ECONOMIC IMPACTS OF SEMI-SUBSISTENCE FARM SUPPORT MEASURE OF LATVIAN RURAL DEVELOPMENT PROGRAMME 2007-2013

Elita Benga
Institute of Agricultural Resources and Economics, Latvia
elita.benga@arei.lv

Abstract
The evaluation of EU Member States' co-founded programmes was assigned particular importance in recent years. The core question to be answered in programme evaluation is whether the stated objectives are accomplished by particular intervention. Until recently, the use of 'naïve' estimates was common in the evaluations of EU Member States Rural Development Programmes. The use of these estimates leads to potentially substantial selection bias resulting from using the outcomes of non-participants as a proxy for the possible outcomes of participants in the case of non-participation. The effectiveness of interventions on outcomes of interest can be better evaluated by propensity score matching (PSM). The objective of the study is to determine the net average effects from semi-subsistence farm support measure of the Latvian Rural Development Programme 2007 – 2013. To reach the research objective, propensity scores based on the most important characteristics were calculated for participants and non-participants of the measure and average treatment effects for gross value added were evaluated by matching methods. The research results show that the positive programme effects evaluated by 'naïve' estimators are overestimated in comparison with the results obtained by more rigorous PSM method.
Key words: policy evaluation, rural development programme, propensity score matching, treatment effects.

Introduction

The evaluation of EU Member States' co-founded programmes was assigned particular importance in recent years. The significance of the monitoring and evaluation has been confirmed in the European Community Agenda in 2000. Periodic evaluation of EU Member States Rural Development Programme (RDP) specific policy interventions is considered crucial in policy development. The main reasons for the evaluation of specific policy interventions are the assessment of the programme's impact, the improvement of programme management and administration, identification of necessary improvements in the delivery of interventions and meeting the accountability. According to the EU definition, programme evaluation is a process that culminates in a judgment (assessment) of policy interventions according to their results, impacts and the needs. In the case of rural development (RDP) programmes, EU regulations distinguish between ex-ante, midterm, ex-post and ongoing evaluations. The existing study is considered a part of an ongoing evaluation which would provide the grounds for the ex-post evaluation of Latvian Rural Development Programme 2007 – 2013.The core question to be answered in programme evaluation is whether the stated objectives are accomplished by particular intervention (support or 'treatment' provided to programme participants). The main problem in the process of evaluation is the assessment of the counterfactual outcome by modelling the situation where treatment is absent. The counterfactual outcome has to be estimated by statistical methods as it is usually not observed, unless there exists a rather costly possibility to use the experimental evaluation with random treatment assignment. Moreover, a random assignment has to be implemented before the policy intervention.

Until recently, the use of 'naïve' estimates was common in the evaluations of EU Member States Rural Development Programmes. These included "before - after" or 'with – without' approaches along with the comparisons with national averages. The "before - after" approach attributes the entire effect of the observed change in a particular indicator to the programme support. Thus the real effects may become understated or overstated. The "with - without" technique assumes that the outcome indicators will be the same both for programme participants and non-participants in the absence of the programme support. This leads to potentially substantial selection bias resulting from using the outcomes of non-participants as a proxy for the possible outcomes of participants in the case of non-participation. Naïve standard DID (difference-in-difference) estimator compares the before-and-after changes of selected result indicators for programme participants with the before-and-after changes of the same indicators for arbitrarily selected non-participants. The crucial assumption justifying this method is that the selection bias remains time invariant, and this is not often the case. If trends in the outcomes are not time invariant, the estimation is not correct.

Materials and Methods

To measure causal effects of programme or policy intervention, an experiment would be designed in which any unit of observation (farm) has an option to participate or not to participate in a programme. In order to measure an effect (result or impact) of a given programme, one has to estimate an appropriate counterfactual. As outcomes for participating units only are usually observed by programme monitoring, a way has to be found how to measure what would have happened to the same unit in a situation of its non-participation. For this purpose, a potential outcome model is appropriate. The model was proposed by Roy (1951) and further developed by Rubin (1974) and Holland (1986). Using the potential outcome model, the causal effect of a given programme on unit can be expressed with a basic evaluation formula:

$$e_i = Y_i(1) - Y_i(0), \quad (1)$$

where:
$Y_i(1)$ – potential outcome for unit i in case of participation in RDP (programme participants),
$Y_i(0)$ – potential outcome for unit i in case of non-participation in RDP (counterfactual),
e_i – the effect of programme participation on unit i, relative to effect of non-participation on the basis of a response variable Y.
In evaluation it is relatively easy to obtain for programme beneficiaries the information about $Y_i(1)$ but it is very difficult to estimate $Y_i(0)$ which for programme beneficiaries is not directly observable.

The outcome for a participating unit can be observed directly and it is expressed by formula:

$$e_i = (Y2 - Y1), \quad (2)$$

where:
$Y1$ – value of the outcome variable at programme starting period for a participating unit,
$Y2$ – value of the outcome variable at programme ending period.

The outcome for the same unit without the participation can be interpreted as a result of other factors which may simultaneously affect observable impact variables and it is expressed by formula:

$$e_i = Y3 - Y1, \quad (3)$$

where:
$Y3$ – value of the outcome variable for the same unit without a participation.

The unit can only be observed in one of two possible situations: being supported (participating) or not-supported (without a participation) which means that the real programme effect can be expressed as a difference between the outcome with a participation and outcome without a participation:

$$e_i = (Y2 - Y1) - (Y3 - Y1) = (Y2 - Y3). \quad (4)$$

Under naïve methodological approaches the whole observed change of a value of a given result indicator $(Y2 - Y1)$ is usually attributed to the programme and is erroneously considered to be a programme effect. The real programme effect $(Y2 - Y3)$ cannot be directly observed.

The effectiveness of interventions on outcomes of interest can be evaluated either by regression methods or propensity score matching (PSM). Multiple regression is the most common method for estimating the programme support effect. However, regression cannot take into account the distribution overlap on selected covariates. In many empirical studies, the causal effects are estimated by regressing variable of the outcome of interest on binary treatment variable. Thus the adjustment for the distribution between the treatment group and control group is not provided. PSM is a rigorous non-experimental method. The data for PSM usually are pooled in a panel both from programme participants and non-participants. The non-participating or ‘untreated’ units constitute the ‘control’ group while participants are included in the ‘treatment’ group. The information from control group is used to assess what would be the outcome of interest for participants in the absence of the programme. The difference in outcomes for both groups is evaluated by comparison of relatively similar units in these groups. This helps to avoid the potential biases that may arise by comparing the units with substantial differences in their characteristics, as these might affect the participation in the programme and outcomes of interest. A simple comparison of the difference between the averages of the outcome variables in two groups might lead to biased estimation, as the distributions of the covariates in the two groups may differ. A subclassification method was proposed by Cochran (1968). The observation variable is split into a number of subclasses. The treatment effect is then estimated by comparing the weighted means of the outcome variable in each subclass. Cochran’s research suggests that stratifying into five subclasses can remove much of the bias. However, as stated by Rubin (1997), subclassification may turn to be complicated if many covariates exist. To successfully mitigate the potential bias, unit matching has to be based not on a single or a few characteristics but on a full range of available covariates that have potential impact. The propensity score is then defined as the probability of receiving the treatment by the given unit. Thus the matching is reduced to a single variable, and matching on entire set of covariates is no longer necessary. The method was developed by Rosenbaum and Rubin (1983). They introduced balancing score as a function of covariates that provides the same distributions of covariates in both groups. Furthermore, they also introduced the assumption of strong ignorability, which implies the same distributions of the covariates in both groups given the balancing scores. They proved that treatment assignment is strongly ignorable if it satisfies the conditions of unconfoundedness and overlap. Unconfoundedness means that conditional on observational covariates, potential outcomes for two groups are not influenced by treatment assignment. The overlap assumption means that with the given covariates, the unit with the same covariate values has positive and equal opportunity of being assigned to the treated group or the control group. As stated by Joffe and Rosenbaum (1999), these assumptions eliminate the systematic, pretreatment, and unobserved differences between the units in treatment group and control group. PSM would provide biased estimation of causal effects when assumption of strong ignorability is violated. As suggested by Imbens (2004), if the treatment assignment is strongly ignorable, PSM can be used to remove the difference in the covariates' distributions between the treatment group and control group. He suggests four-step procedure for implementing the PSM:

1. selection of observational covariates and estimation of propensity scores,
2. stratification of propensity scores and testing of balancing properties in each block,
3. calculation of the Average Treatment on Treated (ATT) by matching,
4. sensitivity test for robustness of estimated ATT effects.

If the balancing properties of covariates are not satisfied in all strata, the test has to be repeated with different number of strata. If the balancing properties are not satisfied again, estimation of propensity scores has to be repeated with modified list of covariates by adding higher order (squared) covariates. After getting all covariates balanced in every stratum, causal effects can be estimated by nearest neighbor matching (NNM), radius matching (RM), kernel matching (KM) or stratified matching (SM).

NN matching computes the ATT by finding the unit in the control group whose propensity score is nearest (absolute value of difference is minimal) for every unit in the treatment group. Larger number of comparison units from the control group decreases the variance of the estimator. At the same time, the bias of the estimator increases. Furthermore, one needs to choose between matching with replacement and matching without replacement (Dehejia & Wahba, 2002). When there are few comparison units, matching without replacement will force us to match treated units to the comparison ones that are quite different in propensity scores. This enhances the likelihood of bad matches (increase the bias of the estimator), but it could also decrease the variance of the estimator. Thus, matching without replacement decreases the variance of the estimator at the cost of increasing the estimation bias. In contrast, because matching with replacement allows one comparison unit to be matched more than once with each nearest treatment unit, matching with replacement can minimize the distance between the treatment unit and the matched comparison unit. This will reduce bias of the estimator but increase variance of the estimator.

In RM, the units in both groups are matched when the propensity scores in the control group fall in the predefined radius of the units in the treatment group. The larger the radius is, the more matches can be found. More matches typically increase the likelihood of finding bad matches, which raises the bias of the estimator but decreases the variance of the estimator.

In KM, all units in the treatment group are matched with the weighted average of all units in the control group. The weights are determined by distance of propensity scores, bandwidth parameter and a kernel function. Choosing an appropriate bandwidth is crucial because a wider bandwidth will produce a smoother function at the cost of tracking data less closely. Typically, wider bandwidth increases chance of bad matches so that the bias of the estimator will also be high. Yet, more comparison units due to wider bandwidth will also decrease the variance of the estimator.

In SM, for each block the average differences in the outcomes of the treatment group and the matched control group are calculated. The ATT is then estimated by the mean difference weighted by the number of treated cases in each block. With respect to organizational research, Li (2012) recommends stratified matching as it does not require choosing specific smoothing parameters. The estimation of the ATT then requires minimum statistical knowledge. He regards SM as producing a reliable ATT while being relatively simple. In general, selection of the matching technique is empirical and it largely depends on the results obtained. As proven by Dehejia and Wahba (2002), similar results with most matching methods are obtained when the overlap in the distribution of propensity scores between the treatment group and control group is substantial. After the estimation of the ATT, the sensitivity test is used to investigate whether the causal effect estimated from the matching is susceptible to the influence of unobserved covariates. In detecting the existence of significant unobservables, Rosenbaum (1987) suggested use

Table 1

Average changes in Gross Value Added, EUR of supported (T=1) and non-supported (T=0) units by farm investment support measure of Latvian RDP during the programme period (2007-2013)

Number of Units	Gross Value Added 2007	Gross Value Added 2013	ATT (difference)
T=1 (263)	7 166	8 164	999
T=0 (156)	25 333	19 357	-5 976
Difference	-18 167	-11 193	6 975

Source: research findings, Latvian FADN database.

of multiple comparison groups. Such groups can be used in matching with the treatment group to calculate multiple treatment effects. Comparison of sizes of these effects would provide a sense of the reliability of the estimated ATT. A number of treatment groups can be compared with each other. Comparison of two control groups is possible, too.

The assumption of strong ignorability can be considered violated if causal effects prove to be statistically different between these two control groups. As multiple comparison groups are usually not available, there are three commonly used approaches with respect to sensitivity testing. The first method proposed by Dehejia and Wahba (1999) is changing the specification of the equation by adding or dropping higher order variables. Propensity scores are then recalculated, and newly obtained causal effect is compared to the originally computed effect. Such comparison reveals the reliability of originally computed causal effect. Instrumental variable (IV) method is another technique to assess the bias of the causal effects from original results. However, this method generally reduces the efficiency of the estimator. The bounding approach proposed by Rosenbaum (2002) assumes testing of possible hidden bias in the estimation of treatment effect. The test results would provide the level of sensitivity to hidden biases related to unobserved covariates. Such biases can influence the odds of treatment assignment.

The PSM method first has been empirically applied by Heckman, Ichimura, Smith and Todd (1998) in the estimations of training programmes on future income in the USA labor market. Subsequently, similar studies on the USA labor market were carried out by Dehejia and Wahba (2002), and a few other researchers.

The modules for calculating propensity scores and matching for use in STATA software were developed by Becker and Ichino (2002). Before running the set of necessary modules they recommend to "clean up" the dataset. It is common first to run the *pscore* module which estimates the propensity scores and tests the satisfying of the balancing properties. If the balancing properties are satisfied then ATT can be estimated with one or more of the *att** modules. The modules *attnd* or *attnw*, *attr*, *attk* and *atts* assume the nearest neighbor, radius, kernel and stratified matching, respectively. After the calculation of ATT, the module *mhbounds* developed by Rosenbaum (2002) provides sensitivity analysis with Rosenbaum bounds with Mantel and Haenszel (1959) test statistic.

Results and Discussion

The data on participants and non-participants of Farm Investment Support Measure of Latvian Rural Development Programme semi-subsistence farm support measure are sourced from FADN database which is not publicly available. The economic data in the database include all relevant information on programme participants and non-participants regarding their structure and performance from 2007 to 2013. First, as the information should cover periods before and after the implementation of the programme, 419 units were selected out of total number of 943 units. The possible overlapping was checked, leaving treated units that participated only in selected measure. There were 263 units in treatment group, leaving 156 units for possible controls. For the evaluation purposes differences in values of Gross Value Added after and before the implementation of the programme were obtained using the "naïve" difference-in-differences estimator. The Gross Value Added as the economic variable was selected because of its importance in the evaluation context as it is one of the main economic indicators that measures the impacts of policy interventions on economic performance of single economic units, sectors and national economics in general.

The values of changes in economic variable and calculated treatment effects are shown in Table 1.

The ATT effect on GVA of programme participants calculated by DiD method is positive. For programme non-participants, the ATT effect is negative. Using the "naïve" difference-in-differences estimator would lead to an erroneous assumption that measure contributes to the growth in Gross Value Added for participating units at EUR 6 975.

With respect to propensity score matching (PSM-DiD method), in total, 31 variables related to unit structure which were considered critical for comparability of economic performance were selected for the use in matching process.

Although only 4 and 6 variables proved statistically significant at 5% and 10% level, respectively, after Logit regression, dropping the variables with lower significance levels caused a loss of balancing properties in one or more blocks. Similarly, adding of higher order covariates caused the loss of balancing properties. Therefore, the original specification of Logit function was preferred. A list of structural variables with their propensity scores obtained with Logit equation is provided in Table 2.

Table 2

Results of estimation of logit function

Variable	Coefficient	Standard deviation	z	P>|z|	95% confidence interval	
Organic farming	-0.338324	0.288963	-1.17	0.242	-0.904682	0.228034
Labor inputs	-0.029574	0.159048	-0.19	0.852	-0.341303	0.282155
Agricultural land	0.004518	0.017689	0.26	0.798	-0.030152	0.039189
Livestock units	-0.008048	0.013007	-0.62	0.536	-0.033541	0.017446
Output	-0.000045	0.000049	-0.91	0.361	-0.000141	0.000051
Output in crop farming	-0.000893	0.001155	-0.77	0.440	-0.003157	0.001372
Output in livestock farming	-0.000825	0.001155	-0.71	0.475	-0.003089	0.001440
Total agricultural output	0.000887	0.001154	0.77	0.442	-0.001376	0.003149
Processing	-0.000046	0.000098	-0.47	0.639	-0.000238	0.000146
Net turnover	-0.000059	0.000031	-1.90	0.057	-0.000119	0.000002
Depreciation	0.001603	0.001610	1.00	0.320	-0.001554	0.004759
External costs	0.000063	0.000058	1.09	0.274	-0.000050	0.000176
Gross value added 2007	-0.001615	0.001608	-1.00	0.315	-0.004767	0.001537
Gross value added 2012	0.000044	0.000020	2.19	0.029	0.000005	0.000084
Net value added	0.001624	0.001609	1.01	0.313	-0.001530	0.004777
Gross margins in crop farming	-0.000121	0.000199	-0.61	0.542	-0.000510	0.000268
Gross margins in livestock farming	-0.000816	0.000747	-1.09	0.274	-0.002280	0.000647
Total assets	0.000003	0.000008	0.40	0.690	-0.000013	0.000020
Buildings	-0.000038	0.000019	-2.01	0.045	-0.000076	-0.000001
Equipment and machinery	0.000013	0.000019	0.69	0.493	-0.000024	0.000050
Total liabilities	0.000022	0.000020	1.09	0.277	-0.000018	0.000061
Long-term liabilities	-0.000020	0.000028	-0.72	0.473	-0.000074	0.000034
Gross investments	0.000016	0.000026	0.60	0.549	-0.000035	0.000066
Total state support	-0.001246	0.001601	-0.78	0.436	-0.004385	0.001892
Area payments	-0.001156	0.000697	-1.66	0.097	-0.002521	0.000210
Less favorable area payments	-0.000253	0.000197	-1.28	0.200	-0.000639	0.000134
Crop subsidies	-0.000518	0.000231	-2.24	0.025	-0.000970	-0.000065
Livestock subsidies	-0.000474	0.000166	-2.86	0.004	-0.000799	-0.000149
Compensated excise tax	-0.000547	0.000484	-1.13	0.259	-0.001495	0.000402
Interest subsidies	0.000937	0.000804	1.17	0.244	-0.000638	0.002512
Subsidies for investments	0.001468	0.001602	0.92	0.359	-0.001672	0.004608
Constant	2.287470	0.496222	4.61	0.000	1.314893	3.260047
Logit regression	Observations	LR chi²(31)	Prob>chi²	Log likelihood		Pseudo R^2
	419	158.05	0.000	-197.59		0.29

Source: research findings, Latvian FADN database.

Table 3

Blocks of propensity scores

Inferior of block of propensity score	T(0)	T(1)	Total
0.0121553	31	4	35
0.2	23	5	28
0.4	23	28	51
0.6	47	110	157
0.8	11	116	127
Total	135	263	398

Source: research findings.

Table 4

Average treatment effects (EUR) on Gross Value Added by method and test statistics

Method	Nearest neighbor	Radius matching (0.001)	Radius matching (0.01)	Radius matching (0.1)	Kernel matching
Treated	263	64	251	263	263
Controls	69	45	93	135	135
ATT	4 853	3 809	4 523	6 676	5 748
t	1.47	2.28	1.73	1.83	2.33

Source: research findings.

The Table 3 shows the inferior bound, the number of treated units and the number of control units for each of iterated five blocks. As the computed z-value does not exceed the critical value for the 5% confidence interval for all three variables, null hypothesis can not be rejected.

The common support option has been selected. This restriction implies that the test of the balancing property is performed only on the observations whose propensity score belongs to the intersection of the propensity scores in both groups. With the given specification the balancing property is satisfied. The results of evaluation of average treatment effects with various matching methods and respective test statistics are shown in Table 4.

The average treatment value with the highest test statistics (kernel matching) were considered the best estimate for economic variable. Sensitivity analysis was carried out using the Rosenbaum bounding approach. The results show that the estimated effects on Gross Value Added of the Measure intervention are rather sensitive. The sensitivity test shows that a hidden bias which increases the odds ratio from 1 to 1.05, would make the obtained results statistically insignificant. The relatively high sensitivity would have been caused by relatively small number of observations in control group. It is recommended to have up to 4 times more observations for potential controls which is not the case. However, the results of sensitivity tests are providing only additional information with respect to the calculated effects stability. The overall validity of the obtained results is not questioned.

Using the PSM-DiD estimator provides statistically rigorous estimation of the contribution of a measure to a growth in Gross Value Added for participating units at EUR 5 748. The value of changes in economic variable obtained by PSM-DiD method is slightly lower than yielded by 'naïve' difference-in-differences estimator (EUR 6 975). This indicates to a possible overestimation of programme effects if 'naïve' method is used.

Conclusions

The use of 'naïve' estimators in evaluation of programme effects on economic variables can lead to the overestimation of changes in economic variables attributed solely to the programme. The rather small difference in results obtained by "naïve" difference-in-differences estimator and propensity score matching is purely accidental as the analysis of the other measures of the programme show the effects estimated can be either negative or positive depending upon the method applied. Propensity score matching has to be considered a more suitable method in establishing a sound counterfactual. The changes in Gross Value

Added estimated by propensity score that can be viewed as direct programme effects on beneficiaries matching are significant and positive.

The direct economic impacts of semi-subsistence farm support measure of Latvian Rural Development programme 2007 – 2013 are significant and positive.

References

1. Becker, S., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal, 2*, pp. 358-377.
2. Cochran, W. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24,* pp. 295-313.
3. Dehejia, R., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*, pp. 1053-1062.
4. Dehejia, R., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*, pp. 151-161.
5. Heckman, J.J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data, *Econometrica, 66(5),* pp. 1017-1098.
6. Holland, P.W. (1986). "Statistics and Causal Inference" (with discussion), *Journal of the American Statistical Association, 81*, pp. 945-970.
7. Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics, 86*, pp. 4-29.
8. Joffe, M.M., & Rosenbaum, P.R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology, 150,* pp. 327-333.
9. Li, M. (2012). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. *Organizational Research Methods, 00(0)* pp. 1-39.
10. Mantel, N., & Haenszel, W. (1959). Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease, *Journal of the National Cancer Institute 22(4)*, pp. 719-748.
11. Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika, 70*, pp. 41-55.
12. Rosenbaum, P. (1987). The role of a second control group in an observational study. *Statistical Science, 2,* pp. 292-306.
13. Rosenbaum, P. (2002). *Observational Studies*, New York: Springer, 2nd edition.
14. Roy, A. (1951). "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, 3*, pp. 135-146.
15. Rubin, D. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology, 66*, pp. 688-701.
16. Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, pp. 757-763.
17. Smith, J., & Todd, P. (2005). Does Matching Overcome LaLonde's Critique of Non Experimental Estimators?, *Journal of Econometrics 125(1-2)*, pp. 305-353.