne – zeitraumfremde Erträge + zeitraumfremde Aufwendungen")

45 Net increase of economic assets = profit + sum of deposits to farms – sum of transfers from the farm + transfers for building of private assets – transfers from private assets

46 Corrected profit (adjusted for taxes and other payments pre-paid) = profit - (v2460i5 + v2461i5 + v2462i5 + v2463i5 + v2489i5 + v2492i5 + v2493i5 + v2494i5 + v2495i5 + v2496i5) + v2870i5 + v2871i5 + v2872i5 + v2873i5 + v2889i5 + v2890i5 + v2891i5 + 0.9*v2894i5 + v2895i5

47 pro_corr_akf_ = profit_corr_ / (v7099i3_01 - v7098i3_01)

48 Corrected profit per fully employed person= profit_corr_ / v7099i3

11.4.4.5. Effects of the AFP programme on direct programme beneficiaries (re-estimation results)

The re-estimated effects of the AFP programme on the above result indicators and respective comparisons with results obtained from using traditional evaluation techniques (e.g. before-after; beneficiaries vs. all (unmatched) non-beneficiaries (1-0); beneficiaries vs. country's averages comprising both beneficiaries and all non-beneficiaries (1- **Ø**), etc. are shown in Tables 22a-22c.

11.4.4.5.1. Leverage effects (re-estimated)

Leverage effects were re-estimated by applying the procedures described above and dropping all non-beneficiary farms that were located in regions NF and S-F (the latter are considered to be affected by the AFP programme, i.e. by taking into account programme substitution effects) from the data set.

The new (re-estimated) results (Table 22a-22c) which are based on a considerable reduction of

Table 22a. Re-estimated leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Money transfer from farm to farm household for living

Calculation basis	Total money transfers from farm to farm household for living		
	2001	**2007**	**D I D (2007-2001)**
Unmatched P=1 (101)	30,072	43,810	13,738
Unmatched P=0 (706)	24,770	32,726	7,956
Average Ø (807)	25,433	34,113	8,680
Difference (1-0)	5,302	11,083	5,781
Difference (1- Ø)	4,639	9,697	5,058
Matched M=1 (99)	30,292	44,161	13,869
Matched M=0 (662)	28,299	37,508	9,209
ATT	1,993	6,652	4,659
ATNT	-3,051	-2,682	369
ATE	-2,395	-1,467	928

Table 22b. Re-estimated leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Money transfer from farm to farm household for building of private assets

Calculation basis	Total money transfers from farm to farm household for building of private assets		
	2001	**2007**	**D I D (2007-2001)**
Unmatched P=1 (101)	18,447	48,302	29855
Unmatched P=0 (706)	11,490	27,973	16,483
Average Ø (807)	12,361	30,517	18,156
Difference (1-0)	6,956	20,329	13,373
Difference (1- Ø)	6,086	17,785	11,699
Matched M=1 (99)	18,541	47,848	29,307
Matched M=0 (662)	15,170	34,952	19,782
ATT	3,370	12,896	9,526
ATNT	2,827	4,736	1,909
ATE	2,897	5,797	2,900

Table 22c. *Re-estimated leverage effects in AFP programme (Schleswig-Holstein). Result indicator: Total money transfer from farm to farm household*

Calculation basis	Total money transfers from farm to farm household		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	75,415	137,886	62,471
Unmatched P=0 (706)	61,205	94,034	32,829
Average Ø (807)	62,984	99,523	36,539
Difference (1-0)	14,210	43,851	29,641
Difference (1- Ø)	12,431	28,363	25,932
Matched M=1 (99)	75,596	138,009	62,413
Matched M=0 (662)	71,449	111,160	39,711
ATT	4,146	26,848	22,702
ATNT	-2,602	6,953	9,555
ATE	-1,724	9,541	11,265

the selection bias (originating from the substitution effects) show the AFP programme to have slightly higher leverage effects in comparison with former outcomes. Indeed, the AFP programme was found to substantially induce private spending among programme beneficiaries, i.e. participation in the AFP programme led to: an increase in money transfers from farm to farm household for living compared to similar non-beneficiaries by approximately +4,659 EUR per farm (Table 21a); an increase in money transfers from farm to farm household for building of private assets by approximately +9,526 EUR per farm (Table 21b); and an increase in total money transfers from farm to farm households by approximately +22,702 EUR (Table 21c).

The above results show also that an extension of the AFP programme to other non-beneficiaries (ATNT) would result in positive leverage effects (inducement of private spending among non-beneficiaries), i.e. an increase in money transfers from farm to farm household for living by approximately +369 EUR per farm (ATNT in Table 22a); an increase in money transfers from farm to farm household for building of private assets by approximately +1,909 EUR per farm (ATNT in Table 22b); and an increase in total money transfers from farm to farm households by approximately +9,555 EUR (ATNT in Table 22c). The leverage effects on a randomly selected agricultural farm, i.e. ATE (from a set consisting of programme beneficiaries and non-beneficiaries) would also be positive: i.e. respective additional money transfers from farms to farm households would be as follows: +928 EUR per farm for money transfers for living, +2900 EUR per farm for money transfers for building of private assets, and +11265 EUR per farm for total transfers (ATE's in respective tables 22 a- 22c).

11.4.4.5.2. Effects of the AFP programme on farm profits

The application of the PSM methodology (conditional ATT-DID) to the assessment of the direct effects of the AFP programme on programme beneficiaries (re-estimated results) shows the positive impact of the programme on both the standard profit (ATT-DID = +9,285 EUR per farm, see: Table 23a) as well as the corrected profit achieved by farms supported by the programme (ATT-DID = 6,455 EUR per farm, see: Table 23b). Should the AFP programme be extended to non-programme beneficiaries its effect (ATNT-DID) would also be positive (+7,634 EUR increase in profits and +9,084 EUR increase in case of corrected profits). The same is also true for the average treatment effects (ATE-DID). The effect of the AFP programme measured in terms of ATE-DID on profits and corrected profits was found to be positive (+ 7,848 EUR and + 8,743 EUR respectively).

Table 23 a. Standard profit per farm (profit)

Calculation basis	Profits per farm in EUR		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	54,629	116,777	62,148
Unmatched P=0 (706)	40,518	82,983	42,465
Average Ø (807)	42,284	87,213	44,929
Difference (1-0)	14,111	33,793	19,682
Difference (1- Ø)	12,345	29,564	17,219
Matched M=1 (99)	54,634	115,908	61,274
Matched M=0 (662)	52,292	104,281	51,989
ATT	2,341	11,626	9,285
ATNT	2,032	9,666	7,634
ATE	2,073	9,921	7,848

Table 23 b. Corrected profit per farms (profit_corr)

Calculation basis	Corrected profits per farm in EUR		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	35,728	97,243	61,515
Unmatched P=0 (706)	23,888	67,771	43,883
Average Ø (807)	25,370	71,459	46,089
Difference (1-0)	11,839	29,472	17,633
Difference (1- Ø)	10,358	25,784	15,426
Matched M=1 (99)	35,854	96,354	60,500
Matched M=0 (662)	34,159	88,204	54,045
ATT	1,695	8,150	6,455
ATNT	3,553	12,637	9,084
ATE	3,311	12,053	8,743

11.4.4.5.3. Effects of the AFP programme on own capital formation[49]

An important variable showing economic performance of agricultural farming (including farm and household) is the increase in the value in own economic assets (including farm and household) which is measured in terms of current profits + deposits in farm + net transfers for building of private assets. It may be expected that an important long-term goal of farming (in the case of presence of a farm successor) is to increase this variable over the years. As public support provided to the agricultural sector, inter alia, aims to strengthen the economic viability of agricultural enterprises, it may be expected that a relative increase of the value of own economic assets in farms receiving public support should be higher than in non-supported enterprises. Unfortunately, our results cannot confirm these expectations. Indeed, the value of own economic assets in farms supported by the AFP programme increased over the period 2001-2007 by +35,809 EUR per farm, i.e. it grew by less than in similar agricultural farms that did not receive any support from the AFP programme (the value of economic assets in the control group of agricultural farms increased by +37,045 EUR per farm). This implies that the effect of the AFP programme on this specific variable was negative (-1,237 EUR per farm, see Table 24).

49 (Ger): „Bereinigte Eigenkapitalbildung"

Table 24. Schleswig-Holstein. Effects of the AFP programme on the value of economic assets (2001-2007).

Calculation basis	Increase of the value of economic assets		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	4817	41814	36997
Unmatched P=0 (706)	2168	28842	26674
Average Ø (807)	2499	30466	27967
Difference (1-0)	2649	12971	10322
Difference (1- Ø)	2318	11348	9030
Matched M=1 (99)	4847	40656	35809
Matched M=0 (662)	6284	43329	37045
ATT	-1436	-2673	-1237
ATNT	5304	7347	2043
ATE	4427	6043	1616

Our results differ significantly from those obtained by using traditional evaluation methods (a qualitative difference), see Table 24. For example, a naïve before-after estimator shows an increase of the net value of economic assets by + 39.997 EUR per farm; the comparison of farms supported by the programme with all other farms non-supported from the programme DID in (1-0) shows also a positive effect of the programme (+10,322 EUR per farm), and a similar outcome would be obtained if programme beneficiaries were compared with a country's average (+9,030 EUR per farm). Obviously, the economic performance of programme beneficiaries differed significantly from the economic performance of programme non-beneficiaries and from the country's average. Thus, the application of more sophisticated matching techniques for derivation of relevant counterfactuals is here fully justifiable.

11.4.4.5.4. Effects of the AFP programme on milk production

Our results show that the AFP programme significantly contributed to an increase in milk production among programme beneficiaries, i.e. + 61,276 litres per farm (see table 25). Indeed, due to the AFP programme milk production increased in the examined period in the group of the matched programme beneficiaries by 155,413

Table 25. Schleswig-Holstein. Effect of the AFP programme on milk production (years 2001-2007)

Calculation basis	milk production		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	530973	692428	161455
Unmatched P=0 (706)	407068	478612	71544
Average Ø (807)	422575	505372	82797
Difference (1-0)	123904	2132816	89912
Difference (1- Ø)	108298	187056	78658
Matched M=1 (99)	526623	682036	155413
Matched M=0 (662)	514333	608470	94137
ATT	12290	73566	61276
ATNT	15949	83232	67283
ATE	15473	81974	66501

litres (by 29.5%) while in the control group (i.e. matched non-beneficiaries) it grew by only 94,137 litres per farm (by 18.3 %). Also an extension of the AFP programme to non-supported farms would lead to a significant increase in their milk production (+67,282 l per farm). Furthermore, the estimated ATE effect of the AFP programme on milk production was also found to be positive.

11.4.4.5.5. Effects of the AFP programme on farm employment

Our results show that the AFP programme had only a marginal positive impact on farm employment. In the examined period total farm employment (family and hired labour expressed in full-time equivalents, FTE) on farms that were programme beneficiaries increased by 0.103 FTE (from 1.752 FTE to 1.855 FTE per farm, see Table 26) while in comparable farms which did not receive support from the AFP programme it grew by 0.093 FTEs (from 1.732 to 1.825 FTE per farm). Furthermore, should the AFP programme be extended to other farms (non-beneficiaries) programme participation would bring about a reduction of employment (by -0.054 FTE). Also the ATE effects on farm employment were found to be negative.

Table 26. Schleswig-Holstein. Effect of the AFP programme on farm employment (years 2001-2007)

Calculation basis	Farm employment		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	1.746	1.852	0.106
Unmatched P=0 (706)	1.742	1.787	0.045
Average Ø (807)	1.743	1.795	0.052
Difference (1-0)	0.003	0.064	0.061
Difference (1- Ø)	0.003	0.057	0.054
Matched M=1 (99)	1.752	1.855	0.103
Matched M=0 (662)	1.732	1.825	0.093
ATT	0.019	0.029	0.010
ATNT	-0.0005	-0.054	-0.054
ATE	0.002	-0.043	-0.045

11.4.4.5.6. Effects of the AFP programme on labour productivity at the farm level

Labour productivity at farm level was measured using the following result indicators:

- Standard profit per total fully employed persons (profit/person in EUR/FTE)
- Standard profit per family labour (profit/family labour in EUR/FTE)
- Corrected profit per total fully employed persons (profit/person in EUR/FTE)
- Corrected profit per family labour (profit/family labour in EUR/FTE)
- Extended profit per total labour employed measured in terms of (standard profit + wages/salaries paid for hired labour)/total labour employed on farm (EUR/FTE)
- Production of milk/beef per a fully employed person (production value/person in EUR/FTE)

Our results show that the AFP programme had a positive impact on labour productivity on direct programme beneficiary farms, irrespective of the applied productivity measure. In all six cases (i.e. various productivity measures) the estimated ATT-DIDs appeared to be positive, i.e. productivity measures in the group of programme beneficiaries increased over-proportionally compared to the control group of farms (see: Tables 27a-27f). Furthermore, should the AFP programme be extended to also include other programme non-beneficiaries, the AFP programme would be found to have a positive impact on labour productivity in these farms, irrespective of the applied productivity measure. While both ATT-DID and ATNT-DID were found to be positive the average effect of the AFP programme (ATE-DID) was also positive.

Table 27a. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of profits per total employed (years 2001-2007)

Calculation basis	Profits per total employed		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	34021	64754	30733
Unmatched P=0 (706)	24977	49139	24162
Average Ø (807)	26109	51093	24984
Difference (1-0)	9043	15615	6572
Difference (1- Ø)	7912	13661	5749
Matched M=1 (99)	33944	63992	30048
Matched M=0 (662)	34354	62868	28514
ATT	-410	1123	1533
ATNT	1523	5615	4092
ATE	1271	5030	3759

Table 27b. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of profits per family labour employed (years 2001-2007)

Calculation basis	Profits per family labour employed		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	37762	80396	42634
Unmatched P=0 (706)	27818	55950	28132
Average Ø (807)	29062	59010	29948
Difference (1-0)	9944	24446	14502
Difference (1- Ø)	8700	21386	12686
Matched M=1 (99)	37726	79792	42066
Matched M=0 (662)	37290	71930	34640
ATT	435	7861	7426
ATNT	1164	8223	7059
ATE	1070	8176	7106

Table 27c. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of corrected profits per total employed (years 2001-2007)

Calculation basis	Corrected profits per total employed		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	23060	55234	32174
Unmatched P=0 (706)	15124	40653	25529
Average Ø (807)	16118	42478	26360
Difference (1-0)	7935	14581	6646
Difference (1- Ø)	6942	12756	5814
Matched M=1 (99)	23121	54510	31389
Matched M=0 (662)	23000	54017	31017
ATT	121	492	371
ATNT	2300	7235	4935
ATE	2016	6358	4342

Table 27d. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of corrected profits per family labour employed (years 2001-2007)

Calculation basis	Corrected profits per family labour employed		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	25460	65426	39966
Unmatched P=0 (706)	15965	45811	29846
Average Ø (807)	17153	48266	31113
Difference (1-0)	9494	19614	10120
Difference (1- Ø)	9307	17160	8853
Matched M=1 (99)	25554	64772	39218
Matched M=0 (662)	24923	61450	36527
ATT	631	3321	2690
ATNT	2420	9999	7579
ATE	2187	9130	6943

Table 27 e. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of extended profit per total farm employment (EUR/farm)

Calculation basis	Extended profit per total farm employment		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	31499	60290	28781
Unmatched P=0 (706)	22979	45944	22965
Difference (1-0)	8520	14345	5825
Matched M=1 (99)	31445	59545	28100
Matched M=0 (662)	31933	58541	26608
ATT	-487	1004	1491
ATNT	1665	5839	4174
ATE	1385	5210	3825

Table 27 f. Schleswig-Holstein. Effect of the AFP programme on labour productivity measured in terms of production value milk and beef per total employed (years 2001-2007)

Calculation basis	Production value of milk and beef per total employed		
	2001	2007	D I D (2007-2001)
Unmatched P=1 (101)	141977	198320	56343
Unmatched P=0 (706)	105289	140862	35573
Average Ø (807)	109880	148053	38173
Difference (1-0)	36687	57458	20771
Difference (1- Ø)	32097	50267	18170
Matched M=1 (99)	140245	195161	54916
Matched M=0 (662)	139968	185684	45716
ATT	277	9476	9199
ATNT	8213	32380	24167
ATE	7180	29400	22220

11.4.4.5.7. Re-estimated deadweight loss effect of the AFP programme

Deadweight loss effects were calculated by comparing relevant outcomes (result indicators) in the group of beneficiary farms with similar non-beneficiary farms (control group) that undertook in the examined period (2002-2007) analogous investment (i.e. modernization of buildings). Due to the dropping of all programme non-participants located in regions with the highest programme intensity from the data base, i.e. regions NF and S-F, the number of non-beneficiary farms remaining in the data base which undertook similar investments also changed (i.e. out of 706 non-beneficiary farms used for re-estimation of direct programme effects only 161 farms could be used to re-estimate deadweight loss effects). Consequently, a different structure of the data base (compared data base used to derive other re-estimated results) necessitated a new search for an optimal matching algorithm and the performance of all other steps as described in section: Stage 4.

The major steps carried out to re-estimate the effect of the programme deadweight loss effects were consistent with those described under Stage 4, and included:

- **Selection of a new relevant matching algorithm**. Given previously calculated individual propensity scores for programme beneficiaries and non-beneficiaries, and after imposing restrictions on the common support region, a new relevant matching technique was selected (a truncated data base consisted of 244 observations of which 83 observations were on programme beneficiaries and 161 on programme beneficiaries). This was carried out using three independent criteria mentioned above: i) standardized bias (Rosenbaum and Rubin, 1985); ii) t–test (Rosenbaum and Rubin, 1985); and iii) joint significance and pseudo R^2 (Sianesi, 2004) by applying methodology described in the section: 10.4. As a result a kernel (normal kernel, b.w. 0.08) was found to be the "best" matching technique and was selected for calculation of the deadweight loss effects of the AFP programme.

- **Statistical verification of the "similarity" of both groups (programme beneficiaries vs. control group) prior to their participation in the programme** (e.g. by performing balancing property tests on the most important farm characteristics) was performed

Table 28. Re-estimated deadweight loss effects of AFP programme on milk farms (Schleswig-Holstein, Germany)

Calculation basis	Value of inventories in EUR		
	2001	2007	DID (2001-2007)
Participants (P=1) (83)	80,058	153,545	73,487
Non-participants (P=0) (161)	51,607	107,265	55,658
Matched participants (M=1) (78)	77,609	149,938	72,329 (+93.2%)
Matched non-participants (M=0) (155)	56,704	128,643	71,939 (+126.8%)
Deadweight loss (M)			99%

- **Calculation of a change in relevant result indicator** (value of assets) over the examined period in the group of programme beneficiaries and comparable non-beneficiaries

The application of the above procedure resulted in new estimates of the deadweight loss of the AFP programme. Our results show that the re-estimated deadweight loss effect was huge (close to 100%, see: Table 28). In fact, in the control group of the matched programme non-beneficiaries the value of inventories over the period of 2001-2007 increased over proportionally (i.e. by 126.8%) compared with the group of farms supported by the AFP programme (+93.2%). This means that, due to prevailing economic conditions affecting performance of all milk producers (i.e. increase in milk prices) similar investments in the examined period would have been undertaken even without the programme support.

11.4.4.5.8. Estimation of programme displacement effects

As described in Section 6.2.2. spatial displacement effects can generally be measured by applying a similar methodology to in the case of direct programme effects, yet comparing two relationships: a) the performance of programme supported units (j) with similar non-supported units (m) both located in regions characterised by a high programme intensity, and b) the performance of programme supported units (j) located in regions characterized by high programme intensity with similar non-supported units (k) located in regions characterised by a low programme intensity before and after the RD programme. The lack of displacement effects would result in similar differences in DID-ATT between a) and b) (i.e. location of units would be considered as irrelevant).[50] The applicability of this methodology is however restricted only to the case of no substantial substitution effects.

In our analysis we found, however, considerable substitution effects in Schleswig-Holstein regions characterized by a high intensity of the programme (high programme exposure). This means that non-supported farms in regions with high programme intensity were also affected by the AFP programme. The basic methodological problem arises from the fact that a shift of employment from non-supported farms in regions with a low programme intensity could take place both to programme supported farms (in regions with high programme intensity) as well as to non-supported but programme affected farms (in regions with high programme intensity).

50 Generally speaking, and assuming no other general equilibrium effects (e.g. substitution effects), the bigger the difference in DID-ATT between both groups (j-k) and (j-m) after the programme is (the result of a shift of employment and a "shift" of GVA from units k to units j and m), the higher is the probability that the better performance of units j and m located in area aj occurred at detriment of units k located in non-supported areas ai.

Table 29. Schleswig-Holstein. Estimated effects of the AFP programme on employment per farm

Change on employment per farm in regions with the highest programme intensity (NF and S-H) in FTE units				Change on employment per farm without non-beneficiary farms located in regions with the highest programme intensity in FTE units			
Calculation basis	2001	2007	D I D (2007 - 2001)	Calculation basis	2001	2007	D I D (2007 - 2001)
Unmatched 1(59)	1.638	1.763	0.125	Unmatched 1 (101)	1.746	1.852	0.106
Unmatched 0 (489)	1.591	1.678	0.087	Unmatched 0 (706)	1.742	1.787	0.045
Ø (548)	1.585	1.669	0.084	Ø (807)	1.743	1.795	0.052
Difference (1-0)	0.046	0.084	0.038	Difference (1-0)	0.003	0.064	0.061
Difference (1-Ø)	0.053	0.094	0.041	Difference (1- Ø)	0.003	0.057	0.054
Matched M1 (55)	1.598	1.733	0.135	Matched M1 (99)	1.752	1.855	0.103
Matched M 0 (359)	1.600	1.745	0.145	Matched M 0 (662)	1.732	1.825	0.093
ATT	-0.002	-0.012	-0.010	ATT	0.019	0.029	0.010
ATNT	-0.088	-0.112	-0.024	ATNT	-0.0005	-0.054	-0.054
ATE	-0.076	-0.099	-0.023	ATE	0.002	-0.043	-0.045

The analysis of the displacement effect of the AFP programme in regions with high programme intensity (548 observations of which 59 were programme beneficiaries and 489 were programme non-beneficiaries) was carried out by implementing all steps described in Stage 1 applied to observations on farms located in these two regions only (i.e. NF and S-F). The estimation procedure lead to the selection of 55 programme beneficiaries and 359 similar programme non-beneficiaries). The effects on employment per farm are shown in Table 29.

The results (Table 29) show that in regions with the highest programme exposure, i.e. NF and S-H the employment per farm in the examined period (2002-2007) increased in programme non-beneficiary farms more (i.e. by 0.145 FTE units per farm) than in farms which were direct programme beneficiaries (i.e. 0.135 FTE units per farm), i.e. the direct effect of the AFP programme on the employment was negative. The comparison of these results with the effects of the AFP programme on employment per farm calculated without non-beneficiary farms located in regions with the highest programme intensity shows that employment on non-beneficiary farms located in other regions (i.e. low programme intensity) increased at a **lower** rate (i.e. +0.093 FTE per farm) than employment in the group of direct programme beneficiaries located in regions with the highest programme support (i.e. +0.135 FTE per farm, see Table 29) as well as in the group of programme non-beneficiaries located in regions with the highest programme support (i.e. +0.145 FTE per farm, see Table 29). This may imply that a part of employment in the group of farm non-beneficiaries in regions characterized by a low programme intensity “went” to farms (direct programme beneficiaries as well as programme non-beneficiaries) located in the regions with the highest programme exposure, i.e. thus indicating **slight programme displacement effects**.

11.5. Sensitivity of obtained results

Sensitivity analysis was carried out using the Rosenbaum bounding approach methodology described in Chapter: 5.6. The results show that the estimated effects of the AFP programme (Schleswig-Holstein) appeared to be rather sensitive. For example, in the case of the estimated effect of the AFP programme on milk production, the performed sensitivity analysis shows that a presence of a hidden bias of the magnitude of 5-10%, i.e. increasing the odds ratio from 1 to 1.05-1.10, would make the obtained results statistically insignificant. The relatively high sensitivity of the obtained results could have been caused by a relatively small number of observations used in these tests (99 matched pairs). Yet, sensitivity tests provide only additional information regarding effects' stability and do **not** question the overall validity of the obtained results.

Table 30. Rosenbaum bounds for milk production (2007) (N = 99 matched pairs)

Gamma	sig+	sig-	t-hat+	t-hat-	CI+	CI-
1	0.069919	0.069919	38323.8	38323.8	-12675.9	100171
1.05	0.102743	0.045745	32667.5	45247.6	-16715	105753
1.1	0.143156	0.029386	26535.7	50671.1	-23047.3	111138
1.15	0.190558	0.018573	20494.7	56804.8	-28464.2	118174
1.2	0.243857	0.01157	15767.1	63806.7	-32436.1	123938
1.25	0.301608	0.007115	11303.5	69335.3	-36879	129455
1.3	0.362176	0.004325	7544.93	74078.8	-42561.2	135367
1.35	0.423889	0.002602	4106.96	78950.9	-47675.1	140823
1.4	0.485175	0.001551	837.711	83388.3	-51330.1	146999
1.45	0.544657	0.000916	-3441.52	87391.5	-55648	151453
1.5	0.601211	0.000537	-7664.68	91732.8	-59843.9	156474
2	0.931652	1.90E-06	-35916	128711	-94189.1	207359
2.05	0.944644	1.10E-06	-38844.5	131215	-98107.4	212718
2.2	0.971403	1.80E-07	-48006.8	141362	-105729	226869
2.5	0.993121	5.00E-09	-62006	158358	-117343	246818
2.55	0.994635	2.70E-09	-65351	161662	-119505	249272
2.95	0.999327	2.00E-11	-79927.9	183363	-134223	277348
3	0.999486	1.10E-11	-81039.1	187673	-137031	280889

** gamma - log odds of differential assignment due to unobserved factors*
sig+ - upper bound significance level
sig- - lower bound significance level
t-hat+ - upper bound Hodges-Lehmann point estimate
t-hat- - lower bound Hodges-Lehmann point estimate
CI+ - upper bound confidence interval (a= .95)
CI- - lower bound confidence interval (a= .95).

12. Conclusions

The main objective of this study was to show how various micro-economic *direct/indirect effects* (e.g. *deadweight loss, leverage effects*, etc.) and selected *general equilibrium effects* (e.g. *substitution and displacement effects)* of EU RD programmes can be calculated using recently developed advanced econometric semi-parametric evaluation methodologies. Answers to EU *Common Evaluation Questions* (CEQ) regarding the effects of an RD programme on programme beneficiaries at farm level (including deadweight loss and leverage effects) were provided by comparing changes in specific *result indicators* collected at a farm level (e.g. profits, employment, gross-value added, labour productivity, etc.) in the group of programme beneficiaries with an appropriately selected control group (counterfactual analysis - based on matching). Direct programme effects were calculated on the basis of Average Treatment on Treated (ATT) indicators (for programme beneficiaries), Average Treatment Effects on Non-Treated (ATNT) indicators (for programme non-beneficiaries) and Average Treatment Effects (for both groups) using a combination of propensity score matching (PSM) and difference in differences (DID) methods. A modification of combined propensity score and difference in differences methodology (modified PSM-DID) was applied to derive various general equilibrium effects (e.g. substitution effects). The empirical analysis was focused on evaluation of effects of the SAPARD programme in Slovakia (years 2002-2005) and the Agrarinvestitionsförderungsprogramm (AFP) in Schleswig Holstein, Germany (2000-2006) using micro-economic data (balanced panels) of bookkeeping farms (including programme participants and non-participants) in respective countries. The methodology described in this study appeared as highly applicable to estimation of impacts of EU RD programmes. Using combination of propensity score matching with difference in differences estimator (PSM-DID) as the basic evaluation technique improved significantly representativeness of control groups and allowed to estimate much more precisely the direct, and indirect (general equilibrium) effects of a given RD programme. Our results show significant differences in estimated effects of a given RD programme in dependence on whether traditional (naïve techniques) or advanced evaluation methods were applied. Comparisons of advanced ex post impact evaluation methods (e.g. combined propensity score matching and difference in differences estimator) with numerous traditional approaches (e.g. "naïve" techniques: before-after, or all participants vs. all non-participants, etc.) clearly demonstrate that "naïve" evaluation techniques usually lead to biased policy conclusions, irrespectively on the selected result indicator. Clearly, application of advanced evaluation methodologies can lead to quite different (compared with traditional techniques), yet more reliable results. On the other side, the use of more sophisticated evaluation techniques is especially demanding in terms of data (number of observations and quality) but it requires also more technical skills and extensive capacity building on side of programme evaluators. While quantitative methods are advantageous for estimating and comparing net-impacts of various RD programmes they should be complemented with qualitative methods that are very helpful to answer questions: WHY? these effects occurred/ not occurred in a given magnitude. A right combination of those both approaches appears therefore decisive for improving the quality of evaluation studies.

References:

Aakvik, A. (2001), "Bounding a Matching Estimator: The Case of a Norwegian Training Program," *Oxford Bulletin of Economics and Statistics*, Department of Economics, University of Oxford, 2001, 63(1), pp. 115-43.

Aakvik, A. and Heckman, J. J. and Vytlacil, E. J. (2000), "Treatment Effects for Discrete Outcomes when Responses to Treatment Vary among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs", 2000, NBER Technical Working Paper 262, Cambridge, Mass.

Abadie, A. and Imbens, G. (2004), "Large sample properties of matching estimators for average treatment effects." *Working paper*, 283. http://elsa.berkeley.edu/ ,2004.

Asian Development Bank, (2006), Impact Evaluation: Methodological and Operational Issues, Guidelines, Handbooks, and Manuals; September 2006

Baker, J. H. (2000), "Evaluating the impact of development projects on poverty. A Handbook for practitioners", World Bank, 2000.

Becker, S. and Caliendo, M. (2007), "Sensitivity Analysis for Average Treatment Effects," *Discussion Papers of DIW Berlin 659*, DIW Berlin, German Institute for Economic Research, 2007.

Björklund, A. and Moffitt, R. (1987), "The Estimation of Wage and Welfare Gains in Self-Selection Models", *Review of Economics and Statistics*, 1987, 69, pp. 42-49.

Black, D. and Smith, J. (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching?", *Journal of Econometrics*, 2004, 121(1), pp. 99-124.

Blundell, R. and Costa Dias, M. (2002), "Alternative Approaches to Evaluation in Empirical Microeconomics", *cemmap working paper CWP10/02*, The Institute for Fiscal Studies, Department of Economics, UCL, 2002.

Bourguignon, F. and Pereira da Silva, L. A. "Evaluating the poverty and distributional impact of economic policies: a compendium of existing techniques," In: World Bank (eds.), *The impact of economic policies on poverty and income distribution*. World Bank: Washington, DC [u.a.], pp. 1-24, 2003.

Bryson, A. et al. (2002), "The use of propensity score matching in the evaluation of active labour market policies.", Policy Studies Institute, U.K. Department for Work and Pensions Working *Paper No. 4*. http://www.dwp.gov.uk/asd/asd5/wp-index.html, 2002.

Burtless, G. (1995), "International Trade and the Rise in Earnings Inequality", Journal of Economic Literature, 1995, 33, pp. 800-816.

Caliendo, M. and Hujer, R. (2005), "The Microeconometric Estimation of Treatment Effects - An Overview," *IZA Discussion Papers 1653*, Institute for the Study of Labor (IZA), 2005.

Caliendo, M. and Kopeinig, S. (2005), "Some Practical Guidance for the Implementation of Propensity Score Matching. ", *DiscussionPaper 485,* DIW Berlin, IZA Bonn, 2005.

Calmfors, L. (1994), "Active Labour Market Policy and Unemployment: A Framework for the Analysis of Crucial Design Features", *OECD Economic Studies*, 1994, 22, pp. 8-47.

CEAS (2003), Ex-post evaluation of measures under regulation (EC) No 951/97 on improving the processing and marketing conditions of agricultural products, Agra CEAS Consulting, September.

Cochran, W. and Rubin, D. (1973), "Controlling Bias in Observational Studies: A Review" Sankhya, 1973, 35, pp. 417-46.

Dehejia, R. H. and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association,* 1999, 94, pp. 1053–1062.

Dehejia, R. H. and Wahba, S. (2002), "Propensity Score-Matching Methods For Nonexperimental Causal Studies," *The Review of Economics and Statistics,* 2002, 84, pp. 151-161.

DiPrete, T. and M. Gangl, M. (2004), "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments." *Sociological methodology*, 2004, 34, pp.271 – 310.

EC, European Commission (1999), "Evaluation of Agri-environment programmes", *European Union, Working Document*. 1998. State of application of regulation (EEC), No 2078/92, EU publication VI/7655/98, http://europa.eu.int/comm/agriculture/envir/programs/evalrep/.

EC, European Commission (2002a), Guidelines for the mid-term evaluation of rural development programmes 2000-2006 supported from the European Agricultural Guidance and Guarantee Fund. DOC. STAR VI/43517/02.

EC, European Commission (2002b), Guidelines for the mid-term evaluation of rural development programmes funded by SAPARD 2000-2006, DG AGRI.

EC, European Commission (2006), CMEF, Handbook on, Guidance document (including all Guidance Notes A-O); http://ec.europa.eu/agriculture/rurdev/eval/guidance/document_en.pdf

EC, European Commission (2011), Monitoring and Evaluation for CAP post 2013. Background document for stakeholder conference held in Brussels on 20-21 September 2011.

Ederveen, S. et al. (2003), "Funds and games: the economics of European cohesion policy", *Occasional paper*, No. 3, ENEPRI, 2003.

Essama-Nssah, B. (2006), "Propensity score matching and policy impact analysis: a demonstration in EViews." Washington, DC: World Bank, Poverty Reduction and Economic Management Network, Poverty Reduction Group, 2006.

EENRD (European Evaluation Network for Rural Development) (2010), "Approaches for assessing the impacts of the Rural Development Programmes in the context of multiply intervening factors", March 2010, pp 1-225.

FAL, 2006, "Proceedings of the European workshop on the evaluation of farm investment support", FAL Federal Agricultural Research Centre, Working papers in agricultural economics, No 3, 2006.

Forstner, B. and Plankl, R. (2004), "Mid-term evaluation of single farm assistance under the Regulation", (EC) No. 1257/1999: methodological problems, data shortcomings and possible solutions, 2004.

Froelich, M. (2004), "Finite sample properties of propensity-score matching and weighting estimators", Review of Economics and Statistics, 86(1), 77-90)

Gu, X. and Rosenbaum, P. R. (1993), "Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms." *Journal of Computational and Graphical Statistics*, 1993, 2 (4), pp. 405–420.

Heckman, 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*, 1976, 5, pp. 475-492.

Heckman, J. (1996), "Randomization as an Instrumental Variable Estimator", *Review of Economics and Statistics*, 1996, 56, pp. 336-341.

Heckman, J. (1997), (eds.) "Characterizing selection bias using experimental data", Cambridge, MA, 1997.

Heckman, J. (2001), "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture", *Journal of Political Economy*, 2001, 109, pp. 673-748.

Heckman, J. et. al, (2003) – Simple estimates for treatment parameters, The Review of Economics and Statistics, August 2003, 85(3)).

Heckman, J. (2005), "The Scientific Model of Causality", *Sociological Methodology*, 2005, 35 (1), pp. 1-98.

Heckman, J. and Honore, B. (1990), "Empirical content of the Roy Model", *Econometrica*, 1990, 58, pp. 1121-1149.

Heckman, J. and Robb, R. (1985), "Alternative Methods for Estimating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, New York: Cambridge University Press, pp. 156-245.

Heckman, J. and Vytlacil, E. (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, pp. 4730 - 4734.

Heckman, J. and Vytlacil, E. (2000a),. "Local Instrumental Variables," *NBER Technical Working Papers 0252*, National Bureau of Economic Research, Inc.

Heckman, J. and Vytlacil, E. (2000b), "The relationship between treatment parameters within a latent variable framework," *Economics Letters, Elsevier*, vol. 66(1), pages 33-39, January.

Heckman, J. and Vytlacil, E. (2000c), "Econometric Evaluations of Social Programs," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Volume 5, (North-Holland: Amsterdam).

Heckman, J. and Vytlacil, E. (2000d), "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," *NBER Technical Working Papers 0259*, National Bureau of Economic Research, Inc.

Heckman, J. and Vytlacil, E. (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica,* 2005, 73 (3), pp. 669–738.

Heckman, J. et al. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," Review of Economic Studies, Blackwell Publishing, 1997, 64(4), pp. 605-54.

Heckman, J., Ichimura, H. and Todd, P. (1998), "Matching as an Econometric Evaluation Estimator" Review of Economic Studies, 1998, 65(2), pp. 261-294.

Heckman, J. et al. (1996), "Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method," *Proceedings of the National Academy of Sciences*, 1996, 93, pp. 13416–13420.

Heckman, J. et al. (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, 66 (5), pp. 1017–1098.

Heckman, J. et al. (2003), "Simple estimates for treatment parameters", *The Review of Economics and Statistics*, 853 (3), 2003.

Heckman, J., LaLonde, R. and Smith, J. (1999), "The Economics and Econometrics of Active Labour Market Programs." In: *Handbook of Labour Economics* , ed. Orley Ashenfelter and David Card, 3A, pp. 1865–2097. Amsterdam: Elsevier Science.

Heckman, J. and Navarro-Lozano, S. (2003), "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models", *IZA Discussion Papers 768*, Institute for the Study of Labor (IZA).

Henning, C. H. and Michalek, J. (2008a), "Ökonometrische Methoden der Politikevaluation: Meilensteinfür eine sinnvolle Agrarpolitik der II. Säule oder akademische Fingerübung?", *AGRARWIRTSCHAFT*, 57, pp. 232-243.

Henning, C. H. and Michalek, J. (2008b), " Theoretische Grundlagen und die Rolle parametrischer bzw. Nicht-parametrischer Modellansätze in der Politikevaluation. Replik zu Anne Margarian Johann Heinrich von Thünen-Institut, Braunschweig", *AGRARWIRTSCHAFT*, 57, pp. 348-351.

Holland, Paul. W. (1986), "Statistics and Causal inference (with discussion"*), Journal of the American Statistical Association,* 1986, 81, pp. 945-970.

Imai, K., G. King and E. A. Stuart (2006), "The Balance Test Fallacy in Matching Methods for Causal Inference", http: *polmeth.wustl.edu/retrieve.php?id=632*

Imbens, G. (2003), "Sensitivity to Exogeneity Assumptions in Program Evaluation," *American Economic Review*, Papers and Proceedings, 2003, 93, pp.126-132.

Imbens, Guido and Wooldridge, J. (2007), "Recent Developments in the Econometrics of Program Evaluation," *unpublished manuscript, department of economics,* Harvard University, 2007.

Jalan, J. and Ravallion, M. (2002), "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching", *Journal of Business and Economic Statistics*, 21: 19-30.

Larsson, L. (2000), Sweden "Evaluation of Swedish youth labour market programmes," *IFAU Working Paper 2000:1*, Office of Labour Market Policy Evaluation, Uppsala.

Lechner, M. (2001), "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in Lechner, M., Pfeiffer, F. (eds), Econometric Evaluation of Labour Market Policies, Physica/Springer, Heidelberg, 43-58.

Lechner, M. (2002a), "Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies," *Review of Economics and Statistics*, 2002a, 84, pp. 205–220.

Lechner, M. (2002b), "Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods," *Journal of the Royal Statistical Society, Series A*, 2002b, 165, pp. 659–682.

Leuven, E. and Sianesi, B., (2009), Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Statistical Software Components, Boston College Department of Economics, http://econpapers.repec.org/RePEc:boc:bocode:s432001, 2009.

Mantel, N. and Haenszel, W. "Statistical aspects of the analysis of data from retrospective studies of disease". *Journal of the National Cancer Institute*, 1959, *22*, pp. 719-748.

Michalek, J. (2007), "Quantitative tools for ex-post evaluation of EU RD programmes", Advanced-Eval. Working Paper Series, University of Kiel, April 2007. http://www.advanced-eval.eu

Michalek, J. (2009), "*"Application of advanced methodologies to ex post evaluation of RD programmes (micro-level)"*, paper presented at the final workshop ADVANCED-EVAL, Brussels, 27 October 2009.

Newman et al., (2002), Bolivia An impact evaluation of education, health, and water supply investments by the Bolivian social investment fund, *World Bank Economic Review* **16** (2002) (2), pp. 241–274.

NONIE (2009), Impact evaluations and development, NONIE Guidance on Impact Evaluation, by Frans Leeuw and Jos Vaessen, Network of Networks for impact evaluation.

Pradhan, M. and Rawlings, L. (2002), Nicaragua, "The impact and targeting of social infrastructure investments: Lessons from the Nicaraguan social fund," *World Bank Economic Review* **16** (2002) (2), pp. 275–296.

PCM (2007), Ex-post evaluation of the SAPARD programme in the Slovak Republic. Final Report. P.C.M. Group. December 2007.

Pufahl, A. and Weiss, Ch. (2007), "Evaluating the effects of farm programs : results from propensity score matching", *Department of Economics Working Paper Series 113*, Wien : Vienna University of Economics and Business Administration, 2007.

Quandt, R.E. (1972), "A New Approach to Estimating Switching Regression", Journal of the American Statistical Association, 1972, 67.

Ravallion, M. (2004), "Pro-Poor Growth: A Primer", *World Bank Policy Research Working Paper 3242*, World Bank, Washington, D.C., 2004.

Ravallion, M. (2006?) „Evaluating Anti-Poverty Programs", in *Handbook of Development Economics Volume 4*, edited by T. Paul Schultz and John Strauss, Amsterdam, North-Holland

Rawlings, L. B. and Schady, N. R. (2002), "Impact evaluation of social funds: an introduction," *The World Bank economic review*, 2002, 16(2), pp. 213-217.

Rosenbaum, P. R. and Rubin, D. B. (1983a), "The central role of the propensity score in observational studies for causal effects." *Biometrika,* 1983a, 70(1), pp. 41-55

Rosenbaum, P.R. and Rubin, D. B. (1983b), "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome", *Journal of the Royal Statistical Society*, 1983b, 45, pp. 212.

Rosenbaum, P. R. and Rubin, D. B. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, 1985, 39, pp. 33 - 38.

Rosenbaum, P. R. (2002), "Observational Studies", New York: Springer-Verlag. 2nd Edition, 2002.

Rossi, P. H. and Freeman, H. E. (1993), (eds.) "Evaluation. A Systematic Approach" , SAGE Newbury Park London, 1993.

Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 1951, New Series, 3(2), pp. 135-146.

Rubin, D. B. (1974a), "Characterizing the Estimation of Parameters in Incomplete Data Problems", *The Journal of the American Statistical Association*, 1974, 69(346), pp. 467- 474.

Rubin, D. B. (1974b), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 1974, 66(5), pp. 688 – 701.

Rubin, D. B. (1975), "Bayesian inference for causality : The importance of randomization." *Proc.Social Statistics Section, Am. Statist. Assoc.*, 1975, pp. 233-239.

Rubin, D. B. and Thomas, N. (1996), "Matching Using Estimated Propensity Scores: Relating Theory to Practice", *Biometrics*, 1996, 52, pp. 249-264.

Schmitt B. et al. (2004); "Impact evaluation of the EU programme for rural development in Burgundy: Correcting for selection bias", http://www2.dijon.inra.fr/esr/pagesperso/schmitt/BS_PL_MBS_DL_NewV.pdf

Sianesi, B. (2004), "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *The Review of Economics and Statistics*, MIT Press, 2004, 86(1), pp.133-155, 09.

Smith, J. (2000), " A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies", *Schweiz. Zeitschrift für Volkswirtschaft und Statistik*, 2000, 136, pp. 1-22.

Smith, J. and Todd, P. (2003), "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?," CIBC Human Capital and Productivity Project Working Papers 20035, University of Western Ontario, CIBC Human Capital and Productivity Project, 2003.

Smith, J. and Todd, P. (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators? ", *Journal of Econometrics*, 2005, 125(1-2), pp. 305-353.

Tissen, G. and Schrader, H. (1998), "Importance of measure studies as part of the evaluation approach," Conference of evaluation practices in the field of structural policies, Sevilla, 1998.

Todd, P. (2006, 2008), "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated", in: Handbook of Agricultural Economics", Volume 4, North Holland, (eds.) R. E. Evenson and T. P. Schultz, 2006; also in: Handbook of Development Economics, Volume 4, Chapter 60, Elsevier B.V. 2008.

Todd, P. (2006), "Matching Estimators", October, 2006; http://athena.sas.upenn.edu/~petra/papers/mpalgrave2.pdf

Toulemonde, J. et al. (2002), "Three layers of quality assurance: would this help provide EU policy makers with the evaluative information they need?", *paper to the EES Conference*, Seville. (www.europeanevaluation.org), 2002.

Vanhove, N. (1999), "Regional Policy: A {European} Approach", Aldershot, UK, 1999.

Venetokis, T. (2004), Finland, "An Evaluation of wage subsidy programs to SMEs utilizing propensity score matching." Helsinki: Government Institute for Economic Research. (VATT Research Reports, 104), 2004.

Van de Walle, D., and Cratty, D. (2002), "Impact Evaluation of a Rural Road Rehabilitation Project." World Bank, Washington D.C. Processed, 2002.

Watson, I. (2005), "The earnings of casual employees: the problem of unobservables." Paper presented to HILDA Survey Research Conference, University of Melbourne, 2005.

Winship C. and Morgan C. (1999), “The estimation of causal effects from observational data.” *Annual Review of Sociology*, 1999, 25, pp.659–707.

World Bank, (1999), “Impact Evaluation: A Note on Concepts and Methods”, by Kene Zemenari, Anders Rudqvist and K. Subbarao, PRMPO Poverty Reduction and Economic Management Network The World Bank, January 1999

World Bank, (2002), “Review of Impact Evaluation Methodologies Used By The Operations Evaluation Department Over Past 25 Years”, by Anju Gupta Kapoor, THE WORLD BANK OPERATIONS EVALUATION DEPARTMENT

World Bank, (2002), “DIRECTIONS IN DEVELOPMENT. Evaluating the Impact of Development Projects on Poverty, A Handbook for Practitioners, by Judy L. Baker,

Zhao, Z. (2005), “Sensitivity of Propensity Score Methods to the Specifications”, *IZA DP No. 1873*, 2005, Institute for the Study of Labor , Bonn.

European Commission
EUR 25421 – Joint Research Centre – Institute for Prospective Technological Studies

Title: Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States.
Volume 1: A micro-level approach

Author: Jerzy Michalek

Luxembourg: Publications Office of the European Union

2012 – 91 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424

ISBN 978-92-79-25681-3

doi:10.2791/82592

Abstract

The main objective of this study is to show how various direct and indirect effects (e.g. deadweight loss, leverage effects, substitution and displacement effects) of EU RD programmes can be empirically estimated using recently developed advanced econometric evaluation methodologies. Answers to EU Common Evaluation Questions (CEQ) regarding effects of the RD programme at a farm level are provided by comparing changes in respective result indicators (e.g. profits, employment, gross-value added, labour productivity, etc.) in the group of programme beneficiaries with a control group (counterfactual analysis). Policy relevant direct programme effects are calculated on the basis of Average Treatment on Treated (ATT) indicators (for programme beneficiaries), Average Treatment Effects on Non-Treated (ATNT) indicators (for programme non-beneficiaries) and Average Treatment Effects (for both groups) using a combination of propensity score matching (PSM) and difference in differences (DID) methods. Furthermore, a modified conditional DID estimator (PSM-DID) is applied to confer about specific programme's general equilibrium effects (e.g. substitution and replacement effects). Robustness of obtained results (potential effect of a hidden bias) is analysed by applying a sensitivity analysis (Rosenbaum bounds). The empirical analysis - focused on the evaluation of an impact of two RD programmes implemented in one new and one old member states, i.e. the SAPARD programme in Slovakia (years 2002-2005) and the Agrarinvestitionsförderungsprogramm (AFP) in Schleswig Holstein, Germany (2000-2006) - is based on micro-economic data (balanced panels) of bookkeeping farms (including programme participants and programme non-participants). Furthermore, methodological and policy recommendations are provided on a general applicability of conventional and advanced evaluation methods, selection of variables and matching techniques, and the use of various data-bases for evaluations of RD programmes carried out at micro-economic levels.

LF-NA-25421-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-25681-3