Evaluating measures for improving farm competitiveness in the European Rural Development programme: a comparison of different matching approaches

Stefan Kirchweger, Michael Eder, Martin Kapfer and Jochen Kantelhardt

Stefan Kirchweger, Michael Eder, Martin Kapfer and Jochen Kantelhardt are with the Institute of Agricultural and Forestry Economics at the University of Natural Resources and Life Sciences, Vienna, Feistmantelstrasse 4, A-1180 Vienna, Austria. E-Mail: stefan.kirchweger@boku.ac.at for correspondence.

Selected Poster prepared for presentation at the International Association of Agricultural Economists (IAAE) Triennial Conference, Foz do Iguaçu, Brazil, 18-24 August, 2012.

Abstract

The increasing demand for agricultural policy evaluation and the complexity of identifying the effects of EU Rural Development programmes are two challenges to agricultural economics. In this context the following paper compares *Direct Matching* with *Propensity-Score Matching* as suitable quantitative methods, as exemplified by the agricultural investment support programme for Austria. The results show no statistical differences in matching quality and resulting effects. The conclusion is that the more sophisticated approach, using the Propensity Score, should only be used when necessary. More work should be done on collecting information prior to the analysis.

Key words: Rural Development programmes, causal effects, *Direct Matching*, *Propensity-Score Matching, difference-in-difference estimator*

JEL classifications: Q10, Q12, Q18

Introduction

The increasing importance of Rural Development (RD) measures in the Common Agricultural Policy (CAP) requires a consistent evaluation of causal effects to verify those expenditures for the rest of the society. Considering the complexity of RD programmes, this presents a big challenge for policy evaluators. To overcome this, a framework of consistent evaluation including quantitative methods, multiple perspectives (internal/external, different stakeholders) and different world views (scientific/non-scientific) is vitally important (Lukesch and Schuh, 2010). This has also been recognised by the EU and guidelines for a consistent evaluation have been set up. But practical work still fails regarding consistency and validity mainly because of the lack of data and appropriate methodology. Current evaluation of RD programmes is often based on naive *before/after analysis and with/without treatment analysis.*

In *before/after analysis* the outcome of participants before and after the treatment is compared. This approach has its main drawback in evaluating gross effects instead of net effects. Net effects can be assessed by using *with/without treatment analysis.* These are mainly done using empirical data, as experiments are not suitable in agricultural policy evaluation. When empirical data for evaluating the effects of farm programmes is used, this methodology causes further problems (Pufahl and Weiss, 2009a). One of the major difficulties is the identification of an adequate control group, which is required for measuring causal effects. Rural Development measures in particular show systematic differences in participants and non-participants (selection bias), caused by voluntary selection to programme participation. Salhofer und Streicher (2004) illustrate the evidence of selection bias in the Austrian agro-environmental measure. The selection bias can be seen as a fundamental evaluation problem.

To overcome this problem, parametric and non-parametric approaches are applied and discussed in evaluation literature. When a large dataset is available, the non-parametric method *Matching* is preferable over another (Reinowski, 2008). This method is used, generally in combination with other methods, in almost any situation where there is a treatment, participants and a big group of non-participants.[1] Next to the applications in medicine and in the labour market, which are the main fields, it has also been applied in environmental economics (Greenstone and Gayer, 2009) and agricultural farm-policy evaluation (Godtland et al., 2005, Henning and Michalek, 2008; Pufahl and Weiss, 2009a; Pufahl and Weiss, 2009b). *Matching* is built upon selection on observables and a quasi-experimental design, as it deals with the fundamental evaluation problem by identifying an adequate control group for participants based on observable covariates (X). Therefore it is of

[1] Caliendo and Kopeinig (2008) offer a short overview of settings where matching is used.

paramount importance to choose those X which influence the decision to participate in the programme and the operative outcome at the same time. The most straightforward and non-parametric way is to compare the participants with the potential non-participants directly on their X (*Direct Matching*). But this might lead to dimension problems when the number of X is high. Therefore the applications of the Propensity Score (PS), as the probability of receiving treatment, have increased in recent years. This is, in some cases, indispensable but has lead to complex models being especially obscure for (non-)scientific stakeholder and partners.

The basic objective of this study is to prove the necessity of the PS in the evaluation of RD programmes. We therefore apply *Direct Matching* and test the robustness of this approach by comparing it with *Propensity-Score Matching* as regards matching quality and results. The *matching* approaches are combined with a *difference-in-difference estimator (DiD), as* recommended by Smith and Todd (2005) for measuring the *Average Treatment Effect on the Treated (ATT)* in the Austrian farm-investment programme. The application is done for the Austrian farm-investment support programme, as this has seen a particular increase in expenditures in the last few years and the evaluation so far has been mainly based on only qualitative or naive quantitative methods, as mentioned above (Beck und Dogot, 2006; Dirksmeyer et al., 2006; Pfefferli, 2006; Striewe et al., 1996). But Lukesch and Schuh (2010) point out that in particular Axis One of the RD is the main field for applying the match method. Nonetheless, evaluating causal effects of the farm-investment support programme is still challenging, as those payments are always in combination with an investment. A control group of those farms which have invested but are not supported by the programme can hardly be found (Dirksmeyer et al., 2006). This study therefore considers the causal effects of support and investment jointly.

In Section 2, which follows, there is a brief description of the farm-investment support programme in Austria, in which we illustrate the distribution of payments and participants as well as the structural characteristics of participants. Section 3 explains the evaluation problem and the method used in detail and database. The empirical application and the results applying the *Direct Matching* and *Propensity-Score Matching* approach and the comparison of those are displayed in Section 4. In Section 5 a conclusion is drawn.

The farm-investment support programme in Austria

The farm-investment programme is part of the second pillar of the Common Agriculture Policy and basically concerns improving competitiveness, work conditions, animal welfare and environmental conditions. To achieve these goals 576 million Euros have been spent in Austria (Dantler et al., 2010). The number of fostered farms during this period is slightly above 37,000, all mainly located in mountainous regions (see Figure 1). Consequently, forage farms (including mainly dairy and suckler-cow farms) are the main beneficiary of farm-investment payments, with a share of more than 56%. In contrast, in the distribution of farm type of all farms in Austria, forage farms have only a share of 37% (BMLFUW, 2010). In addition, there is an over-representation of granivore farms in contrast to field-crop farms. It is therefore not surprising that more than 50% of these funds foster the construction of barns mainly for dairy farming. Even though participants are mainly mountainous farms, it illustrates a low share of participants in the western federal states of Tyrol and Vorarlberg. This might be due to specific achievements by the federal states.

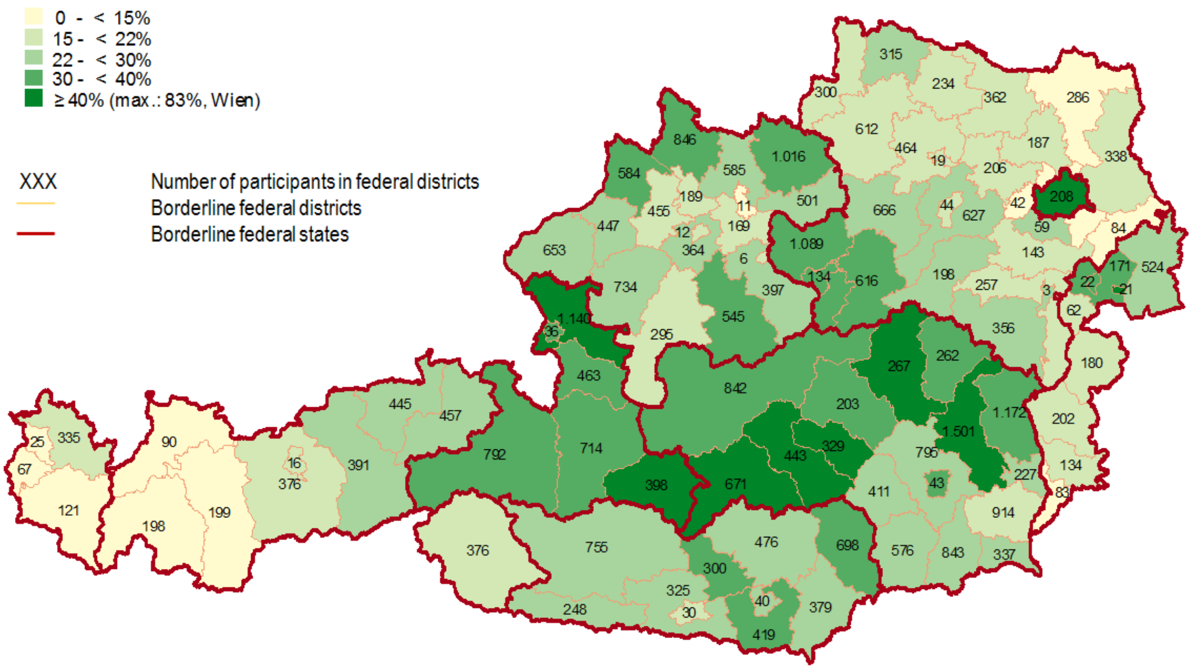

Figure 1: Share of participating farms in Austrian farm-investment programme (Source: Dantler et. Al, 2010)

Furthermore, on average the share of participating farms increases for bigger farms. Hence the means of participants and non-participants differ, especially for the utilized agricultural area (UAA), total livestock units (LU) and milk quota. It is evident, therefore, that there has been a selection for participation based on structural and regional variables such as region, farm type and farm size.

Method and database

In this chapter, we begin with theoretical considerations of microeconomic evaluation and continue with the description of the *Matching* and *Difference-in-Difference* approach. Afterwards we specify our methodological applications and present the used database and assumptions.

1.1 Microeconomic evaluation

Causal inferences have created a big debate in the social science. One attempt for the calculation of causal effects is to rely on the framework of potential outcome, also named as the Neyman-Rubin (Sekhon, 2009), Neyman-Rubin-Holland (Brady, 2008) or Roy-Rubin (Caliendo and Hujer, 2006) model. The model was first introduced by Neyman (1923) and is nowadays used in a wide range of topics for microeconomic evaluation (Sekhon, 2009). The underlying assumption is that the causal effect (Δ_i) for one individual (i) can be computed by comparing the outcome under the situation of participation (Y_i^1) and the outcome under the situation without participation (Y_i^0).

$$\Delta_i = Y_i^1 - Y_i^0 \tag{1}$$

With this, a fundamental problem of microeconomic evaluation arises: it is not possible to observe the outcome of participants without participation or the outcome of non-participants under participation. Therefore we cannot observe both potential outcomes (Y_i^1, Y_i^0) simultaneously. Thus, in our case we know the income of a participating farm, but we do not know the income of this farm if it had not invested in new technology or buildings and received payments. To deal with the fundamental evaluation problem in experimental studies, the counterfactual outcome can be replaced by the outcome of non-participants. For mainly ethnical and methodical reasons, experimental studies cannot be used for ex-post evaluation of policy measures; we have to rely on an empirical framework (Henning and Michalek, 2008). Compared to experiments, where participation is randomly assigned, empirical studies carry the problem that participants and non-participants voluntarily decide on participation. This might result in selection biases, which are systematic differences of variables between both groups. These differences must be controlled in order to identify the causal effects of political programmes. There are several approaches available for solving these problems. Whereas the estimators *matching* and *regression* construct the counterfactual outcome based on observables, *difference-in-difference estimator*, *instrumental variables* and *selection models* allow for selection on unobservable variables (Caliendo and Hujer, 2006).

For assessing treatment effects, the Average Treatment Effect on the Treated (ATT) is most commonly used in empirical evaluation. This parameter focuses directly on the effects of the participants (P=1) and is defined as

$$ATT = (Y_i^1 - Y_i^0 \mid P=1) \tag{2}$$

The outcome of this parameter might help to decide whether the programme is successful or not by comparing it to the programme costs (Heckmann et al., 1999).

In this paper we look for the ATT by applying a combination of *matching* and *difference-in-difference estimator*, the so-called "*conditional difference-in-difference estimation*" which will be explained in the following sections.

1.2 Matching approach

Matching is a non-parametric approach and follows the Conditional Independence Assumption (CIA) in order to find an adequate control group. Based on Rubin (1977), the CIA assumes that under a given set of observable covariates (X), the outcome of one individual is independent of treatment or non-treatment. The probability of receiving treatment should therefore be dependent on X but independent of the outcome and unobserved covariates (Rosenbaum, 2010). This requires the identification of those X which influence the outcome and the probability of participation but are not influenced by treatment. Blundell et al. (2005) point out, that the selection of X can be seen as the central issue in the matching method and call it a decision on the knife-edge as there can be too many, as well as too few, covariates. As the chosen covariates should describe the probability of participation as accurately as possible, too few and inaccurate covariates would violate this assumption. It is therefore of great importance to acquire the theoretical and practical background in order to choose the appropriate covariates. This can be done by analysing the distribution of the selected measure payments.

Another major assumption which needs to be fulfilled but is often harmed is the so-called Common Support Assumption (Lechner, 2001). This basically requires the existence of non-participants having similar X to all participants. Violation arises especially when covariates are used which predict too well the probability of treatment, but this is simply detected by visual control. Losing observations because of missing common support, is usually not a problem when there are not too many.

On these conditions, pairs consisting of participants and controls are built, and a control group which is similar to the participant group is generated. This should lead to a reduction of systematic mean differences between these groups. Therefore the ATT with the reduced bias can be computed, as the difference of the mean outcome of participants and controls:

$$ATT = \sum_{i=1}^{n} (Y_i^1 \mid X)/n_i - \sum_{j=1}^{n} (Y_i^0 \mid X)/n_i \qquad (3)$$

For *matching* a distance function and an algorithm are needed to identify similar controls for participants (Augurzy and Kluve, 2004). Distance functions are used to condition X of individuals, which can be done by approaches matching directly on covariates as well as using single or aggregated distance functions. The most straightforward and non-parametric way to condition X is to match them exactly (Sekhon, 2007). The *Direct Matching* (DM) approach is favourable when a small number of covariates are used, but it has its drawbacks when too many covariates are needed to describe the probability of treatment (Gensler et al., 2005). Matching many covariates leads to a reduced success in finding controls and an increase in time and effort. One attempt to overcome this problem is to use the Propensity Score (PS) of each individual instead of a big bulk of covariates. The PS reduces the matching dimensionality to one and is the probability of participation for one individual given by X, independent of observed participation. Rubin and Rosenbaum (1983) prove that matching on the propensity score is sufficient. *Propensity-Score Matching* (PSM) differentiates from DM as the values of covariates are usually different within the pairs with the same propensity score but are balanced in the treated and control group (Rosenbaum, 2010). The estimation of the PS is commonly based on a parametric logit or probit model, using observed treatment assignment (yes/no) as the explained and X as the explanatory variable. The model must not be linear but may include interactions, polynomials and transformations of the covariates.

Specifying the distance function is not enough; furthermore, the right algorithm needs to be selected. There are many matching algorithms available (see Caliendo and Kopeinig, 2008, Pufahl, 2009), where we choose the *Radius Matching* with replacement approach. Similarity for metric covariates is established by using a self-defined calliper. Several non-participants which are found within the calliper can serve as control for one participant and one non-participant and can also be used as control for more than one participant. Through this the quality of matching rises, as the controls are much more similar in contrast to *Nearest Neighbour Matching* or *Matching without replacement (*Caliendo and Kopeinig, 2008). By setting up a calliper the condition of common support can be fulfilled, but it has one main drawback: the calliper is hard to specify *a priori*. A misspecified calliper might either lead to dissimilarity or a significant loss of controls and participants. Rosenbaum (2010) proposes the strategy to start with a calliper width of 20% of the standard deviation of the covariates, adjusting the calliper if needed to obtain balance within the covariates. Augurzky und Kluve (2004) argue that callipers which are not too narrow are preferable when the heterogeneous effects of treatment are expected.

1.3 Difference-in-Difference Estimator

One of the drawbacks of the matching method is that it only conditions on observable covariates and leaves out hidden biases from unobservable covariates (Ankarali et al., 2009). To overcome this problem, Smith and Todd (2005) recommend the implementation of a *difference-in-difference (DiD)* estimator. The *DiD* relies on the assumption that the differences of participants and non-participants are similar at every time. It is computed as the difference of the progress of the participant and the non-participant from one point before (t') to one point after (t) the time of treatment (t_T) (Heckmann et al., 1998). By implementing the factor time and the before- and after-estimation in the analyses, we can monitor for unobservable, linear and time-invariant effects such as price fluctuations (Gensler et al., 2005). The combination of matching and *DiD* results in the *Conditional difference-in-*difference (CDiD) estimation and the used formula can be written as

$$ATT = \sum_{i=1}^{n} (Y_{i,t} - Y_{i,t'})/n_i - \sum_{j=1}^{n} (Y_{j,t} - Y_{j,t'})/n_j \qquad t' < t_T < t \qquad (4)$$

1.4 Variable Specification and Database

We compare *DM* using a small number and *PSM* using a larger number of covariates. We apply *Radius Matching* with replacement in both approaches. The *Direct Matching* approach matches participants and non-participants on the following variables: farm type, region, organic farming, livestock dairy production, output, non-farm income and age. These covariates are chosen on statistical grounds (high z-Value in logit-regression) and the specific distribution of farm-investment payments focusing on mountainous dairy farms (see Section 0). Whereas the dummy variables are set to equal, for metric variables a specific calliper is used (see Table 1).

Table 1: Used covariates and callipers

Covariates (2003)	Callipers
Forage farms	Dummy
Cash crop farms	Dummy
Granivore farms	Dummy
Region west	Dummy
Region south	Dummy
Livestock	+/30%, or +/-10 LU
Dairy	Dummy
Total farm output	+/-30%, or +/-25,000 Euro
Nonfarm income	+/-50%, or +/-1,000 Euro
Age of farm manager	+/-15 years

Because of the possibility of omitted variables, we include more covariates in our model. This would lead to a severe matching problem which is discussed earlier (see Section 1.2.) Therefore we compute the PS based on the covariates listed in Table 7 using a logit-regression. Table 7 also includes the estimates and z-values from the logit regression. The variables used in the model do not describe the decision to participate very well (Pseudo-R^2=0.11) but Lechner (2001) argues that too effective a set of predictors of participation might be a source for violating the Common Support Assumption. Furthermore we not only match on the estimated PS but on the farm types forage, granivore and cash crop as well.

The farm type is included in the matching process, because it plays a major role in the probability of receiving treatment and in the economic outcome. We should not match different farm type, which is not given by the propensity score.

For our analysis, the pre-treatment situation is in 2003, post-treatment is 2009 and the treatment itself took place between 2005 and 2008. The applied pre-treatment estimation for *matching* ensures that the *matching* covariates are not influenced. The two-year gap before treatment is necessary, since the year of treatment is the year of payment and the investment usually happens a year or two before payment. We use a panel data from 2000 to 2009 of 1,636 voluntary bookkeeping farms in Austria. We found 195 farms who have only participated in the farm-investment support programme at least once between 2005 and 2008 and 845 farms who did not participate between 2000 and 2009. Participants and non-participants are matched with data based on the year 2003.

Empirical Results

1.5 Results using *Direct Matching*

The D*irect Matching* approach identified 122 pairs of participants and controls, whereas the controls consist on average of 5 non-participating farms. Table 1 shows the mean differences of selected variables for potential participants and non-participants (controls) before and after the matching. Before matching, most of the variables differ between the two groups significantly. This is especially obvious for the variables Livestock and total farm income. Matching reduces the mean differences. Thus they are not statistically significant, which can be considered as successful matching (Diwish et al., 2009). The last row in Table 2 illustrates the means of the 73 participants not selected. This group of farms is not selected because no control was found. The group mainly consists of big forage and granivore farms. The 122 pairs are used for computing the ATT based on the CDiD. A positive (negative) and statistically significant value for ATT shows that the farm-investment programme has a positive (negative) effect on structural farm growth.

The results in Table 3 show different developments between selected participants and controls. While participants increase their livestock by 1.9 livestock units (LU), controls reduce livestock by 0.7 LU. This results in a positive effect of 2.6 LU caused by investment and investment support. The relatively big structural growth of participating farms enables them, on the one hand, to increase their total farm output by roughly 21,000 Euros but enhances, on the other, their total farm inputs by roughly 17,000 Euros. When this is compared with selected controls, it results in an effect of roughly 15,000 and 9,000 Euros respectively. This leads to an effect of more than 5,000 Euros in farm income. Furthermore, farm investment including support has the effect on those selected farms of increasing depreciation, equity and especially debts. The computed ATT is, in almost all variables which are shown in Table 3, positive and statistically significant at the 5% level. Even though those effects are of positive value, mean values do not necessarily count for all participating individuals.

Table 2: Mean values of variables for participants and controls before and after matching

	before matching			after matching			
	Potential participants	Potential controls	t^1	Selected controls	Selected participants	t^1	Not selected participants
Number of Farms	195	845		122	122		73
Forage farms (share)	0.43	0.35	**-2.15**	0.49	0.49	0.00	0.33
Cash crop farms (share)	0.12	0.28	**5.68**	0.12	0.12	0.00	0.11
Granivore farms (share)	0.18	0.10	**-2.99**	0.12	0.12	0.00	0.29
Dummy variable 'west'	0.08	0.09	3.74	0.59	0.59	0.00	0.38
Dummy variable 'south'	0.41	0.25	**-4.16**	0.34	0.34	0.00	0.52
Dummy variable 'north'	0.51	0.66	**3.74**	0.07	0.07	0.00	0.10
Age of farm manager	50.98	53.29	**3.35**	50.39	51.75	1.62	51.97
Livestock (LU)	26.57	18.29	**-5.59**	23.59	22.27	0.67	36.59
Livestock density (LU/ha)	1.15	0.88	**-0.06**	1.09	1.08	-0.11	1.28
Milk sold (kg)	41,399	27,344	**-3.14**	43,287	40,525	-0.44	38,244
Equity (Euro)	346,642	296,673	**-3.27**	333,931	312,288	-1.02	367,885
Debts (Euro)	41,219	29,294	-1.50	27,385	35,452	-0.61	64,338
Total farm input (Euro)	78,336	54,682	**-3.91**	57,808	54,242	-0.79	112,645
Total farm output (Euro)	106,426	78,829	**-3.88**	81,015	78,771	-0.37	148,892
Depreciation (Euro)	14,544	12,333	**-3.86**	13,693	12,930	-0.90	15,966
Non-farm Income (Euro)	8,458	8,752	0.33	6,699	5,954	-0.67	11,397
Farm income (Euro)	28,089	24,147	**-2.05**	23,208	24,529	0.53	36,247
Programme payments (Euro)[2]	15,844	-		13,368	-		19,981

[1)] t-test for equality of means: Bold numbers indicating significantly different means between observations from the potential (selected) participants group and from the potential (selected) control group at the 5 per cent level

[2)] Payment of farm-investment programme (measure 121) from 2005 to 2008

Table 3: Mean developments of selected participants and controls from 2003 to 2009 for selected variables and the ATT-effect

	Selected participants		Selected controls			
	Mean	*sd*	Mean	*sd*	*ATT*	t^1
Livestock (LU)	1.92	*11.44*	-0.71	*4.85*	2.64	**-2.34**
UAA (ha)	2.51	*6.40*	1.02	*6.26*	1.49	**-1.83**
Livestock density (LU/ha)	0.71	*37.54*	-3.06	*45.12*	3.77	-0.70
Total farm input (Euro)	17,234	*28,002*	7,883	*14,279*	9,351	**-3.28**
Total farm output (Euro)	21,400	*33,980*	6,730	*17,963*	14,670	**-4.21**
Farm income (Euro)	4,165	*19,185*	-1,153	*15,617*	5,319	**-2.37**
Equity (Euro)	49,942	*110,999*	25,768	*58,481*	24,174	**-2.12**
Debts (Euro)	35,236	*114,664*	-1,758	*31,576*	36,994	**-3.43**
Depreciation (Euro)	2,973	*6,277*	452	*2,586*	2,522	**-4.10**

[1)] t-test for equality of means: Bold numbers indicating significantly different means between observations from the potential (selected) participants group and from the potential (selected) control group at the 5 per cent level

sd=Standard deviation

1.6 Results using *Propensity-Score Matching*

Out of 195 potential participants, *Propensity-Score Matching (PSM)* develops a new sample with 165 pairs. For each of the selected participants on average can be found. Through this, the sample increased its balance between the two groups (participants and controls) for all variables (see Table 4). Even when PSM is used, not all participants are selected. This might be because there is a group of farms where everyone has participated and this is especially the case for big granivore farms with high propensity score (see Table 4).

With the new sample of 165 pairs gained from PSM the ATT is computed by comparing the mean developments from 2003 to 2009 of participants and controls. The development in selected variables of both groups and the ATT is displayed in Table 5. Participants extended with the supported investment their production in livestock and raised the total farm output and equity. But they also increased the total farm input, debts and depreciation, as well as their farm income. When it is subtracted from the development of controls, a positive effect for all variables result. With the exception of farm income, the results shown in Table 4 are statistically significant for all of them. The results for farm income are expected still to be underestimated, as the big granivore farms with investment were mainly lost through the application of the caliper.

Table 4: Mean values of variables for participants and controls before and after *Propensity-Score Matching*.

	before matching			after matching			
	Potential participants	Potential controls	t^1	Selected participants	Selected controls	t^1	Not selected participants
Number of farms	195	845		165	165		30
Forage farms (share)	0.43	0.35	**-2.15**	0.49	0.49	*0.00*	0.10
Cash crop farms (share)	0.12	0.28	**5.68**	0.10	0.10	*0.00*	0.20
Granivore farms (share)	0.18	0.10	**-2.99**	0.13	0.13	*0.00*	0.47
Dummy variable 'west'	0.08	0.09	3.74	0.09	0.11	*0.87*	0.00
Dummy variable 'south'	0.41	0.25	**-4.16**	0.36	0.32	*-0.82*	0.70
Dummy variable 'north'	0.51	0.66	**3.74**	0.55	0.57	*0.31*	0.30
Age of farm manager	50.98	53.29	**3.35**	51.47	51.36	*-0.14*	48.27
Livestock (LU)	26.57	18.29	**-5.59**	25.19	26.92	*0.97*	46.43
Livestock density (LU/ha)	1.15	0.88	**-5.6**	1.09	1.09	*-0.02*	1.52
Milk sold (kg)	41,399	27,344	**-3.14**	43,867	46,384	*0.40*	27,828
Equity (Euro)	346,642	296,673	**-3.27**	329,943	334,539	*0.25*	438,486
Debts (Euro)	41,219	29,294	-1.50	37,086	36,864	*-0.03*	63,950
Total farm input (Euro)	78,336	54,682	**-3.91**	62,881	64,183	*0.32*	163,342
Total farm output (Euro)	106,426	78,829	**-3.88**	88,089	90,447	*0.44*	207,277
Depreciation (Euro)	14,544	12,333	**-3.86**	13,719	14,425	*1.07*	19,081
Non-farm Income (Euro)	8,458	8,752	0.33	8,152	8,146	*-0.01*	10,179
Farm income (Euro)	28,089	24,147	**-2.05**	25,208	26,264	*0.52*	43,935
Propensity score	0,27	0,17	**-9,29**	0,24	0,24	*0,00*	0,46
Programme payments (Euro)[2]	15,844	-		15,061	-		20,153

1) t-test for equality of means: Bold numbers indicating significantly different means between observations from the potential (selected) participants group and from the potential (selected) control group at the 5 per cent level

2) Payment of farm-investment programme (measure 121) from 2005 to 2008

Table 5: Mean developments of selected participants and controls from 2003 to 2009 for selected variables and the ATT-effect

	Selected participants		Selected controls			
	Mean	sd	*Mean*	sd	*ATT*	t^1
Livestock (LU)	2.65	*12.32*	-0.67	*5.62*	3.32	***-3.10***
UAA (ha)	3.31	*7.92*	1.23	*5.28*	2.08	***-2.80***
Livestock density (LU/ha)	0.02	*0.35*	-0.08	*0.23*	0.10	***-2.97***
Total farm input (Euro)	21,944	*36,615*	10,162	*12,694*	11,783	***-3.91***
Total farm output (Euro)	26,319	*52,453*	11,368	*17,214*	14,952	***-3.48***
Farm income (Euro)	4,375	*25,032*	1,206	*12,556*	3,169	*-1.45*
Equity (Euro)	63,692	*138,845*	23,450	*56,589*	40,242	***-3.45***
Debts (Euro)	38,261	*111,484*	-1,586	*24,777*	39,847	***-4.48***
Depreciation (Euro)	3,600	*6,457*	964	*2,819*	2,636	***-4.81***

[1)] t-test for equality of means: Bold numbers indicating significantly different means between observations from the potential (selected) participants group and from the potential (selected) control group at the 5 per cent level

1.7 Comparison of Approaches

So far we have only looked at the results of *Direct Matching (DM)* and *Propensity-Score Matching (PSM)* separately. The following chapter compares them, looking in particular at the quality of matching and the results. There is an ongoing discussion about how the quality of matching can be measured and what the best method is. An overview of this is given in Reinowski (2008) or Caliendo and Kopeining (2008). Smith and Todd (2005) argue that there is no statistically assured methodology for judging the quality of matching. Therefore we simply rely on the t-test to compare the means before and after the matching which are already displayed in Table 2 Table 4. They show that both approaches manage to balance all mean values of the variables after matching.

Using the *Propensity-Score*, instead of all covariates separately, reduces the matching dimension and therefore increases the chance of identifying controls for participants. Through this the PSM sample contains 23 farms more than the DM sample. Our results show that including more farms in the analysis does not result in dissenting effects between these two approaches. When we look again at Table 3 and Table 5 we find statistically significant effects for the same variables in both approaches similarly. Whereas Figure 2 shows the mean ATT-values for selected variables, Table 6 displays standard deviation and the t-value for testing differences in means between the effects using DM and PSM additionally. Even though there are no significant effects for a few variables or relatively big differences in absolute mean values, the t-test indicates no statistical significant difference when the two approaches (DM and PSM) are compared for differences in mean effects (see Table 6).

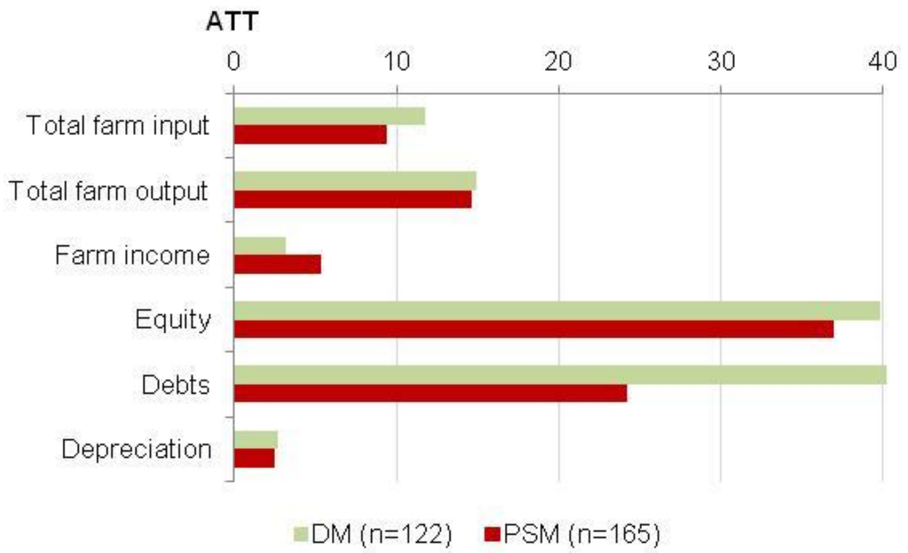

Figure 2: ATT-values for selected variables using *Direct Matching* (DM) and *Propensity-Score Matching* (PSM)

Table 6: Comparison of mean standard deviation (sd) of ATT-values for *Direct Matching* (DM) and *Propensity-Score Matching* (PSM)

	DM (n=122)		PSM (n=165)		
	Mean	sd	*Mean*	sd	*t*[1]
Livestock (LU)	2.64	11.70	3.32	13.81	-0.45
UAA (ha)	1.49	8.96	2.08	8.87	-0.55
Livestock density (LU/ha)	0.04	0.64	0.10	0.40	-0.89
Total farm input (Euro)	9,351	30,902	11,783	39,778	-0.58
Total farm output (Euro)	14,670	34,147	14,952	58,573	-0.50
Farm income (Euro)	5,319	23,808	3,169	29,460	0.68
Equity (Euro)	36,994	118,040	39,847	113,495	-0.21
Debts (Euro)	24,174	120,475	40,242	142,062	-1.03
Depreciation (Euro)	2,522	7,464	2,636	6,989	-0.13

[1)] t-test for equality of means: Bold numbers indicating significantly different means between observations from the potential (selected) participants group and from the potential (selected) control group at the 5 per cent level

Discussion and conclusions

Because of an increasing need for a quantitative evaluation of farm policy measures, we analyse the application of the matching method to evaluate Rural Development measures. Whereas the matching method is commonly applied in medicine and labour-market analysis, up to now there have been only a few studies concerning agricultural policy (e.g. Pufahl and Weiss, 2009b). Basically, matching creates a new sample by identifying similar controls for each participating individual based on observed covariates. The selection of these covariates is a central issue and of high sensitivity. It is necessary to identify those variables which have the greatest influence on the decision to participate and on the outcome. Within matching, there is a multitude of different approaches varying by distance functions and selection algorithm. In this paper we apply *Direct Matching (DM)* and *Propensity-Score Matching* (PSM) in combination with the *difference-in-difference estimator* to assess causal effects of the farm-investment programme in Austria. It is not our aim to focus on the measured effects of farm-investment support, but on the differences of the approaches mentioned above. Whereas the first approach directly matches covariates, the latter on the *Propensity-Score,* an aggregate distance function. Using DM, the number of covariates used is restricted, since matching success decreases and time and effort increase with each variable, to a dramatic extent. With PSM this problem can be overcome, as an aggregate distance function is applied and matching dimension is reduced to one. For both cases a radius algorithm was used. The radius algorithm implies a calliper on each participant to define similarity with controls. This might lead to a loss of participants, when no control is found.

When DM is applied to a dataset consisting of 195 participating and 845 non-participating farms in Austria, 122 pairs are identified. This leads to a well balanced dataset, but also to a loss of 73 individuals. These farms tend to be bigger and have a slightly higher share of granivore farms than the selected sample. The mean results show a positive significant effect in total farm output and farm income but also in total farm input and debts. This indicates that it is hard for farms to gain income right after the investment. Given the reality of a heterogeneous Austrian farming structure and the heterogeneous goals of the programme, heterogeneous effects are expected. This also means that each participant carries important information about the treatment effect and each loss of participant might increase the bias (Augurzky und Kluve, 2004). Therefore the relatively low effect on farm income might be also due to the fact of high share of forage farm in the matched sample. Forage farms have by trend lower farm income (BMLFUW, 2010) than cash crop and granivore farms and need more time to take full advantage of their investment.

The PSM approach increases the number of pairs in the selected sample to 165. In this case the lost individuals are mainly big granivore farms with more livestock and high total farm output and equity. This new sample can achieve a statistically perfect balance between participants and controls in all variables as well. Comparing these findings with the results from DM, we can conclude that even when PSM is used and more variables and more participants are included, the matching quality does not vary. Furthermore, the computed ATT has no statistically significant difference between these two approaches. We observe no differences in the outcome of farm-investment support evaluation when the most non-parametric approach is compared with PSM using a parametric regression. Ravallion (2005) notes as well that seemingly sophisticated non-experimental methods will not always perform better. This is especially the case when the inclusion of more covariates cannot describe the decision to participate better, but there might be also cases when this is applicable. Therefore, a large amount of work should be put into pooling information and applying covariates which are plausible for the institutional environment, in which the study is carried out (Lechner, 2002).

Even though the DM approach is dependent on several assumptions, next to individual adjustments it allows transparency for non-scientific stakeholders in the evaluation process. This is particular necessary as practical information is important to find covariates. Furthermore, it shows the advantage of easily communicated results. We would like to stress that policy evaluation must be carried out with and for stakeholders and not only for scientists.

We acknowledge that further research has to be done on identifying covariates and their influence on participation as well as sensitivity analysis. The model and results can be improved by using qualitative data, but hidden bias might still remain, as the decision to participate in the farm-investment support programme is often due to the need for investment. We would also point out that we never know if a farmer would have invested in, for example, new building without federal support. This illustrates the complex effects of this measure and challenges for evaluation. However, we find that the approach used, in combination with pre-studies and stakeholder information, can help towards a consistent farm-policy evaluation in Rural Development programmes.

Acknowledgements

We are grateful to the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (Division II/5) for funding the project on farm-investment support programme evaluation.

Literature

ANKARALI, H.C., V. SUMBULOGLU, A.C. YAZICI, I. YALUG , M. SELEKLER (2009): Comparison of Different Matching Methods in Observational Studies and Sensitivity Analysis: The Relation Between Depression and STAI-2 scores. In: Expert Systems with Applications 36, 1876-1884

AUGURZKY, B., J. KLUVE (2004): Assessing the Performance of Matching Algorithms - When Selection into Treatment Is Strong. Forschungsinstitut zur Zukunft der Arbeit, Diskussionspapier No. 1301

BECK, M., T. DOGOT (2006): The Use of Impact Indicators for the Evaluation of Farm Investment Support – A Case Study Based on the Rural Development Programme for Wallonia (2000 – 2006). In: BERGSCHMIDT A., W. DIRKSMEYER, J. EFKEN, B. FORSTNER, I. UETRECHT (ed.) Proceedings of the European Workshop on the Evaluation Farm Investment Support, Investment Support for Improvement of Processing and Marketing of Agricultural Products. Arbeitsberichte des Bereichs Agrarökonomie 03/2006, 69-77

BLUNDELL R., L. DEARDEN, B. SIANESI (2005): Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. In: J.R. Statist. Soc.A (2005) 168, Part 3, 473-512

BMLFUW (2010): Grüner Bericht 2010. Wien

BRADY, H. E. (2008). Causation and explanation in social science. In: BOX-STEFFENSMEIER, J. M., H. E. BRADY, D. COLLIER (ed.), The Oxford handbook of political methodology. Oxford: Oxford University Press, pp. 217–270

CALIENDO, M., R. HUJER (2006): The Mircoeconometric Estimation of Treatment Effects – An Overview. In: Allgemeines Statistisches Archiv 90, 199-215

CALIENDO, M., S. KOPEINING (2008): Some Practical Guidance for the Implementation of Propensity Score Matching. In: Journal of Economic Surveys (2008) Vol. 22, No. 1, pp. 31–72

DANTLER, M., S. KIRCHWEGER, M. EDER, J. KANTELHARDT (2010): Analyse der Investitionsförderung für landwirtschaftliche Betriebe in Österreich. Universität für Bodenkultur, Institut für Agrar- und Forstökonomie, Wien.

DIRKSMEYER, W., B. FORSTNER, A. MARGARINA, Y. ZIMMER (2006): Aktualisierung der Zwischenbewertung des Agrarinvestitionsförderprogramms (AFP) in Deutschland für den Förderzeitraum 2000 bis 2004. Länderübergreifender Bericht. Bundesanstalt für Landwirtschaft (FAL), Braunschweig

DIWISCH, D.S., P. VOITHOFER, CH.R. WEISS (2009): Succession and Firm Growth: Results from a Non-parametric Matching Approach. In: Small Business Economics 32:45-46

GENSLER, S., B. SKIERA, M. BÖHM (2005): Einsatzmöglichkeiten der Matching Methode zur Berücksichtigung von Selbstselektion. In: Journal für Betriebswirtschaft 55, 37-62

GODTLAND, E.M., E. SADOULET, A. DE JANVRY, R. MURGAI, O. ORTIZ (2005): The Impact of Farmer Field Schools on Knowledge and Productivity: A Study of Potato Farmers in the Peruvian Andes. In: Economic development and cultural change, Vol. 53, No. 1 (October 2004), pp. 63-92

HECKMAN, J. J., H. ICHIMURA, J. A. SMITH, P. E. TODD (1998): Characterizing Selection Bias Using Experimental Data. In: Econometrica 66 (5), 1017–1098

HECKMAN, J. J., R. J. LALONDE, J. A. SMITH (1999): The Economics and Econometrics of Active Labor Market Programs. In: ASHENFELTER, O., D. E. CARD (ed.): Handbook of Labor Economics. Band III, Amsterdam: Elsevier Science B.V., 1865–2097

HENNING, C.H.C.A., J. MICHALEK (2008): Ökonometrische Methoden der Politikevaluation: Meilensteine für eine sinnvolle Agrarpolitik der 2. Säule oder akademische Finderübung. In: Agrarwirtschaft 57(3/4), 232-243

LECHNER, M. (2001): A Note on the Common Support Problem in Applied Evaluation Studies. University of St. Gallen, Department of Economics, Discussion Paper no. 2001-01

LECHNER, M. (2002): Mikroökonomische Evaluation arbeitspolitischer Maßnahmen. University of St.Gallen, Department of Economics, Discussion Paper No. 2002-20

LUKESCH R., B. SCHUH (2010): Working Paper on Approaches for Assessing the Impacts of the Rural Development Programmes in the Context of Multiple Intervening Factors. Findings of a Thematic Working Group established and coordinated by the European Evaluation Network for Rural Development

NEYMAN, J. [1990 (1923)].On the Application of Probability Theory to Agricultural Experiments Essay on Principles. In: Sec. 9 Stat. Sci. 5(4):465–72. Transl. DM Dabrowska, TP Speed

PFEFFERLI, S. (2006): Impact Analysis of Investment Support for Agricultural Buildings in Switzerland. In: BERGSCHMIDT A., W. DIRKSMEYER, J. EFKEN, B. FORSTNER, I. UETRECHT (ed.) Proceedings of the European Workshop on the Evaluation Farm Investment Support, Investment Support for Improvement of Processing and Marketing of Agricultural Products. Arbeitsberichte des Bereichs Agrarökonomie 03/2006, 69-77

PUFAHL, A. (2009): Empirische Wirkungsanalyse direkter Transferzahlungen – am Beispiel von Agrarumweltmaßnahmen und der Ausgleichszulage für benachteiligte Gebiete. Dissertation, Georg-August Universität Göttingen

PUFAHL, A., CH.R. WEISS (2009a): Farm Structure and the Effects of Agri-Environmental Programs: Results from a Matching Analysis for European Countries. 111 EAAE-IAAE Seminar ‚Small Farms: decline or persistence'. University of Kent, Canterbury;UK 26th – 27th June 2009

PUFAHL, A., CH.R. WEISS (2009b): Evaluating the Effects of Farm Programmes: Results from Propensity Score Matching. In: European Review of Agricultural Economics Vol 36 (1) (2009) pp. 79–101

REINOWSKI, E. (2008): Matching kleiner Stichproben. Ein Vergleich verschiedener Verfahren. Dissertation, Martin-Luther-Universität Halle-Wittenberg

ROSENBAUM, P.R. (2010): Design of Observational Studies. New York: Springer

ROY, A. (1951): Some Thoughts on the Distribution of Earnings, Oxford Economic Papers, 3, 135-145

RUBIN, D.B. (1974): Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies. In: Journal of Educational Psychology, 66, 688-701

RUBIN, D.B. (1977). Assignment to Treatment Group on the Basis of a Covariate. In: Journal of Educational Statistics, 2, 1–26

RUBIN, D.B., P.R. ROSENBAUM (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects. In: Biometrika 70(1):41–55

SALHOFER, K., G. STREICHER (2005): Self-selection as a Problem in Evaluation Agri-environmental Programs. In: ORTNER, K.M. (ed.): Assessing Rural Development Policies of the Common Agricultural Policies. Selection of Papers from the 87th EAAE-Seminar, Kiel: Vauk, 203-213

SEKHON, J.S. (2008): The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. In: BOX-STEFFENSMEIER, J. M., H. E. BRADY, D. COLLIER (ed.), The Oxford handbook of political methodology. Oxford: Oxford University Press, pp. 271–299

SEKHON, J.S. (2009): Opiates for the Matches: Matching Methods for Causal Inference. In: Annu. Rev. Polit. Sci. 2009.12:487-508

SMITH, J.A., P.E. TODD (2005): Does Matching Overcome LaLonde's Critique of nonexperimental Estimators? In: Journal of Econometrics 125, 305-353

STRIEWE, L., J.P. LOY, U. KOESTER (1996): Analyse und Beurteilung der einzelbetrieblichen Investitionsförderung in Schleswig-Holstein. In: Agrarwirtschaft 45 (12), 423-434

Appendix

Table 7: Parameter estimates of logit-models explaining programme participation

Variable	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-13.032268	2.319103	-5.620	**1.91e-08**
Dummy variable ' forage farm '	0.559873	0.264247	2.119	**0.03411**
Dummy variable ' granivore farm '	0.356238	0.333994	1.067	0.28615
Dummy variable ' cash crop farm '	-0.818824	0.343969	-2.381	**0.01729**
Dummy variable 'west'	0.888726	0.212176	4.189	**2.81e-05**
Dummy variable 'south'	0.007920	0.351632	0.023	0.98203
Dummy variable 'north'	0.025015	0.226506	0.110	0.91206
Log (livestockdensity)	0.215732	0.512586	0.421	0.67385
Log (livestock)	-0.017511	0.146198	-0.120	0.90466
Log (milk sold)	-0.008988	0.023917	-0.376	0.70706
Log (debts)	0.010552	0.023719	0.445	0.65642
Log (share of equity)	1.030567	1.020374	1.010	0.31250
Log (total farm output)	0.850258	0.192395	4.419	**9.90e-06**
Log (non-farm income)	0.114204	0.040893	2.793	**0.00523**
Log (assets for farm buildings)	0.120519	0.146115	0.825	0.40947
Log (assets for fruit production)	-0.005043	0.038708	-0.130	0.89633
Log (assets for wine production)	0.059635	0.033371	1.787	**0.07393**
Age of farm manager	-0.027472	0.009943	-2.763	**0.00573**
Pseudo R^2	0.11			

[1)] Bold numbers indicating significant influences at the 10 per cent level