Programme evaluation with micro-data: the use of FADN data to evaluate effects on the market situation of programme participants

Franz Sinabell and Gerhard Streicher[1]

Abstract

Community legislation requires that the programme for rural development is evaluated. A detailed set of indicators was developed by the commission and member states are required to use them to measure program success. To make results comparable requires that the methodology by which indicators are measured is well defined. However, this is not the case in this evaluation task. We take a small set of indicators and use two different approaches using the same sample and show that the results are deviating. We conclude that the success of this programme can only be compared across measures/regions when both, indicators and method are the same.

Keywords: rural development, microeconometrics, socio-economic evaluation

[1] Franz Sinabell: WIFO - Austrian Institute of Economic Research Vienna, Arsenal Obj. 20, P.O.Box 91, A-1103 Vienna, franz.sinabell@wifo.ac.at; Gerhard Streicher, Joanneum Research, Institut für Regional- und Technologiepolitik, Wiedner Hauptstr. 76, A-1040 Vienna, gerhard.streicher@joanneum.at; the authors equally share responsibility for this paper. Efficient research assistance by Dietmar Weinberger is gratefully acknowledged.

1. Introduction

The EU is committed to rational policy making and accountability. Tight public budgets and concurrent uses of public funds for different policy fields is stimulating the to strive for evaluation of the effectiveness of policies. Detailed rules for monitoring of the application of the Rural Development Regulation (No. 1257/99) are outlined in a monitoring regulation (No. 445/02), which specifies that an evaluation has to be carried out by reference to pre-agreed specific physical and financial indicators. Monitoring of this program is a continuous process, carried out during its execution, with the intention of correcting any deviation from operational objectives and of improving programme performance. Member states are required to submit ex ante, mid-term and ex post evaluations for measures implemented under that regulation for the period 2000-2006. By the end of 2003, the mid-term evaluation reports were submitted to the Commission. The final evaluation report is due in 2006.

In the mid term review, an assessment was made about programme performance so far and lessons learned during that phase will be an input for the design of the follow up programme which will be drafted in 2005. In order to facilitate comparability of findings from different authorities and/or member states on the same measures, projects evaluators had to use the same benchmarks, sets of criteria and indicators which were grouped together in nine chapters and six cross-cutting questions (see European Commission, 1999a, 2000, 2002).

In 2002, subsidies on agricultural products amounted to EUR 535 millions in Austria (Sinabell, 2003). This amount represents approximately 10 % of the total value of output of the agricultural sector at producer prices. Other support measures are even more important in Austria (EUR 1.2 billions - equivalent to 23 % of the total value of output). Most of these transfers are financed by the programme for rural development. Its budget is 6,7 bill.€ - including an EU-contribution of 49 % - over a period of seven years. The Austrian agri-environmental programme and compensatory allowances for farms in less favoured areas (in total, EUR 900 millions in 2002) are the two most important components of this program.

Several studies were carried out in Austria to evaluate socio-economic aspects of this programme. Beyond qualitative assessments (Pfusterschmid and Kamelott, 2003) and case-study analyses (e.g. Tamme, 2003, Wagner and Parizek, 2003a and 2003b) evaluators used a plethora of evaluation techniques: analysis of accounting data (Pistrich and Preinstorfer, 2003), surveys among representative samples of participants (Ortner and Simon, 2003), statistics of administrative files (Janetschek, 2003a; Müller and Pröll, 2003), and model calculations (Janetschek, 2003b).

Some of these studies had to evaluate the same topic with the same or rather similar indicators. However, in order to truly obtain comparable results across regions (and/or programs) it is necessary to use the same methods as well. Compared to the scrutiny of the EU Commission with respect to the relevant indicators, evaluators are given considerable leeway to choose the methodology they deem appropriate and operational. An EU handbook on the evaluation of socio-economic programmes is not very specific on subtle methodological issues either (European Commission, 1999) and therefore deviating results on the same evaluation topic are to be expected even if the same indicator is used.

Our paper addresses this issue by looking at the implications of the choice of methodology on evaluation results. Our results are based on book-keeping data of a sample of over 2,000 farms. We are using observations over four years from the Austrian FADN (farm accountancy data network) and identify participants in the programme for rural development among these farms. Thus we are able to employ micro-econometric tools for the evaluation of the economic effects of single measures from this program.

The remainder of the paper is structured as follows: In the next chapter we compare different approaches to the problem of programme evaluation at an abstract level. Then we describe the sample which was used in the quantitative analysis of the market position of participant of several measures or the programme for rural development. Findings from two different ap-

proaches (comparison of participants with non-participants versus results from a difference-in difference estimate) are presented before discussing our results.

2. The programme for rural development and the evaluation problem

Even if not explicitely stated, the principal problem in any evaluation is the question of „what if". This question is most intuitively asked from a participant's point of view: what would have been the post-programme state of some target variable, if this participant had not been part of the programme under consideration. Of course, this state of the world is not directly observable; it is called the *counter-factual*. For any meaningful evaluation, it is however necessary to estimate this counter-factual value: the effect of the programme is nothing else but the difference between the (observable) actual and the (unobservable) counter-factual value of the target variable.

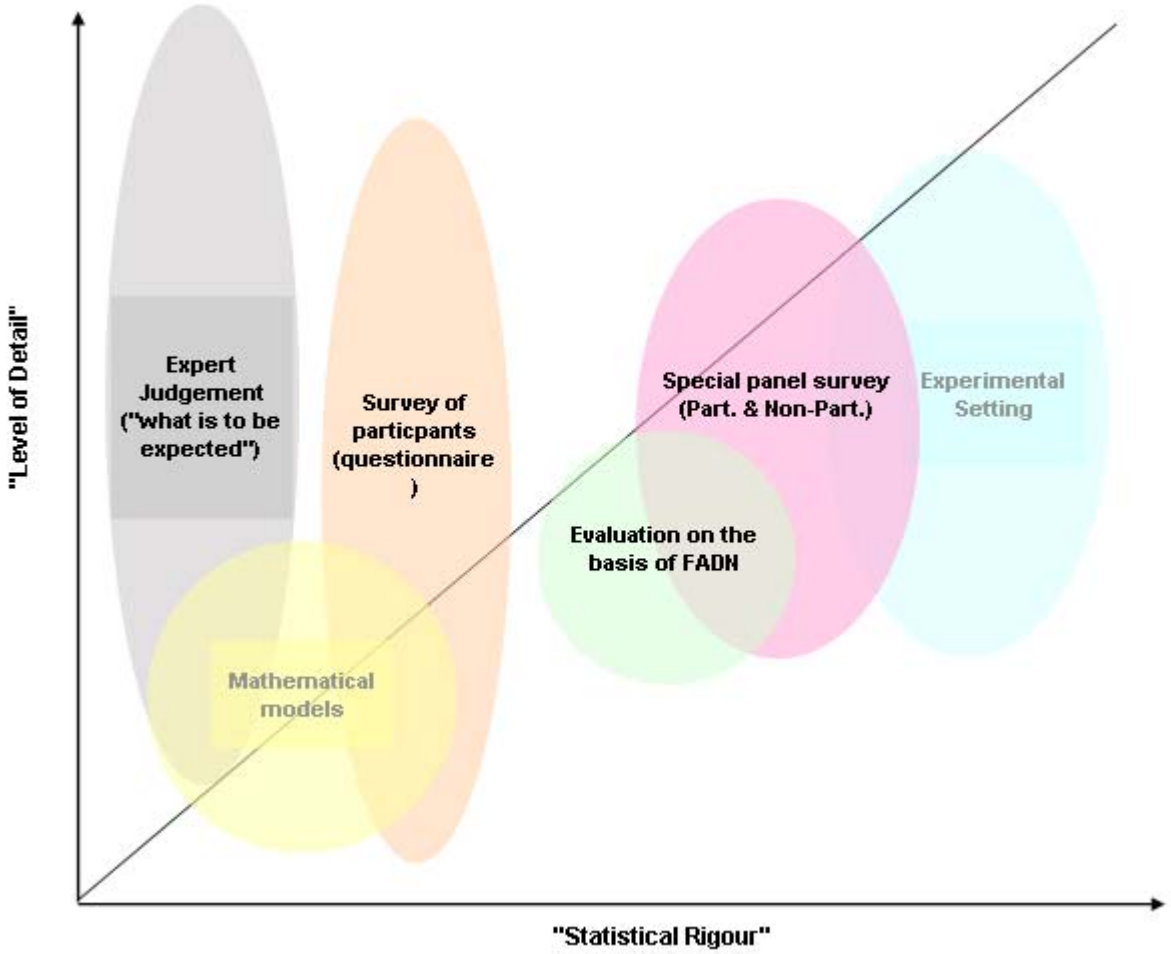
Not being directly observable, the counter-factual has to be estimated. Numerous approaches exist, ranging from more qualitative to expressly quantitative methods. In Fig. 1, we have attempted a critical appraisal of some of them, assessing their relative position along two dimensions: "level of detail" and "statistical rigour".

In the diagram, "level of detail" pertains to the kind of questions which can be answered: the higher up on this axis, the more questions can meaningfully be answered without pushing the inherent limits of each approach too far. "Expert judgement", in our perception, can go a long way in answering all sorts of questions. Its main drawback, on the other hand, is a lack of "statistical rigour": it is, after all, "only" expert judgement, possibly reflecting dearly held beliefs. A survey of participants, using questionnaires, can add some statistical rigour while probably losing some attainable level of detail.

To score high on "statistical rigour", mathematical methods are called for. We can envisage at least three basic approaches: a "dedicated panel survey", which presents a standardized questionnaire to both participants and non-participants (probably a full survey, but more likely a randomized survey), both before and after the implementation of some programme[2]. Such dedicated surveys are expensive. If, however, existing data, which are probably collected for some completely different reason, can be harnessed, a more cost-conscious approach can be pursued: for our purpose, FADN data might constitute just such a data base. The attainable level of detail is typically more limited, as FADN data do not necessarily contain all relevant variables. Also, FADN farms most likely do not constitute a statistically very sound sample (participation in such a network being voluntary and not entirely "cost free" (some time has to be devoted to this exercise on the farmer's part), farms which actually do take part conceivably exhibit some bias towards "sophistication"). On the other hand, the fact that FADN data are collected for different purposes might eliminate (or, at least, mitigate) one potentail problem with a dedicated survey, viz. strategic answering: in so far as farms perceive the survey as "supervision" they might be tempted to (probably unconsciously) tailor their answers somewhat to better conform to "expectations". FADN data have an additional advantage: as often happens with public programmes, evaluators' necessities are not considered in the design phase; rather, they pop up later in the process, probably not before the end of the programme. In such cases, any chances to use a dedicated panel survey are lost. FADN is much more flexible in this respect.

[2] The "experimental setting" is depicted as a purely fictitious approach: it would improve on statistical rigour by randomizing not only on the participants in the survey, but on the participants in the programme as well. In a real world setting, this is certainly not feasible: it would involve the random selection of participants which would then be required to take part in the programme. On the other hand, candidates not chosen by the random selection process would be forbidden to take part or even to change their ways of doing business. As an example: an experimental setting to assess the impact of bio-dynamic husbandry would involve forcing randomly chosen farms to switch modes whereas all other farms would be required to continue farming as usual – an unacceptably totalitarian approach.

Fig. 1: Assessment of approaches to the evaluation of rural development programmes



Source: own graph

The “odd method out” are mathematical models. Their limitations are set by their mathematical and statistical foundations: “optimizing models” share with expert judgement a probably rather low score on statistical rigour, whereas “econometric models” are probably rather limited in the scope of questions they can sensibly be asked. It might however be argued that the “natural place” of mathematical models is not so much within an ex post-evaluation setting, i.e. at the end of a programme, but rather at the beginning: their ability to model reactions to external stimuli should lend itself to ex ante-evaluations of proposed programme schemes, thus assigning such instruments a natural places in the design phase.

Mathematical models share with expert judgement a probably important drawback: their reliance on “what should be the case”, i.e. their leaning towards a normative approach. In estimating a programme’s effect on some target, they more or less assume that the rules of the programme lead to predictable changes in behaviour and outcome – and they tend to assume full compliance with these rules. Both assumptions, however, need not be borne out by reality. To varying degrees, the other approaches do not need those assumptions.

In the next chapters we propose a mixture of two approaches: expert judgement combined with the statistical analysis of FADN data. We were however not completely free in this choice: as no provisions for future evaluatory needs were taken in the design phase (except the stipulation for such a future evaluation), a dedicated panel survey was not provided for, rendering this approach unfeasible. As a result, typical examples of this midterm revision exercise relied heavily (or exclusively) on expert judgement and case studies. Instead, we want to argue that the supplement of these methods with quantitative methods based on FADN data is a worthwhile – and still rather cost-conscious – exercise. As we will also see,

the quantitative approach can serve to slightly dampen expectations as to the effects of the programmes in question.

3. The research question, the sample, and the method

The question addressed in this paper is: how would the market position of farmers be had the programme not been in place? A large body of literature is dealing with adequate approaches to address this counter-factual (see e.g. Blundell and Costas, 2000, Frölich, 2003, Wooldridge, 2002).

In the simplest case indicators measuring the market position of participants and non-participants are compared after the programme period. However, only under certain very specific conditions differences of indicators can be interpreted causally (as effected by the programme). A necessary precondition for such an approach is that participants and control group are statistically identical in all aspects apart from programme participation. Such ideal conditions are only given in an experiment where participation is the result of a random selection process, a social experiment. Programme participants are selected at random and the effects of programmes can be estimated by measuring the difference to non-participants.

The programme for rural development does not allow to employ such a mechanism because participation is voluntary and non-participants can not be forced to provide the necessary information. An alternative to the social experiment is the *before-after* comparison. In such an approach participants are asked to give information on the level of an indicator at two different points in time, before and after the programme. However, to conclude that differences are due to participation is problematic. In many cases factors completely unrelated to the programme in question may (and will) exert their influence on participants and non-participants alike. In a dynamic environment the framework of decisions and economic behaviour is changing continuously and therefore many different causes could explain a given outcome.

A method which combines these two approaches is the *difference-in-differences* estimator. This is quite data demanding: observations for two points in time (before and after the treatment) as well as two groups (participants and control group) are necessary. For both groups a before-and-after-difference is estimated. The comparison of these differences can give an indication of the effect of the treatment.

This method is not free from biases either. Structural characteristics can be the explanations for a development that could be falsely attributed to the programme. Such biases can be identified by measuring the correlation between the incidence of participation and structural variables. Econometric tools (multiple regression) can be used to separate programme effects from structural effects.[3]

[3] A non-parametric approach, pioneered inter alia by Heckman (see. for example, Heckman et al. (1999)), is called Matching Pairs analysis. Basically this involves identifying, for each participant, non-participant(s) which are "as similar as possible" (same size, same location, same products, etc.), Then, the differences between each participant and its non-participating twin are statistically evaluated. If conditioned on the "right" set of structural variables, this method can effectively mitigate selection bias.
For our purposes, however, this elegant approach seemed somewhat less intriguing: the whole programme of rural development consists of a range of measure, participation in more than one of which is not uncommon. For such a multi-line programme, however, the definition of "participants" and "non-participants" is by far not as straightforward as in a "single-purpose programme"

Tab. 1: Participants of the programme for rural development and sample characteristics during the period 2000 – 2002

programme	program participants	of which		
		IACS-Farms (not in FADN sample)	IACS-Farms (in FADN sample)	non-IACS recipients
	number of participants			
adaptation and development	3,225	1,684	46	1,495
training	9,638	8,327	384	927
forestry	4,450	3,188	89	1,173
investment	14,655	14,195	412	48
setting up young farmers	4,656	4,493	155	8
processing and marketing	142	18	0	124
wine	4,831	4,650	162	19
Total	41,597	36,555	1,248	3,794
	%-share per group of recipients			
adaptation and development	7.8	4.6	3.7	39.4
training	23.2	22.8	30.8	24.4
forestry	10.7	8.7	7.1	30.9
investment	35.2	38.8	33.0	1.3
setting up young farmers	11.2	12.3	12.4	0.2
processing and marketing	0.3	0.0	0.0	3.3
wine	11.6	12.7	13.0	0.5
Total	100.0	100.0	100.0	100.0
	%-share per measure			
adaptation and development	100.0	52.2	1.4	46.4
training	100.0	86.4	4.0	9.6
forestry	100.0	71.6	2.0	26.4
investment	100.0	96.9	2.8	0.3
setting up young farmers	100.0	96.5	3.3	0.2
processing and marketing	100.0	12.7	0.0	87.3
wine	100.0	96.3	3.4	0.4
Total	100.0	87.9	3.0	9.1

Source: BMLFUW, LFBIS, ASBIS, own calculations.

For our analysis we used data from the Austrian FADN (LBG and BMLFUW, 2003) over a period from 1998-2002. This large data set allows us to compare the simple approach of a participant-control-group comparison with the difference-in-differences method. We compare several indicators from the years 1998 and 2002. Using information from the IACS and from administrative records of programme participants we can identify which participated in which measure (in either 2001 or 2002). Two untested hypotheses are underlying our analysis: a) the effect of a measure can still be identified, even if a person participated in the year 2001 (the introduction of the program) and b) the effects of a very recent programme participation (2002) can already be identified in the indicators.

Austrian FADN-data do not meet all the criteria of a random experiment. The most important reason is that only volunteers are willing to make the efforts for providing a public good for a comparably small private benefit. Given that bias by self-selection, participants in this network probably are on average better educated than the whole farm population, and do not represent the size of Austrian farms very well because the sample distribution has truncated tails (the smallest and biggest farms are under-represented, see Bradler and Schneeberger, 1999).

More details on the sample and programme funds are presented in Tab. 1 and Tab. 2:

- the share of farms which are registered in the IACS relative to the population of programme participants
- the number of non-farm participants (among them educational institutions providing training)
- the number of participants from the FADN-sample who participated in some measure of the program

Tab. 2: Absorption of the rural development programme and sample characteristics during the period 2000 – 2002

	funds	allocated to		
		IACS-Farms (not in FADN sample)	IACS-Farms (in FADN sample)	non-IACS receipients
	mil. EUR			
adaptation and development	71.0	20.9	0.4	49.7
training	15.8	2.8	0.1	12.9
forestry	16.1	8.6	0.2	7.3
investment	107.9	103.6	3.3	0.9
setting up young farmers	42.9	41.4	1.5	0.1
processing and marketing	20.2	1.3	0.0	18.9
wine	24.6	23.5	0.7	0.4
Total	298.5	202.2	6.2	90.2
	%-share per group of recipients			
adaptation and development	23.8	10.4	6.0	55.1
training	5.3	1.4	1.4	14.3
forestry	5.4	4.3	2.7	8.1
investment	36.1	51.3	54.2	1.0
setting up young farmers	14.4	20.5	23.7	0.1
processing and marketing	6.8	0.6	0.0	21.0
wine	8.2	11.6	12.0	0.4
Total	100.0	100.0	100.0	100.0
	%-share per measure			
adaptation and development	100.0	29.5	0.5	70.0
training	100.0	17.6	0.5	81.9
forestry	100.0	53.6	1.0	45.3
investment	100.0	96.1	3.1	0.8
setting up young farmers	100.0	96.4	3.4	0.2
processing and marketing	100.0	6.5	0.0	93.5
wine	100.0	95.5	3.0	1.5
Total	100.0	67.7	2.1	30.2

Source: BMLFUW, LFBIS, ASBIS, own calculations.

Many farmers participated in more than one measure of the programme. A participation matrix is presented in Tab. 1. The most popular measure of those investigated is organic farming, followed by recipients of investment aids, and participants in training courses. In the econometric analysis we are taking account of multiple participation by evaluating the combined effect of e.g. "adaptation and development" plus "organic farming".

The sample size is 2,288 farms. Among them, 848 farms participated in at least one of the investigated measures (support for farms in less favoured areas and agri-environmental measures apart from organic farming are not analysed in our study). Structural variables of the participants in the sample are rather similar to non-participants in the sample (data not reported). However some of the variables have significantly higher levels among participants in measures: educational level, share of labour of total labour spent for farming, size of forests. Apart from these structural variables participants are not distinguishable from non-participants.

From the FADN data set indicators were selected that were used in the econometric analysis along with structural variables. The evaluation of the p-values of the differences between participants and the control group is the measure of programme success in the first method. A p-value of 0.1 means that the difference of an indicator of participating farms to non-participants is significantly different from zero at a significance level of 90%.

Tab. 3: Sample participants in the identified measures

	adaptation & development	training	forestry	invest-ment	young farmers	organic farming
adaptation and development	40	5	2	13	2	14
training		332	15	64	45	61
forestry			76	15	4	22
investment				350	69	78
setting up young farmers					151	34
organic farming						416

Source: LBG, ASBIS, BMLFUW, Joanneum Research, WIFO

In our second approach we measure the treatment effect with the difference-in-differences method: The general equations are:

$(I_{2002}-I_{1998}) = f(X_{2002}-X_{1998}, D_{measure}, D_{subregion})$

I_{2002}, I_{1998} .... level of indicator in 2002 and 1998;

X_{2002}, X_{1998} .... level of structural variable in 2002 and 1998[4];

$D_{measure}$ .... participation dummies.[5]

$D_{subregion}$ .... dummies of 89 bio-climatological production units.

In the equations the differences of the levels of indicators are regressed against the differences of structural variables and dummies of the bio-climatological geographical units[6] and the dummies of programme participation.

This approach is a sort of *fixed-effects* model: we are not looking for the absolute level of a given indicator, but we are interested in the temporal development. Using that method we can test e.g. the following hypothesis: did participants receive better milk prices? A given measure has a positive effect if participants experience a more favourable development in their milk prices than non-participants (it is not necessarily the case that participants face absolutely higher prices than non-participants). If the general prices change both, participants and non-participants are exposed to this development which is independent of the program. Thus temporal structural effects cancel out. The logic is similar concerning (more or less) time-invariant structural variable (attitude of farm, size of farm,...).

4. Results

Given the fact that our sample is not fully representative for Austrian agriculture, our analysis is not measuring the programme effect for the whole farm community. We measure the effect of the programme among farms that have characteristics similar to the sample of farms in the FADN data set.

The indicators that were used in the econometric analysis are: farm income and household income (including social transfers and off-farm income), farm revenues and farm expenditures and milk prices (calculated as average revenue). The evaluation of the p-values of the differences between participants and the control group is the measure of programme success in the first evaluation approach. The results are listed in Tab. ??. A '+' indicates that a p-value of 0.1 was calculated. This means that the difference of an indicator of participating

4 Typical variables are: acreage, education, part-time/full-time farming, production structure.

5 The value is 1 if a farm participated in a measure during 2000-2002; the dummy is 0 in case of non-participants. Organic and conventional farms were identified by a separate dummy variable.

6 Of course, geographical location is time-invariant and as such should be captured by the fixed effect. Our reasoning, however, was to include these geographical dummies as proxy for differences in reogional-specific characteristics, most important weather conditions: conceivably, favourable or adverse weather conditions can influence any agri-economic variable – price, quantity, ... – more effectively than the most efficient programme. The geographical dummies are included to allow for spatial variation in weather conditions.
The necessity to include such regional-specific influences would also pose a further problem in using matching methods (see also footnote 3): their use would drastically diminish the group of available "twins", as now the group of non-participating comparison farms would not only be required to share with participants a set of structural characteristics, but also to be located in the same geographical area.

farms to non-participants is significantly different from zero at a significance level of 90%. As outlined above, we controlled for structural differences.

Using this method we find that participants in almost all measures have a higher household income. Only the household income of organic farms is not statistically different from the rest of the sample. Participants in the measures investment, setting up of young farms, and training have higher farm incomes, as well. In this group farm revenues are higher too, as are expenditures. The figures in Tab. ?? also show that organic farms have lower farm expenditures along with lower revenues, a result that is consistent with our prior assumptions.

Tab. 4: Significant deviations of participants from non-participants in the sample in the year 2002

measure	1257/99 article	indicator of market position			
		income		farm	
		household	farm	revenues	expenditures
investment	4	+	+	+	+
setting up young farmers	8	+	+	+	
training	9	+	+	+	+
organic farming	22			-	-
forestry	29	+			
adaptation and development	33	+			
any measure		+	+	+	+

Source: own estimates base on LBG, ASBIS, BMLFUW

Given these results, can we conclude that the programme is a success or failure? Probably not because there may be factors we could not account for when we compared participants and non-participants. This can be demonstrated when we look at organic farming. This type of management was already booming in Austria before the programme was implemented. If we observe lower revenues as indicated in Tab. 3 we can not be sure that this is due to programme. There are good reasons to assume that organic farms had already lower revenues before the programme period started.

This assumption is corroborated when we use the difference-in-differences method to estimate the programme effects (see Tab. 4 and Tab. 5). Compared to the method used to obtain the results reported in Tab. 3 the programme effects are less explicit. Investment aid and support for young farmers (put together to get enough observations) seem to be the most effective measures to boost income. Forestry measures are effective as well. As would be expected from an investment support, obviously the capacity is increased with the effect that both expenses and revenues increase.

Farm expenses are increasing among participants of training courses and farms enrolled in adaptation and development measures. Our estimates do not indicate that income is increasing as well. Given this finding one might be tempted to conclude that these measures are actually counter-productive. Such a conclusion would be flawed because of two reasons: a) farm income is not increasing, but it is not decreasing either, therefore b) it may well be that the effects of these measures can only be observed in the longer term. It seems that participants have already changed their production pattern but were not yet able to benefit from increased revenues.

Tab. 5: Effects of programme measures on the market position of participants

	1257/99 article	indicator of market position				
		income		farm		
		house-hold	farm	revenue	expenses	revenue/ expenses
		€	€	€	€	ratio
investment, setting up young farmers	4, 8	4773 (0,00)	**24,4 (0,00)**	**12268 (0,00)**	**7794 (0,00)**	-0,011 (0,36)
training	9	1313 (0,34)	-4,1 (0,62)	2244 (0,28)	**2660 (0,05)**	-0,016 (0,24)
organic farming	22	411 (0,75)	10,2 (0,21)	1323 (0,50)	693 (0,59)	0,004 (0,76)
forestry	29	**4655 (0,07)**	**27,8 (0,08)**	2992 (0,44)	1403 (0,59)	0,012 (0,64)
adaptation and development	33	2172 (0,56)	-10,0 (0,66)	7804 (0,17)	**6373 (0,09)**	-0,026 (0,49)
observations		1937	1884	1968	1968	1968
R2		0,07	0,08	0,14	0,15	0,07

Source: own estimates base on LBG, ASBIS, BMLFUW

A closer look at a set of some additional indicators of the market position is corroborating this view. Farmers who have made training courses get higher milk revenues per kg. Participants in adaptation and development measures offer more beds and obtain more revenues from direct sales than non-participants. Organic farms get higher milk prices, as well and are able to achieve a better utilisation of their tourist branch (measured as nights rented per bed). As shown in Tab. 5 recipients of investment aids are able to boost milk output but they get slightly lower prices. This may be due to lower direct sales (not significant) and quality problems during the investment period.

Tab. 6: Effects of programme measures on the market position of participants (cont.)

	1257/99 article	indicator of market position				
		milk price	milk sales	beds	nights/bed	direct sales
		€/kg	kg	number	numbers	€
investment, setting up young farmers	4, 8	**-0,0051 (0,07)**	**9608,3 (0,00)**	-0,07 (0,24)	-10,8 (0,26)	-256 (0,33)
training	9	**0,0072 (0,03)**	1663,3 (0,43)	-0,09 (0,19)	9,5 (0,45)	-101 (0,74)
organic farming	22	**0,0155 (0,00)**	-857,8 (0,64)	0,03 (0,62)	**15,7 (0,09)**	151 (0,60)
forestry	29	-0,0028 (0,65)	-586,8 (0,88)	**-0,53 (0,00)**		-288 (0,61)
adaptation and development	33	-0,0013 (0,87)	**17035,4 (0,00)**	**0,37 (0,05)**	5,5 (0,79)	**2837 (0,00)**
observations		975	975	1968	145	1968
R2		0,16	0,23	0,11	0,27	0,05

Source: own estimates base on LBG, ASBIS, BMLFUW

5. Discussion

A detailed evaluation procedure is part of the programme for rural development which constitutes the 'Second Pillar' of the Common Agricultural Policy (CAP). In order to assess the effectiveness of this programme, Commission Regulation (EC) No 1750/1999 requires Member States to evaluate a detailed set of criteria and to report on a large number of indicators. In order to facilitate comparability of findings from different authorities and/or Member States on the same measures or projects it is necessary to use the same benchmarks.

However, in order to truly obtain comparable results across regions (programs) it is necessary to use the same methods as well. Compared to the scrutiny with respect to the relevant indicators, evaluators are given considerable leeway to choose the methodology they deem appropriate and operational. The paper shows some of the implications, the choice of method has on the findings of the analysis of one concrete criterion.

The data for the analysis presented in this paper are single farm observations of the Austrian farm accountancy data network (FADN) which are combined with data from the integrated administration and control system (IACS). The results show that in the mid-term evaluation not all measures of the programme have significant effects on the investigated indicators. It is depending on the method employed for measuring the effect which of the measure is "successful". When we simply compare indicators of participants with those of non-participants the program can be classified to be effective. However, if we use a more adequate approach (the difference in differences method) the programme seems to be less effective.

Several reasons can explain this result: a) we used data which were not specifically collected for programme evaluation and therefore some information may have gone lost when we transformed data into the required indicator; b) the programme period is longer than the observation period and therefore not all relevant effects may have yet materialised, c) some effects actually cannot be measured because participants were enrolled in a similar measure of the previous program and therefore no further improvement can be expected (e.g. lower yields of organic farms). However, although a small programme effect was estimated in this case, the correct conclusion would not be that the programme was inefficient; it might well be the that the programme managed to induce the continuation of desirable behaviour, whose effects unfortunately did not materialize within the period under consideration, but before. In such a situation, the difference-in-differences estimates would be misleadingly small.

The system of voluntary book-keeping which is widely established in EU member states is a valuable source of information. The focus of this information system is on farm income indicators, reflecting the income goal of European farm policy as laid down in Art. 33 of the Treaty. Recent policy reforms added new dimensions to the traditional goals of agriculture. It seems to be necessary that information systems are established to reflect these changes. The FADN-data are in principal well suited to be the primary source for a wide range of indicators. However, not in the current design. Therefore it seems to be necessary that there are moves being made to reform this system.

6. Literature

BRADLER, H. SCHNEEBERGER, W. (1999)
Untersuchung landwirtschaftlicher Betriebe mit einem Standarddeckungsbeitrag unter 90.000 ATS.. Endbericht zum Forschungsprojekt L 987/95 an das Bundesministerium für Land- und Forstwirtschaft, 75p.

BMLFUW (BUNDESMINISTERIUMF FÜR LAND- UND FORSTWIRTSCHAFT),
Farm accountancy data and IACS-data
electronic data set, Vienna 2003.

BLUNDELL, R. AND M. COSTA DIAS:
Alternative Approaches to Evaluation in Empirical Microeconomics,
University College London and Institute for Fiscal Studies, London, 2000.

EUROPEAN COMMISSION:
Evaluating socio-economic programmes, MEANS Collection,
Office for Official Publications of the European Communities, Luxembourg, 1999.

EUROPEAN COMMISSION, AGRICULTURAL DIRECTORATE-GENERAL,
Evaluation of rural development programmes 2000-2006 supported from the European Agriculture Guidance and Guarantee Fund (1999)
Brusseles, 1999
http://europa.eu.int/comm/agriculture/rur/eval/index_en.htm

EUROPEAN COMMISSION, AGRICULTURAL DIRECTORATE-GENERAL,
Common Evaluation Questions with Criteria and Indicators" (12/2000)
Brusseles, 2000
http://europa.eu.int/comm/agriculture/rur/eval/index_en.htm

EUROPEAN COMMISSION, AGRICULTURAL DIRECTORATE-GENERAL,
Guidelines for the mid term evaluation of rural development programmes 2000-2006 (05/2002)
Brusseles, 2002
http://europa.eu.int/comm/agriculture/rur/eval/index_en.htm

FRÖLICH, M., :
Programme Evaluation and Treatment Choice,
Lecture Notes in Economics and Mathematical Systems, Vol. 524, Heidelberg: Springer, 2003.

HECKMAN, J. J., R. J. LALONDE AND J. A. SMITH:
The economics and econometrics of active labour market programs', in: A. Aschenfelter and D. Card *Handbook of Labour Economics*, Amsterdam, 3, 1866-2097, 1999.

JANETSCHEK, H.:
Halbzeitevaluierung (MTE) des Kapitels I des ländlichen Entwicklungsplans (Art. 33 früher 5b-Programme), Energie aus Biomasse sowie anderer Energiealternativen, Investition in Landw. Betrieben,
Rohbericht, Bundesanstalt für Agrarwirtschaft, Wien, 2003.

JANETSCHEK, H.:
Halbzeitevaluierung (MTE) des Kapitels IX des ländlichen Entwicklungsplans (Art. 33 früher 5b-Programme), Energie aus Biomasse sowie anderer Energiealternativen, Kapitel Diversifizierung sowie Neuausrichtung, Innovation und Kooperation im landwirtschaftlichen und landwirtschaftsnahen Bereich,
Rohbericht, Bundesanstalt für Agrarwirtschaft, Wien, 2003b

LBG (LANDWIRTSCHAFTLICHE BUCHFÜHRUNGSGESELLSCHAFT)
BMLFUW (BUNDESMINISTERIUMF FÜR LAND- UND FORSTWIRTSCHAFT),
Farm accountancy data and IACS-data
electronic data set, Vienna 2003.

MÜLLER, F. AND W. PRÖLL:
Kapitel VIII Forstwirtschaft,
Endbericht, Bundesamt und Forschungszentrum für Wald, Wien, 2003.

ORTNER, K. AND L. SIMON:
Halbzeitevaluierung des Kapitels IX des ländlichen Entwicklungsplans Anpassung und Entwicklung von ländlichen Gebieten (Art. 33 der VO 1257/99), Vermarktung landwirtschaftlicher Qualitätsprodukte,
Bundesanstalt für Agrarwirtschaft, Wien, 2003.

PFUSTERSCHMID, S. AND S. KAMELOTT:
Halbzeitevaluierung der Maßnahme Kapitel III Berufsbildung des Österreichischen Programms für die Entwicklung des ländlichen Raums (Art. 9 der VO (EG) 1257/99),
Endbericht, Bundesanstalt für Agrarwirtschaft, Wien, 2003.

PISTRICH, K. AND C. PREINSTORFER:
Zwischenevaluierung der Fördermaßnahme "Verbesserung der Verarbeitung und Vermarktung landwirtschaftlicher Erzeugnisse" des Österreichischen Programms für die Entwicklung des ländlichen Raums,
Bundesanstalt für Agrarwirtschaft, Wien, 2003.

SINABELL, F.:
Einkommensrückgang in der Landwirtschaft nach der Erholung im Vorjahr. in Scheiblecker, M., et al., Österreichs Wirtschaft im Jahr 2002: Neuerlich ungenügendes Wachstum,
WIFO-Monatsberichte, 2003, 76(4), pp. 315-320.

TAMME, O.:
Zwischenevaluierung der Förderungsmaßnahme "Verkehrserschließung ländlicher Gebiete" (Artikel 33) im Rahmen des Programmes über die Förderung der Anpassung und Entwicklung des ländlichen Raumes gemäß VO (EG) Nr. 1257/1999 des Rates vom 27. Mai 1999,
Endbericht, Bundesanstalt für Bergbauernfragen, Wien, 2003.

WAGNER, K. AND TH. PARIZEK:
Halbzeitevaluierung des Kapitels IX des ländlichen Entwicklungsplans Anpassung und Entwicklung von ländlichen Gebieten (Art. 33 der VO 1257/99),
Bereich Bewirtschaftung der landwirtschaftlichen Wasserressourcen, Endbericht, Bundesanstalt für Agrarwirtschaft, Wien, 2003a.

WAGNER, K. AND TH. PARIZEK:
Halbzeitevaluierung des Kapitels IX des ländlichen Entwicklungsplans Anpassung und Entwicklung von ländlichen Gebieten (Art. 33 der VO 1257/99), Bereich Schutz der Umwelt, Landschaftspflege, Naturschutz,
Endbericht, Bundesanstalt für Agrarwirtschaft, Wien, 2003b.

WOOLDRIGE J. M:
Econometric Analysis of Cross Section and Panel Data,
MIT-Press, Cambridge MA, 2002.