INTERNATIONAL CONFERENCE OF AGRICULTURAL ECONOMISTS
ICAE
29th | Milan Italy 2015
UNIVERSITÀ DEGLI STUDI DI MILANO AUGUST 8 - 14
AGRICULTURE IN AN INTERCONNECTED WORLD

MATCHING EFFICIENCY RESULTS OF ORGANIC FARMS

Jochen Kantelhardt and **Stefan Kirchweger**

University for Natural Resources and Life Science Vienna (BOKU), Austria

Abstract.

Organic farms work under very heterogeneous natural-site and socio-economic conditions. This heterogeneity is of clear relevance for economic efficiency and for the decision of farms to convert to organic farming. In order to produce proper results efficiency analysis must consider such heterogeneity and self-selection aspects. This applies in particular to data envelopment analysis, since this technique does not calculate error terms, but include heterogeneity into efficiency results. One way to control for such effects is matching. Matching is based on the assumption that under a given vector of observable variables, the outcome of one individual is independent of the adoption of a specific treatment. In our paper we present how to implement matching into efficiency analysis of organic farms. We give a brief overview on literature applying this technique and we discuss which insights the application of matching might contribute to the current discussion on organic farming.

Keywords: Dairy Farming, Farm Competitiveness, Low-input farming, Cluster Analysis, Matching Method

IAAE

1. Introduction

Efficiency of organic farms is mostly analysed by comparing organic and conventional farms (e.g. Tzouvelekas et al., 2001; Lansink et al., 2002; Kumbhakar et al., 2009; Mayen et al., 2010; Breustedt et al., 2011). A major discussion point in literature with regard to such a comparison is to what extent the production technology of organic and conventional farms is directly comparable. Kumbhakar et al. (2009) argues that, although most of the machinery can be used in both systems, production practices are due to the ban of applying synthetic fertilizers and plant protection and other regulations in organic farming too different to allow a direct comparison. Many authors take a similar position and claim that efficiency analysis has to be executed in different subsamples (Tzouvelekas et al., 2001; Lansink et al., 2002; Breustedt et al., 2011). Further proposals are to use a joint technology, but to apply a Heckman-type sample selectivity correction to correct for endogeneity in the technology choice decision (Sipiläinen and Lansink, 2005). And finally, a variety of authors uses both, a joint and a separated frontier, and compares the results in the discussion (Mayen et al., 2010).

A further important aspect in analysing efficiency of organic farms is triggered by the fact that agricultural work is shaped by natural-site conditions. Bowman and Zilberman (2013) emphasize in this context the importance of biological and geophysical factors, which impact beside input and output market conditions farmer decision-making and adoption of land use practices or technologies. Site conditions and their heterogeneity are of clear relevance for the decision of farms to convert to organic farming; for instance, farms on land of low quality might rather switch to organic farming (Lansink et al., 2002). As land is an important input variable in agriculture, site conditions also influence economic efficiency of farms as well as the result in efficiency. In order to produce proper results efficiency analysis must consider such heterogeneity in site conditions and self-selection aspects (Mayen et al., 2010). This applies in particular to data envelopment analysis, since this technique does not calculate error terms, but include heterogeneity into efficiency results (Lansink et al., 2002).

Literature shows that there are several ways to cope with heterogeneity and self-selection. Breustedt et al. (2011) include an indicator for site conditions, the so-called EMZ, as an additional non-discretionary input into their efficiency analysis. (Mayen et al., 2010) apply a separate frontier and match data sets on site conditions before running the efficiency analysis in order to compare the efficiency and productivity of organic and conventional dairy farms in the United States. As Kellermann and Salhofer (2014) argue, this procedure has the advantage that “the measured differences are directly and solely attributable to the difference in technology”. Whereas in other study areas authors use matching after executing efficiency analysis (Sauer *et al.*, 2014).

In our paper we aim to analyse the impact of site conditions on efficiency results of organic farms. Therefore we apply a data envelopment analysis and execute this procedure in two separate subsamples. In order to measure the influence of site conditions we use a matching procedure. Matching is either applied before or after DEA analysis, in order to use the explanatory power of both approaches. The remainder of our paper is organized as follows: In Section 2 we introduce briefly the used methods: data envelopment analysis and genetic matching. In Section 3 we present the case study, define the required economic input and output variables and introduce our data basis. The results of our calculations are displayed in Section 4. Finally, in Section 5, we discuss our results and draw conclusions for the further development of our model.

2. Methods

The first part of the following section contains a description of the methods applied in our empirical analysis, namely data envelopment analysis and matching. We introduce the applied methods only very briefly, since all methods itself are well-known and well-described in literature. However, further information on the applied methods can be found in the indicated references.

DEA is a non-parametric mathematical programming approach. It enables the comparison of production performances of so-called Decision-making Units (DMU). In our case these DMUs are farms deciding on the use of production factors in order to minimize farm input. The performance of each farm is rated by calculating the output-to-input ratio of the respective production processes; the less input a farm requires for producing a given output or the more output it produces with a given input, the higher is the productivity of the farm. The final efficiency score is derived by benchmarking the output-to-input ratio of an individual farm against the output-to-input ratio of all best-practice farms.

The linear programming problem to be solved for each farm is as follows:

$$min_{\emptyset,\lambda,\delta}\,\theta\ s.t.$$
$$-y_i + Y\lambda \geq 0$$
$$\theta x_i - X\lambda \geq 0$$
$$N1'\lambda = 1$$
$$\lambda \geq 0$$
$$0 < \delta \leq 1$$

where θ is the overall technical efficiency score for the ith firm, yi is a Mx1 vector of output quantities for the i-th farm, Y is a NxM matrix of output quantities for all N farms, λ is a Nx1 vector of weights, xi is a Kx1 vector of input quantities for the i-th farm and X is a NxK matrix of input quantities for all N farms. In DEA the relevance of input (X) und output variables (Y) is expressed by weights (η in the input case, μ in the output case), which are determined in a way that the assessed DMU achieves the highest possible level of efficiency. In order to derive input weights η and output weights μ, additional to the above described envelopment model the multiplier model has to be solved (Cooper *et al.*, 2007). The constraint $N1'\lambda = 1$ implies the sum of the lambdas equals one and allows for a variable returns to scale (VRS) technology.

Matching follows the Conditional Independence Assumption (CIA) and goes back to the works of Rubin (1977) and Rosenbaum and Rubin (1983). Matching basically controls for observable variables assuming that under a given vector of observable variables (Z), the outcome (Y) of one individual is independent of treatment or technology (T):

$$\{Y^0, Y^1 \perp\!\!\!\perp T\} \,|Z \tag{2}$$

where ⫫ denotes independence (Sekhon, 2009). Therefore pairing participating and non-participating farms based on observable variables (covariates) allows the interpretation of differences with regard to respective outcome variables as unbiased effect estimates.

As matching is performed in a non/semi-parametric way, it has the considerable advantage of requiring fewer functional forms than regression-based analyses (Lechner, 2002b; Smith and Todd, 2005; Imbens and Wooldridge, 2009). Further advantages of matching are its allowance for arbitrary heterogeneity of the effects, its simplicity and its intuitive appeal (Lechner, 2002b, a).

This requires the identification of those Z which influence the outcome and the probability of participation but are not influenced by treatment. Distance functions are used to control for Z of individuals, which can be done by approaches which match directly on covariates as well as using aggregated distance functions. Whereas the first is referred as a direct matching (DM) approach, for aggregated distance functions approaches like propensity score matching, mahalanobis matching and genetic matching are used. For our analysis we apply a genetic matching approach which is nonparametric and basically a generalization of propensity score and mahalanobis distance matching. Genetic matching optimizes the balance of observed covariates between participating and non-participating groups using a genetic algorithm (Sekhon, 2011).

The variety of matching algorithms is big and includes nearest-neighbour matching, calliper matching, radius matching, stratification matching, interval matching, kernel matching and local linear matching[1]. Literature gives almost no advice on the superiority of any one of these algorithms over another. The selection of the appropriate algorithm should rather be done individually, depending on the structure of data (Zhao, 2004). In our study, we apply a nearest-neighbour algorithm. This algorithm pairs each organic (treated) farm with this conventioal (control) farm, which shows the smallest distance with regard to the applied matching covariates (Caliendo and

[1] See Caliendo and Kopeinig (2008)) for detailed descriptions of matching algorithms.

Kopeinig, 2008). Matching can be considered successful when the mean of the covariates between treated and control group is balanced. Balance can be judged by conventional testing.

3. Case study definition

We apply the model on an Austrian farm panel data set consisting of data for 647 voluntary bookkeeping farms for the year 2012. In order to ensure a principal comparability of farms, we consider only forage farms (output from grazing animals is greater than 66% of total output) and exclude all other farm types.

Using DEA for the assessment of farms, we have to define appropriate input and output factors. A fundamental requirement doing this is that the factors have to cover the full range of resources used. Moreover, all relevant activity levels and performance measures have to be captured (Dyson *et al.*, 2001). As input variables we use the cultivated area (ha; include agricultural and forested area), labour (WU; include family members and employees), capital (EUR; depreciation) and intermediate inputs (EUR). As output variables we use the total farm revenue (EUR), including agricultural services and all payments for agri-environmental programmes. The resulting technical efficiency measure expresses the economic success of the farm and therefore represents the performance of farmers. In this case study we apply a input-oriented DEA, which results in efficiency scores explaining the individual potential of minimizing the input, at a given output level, to increase productivity. The DEA in our analysis is executed with the R-package “Benchmark”.

In order to control for site conditions we use the following matching covariates (Z): the main production region, the altitude of the farm, the mountain farm cadaster, the value for taxing real-estate based on government valuation (“Einheitswert”) per hectare land, a dummy variable for alpine farming, ,the share of forested area the reduced utilized agricultural area (alpine and extensive grassland measured as 1/2 of 1/3), the share of grassland, the share of extensive pastures as well as the share of extensive grassland, and. The matching algorithm in our analysis is run with the R-package “Matching” by J.S. Sekhon (Sekhon, 2011).

Table 1: Mean (and standard deviation) of matching covariates, input and output variables with and without matching

	With matching			Without matching		
	Organic	Conventional		Organic	Conventional	
Number of farms	170	477		170	170	
Main production region 1 (%)	28.82 (45.43)	10.69 (30.93)	***	28.82 (45.43)	26.47 (44.25)	
Main production region 2 (%)	21.18 (40.98)	10.06 (30.12)	***	21.18 (40.98)	18.82 (39.21)	
Main production region 3 (%)	9.41 (29.29)	18.45 (38.83)	***	9.41 (29.29)	14.12 (34.92)	
Main production region 4 (%)	25.88 (43.93)	21.17 (40.9)	***	25.88 (43.93)	25.88 (43.93)	
Main production region 5 (%)	1.18 (10.81)	3.56 (18.56)	***	1.18 (10.81)	1.18 (10.81)	
Main production region 6 (%)	11.18 (31.6)	30.61 (46.13)	***	11.18 (31.6)	12.35 (33)	
Main production region 7 (%)	1.18 (10.81)	4.61 (21)	***	1.18 (10.81)	0.59 (7.67)	
Main production region 8 (%)	1.18 (10.81)	0.84 (9.13)	***	1.18 (10.81)	0.59 (7.67)	
Altitude (m)	692.02 (227.34)	595.91 (225.5)	***	692.02 (227.34)	702.64 (226.98)	
Mountain Farm Cadastre (Pt.)	105.38 (76.62)	74.17 (73.12)	***	105.38 (76.62)	107.58 (73.92)	
Einheitswert (€)	446.16 (227.75)	582.33 (312.34)	***	446.16 (227.75)	444.86 (239.75)	
Alpine farming (%)	31.18 (46.46)	14.47 (35.21)	***	31.18 (46.46)	27.06 (44.56)	
Share of forested area (%)	28.98 (16.61)	25.55 (16.98)	*	28.98 (16.61)	28.92 (16.22)	
Reduced agricultural area (ha)	28.16 (15.53)	28.54 (16.4)		28.16 (15.53)	27.22 (15.37)	
Share of grassland (%)	79.53 (25.59)	61.31 (28.91)	***	79.53 (25.59)	78.68 (26.21)	
Share of extensive pastures (%)	6.62 (11.23)	4.82 (9.1)		6.62 (11.23)	6.19 (10.52)	
Share of extensive grassland (%)	1.06 (3.54)	0.55 (2.3)		1.06 (3.54)	0.93 (2.92)	
Cultivated area (ha)	57.46 (47.65)	44.86 (28.75)	**	57.46 (47.65)	50.31 (32.63)	
Labour (WU)	1.67 (0.59)	1.68 (0.59)		1.67 (0.59)	1.67 (0.57)	
Capital (€)	47480 (28631)	65361 (45584)	***	47480 (28631)	60267 (40569)	**
Intermediate inputs (€)	18458 (10362)	19549 (11232)		18458 (10362)	19632 (11195)	
Output (€)	84932 (51767)	101412 (67226)	**	84932 (51767)	93283 (57882)	

Number in parentheses show standard deviations; t-test, Chi square test and McNemar test are used for equality of means: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1;
Source: Own calculation

The means of matching covariates (Z) as well as in- and output variables for organic farms and all conventional farms without matching are displayed in Table 1. The table shows that organic forage farms significantly differ from conventional farms with regard to the location in main production regions. So are organic farms more likely located in the rather alpine main production regions 1, 2, 4 and 8. Also the mean altitude and the mountain farm cadastre points are significantly higher and the "Einheitswert" lower which indicates in average less favourable conditions for organic farms. Furthermore organic farms show a higher share of farms with alpine farming and a higher share of forested area per farm. Whereas the farm groups show similar mean values for "reduced" agricultural area, share of extensive pastures and grassland, they differ in the share of grassland, which is higher on organic farms. The input variable cultivated area, which includes next to agricultural area, the total alpine and extensive grassland and forested area, is significantly higher on organic forage farms. Labour and intermediate inputs are similar in both groups, but capital and output is higher on conventional farms.

The matching procedure reduces the number of conventional farms and balances the mean values of all matching covariates (see Table 1). With regard to in- and output variables the mean value on conventional farms increases for cultivated area and decreases for capital and output. The mean values for labour and intermediate inputs remain similar for the two groups. This indicates that farms with rather more land, but lower capital and output remain in the sample.

4. Presentation of technical efficiency results

In the following chapter we present preliminary results from a first model run, where 6 different analysis where made: (1) DEA without matching for organic farms; (2) DEA without matching for conventional farms; (3) DEA after matching for organic farms; (4) DEA after matching for conventional farms; (5) DEA prior matching for organic farms; (6) DEA prior matching for conventional farms. An overall view on the results shows that the variance of efficiency scores in all analysis is quite high. So, the minimum value is about 0.4 and the maximum value 1.0 in almost all analysis (Table 2). However, mean technical efficiency values distinguish statistically significant between organic and conventional farms in case of the without matching (Analysis 1 and 2). So is the mean value of organic farms with 0.812 statistically significant higher than the 0.765 of conventional farms (Table 2). The results of other studies applying this procedure vary: Lansink et al. (2002) and Tzouvelekas et al. (2001) also find higher mean efficiency for organic farms (0.93 vs. 0.69; 0.69 vs. 0.58 (output-oriented)), but Mayen *et al.* (2010) as well as Sipiläinen and Lansink (2005) detect lower mean efficiencies of organic farms.

Whereas the mean efficiencies of organic farms are constant across all models (Analysis 1, 3 and 5), matching influences the efficiency results of conventional farms as farms are dropped from the sample (see Table 2). When DEA is applied after matching (Analysis 4) the average technical efficiency of conventional farms increases significantly to 0,831, which is slightly higher than comparable organic farms (0,812), but not statistically significant different. Fairly similar result with regard to absolute height of mean efficiencies and differences between both groups is found in Mayen *et al.* (2010). Furthermore, we find that the application of this procedure leads to a loss of formerly efficient conventional farms and the new efficiency frontier moves closer to the remaining farms. However, when DEA is carried out prior matching (Analysis 6) the efficiency frontier does not change but conventional farms are dropped after the efficiency analysis. Thus the same efficiency scores as in analysis 2 are estimated but with only those farms which are similar to the organic farms. This leads in our case to a small decrease in mean technical efficiency of conventional farms (0.747) in comparison to Analysis 1 (0.765). Consequently, it is to conclude that rather more efficient farms are deleted by matching, so that in average less efficient farms remain in the sample.

Table 2: Results of technical efficiency

	Organic	Conventional	
Without matching			
Number of farms	170	477	
Mean	0.812	0.765	***
SD	0.151	0.135	
Min.	0.422	0.368	
Max.	1.00	1.00	
DEA after matching			
Number of farms	170	170	
Mean	0.812	0.831	
SD	0.151	0.140	
Min.	0.422	0.387	
Max.	1.00	1.00	
DEA prior to matching			
Number of farms	170	170	
Mean	0.812	0.747	***
SD	0.151	0.138	
Min.	0.422	0.368	
Max.	1.00	1.00	

Technical efficiencies are calculated using an input-oriented and variable return zo scale model. Number in parentheses show standard deviations; t-test isused for equality of means: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1;
Source: Own calculation

5. Discussion and conclusions

The aim of our paper is to analyse the impact of site conditions on efficiency results of organic farms. To do so we apply data envelopment analysis and estimate efficiency scores for Austrian forage farmers. In a second step we combine this approach with a matching procedure to control for different site conditions. Our study finds, when matching is applied the mean efficiency results alter dramatically. This is also found in Mayen *et al.* (2010) who apply a similar approach. This leads us to the conclusion to emphasize the necessity of establishing comparable data sets in order to allow accurate cross conventional organic farm comparisons. This applies in particular, when a separate frontier approach is applied and the influence of site conditions cannot be considered by an implementation of covariates in the productivity model (as it is regularly done in stochastic frontier analysis) or by a second stage analysis (as it is regularly practised in connection with DEA models). We therefore recommend applying a matching procedure when establishing an organic conventional farm comparison.

As we also applied the matching procedure after the DEA we find a loss of efficient conventional control farms and therefore a move of the frontier toward the remaining farms. This explains the increasing mean efficiency of conventional farms in Analyses 4. Which leads to the conclusion that we have to consider that we can not be sure about their actual difference in productivity (Tzouvelekas et al., 2001). This is because efficiency scores have to be seen as relative values to the frontier which can also be different in the individual technology. Future works should aim to consider this aspect and should at least simultaneously run a joint frontier model in order to get an idea on the "real" distance between organic and conventional farms.

We also want to stress, that we solely controlled for heterogeneity in site conditions and not on the entire selection of organic farming. This is done e.g. in Mayen *et al.* (2010) but requires the availability of more data. Especially data of organic farms prior to their decision of adopting organic farming would be helpful to cover the whole selection process of organic farming. If this is not the case, production variables are often influenced by organic farming itself and therefore violate the independence assumption in the matching procedure.

References

Bowman, M.S., Zilberman, D., 2013. Economic factors affecting diversified farming systems. *Ecology and Society* **18**.

Breustedt, G., Latacz-Lohmann, U., Tiedemann, T., 2011. Organic or conventional? Optimal dairy farming technology under the EU milk quota system and organic subsidies. *Food Policy* **36**, 223-229.

Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* **22**, 31-72.

Cooper, W.W., Seiford, L.M., Tone, K., 2007. *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software: Second edition*.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., Shale, E.A., 2001. Pitfalls and protocols in DEA. *European Journal of Operational Research* **132**, 245-259.

Imbens, G.W., Wooldridge, J.M., 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* **47**, 5-86.
Kellermann, M., Salhofer, K., 2014. Dairy farming on permanent grassland: Can it keep up? *Journal of Dairy Science* **97**, 6196-6210.
Kumbhakar, S., Tsionas, E., Sipiläinen, T., 2009. Joint estimation of technology choice and technical efficiency: an application to organic and conventional dairy farming. *Journal of Productivity Analysis* **31**, 151-161.
Lansink, A.O., Pietola, K., Bäckman, S., 2002. Effciency and productivity of conventional and organic farms in Finland 1994–1997. *European Review of Agricultural Economics* **29**, 51-65.
Lechner, M., 2002a. Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies. *The Review of Economics and Statistics* **2**, 205-220.
Lechner, M., 2002b. Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **1**, 59-82.
Mayen, C.D., Balagtas, J.V., Alexander, C.E., 2010. Technology Adoption and Technical Efficiency: Organic and Conventional Dairy Farms in the United States. *American Journal of Agricultural Economics* **92**, 181-195.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.
Rubin, D.B., 1977. Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics* **2**, 1-26.
Sauer, J., Walsh, J., Zilberman, D., 2014. Agri-Environmental Policy Effects at Producer Level - Identification and Measurement. *Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.* **49**, 16.
Sekhon, J.S., 2009. Opiates for the Matches: Matching Methods for Causal Inference. *Annual Review of Political Science* **12**, 487-508.
Sekhon, J.S., 2011. Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software* **42**, 1-52.
Sipiläinen, T., Lansink, A.O., 2005. Learning in Organic Farming – an application on Finnish dairy farms. Accessed available at http://purl.umn.edu/24493
Smith, J.A., Todd, P.E., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* **125**, 305-353.
Tzouvelekas, V., Pantzios, C.J., Fotopoulos, C., 2001. Technical efficiency of alternative farming systems: the case of Greek organic and conventional olive-growing farms. *Food Policy* **26**, 549-569.
Zhao, Z., 2004. Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence. *The Review of Economics and Statistics* **86**, 91-107.