https://doi.org/10.61308/DWJS6911

Technical efficiency and assessment of input excess in all European farms by a non-parametric methodology and the conditional inference tree

Nicola Galluzzo
Association of Geographical and Economic Studies in Rural Areas (ASGEAR) – Rieti, Italy
E-mail: asgear@libero.it

Citation: Galluzzo, N. (2023). Technical efficiency and assessment of input excess in all European farms by a non-parametric methodology and the conditional inference tree. *Bulgarian Journal of Agricultural Economics and Management,* 68(4), 3-16.

Abstract

The technical efficiency is a fundamental tool in order to evaluate the performances of farms. The main purpose of this study was to investigate the technical efficiency in all European farms part of the FADN dataset using a non-parametric approach such as the DEA. The estimation of the technical efficiency has pointed out a dichotomy between new and older member states of the European Union. The further stage of this research has been addressed in evaluating the excess of input able to impact to technical inefficiency patterns using a new approach such as the Multi-directional Efficiency Analysis. The Multi-directional Efficiency Analysis is a novelty in the literature because it estimates the percentage of excess in input involved in the inefficiency overcoming the main bottleneck in the estimation on efficiency using the DEA. By the estimation of the technical efficiency, it has been possible also to assess the impact of financial subsides allocated by the CAP. Results have corroborated as CAP subsidies have reduced the technical efficiency in farms. In the third stage of this research, it has been possible to assess by the machine learning using the conditional inference tree the excess of input and how the excess impacts on the technical efficiency with an accuracy of more 90%. The conditional inference tree is also a novelty in the literature about the technical efficiency estimation in farms combining the estimation of the technical efficiency with the patterns and reasons of inefficiency.

Key words: Multi-directional Efficiency Analysis; machine learning; type of farming; DEA; conditional inference tree

Introduction

The agriculture in European Union is a typical mosaic where it is possible to find different type of farmings and different level of production, which have some effects to the technical efficiency and other performances of farms. In general, between farms size and technical efficiency there is a nexus (Galluzzo, 2013). The management of farms is directly influenced by the size of farms and type of farmings. The land capital is one of the main factors able to impact to the farmer's income and to the management of farm. The most recent studies have pointed out as there is a not a homogenous allocation of agricultural land among EU countries (Popescu et al., 2019). According to these authors, over the decade 2009–2018 the EUROSTAT has assessed as 71.46% of the EU–28 agricultural land is used by France, Spain, United Kingdom, Germany, Poland, Romania, and Italy able to produce about 77.58% of the total output in the primary sector.

Focusing the attention to the main EU countries able to produce the highest agricultural output per hectare of land capital, Popescu et al. (2019) have argued as Malta, Cyprus, Belgium, Italy, Denmark, Germany, Luxemburg, Slovenia, Austria and France have had the best results; on the contrary, the countries with the highest level of economic efficiency of land endowment use have been found in Malta, Cyprus, Italy, Belgium, Germany, Denmark, Austria, Greece, Slovenia, Luxemburg, France, Spain, Netherlands, Portugal and Croatia.

The study proposed by Popescu et al. in 2019 has investigated only the statistical data proposed by EUROSTAT comparing all EU states. The further stage of this research has been addressed in evaluating as the differences in farms among EU countries are correlated to the endowment in some inputs as land, capital, and labour (Guth and Smędzik-Ambroży, 2020; Guth et al., 2020). According to Guth and Smędzik-Ambroży (2020), these inputs can influence the technical efficiency of farms as regards both the type of agricultural production and among European regions. The highest level of technical efficiency has been found in some EU–15 regions and not in new member states of the European Union and this has corroborated as some input endowments such as usable agricultural areas play a significant role in the technical efficiency, in productivity due to a different transformation of production factors in output.

This dichotomy between old and new member states of the EU in terms of technical efficiency has been widely investigated. One of the most recent studies in the estimation of the technical efficiency among European Union countries has pointed out significant differences among countries (Nowak et al., 2015; Guth and Smędzik-Ambroży, 2020). According to Nowak et al. (2015), the lowest values of technical efficiency have been found predominately in some new member states of the EU as Czech Republic, Lithuania, Hungary, Latvia, and Slovakia and only in Ireland an old EU member state. However, in some new member states of the EU during the transition phase the degree of technical efficiency is improved (Bojnec and Latruffe, 2013; 2009) even if the type of farming has had an important effect on the technical efficiency (Latruffe et al., 2004). Nowak et al. (2015) have investigated in depth which factors have been able to influence the technical efficiency such as the soil quality, the age of the head of the household and investments; on the (contrary,) the land capital endowment in terms of farms size has been irrelevant. A literature review has pointed out as many factors are able to influence the technical efficiency as argued by Nowak et al. in 2015 such as capital, soil quality and educational and skills that have had a positive effect while size of farms and age of farmers have had a mixed effect on the technical efficiency. In the latest new comers EU members states as Croatia the estimation of the technical efficiency in some specialized farms has been higher than in Hungarian farms due to a different dimension of farms in terms of usable agricultural areas and farm's specialization (Kovács et al., 2022). In this study emerged, as the technical efficiency in small farms has been higher than in medium-size farms (Kovács et al., 2022). These results have been in line with other investigations proposed in literature about the nexus between dimension of farms and technical efficiency as argued by other studies carried out in several European countries (Galluzzo, 2013; 2020; Kovács and Emvalomatis, 2011, Alvarez and Arias, 2004; Bojnec and Fertő, 2009; 2013; Bojnec and Latruffe, 2013; Minviel and Latruffe, 2017; Garrone et al., 2019).

The role and effect of financial subsidies allocated by the Common Agricultural Policy to the technical efficiency is unclear (Minviel and Latruffe, 2017). Hence, it is fundamental to clarify the effect of CAP financial supports to technical efficiency in farms (Bojnec and Fertő, 2009; 2013; Bojnec and Latruffe, 2013; Minviel and Latruffe, 2017; Garrone et al., 2019; Galluzzo, 2019). The effects of subsidies on the performance of farms such as technical efficiency and productivity although have had great interest in the economic literature are sometimes mixed and inconclusive (Barath et al., 2020; Minviel and Latruffe, 2017; Garrone et al., 2019; Galluzzo, 2019). In non-EU member states such as Serbia some subsidies as area payments and input subsidies have had important impacts on the technical efficiency of arable farms but investments and other subsidies do not impact to the farm technical efficiency (Todorović et al., 2020). Through a quantitative approach in the assessment of the technical efficiency by the Data Envelopment Analysis (DEA) in a sample of Italian farms, it has been possible to investigate, as farms were heavily dependent on public support, which affects their technical efficiency (Galluzzo, 2019). Hence, the first and second pillar subsidies and payments of the Common Agricultural Policy represent fundamental forces for increasing the competitiveness and technical efficiency in farms (Galluzzo, 2019; Garrone et al., 2019; Minviel and Latruffe, 2017; De Castris and Di Gennaro, 2017).

Furthermore, considering the huge amount of data investigated it has been useful to do some predictions of technical efficiency and defining the variables involved in the inefficiency using the machine learning. Machine learning (ML) has been proposed for the first time in 1959 by Samuel. ML is a branch of artificial intelligence that, using statistical methods, can improve the performance of an algorithm based on a large quantity of data and, hence, expanding the results through a process of independent learning (De Mauro, 2019; Bishop, 2006; Samuel, 1959; Galluzzo, 2021). One of the most common definitions of machine learning has been proposed by Samuel in 1959, according to which machine learning represents the ability of a machine to learn without any actions of programming. A recent literature review has argued as the concept of machine learning has been introduced because of the elaboration of huge amounts of data, and it offers a significant possibility to investigate in depth the relevance of the data without having to perform any programming actions (Liakos et al., 2018; Galluzzo, 2021). Liakos et al. (2018) carried out a complete and wide literature review arguing, as there are several approaches in this field of research and opportunities in using machine learning in agriculture. One year later, Storm et al. (2020) published a research with the aim of assessing challenges and opportunities in using machine learning in agriculture and in applied economics. Yu and Maruejols (2023) have argued in agricultural economics, as machine learning is a powerful tool to test hypotheses and identify if there are causal relations between variables even if only few studies have applied machine learning algorithms.

Machine learning models can investigate the complexity and diversity of data giving in the primary sector an adequate response to the issues about the optimization and profit maximization that in other field of economic studies are one of the major uses of machine learning (Pallathadka et al., 2023). Furthermore, other researches have been conceived to use ML in analysing big data in agriculture that represents a new, vast, and important challenge for investigations in the primary sector (Coble et al., 2018).

A specific use of machine learning method has been carried out in order to assess if a kind of CAP support such as agri-environmental subsides (AES) have had environmental effectiveness in German farms (Stetter et al., 2022). Despite these studies in machine learning literature, until now, it is not easy to find studies that utilize machine learning to investigate the impact of financial subsidies allocated by the Common Agricultural Policy in Italy and make predictions regarding the evolution of those subsidies for the medium-term. In fact, a recent literature review has investigated a keywords combination of machine learning and some words emphasizing the role of ML in agriculture (Benos et al., 2021). Other authors have underlined as machine learning is a current technology able to minimize the losses by providing some useful recommendations and insights about the crops management (Meshram et al., 2021).

The main purpose of this study was to assess firstly the technical efficiency in all EU farms part of FADN dataset since 2004 to 2020 estimating also if the technical efficiency changes in function of the type of farmings. This paper has tried to complete previous and recent studies in estimating the technical efficiency by a non-parametric approach in all EU countries, clarifying the role of CAP subsidies and filling the lack of scientific studies in agricultural economics assessing in depth which inputs are able to impact to the inefficiency patterns due to an excess in input using the Multi-directional Efficiency Analysis (MEA). The last stage of the research has used the machine learning approach by the conditional inference tree to predict which percentage of excess of input is able to impact to the technical efficiency. The novelty of this study is in the time of investigation (2004–2020) and in the assessment of the technical efficiency among all EU countries. In fact, one of the latest studies in estimating the technical efficiency in all EU countries has been focused in few years (Nowak et al., 2015) without considering CAP subsidies. Another innovative aspect of this research is to clarify the impact of financial subsidies on the technical efficiency for EU farms comparing previous researches proposed in literature (Barath et al., 2020; Minviel and Latruffe, 2017; Garrone et al., 2019; Galluzzo, 2019). Because of the lack in machine learning studies in agricultural economics literature, the conditional inference tree used in this paper has been able to fill the gap in studies that have applied machine learning algorithms in assessing causal relations between variables in a big dataset in agriculture as argued by Yu and Maruejols in 2023 and Coble et al. in 2018.

Methodology

In literature, the estimation of the technical efficiency in enterprises can use two different methodologies: a parametric methodology or a non-parametric method (Coelli et al., 2005; Kumbhakar et al., 2015; Galluzzo, 2018; 2019; 2020; 2021).

Through non-parametric modelling or Data Envelopment Analysis (DEA), that has been used in this study, it is possible to estimate the technical efficiency in a sample of European Union farms part of FADN dataset over the time of investigation 2004–2020 (Coelli et al., 2005; Kumbhakar et al., 2015; Galluzzo, 2021).

A literature review has pointed out as there is a shortage of studies using the DEA in order to estimate the technical efficiency in farms (Bravo-Ureta et al., 2007; Nowak et al., 2015; Galluzzo, 2013) and the results diverge among country and type of farmings. The DEA, is an analytical technique based on a linear programming approach fundamental in evaluating the efficiency of decision-making units in terms of converting inputs into outputs and it does not impose *a priori* constraints and functional forms of production function (Badunenko and Mozharovskyi, 2016). According to these two latter authors, DEA also allows to estimate the technical efficiency using multiple output technologies considering both multi-inputs and multi-output measuring the effectiveness of inputs used in a well-defined productive process (Fraser and Cordina, 1999; Dhungana et al., 2004; Alemdar and Necat Oren, 2006; Chebil et al., 2015; Igwe et al., 2017). Furthermore, according to Coelli et al. (2005), the DEA does not need to define preliminarily the relationship between input and output in order to estimate a frontier of production (Galluzzo, 2021).

Data Envelopment Analysis (DEA) seems to be a not so common method used in measuring the overall technical efficiency in agricultural economic literature in a group of decision-making units (DMU_s,) or farms in this case. The optimal level of efficiency is represented by all the DMU_o placed on the frontier of technical efficiency, that is the optimal combination of inputs and output given a technology set (Coelli et al., 2005; Kumbhakar et al., 2015; Banker et al., 1984; Cooper et al., 2007). On the contrary, the DMU_i placed under this frontier can be considered as inefficient, having a value lower than the optimal threshold that is equal to 1 (Coelli, et al., 2005). In this research it has used an input-oriented approach in the estimation of the technical efficiency by DEA.

The formulae for the minimization in an input-oriented model by the DEA approach are (Cooper et al., 2007; Bravo-Ureta et al., 2007; Coelli et al., 2005; Battese and Coelli, 1995):

$$\min_{\theta_B,\lambda} \theta_B$$

Subject to:

$$\theta_B x_o - X\lambda \geq 0$$

$$Y\lambda \geq y_o$$

$$e\lambda = 1$$

$$\lambda \geq 0$$

θ_B is the scalar value of the technical efficiency; X are the total input used in the productive process; $e\lambda$ is a vector of rows equal to 1; λ is a column vector with non-negative value; Y is the total produced output.

All European farms investigated in this research were part of the FADN dataset and the input variables were: total labour in hours, land capital in terms of usable agricultural area in hectares, specific costs linked to the productive process and total farming overhead costs in euro, total assets in euro and total output made by produced output plus total financial subsidies allocated by the first and second pillar of the Common Agricultural Policy in euro.

However, one of the main points of weakness of the DEA it is the impossibility to identify inefficiency or inefficiency patterns in each variable of the input and output. This bottleneck of the DEA has been overcome and solved in a second stage of this research. At this stage, in fact, it has used a new quantitative approach called Multi-directional Efficiency Analysis, or MEA (Bogetoft and Hougaard; 2003; Asmild et al., 2003; Hansson et al., 2020). According to these authors, MEA has the advantage of simultaneously estimating efficiency in multi-input and multi-output models and assessing inefficiency in each of the inputs used and outputs produced in the production process (Manevska-Tasevska et al., 2021). The most positive advantage of the MEA is to identify deviations from the production frontier, expressed in terms of productivity change, resulting from variables not incorporated in the analysis of previous estimation of the technical efficiency (Bogetoft and Hougaard; 2003, Hansson et al., 2020). Consequently, the MEA scores are in a range between zero, in the case of totally inefficient DMU, and 1, in the case of totally efficient DMUs where there is no excess in input or output. If the score of the MEA is 1 this implies as there was no potential for improvement in the input and/or output variables, while an input efficiency score estimated by the MEA less than one would indicate that the DMU_s should reduce the given input to be efficient.

The estimation of technical efficiency using both the DEA and the MEA approaches has been made using the RStudio software packages deaR, rDEA, and Benchmarking.

The last stage of this research, by the machine learning, has done some predictions on the technical efficiency considering the results of the DEA and the MEA. The results of the DEA have been modified in generating two qualitative categories of variables giving a value of low efficiency in farms having a value of technical efficiency estimated by the DEA under 1. Farms with a value of technical efficiency equal to 1 have been classified as optimal farms. By this assumption, it has been possible to use the conditional inference tree, which is a branch of artificial intelligence as proposed for the first time in 1959 by Samuel (De Mauro, 2019; Bishop, 2006).

This machine learning approach has been able to improve, employing statistical methods the performance of an algorithm based on a large quantity of data and, hence expanding the results through a process of independent learning (De Mauro, 2019; Bishop, 2006; Samuel, 1959).

In this research, it has used the conditional inference tree whose aim is to select and to split recursively a vast majority of predictor variable, as the results in percentage of the excess in input, with a nexus to the outcome that is in this research the technical efficiency estimated by the DEA. As argued by Kassambara in 2017, the advantage of using the conditional interference tree is that the algorithm stops if there is not a relation of dependence between predictor variables (excess in input) and outcome variable (technical efficiency). The estimation of machine learning in the conditional interference tree has been made using the RStudio software packages rpart, rpart.plot, party and caret.

Results

The analysis of descriptive statistics in all EU farms part of the FADN dataset has pointed out as the average value of labour input has been close to 5.000 hours per farm (Table 1) and the average value of land capital in terms of usable agricultural area has been above 62 hectares with significant fluctuations among countries. In fact, the highest land capital endowment has been assessed in Germany, Czech Republic, Sweden, and Estonia with respectively 171, 140, 104 and 100 hectares per farm; on the contrary, the lowest amount of usable agricultural area has been detected in Malta where the average value has been close to 2.79 hectares and Croatia where the average value of usable agricultural area has been equal to 11.75 hectares. The variable specific costs has been almost two times higher than total farming overheads cost and in the same time significant has been the value of total asset per

Table 1. Main descriptive statistics in all EU farms part of FADN dataset over the time of investigation 2004–2020

	Labour (hours)	Land capital (ha)	Specific costs (€)	Farming overhead costs (€)
Average	4916.08	62.78	67257.34	38317.69
Standard deviation	5190.64	107.64	116480.56	59731.97
Range	95236.27	1221.81	1600574.00	800593.00
Min	640.07	0.01	500.00	114.00
Max	95876.34	1221.81	1601074.00	800707.00
Observations (n°)	12183	12183	12183	12183
	Assets (€)	**Total CAP subsidies (€)**	**II pillar CAP (€)**	**Total output (€)**
Average	521685.44	23450.23	4509.14	164189.77
Standard deviation	634353.66	40626.13	10033.39	238717.57
Range	7538252.00	486845.00	185049.00	2616779.00
Min	10036.00	0.00	0.00	4026.00
Max	7548288.00	486845.00	185049.00	2620805.00
Observations (n°)	12183	12183	12183	12183

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

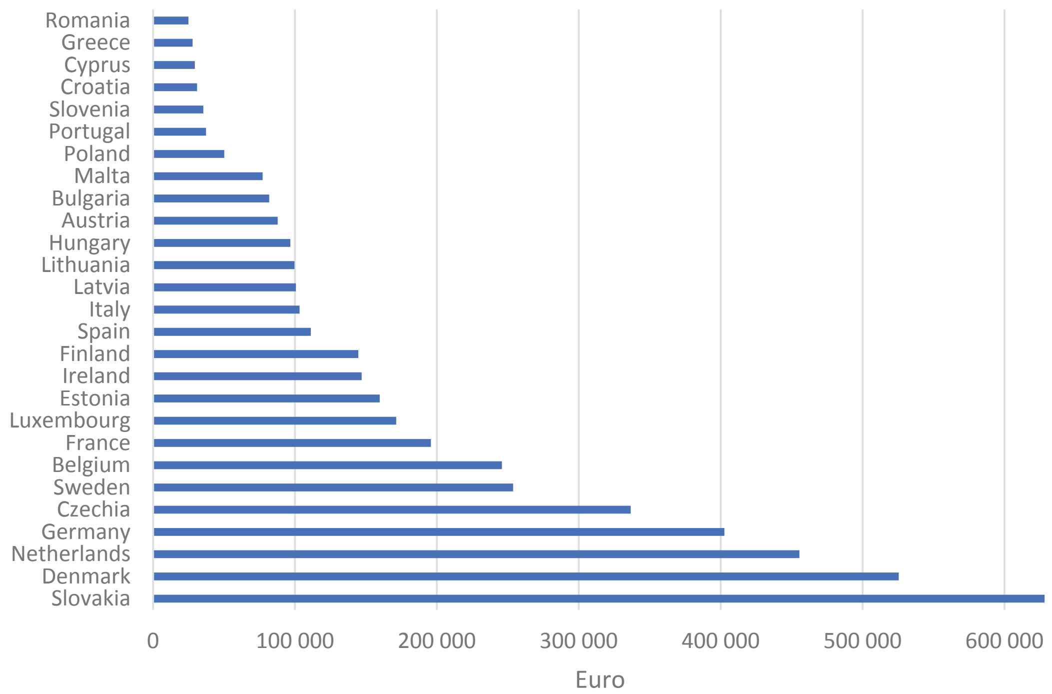

Fig. 1. Total output in average in all EU farms part of FADN dataset

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

farm close to 521.000 euro. The total produced output in average has been close to 164.000 euro per farm with significant unbalances among EU countries (Fig. 1). If Romania farms on average have had output values detected of 25.000 euros, on the contrary the farms of Slovakia, Danish and Dutch have had an output produced per farms that in average was 18–20 times than total output produced by Romanian farms.

Addressing the attention to the financial subsidies allocated by the Common Agricultural Policy (CAP) it emerges as financial subsidies and payments disbursed by the second pillar of the CAP are the one fifth of the total subsidies allocated by the Common Agricultural Policy.

The average value of technical efficiency in all farms part of FADN dataset has been 0.6143 very low compared to other studies (Nowak et al., 2015), which have assessed a value of technical efficiency in a range between 0.812 and 0.848 (Table 2). However, these research's findings seem to be closer to other studies carried out in different EU countries in some specialized farms using parametric and non-parametric approaches in the estimation of technical efficiency (Bravo-Ureta et al., 2007). The farms more technical efficient have been found in Denmark and in Romania where the average value of technical efficiency estimated by the DEA input-oriented method has been equal to 0.78 and 0.70. By contrast, the lowest value of technical efficiency has been found in farms located in Czech Republic and in Slovakia. Nine countries out of 27 have had a maximum value of technical efficiency under the optimal threshold equal to 1 and drawing some conclusions a dichotomy between new member states of the EU and old member states with few exceptions (Luxemburg, Belgium, and Austria) exist. As argued by Nowak et al. (2015) Austria is characterized by the less level of technical efficiency in farms.

Financial subsides and other payments have had some impacts to the technical efficiency in farms. In fact, in studies proposed by Nowak et al. in 2015 the output did not include subsidies as on the contrary proposed in this study. Hence, research findings proposed in this paper are in line with other ones argued in several researches (Bojnec and Fertő, 2013; Bojnec and Latruffe, 2013; Minviel and Latruffe, 2017; Garrone et al., 2019; Galluzzo, 2019; Kumbhakar and Lien, 2010), according to which the CAP subsidies have reduced the technical efficiency in farms. The size of farms and the productive specialization can impact more than financial subsidies the technical efficiency hence, farms with the lowest

Table 2. Main results of technical efficiency in farms part of FADN dataset estimated by DEA input oriented

Member state	Mean	Std. dev.	min	max
Austria	0.4975	0.0641	0.3975	0.6850
Belgium	0.5689	0.1096	0.4111	0.9267
Bulgaria	0.6138	0.1438	0.2408	1.0000
Cyprus	0.6518	0.1462	0.3672	1.0000
Czechia	0.4372	0.1427	0.2549	0.9590
Germany	0.5806	0.1515	0.2722	1.0000
Denmark	0.7881	0.0957	0.5565	1.0000
Estonia	0.4931	0.0888	0.3521	0.9093
Greece	0.6784	0.1045	0.4906	1.0000
Spain	0.6303	0.1272	0.3598	1.0000
Finland	0.6151	0.1736	0.3771	1.0000
France	0.6695	0.1172	0.3678	1.0000
Croatia	0.5910	0.0757	0.4046	1.0000
Hungary	0.5408	0.1178	0.2809	1.0000
Ireland	0.6052	0.1116	0.4563	1.0000
Italy	0.6278	0.1128	0.3686	1.0000
Lithuania	0.5368	0.0874	0.3862	1.0000
Luxembourg	0.5400	0.0482	0.4657	0.6881
Latvia	0.4911	0.0500	0.3178	0.5851
Malta	0.6022	0.0912	0.4144	1.0000
Netherlands	0.6744	0.1289	0.4516	1.0000
Poland	0.4984	0.0641	0.3596	0.8174
Portugal	0.6680	0.1265	0.4346	1.0000
Romania	0.7068	0.1362	0.3022	1.0000
Sweden	0.5778	0.0994	0.4087	1.0000
Slovenia	0.5826	0.1044	0.3942	0.9243
Slovakia	0.4891	0.1855	0.1693	0.8721
Average all EU FADN farms	0.6143	0.1360	0.1693	1.0000

Source: Author's elaboration on data https://agri-data.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

amount of land capital can be less technical efficient than big ones (Zhu and Lansink, 2010). The results have been in line with others proposed by Minviel and Latruffe in 2017 and by Latruffe et al. (2017; 2004), according to which, the type of subsidies such as some of them allocated by the first pillar of the CAP and the total amount of allocated subsidies reduce the technical efficiency in farms. Comparing other studies carried out in EU new entrant member states the results of this research have been almost similar hence, subsidies disbursed towards new member states reduce the technical efficiency in farms (Bakucs et al., 2010).

Comparing the type of farmings the main results have pointed out as farms specialised in granivores and in horticulture have been more technical efficient than farms specialized in milk and in other grazing livestock (Table 3) and this is in line with other studies carried in other countries (Zhu and Lansink, 2010).

Table 3. Technical efficiency in different type of farming in all EU countries

Type of farming	Mean	St. dev.
Field crops	0.6182	0.1273
Horticulture	0.6584	0.1513
Wine	0.6210	0.1433
Other permanent crops	0.6106	0.1434
Milk	0.5776	0.1167
Other grazing livestock	0.5720	0.1125
Granivores	0.6943	0.1433
Mixed	0.6082	0.1284
Total	0.6143	0.1360

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

In order to estimate the patterns of technical inefficiency, overcoming the bottlenecks of the DEA, it has used the Multi-directional Efficiency Analysis (MEA) in all farms part of EU FADN dataset (Table 4). Land capital was the input with the highest excess of input hence, the farms of the FADN sample have had an excess in this input close to 35%. The labour input has had the lowest level of excess that has been equal to 21%. Focusing the attention to the total output the MEA results have pointed out an excess equal to 30% of produced output compared to the optimal level, which has implied some negative effects of the technical efficiency in farms.

The comparison of the MEA results (Table 5) has pointed out as Romania has had the lowest excess of land capital, total farms overhead costs and total asset. Denmark has had the lowest excess in labour input, specific costs, and total output. Slovakia has had the highest excess of labour and land capital equal to 38% and 42% in excess compared to the optimal value. In Austria the results of the MEA have underlined as there is an excess in total asset equal to 35%.

Granivores farms have been characterized by the lowest excess of labour input; on the contrary, milk farms have had the highest level of technical inefficiency due to an excess of 24% of labour input (Table 6). Horticulture farms have been

Table 4. Average value of inefficiency patterns estimated by the MEA in all EU farms

Variable	Obs	Mean	Std. Dev.	Min	Max
MEA Labour	12,183	0.7899476	0.0793168	0.4357442	1.000
MEA Land capital	12,181	0.6477293	0.0805882	0.3676947	1.000
MEA Specific cost	12,183	0.6829993	0.0905781	0.4159821	1.000
MEA Farming overhead costs	12,183	0.692225	0.083881	0.4176181	1.000
MEA Assets	12,183	0.7186001	0.0856952	0.4547878	1.000
MEA Total output	12,183	0.7039512	0.1381292	0.300288	1.000

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

Table 5. Average value of inefficiency estimated by the MEA in all EU countries

Member State	MEA Labour	MEA Land capital	MEA Specific cost	MEA Farming overheads cost	MEA Assets	MEA Total output
Austria	0.7329105	0.5988267	0.6314526	0.6222082	0.6457595	0.5434827
Belgium	0.7614797	0.6314995	0.6591534	0.6974636	0.7049554	0.7270659
Bulgaria	0.7693044	0.6581413	0.6954197	0.7182619	0.7709452	0.7014403
Cyprus	0.8228823	0.6690911	0.7122109	0.6520161	0.6827049	0.5575677
Czechia	0.6555397	0.567947	0.6214817	0.6171129	0.6661285	0.6349576
Germany	0.7598826	0.6243215	0.6807995	0.672497	0.703057	0.7166828
Denmark	0.8902980	0.7139748	0.7965643	0.7755779	0.7478386	0.8865875
Estonia	0.7071969	0.5901754	0.6164368	0.6313587	0.6825535	0.5833416
Greece	0.8358087	0.6866771	0.7241852	0.7271146	0.7458933	0.7122691
Spain	0.8048894	0.656364	0.6978896	0.7058735	0.7168588	0.7110363
Finland	0.787257	0.6699292	0.685874	0.6888368	0.7288628	0.6809344
France	0.8242702	0.6501917	0.6679239	0.695785	0.7628241	0.7738995
Croatia	0.7880202	0.6394732	0.6679591	0.6710585	0.6803717	0.625655
Hungary	0.745056	0.6143395	0.6402626	0.6559048	0.7142498	0.6215266
Ireland	0.7978966	0.6424942	0.6660499	0.6893584	0.6623324	0.6737273
Italy	0.8054621	0.6577199	0.694807	0.6978527	0.6947451	0.7095264
Lithuania	0.734629	0.6158426	0.6624408	0.6735817	0.7108731	0.6339463
Luxembourg	0.7631534	0.6137686	0.6327063	0.6592702	0.6483103	0.6776681
Latvia	0.7049584	0.5862828	0.6167615	0.6259704	0.6888488	0.561423
Malta	0.8009662	0.6730639	0.6787075	0.6683405	0.6841899	0.6260686
Netherlands	0.8283581	0.674231	0.7209274	0.7253131	0.7103715	0.8114213
Poland	0.7349732	0.6113988	0.6437721	0.6440477	0.6711488	0.5562161
Portugal	0.8293378	0.6834064	0.7210088	0.7263054	0.7611626	0.7118482
Romania	0.8358091	0.7257036	0.7534821	0.7767367	0.8024077	0.7390467
Sweden	0.7725881	0.6158561	0.6444153	0.6576147	0.6806219	0.701059
Slovenia	0.7922289	0.6322443	0.6593177	0.6356248	0.6557751	0.5206847
Slovakia	0.6200175	0.5597702	0.6588585	0.6278405	0.6906017	0.7546105
Total	0.7899476	0.6477293	0.6829993	0.692225	0.7186001	0.7039512

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

characterized by the lowest excess in land capital while milk and other grazing livestock farms have had the highest excess in land capital. Focusing the attention to the item costs (specific costs and total farming overhead costs), granivores and horticulture farms have had less excess in terms of costs correlated to the productive process. The input assets have had the highest level of excess equal to 22% in granivores farms. Drawing some brief conclusions, it seems that granivores and horticulture farms have had the best performances with low levels of excess in all used input and produced output.

The conditional inference tree estimated by the machine learning in all European farms part of FADN dataset (Fig. 2) has put in relationship the input in excess estimated by the MEA and the technical efficiency that has been stratified in two qualitative variables (low versus optimal technical efficiency). Low efficiency is a cluster

Table 6. Average value of inefficiency estimated by the MEA in all types of farming

Types of farming	MEA Labour	MEA Land capital	MEA Specific cost	MEA Farming overheads cost	MEA Assets	MEA Total output
Field crops	0.7979995	0.6336298	0.6667976	0.6772077	0.7094625	0.6841334
Horticulture	0.7962405	0.7064881	0.7381782	0.7357733	0.7823261	0.7841878
Wine	0.80636	0.6639059	0.7102384	0.6931997	0.7017685	0.6882033
Other permanent crops	0.7916691	0.6539199	0.7110461	0.6856145	0.7071401	0.6636301
Milk	0.7649329	0.6214858	0.6522151	0.6786611	0.7023561	0.7166518
Other grazing livestock	0.771554	0.6242174	0.6463777	0.6665822	0.6869768	0.6379703
Granivores	0.8216437	0.689252	0.7319219	0.7619718	0.7879806	0.8297781
Mixed	0.7900437	0.6347547	0.6668373	0.6795309	0.7081775	0.6865259

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

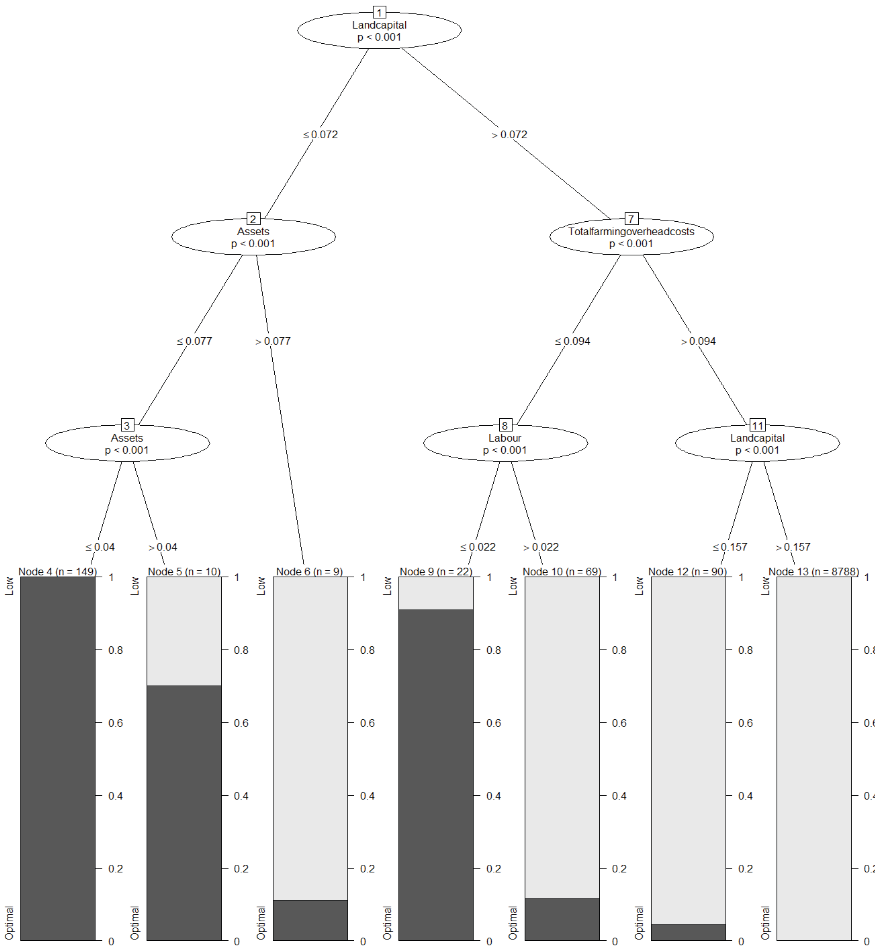

Fig. 2. Conditional inference tree using FADN dataset in order to assess inefficient inputs impacting on the technical efficiency

Source: Author's elaboration on data https://agridata.ec.europa.eu/extensions/FADNPublicDatabase/FADNPublicDatabase.html.

of farms with a score of technical efficiency under 1 while farms stratified as optimal are farms with a score of technical efficiency estimated by the DEA equal to 1.

The p value < 0.001 in all input used in the conditional inference tree has been useful to asses the association between predictor variables (land capital, assets, labour, and total farming overhead costs) and the outcome variable (technical efficiency). All above mentioned predictor variables are mostly associated to the technical efficiency and they can impact to the farm performances. Drawing some conclusions, farms with the optimal level of technical efficiency equal to 1 have been characterized by an excess of land capital under 7.2%, a low excess of asset both under 7.7% and under 4%. On the contrary, farms totally inefficient have had an excess of land capital above 7.2%, an excess in total farming overhead costs above the 9.4% and an excess in land capital above 15.7%. In order to make some predictions in this model, results have underlined as the test model is able to predict the relationships between excess in input and technical efficiency with an accuracy rate of 0.97.

Discussion and conclusions

This study has used lots of data in order to estimate the technical efficiency by a non-parametric approach investigating all type of farmings used in the FADN. In general, in literature it is not so common to find study that have used a large dataset with lots of farms in order to assess the technical efficiency (Bravo-Ureta et al., 2007). Comparing finding's research in this study to others it emerges as the dichotomy in technical efficiency between new and old member states of the EU exists as argued by previous investigations (Nowak et al., 2015; Bakucs et al., 2010; Guth and Smędzik-Ambroży, 2019). The average value of technical efficiency in all EU farms part of the FADN has been lower than the technical efficiency estimated in a limited time of investigation as previously argued in other studies (Nowak et al., 2015). The Multi-directional Efficiency Analysis has filled the gap in literature in order to assess which are the inputs able to be inefficient assessing the percentage of excess of input able to cause this technical inefficiency. In this paper by the MEA, it has been possible to investigate in which EU countries there are specific excesses in some inputs able to impact to the technical efficiency and this could be useful to address a specific policy aimed at reducing some input excess improving the performances of farms. Furthermore, by an analysis in each type of farming it has been possible to observe, which inputs are in excess and in which proportion. In addition, by the machine learning it was possible to make predictions to assess variables with an excess of input may be dumped on the overall technical inefficiency. In this study, the inference tree has corroborated the theoretical hypothesis according to which the machine learning is a good tool in doing some provisions with a good level of accuracy as proposed by other studies and not so commonly used in studies of agricultural economics (De Mauro, 2019; Bishop, 2006; Yu and Maruejols, 2003; Coble et al., 2018).

Comparing this study with other aimed at assessing the technical efficiency in all EU countries the differences in terms of technical efficiency can be attributed in having also included in the total output item the funds and financial aids provided by the CAP (Nowak et al., 2015; Bravo-Ureta et al., 2007). Hence, this study has been remained in the line of previous studies, which argued that financial subsidies allocated by the CAP have modified the technical efficiency in farms and particularly subsidies allocated by the first pillar can directly reduce the efficiency in farms more than subsidies allocated by the second pillar of the CAP (Bojnec and Fertő, 2013; Bojnec and Latruffe, 2013; Minviel and Latruffe, 2017; Garrone et al., 2019; Galluzzo, 2019; Kumbhakar and Lien, 2010). However, the effects of subsidies were quite heterogeneous depending on production specialization, as argued in this study and on farm size in terms of land capital endowment as investigated in different EU countries (Zhu and Lansink, 2010; Galluzzo, 2013).

The novelty of this study was twofold: to assess the input able to impact to the technical inefficiency and to define some patterns of inefficiency where it is possible to evaluate the percentage of excess of input. These two aspects fill the gap in literature about the investigation of the reasons linked to the inefficiency in farms. By the machine learning, it has been possible to carry out some estimations and previsions for the future with a good level of accuracy finding an easy explanatory framework of the relationships between input variables and patterns of technical efficiency. The next lines of research should analyse the role of specific subsidies and payments provided by the first and second pillars of CAP on technical efficiency investigating causes of inefficiency using a dummy variable that can assess whether firm size above or below a certain threshold acts on technical efficiency.

References

Alemdar, T., & Necat Oren, M. (2006). Determinants of technical efficiency of wheat farming in Southeastern Anatolia, Turkey: a nonparametric technical efficiency analysis. *Journal of Applied sciences*, *6*(4), 827-830.

Alvarez, A., & Arias, C. (2004). Technical efficiency and farm size: a conditional analysis. *Agricultural Economics*, *30*(3), 241-250.

Asmild, M., Hougaard, J. L., Kronborg, D., & Kvist, H. K. (2003). Measuring inefficiency via potential improvements. *Journal of productivity analysis*, *19*, 59-76.

Badunenko, O., & Mozharovskyi, P. (2016). Nonparametric frontier analysis using Stata. *The Stata Journal*, *16*(3), 550-589.

Bakucs, L. Z., Latruffe, L., Fertő, I., & Fogarasi, J. (2010). The impact of EU accession on farms' technical efficiency in Hungary. *Post-communist economies*, *22*(2), 165-175.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

Baráth, L., Fertő, I., & Bojnec, Š. (2020). The effect of investment, LFA and agri-environmental subsidies on the components of total factor productivity: the case of Slovenian farms. *Journal of Agricultural Economics*, *71*(3), 853-876.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, *20*, 325-332.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer, Berlin.

Bogetoft, P., & Hougaard, J. L. (2003). Rational inefficiencies. *Journal of Productivity Analysis, 20*(3), 243-271.

Bojnec, Š., & Fertő, I. (2013). Farm income sources, farm size and farm technical efficiency in Slovenia. *Post-Communist Economies*, *25*(3), 343-356.

Bojnec, Š., & Fertő, I. (2019). Do CAP subsidies stabilise farm income in Hungary and Slovenia? *Agricultural Economics, 65*(3), pp. 103-111.

Bojnec, Š., & Latruffe, L. (2009). Determinants of technical efficiency of Slovenian farms. *Post-Communist Economies*, *21*(1), 117-124.

Bojnec, Š., & Latruffe, L. (2013). Farm size, agricultural subsidies and farm performance in Slovenia. *Land use policy*, *32*, 207-217.

Bravo-Ureta, B. E., Solís, D., Moreira López, V. H., Maripani, J. F., Thiam, A., & Rivas, T. (2007). Technical efficiency in farming: a meta-regression analysis. *Journal of productivity Analysis*, *27*, 57-72.

Chebil, A., Frija, A., & Thabet, C. (2015). Economic efficiency measures and its determinants for irrigated wheat farms in Tunisia: a DEA approach. *New Medit*, *14*(2), 32-38.

Coble, K. H., Mishra, A. K., Ferrell, S., & Griffin, T. (2018). Big data in agriculture: A challenge for the future. *Applied Economic Perspectives and Policy*, *40*(1), 79-96.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer science & business media.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. New York: Springer.

De Castris, M., & Di Gennaro, D. (2017). What is below the CAP? Evaluating spatial patterns in agricultural subsidies. In: XXXVIII Annual Scientific Conference of the AIS Re. Italian Association of Regional Science, Cagliari, Italy.

De Mauro, A. (2019). Big data analytics: guida per iniziare a classificare e interpretare dati con il machine learning. Apogeo, Milano.

Dhungana, B. R., Nuthall, P. L., & Nartea, G. V. (2004). Measuring the economic inefficiency of Nepalese rice farms using data envelopment analysis. *Australian Journal of Agricultural and Resource Economics*, *48*(2), 347-369.

Fraser, I., & Cordina, D. (1999). An application of data envelopment analysis to irrigated dairy farms in Northern Victoria, Australia. *Agricultural Systems*, *59*(3), 267-282.

Galluzzo, N. (2013). Farm dimension and efficiency in Italian agriculture: a quantitative approach. *American Journal of Rural Development*, *1*(2), 26-32.

Galluzzo, N. (2018). Impact of the Common Agricultural Policy payments towards Romanian farms. *Bulgarian Journal of Agricultural Science, 24*(2), pp. 199-205.

Galluzzo, N. (2019). A long-term analysis of the common agricultural policy financial subsidies towards Italian farms. *Ukrainian journal of veterinary and agricultural sciences*, *2*(1), 12-17.

Galluzzo, N. (2020). A technical efficiency analysis of financial subsidies allocated by the CAP in Romanian farms using stochastic frontier analysis. *European Countryside*, *12*(4), 494-505.

Galluzzo, N. (2021). A quantitative analysis on Romanian rural areas, agritourism and the impacts of European Union's financial subsidies. *Journal of Rural Studies*, *82*, 458-467.

Garrone, M., Emmers, D., Lee, H., Olper, A., & Swinnen, J. (2019). Subsidies and agricultural productivity in the EU. *Agricultural Economics*, *50*(6), 803-817.

Gunes, E., & Guldal, H. T. (2019). Determination of economic efficiency of agricultural enterprises in Turkey: a DEA approach. *New Medit*, *18*(4), 105-115.

Guth, M., & Smędzik-Ambroży, K. (2020). Economic resources versus the efficiency of different types of agricultural production in regions of the European Union. *Economic research-Ekonomska istraživanja*, *33*(1), 1036-1051.

Guth, M., Smędzik-Ambroży, K., Czyżewski, B., & Stępień, S. (2020). The economic sustainability of farms under common agricultural policy in the European Union countries. *Agriculture*, *10*(2), 34.

Hansson, H., Manevska-Tasevska, G., Asmild, M. (2020). Rationalising inefficiency in agricultural production – the case of Swedish dairy agriculture. European Review of Agricultural Economics, 47(1), 1-24.

Igwe, O. O., Nwaogu, D. C., & Onyegbule, F. (2017). Technical efficiency of poultry enterpreneurs in Abia state: a stochastic frontier approach. *Scientific Papers: Management, Economic Engineering in Agriculture & Rural Development*, *17*(1).

Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). Sthda.

Kovács, K., & Emvalomatis, G. (2011). Dutch, Hungarian and German dairy farms technical efficiency comparison. APSTRACT: Applied Studies in Agribusiness and Commerce, 5(1033-2016-84134), 121-128.

Kovács, K., Juračak, J., Očić, V., Burdiuzha, A., & Szűcs, I. (2022). Evaluation of technical efficiency of Hungarian and Croatian livestock sectors using DEA on FADN data. *Journal of Central European Agriculture*, *23*(4), 909-920.

Kumbhakar, S. C., Wang, H. J., Horncastle, A. P. (2015). A practitioner's guide to stochastic frontier analysis using Stata. Cambridge University Press. Cambridge.

Latruffe, L., Balcombe, K., Davidova, S., & Zawalinska, K. (2004). Determinants of technical efficiency of crop and livestock farms in Poland. *Applied economics*, *36*(12), 1255-1263.

Latruffe, L., Bravo-Ureta, B. E., Carpentier, A., Desjeux, Y., & Moreira, V. H. (2017). Subsidies and technical efficiency in agriculture: Evidence from European dairy farms. *American Journal of agricultural economics*, *99*(3), 783-799.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

Manevska-Tasevska, G., Hansson, H., Asmild, M., & Surry, Y. (2021). Exploring the regional efficiency of the Swedish agricultural sector during the CAP reforms–multi-directional efficiency analysis approach. *Land Use Policy*, *100*, 104897.

Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010.

Minviel, J. J., & Latruffe, L. (2017). Effect of public subsidies on farm technical efficiency: a meta-analysis of empirical results. *Applied Economics*, *49*(2), 213-226.

Nowak, A., Kijek, T., & Domańska, K. (2015). Technical efficiency and its determinants in the European Union. *Agricultural Economics*, *61*(6), 275-283.

Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2023). Impact of machine learning on management, healthcare and agriculture. *Materials Today: Proceedings*, *80*, 2803-2806.

Popescu, A., Dinu, T. A., & Stoian, E. (2019). Efficiency of the agricultural land use in the European Union. *Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development*, *19*(3), 475-486.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), pp. 210-229.

Stetter, C., Mennig, P., & Sauer, J. (2022). Using machine learning to identify heterogeneous impacts of agri-environment schemes in the EU: a case study. *European Review of Agricultural Economics*, *49*(4), 723-759.

Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, *47*(3), 849-892.

Todorović, S., Papić, R., Ciaian, P., & Bogdanov, N. (2020). Technical efficiency of arable farms in Serbia: do subsidies matter? *New Medit: Mediterranean Journal of Economics, Agriculture and Environment = Revue Méditerranéenne d'Economie Agriculture et Environment*, *19*(4).

Yu, X., & Maruejols, L. (2023). Prediction, pattern recognition and machine learning in agricultural economics. *China Agricultural Economic Review*, *15*(2), 375-378.

Zhu, X., & Lansink, A. O. (2010). Impact of CAP subsidies on technical efficiency of crop farms in Germany, the Netherlands and Sweden. *Journal of Agricultural Economics*, *61*(3), 545-564.