Check for updates

Received: 19 June 2020 | Revised: 22 November 2020 | Accepted: 10 March 2021

DOI: 10.1111/agec.12662

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists
IAAE
WILEY

ORIGINAL ARTICLE

Does organic certification make economic sense for dairy farmers in Europe?–A latent class counterfactual analysis

Christian Grovermann[1] | **Sylvain Quiédeville**[1] | **Adrian Muller**[1,2] | **Florian Leiber**[1] | **Matthias Stolze**[1] | **Simon Moakes**[1]

[1] FiBL—Research Institute of Organic Agriculture, Frick, Switzerland

[2] ETHZ–Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Correspondence
Christian Grovermann, FiBL—Research Institute of Organic Agriculture, Ackerstrasse 113, Frick 5070, Switzerland
Email: christian.grovermann@fibl.org

Abstract

Certification in agriculture ensures compliance with tangible standards and should generate economic opportunities for farmers. This study quantifies the variable profit and efficiency impacts of organic certification in dairy farming across Europe, using farm-level FADN data from 25 countries while accounting for heterogeneity through a class splitting model. Four distinct classes with dairy farm enterprises operating under similar production conditions were identified in order to assess gross margin and efficiency differences among certified and non-certified farms. Depending on the nature of the selection bias, treatment effects were estimated either through an endogenous treatment model or through entropy balancing. The results suggest that organic certification increases gross margins for dairy farm enterprises in Europe, while slightly increasing technical efficiency in two out of four classes. These significant effects of certification on efficiency were estimated at 2% and 7%, respectively. As regards variable profit, effects range from to 66 Euros per cow to 234 euros per cow. In relative terms, this implies gains between 38% and 50% for farms classified into more cool or temperate zones and a gain of up to 182% for the farms assigned to the class that designates warmer climatic conditions.

KEYWORDS

efficiency, endogenous treatment model, gross margins, stochastic frontier

JEL CLASSIFICATION

Q12, Q18

1 | INTRODUCTION

Agroecology is increasingly recognized as an important strategy for achieving more sustainable agricultural and food systems (FAO, 2018; HLPE, 2019). To work for farmers, it needs to make economic sense. This requires that agroecology interventions are systematically evaluated in terms of their economic potential. Certification plays a key role in this context. Organic farming by its nature entails the application of many agroecological principles, but is in addition formalized by using standards to certify compliance with these principles (Mockshell & Villarino, 2019). Organic agriculture is globally growing rapidly (Willer & Lernoud, 2019) and, under the right conditions,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Agricultural Economics* published by Wiley Periodicals LLC on behalf of International Association of Agricultural Economists

the prospects for farmers created through organic certification can make an important contribution to sustainable rural livelihoods (Crowder & Reganold, 2015).

Sustainability standards have been evaluated for a wide range of outcomes, such as gender equality (Meemken & Qaim, 2018), adoption of good agricultural practices (Ibanez & Blackman, 2016) or pesticide use reduction (Schreinemachers et al., 2012). Many impact evaluation studies related to sustainability standards seek to examine economic effects (Schleifer & Sun, 2020). One key objective of certification, including organic standards, is to generate economic opportunities for farmers. Increasing case study evidence from a low- and middle-income country context (e.g.; Bolwig et al., 2009; Mendoza, 2004; Ssebunya et al., 2019; Tran & Goto, 2019;) as well as from a high-income country context (e.g., Hoop et al., 2017; Moakes et al., 2016) point towards financial benefits of organic certification. A meta-study of the competitiveness of organic farming showed higher profitability and benefit-cost ratios for organic certified farms as compared to non-certified farms (Crowder & Reganold, 2015). Additionally, a recent review paper compiled empirical evidence on the economic potential of agroecology in Europe, including data from several certified organic cases (van der Ploeg et al., 2019). As regards efficiency, less studies are available though. The literature generally suggests that organic farms are less efficient, especially when measured against the same production frontier as conventional farms (Oude Lansink et al., 2002; Kumbhakar et al., 2009; Mayen et al., 2010).

We add to the existing literature by conducting a cross-country efficiency and gross margin impact analysis using a large representative data set covering dairy farms across 25 EU countries. The dairy industry, both organic and conventional, is highly competitive and, in terms of output value, the second biggest agricultural activity in the EU after vegetable production (EPRS, 2018). Organic cow milk production in Europe has almost doubled between 2008 and 2017, with latest figures indicating an output of 4.7 million metric tons. The organic market share ranges from less than 1% to above 10% in some Central European countries (Willer & Lernoud, 2019). Changing consumer preferences and pressures on milk prices are considered as key drivers behind rising numbers of conversion among dairy farmers (Bouttes et al., 2019). Certification is crucial for managing compliance with agroecological farming requirements and ensuring price premiums, which motivate adoption (Serra et al., 2008).

To account for variable profit and efficiency effects of standard implementation, this study relies on two key performance indicators: Gross margins are calculated to assess impacts on dairy enterprise gross margins, and efficiency scores are estimated to provide insights into the conversion of production inputs into economic output. In the framework of a counterfactual analysis, the two selected outcome measures can explain how organic certification influences farm performance, while taking into account possible trade-offs or synergies between variable profit and efficiency.

The novelty of this research rests upon its scope and methodological approach. To the knowledge of the authors, no rigorous impact studies exist that evaluate the profitability and efficiency effects of organic certification at a regional scale, with other similar studies being location-specific and mostly focusing on farm incomes only. The cross-country impact estimates are of wide relevance for practitioners and policy-makers in Europe and beyond. A key challenge in covering almost the entire EU dairy sector is to account for contextual factors. By combining a latent class model with state-of-the-art impact evaluation methods for observational data, the study quantifies the context-specific economic effects of organic certification. This allows drawing a more precise and comprehensive picture of certification impacts. The broad geographic coverage and the significance of the dairy sector should foster a better understanding of the wider economic implications of organic certification.

The analysis finds that obtaining certification generally makes economic sense for organic dairy farmers in Europe. The next section of the paper explains data sources, the class splitting approach, and the econometric model used to assess certification impacts. Impact estimates for all four classes are then presented in section three and discussed in section four of the paper. The final section also includes some relevant policy implications that can be derived from the analysis.

2 | MATERIAL AND METHODS

2.1 | Data

The study utilizes detailed economic data from the Farm Accountancy Data Network (FADN, 2018) database for farm types that include cattle systems for the period 2011 until 2013 (the latest year that data was available for analysis at point of data application). It covers 25 EU member countries, representing a large diversity of farming across Europe. The organic certification variable in the dataset is specified in line with the regulations that govern the EU organic standard (Reg. EC No. 834/2007 and Reg. EC No. 889/2008). This implies for instance more adequate space for cows, no preventive use of antibiotics and livestock feed of organic origin. Farms defined as organic in the dataset have fully converted and obtained official accreditation.

To identify farms with a significant dairy enterprise, only those farms with a minimum of 35% of economic output from the dairy enterprise (specialization rate) are

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists
IAAE WILEY

included in the dataset, as per FADN recommendations (FADN, 2018). Different specialization rates (25% and 45%) were applied as part of the robustness tests performed in the analysis. Different rates did not change results. The FADN dataset does not represent a complete balanced panel data structure, as there is no exact overlap among farms from 1 year to another. While certain farms remain in the sample, some drop out and others are added. For the 10-year period between 2004 and 2013 only 22 % of farms are identical (between 1% and 56% depending on the country). Each year ca. 18% of farms dropped out and were replaced. This complicates time-series analysis, as impact studies need to compare the same farms across time. Therefore, the present study relies on a pooled subset of data including observations from the three most recent years in the entire dataset, that is, 2011, 2012, and 2013. Year-dummies are included in the analyses where needed to account for temporal effects. In line with the classification explained in the following section, summary statistics for key variables used in the analysis are provided in Table 1.

All inputs, aside from land and labor, are measured in constant monetary terms using 2013 as base year. Output is used in the analysis both as a physical and as a financial parameter (for gross margin calculations). All indicators have been computed according to the proposed FADN calculation methodology (FADN, 2018). The monetary output variable includes direct revenues from milk sales, but excludes subsidies. The inflation adjustment of input values is based on the Harmonized Index of Consumer Prices (HICP) and the price indices of the means of agricultural production provided by Eurostat (2018, 2019). We made use of the latter to deflate feed and forage costs, while the other monetary variables, less associated to fertilizer markets, were adjusted by the HICP. Feed costs comprise coarse fodder, non-fodder, and concentrate expenditures, while forage costs were estimated based on seed, fertilizer, and pesticide expenditures. The variable labeled as other direct costs sums up expenditures for herd renewal, veterinary services, and contract work. Land refers to forage area for dairy cows. Labor input includes hours from hired and own labor. Various data on farm characteristics were used in different parts of the analysis. As outlined in Table 1 this includes information on the area of the farm, economic size of the farm, stocking density, subsidy payments, farm assets, available forage area, degree of specialization, share of family labor used on the farm, rented land, and position in a less favored area. The *bacon* routine was used to identify outliers for the production variables shown in Table 1 (Weber, 2010). After removing seven outliers, the resulting subset for this study contained 41,903 dairy farm enterprise observations. Stata 15.1 was used for the entire analysis.

2.2 | Farm classification

When assessing economic farm performance, it is important to compare and benchmark producers that operate under similar circumstances. In the context of efficiency analysis for example, different production frontiers, reflecting different production technologies, may apply to different sets of farms. Efficiencies of various producers need to be estimated with respect to the appropriate technology (Kumbhakar et al., 2015). Farming conditions and technological choices vary within and between countries. For context-specific data analysis, this heterogeneity requires some kind of classification or grouping. One simple option to address the issue is to estimate production frontiers and calculate economic performance indicators for distinct countries or regions. However, in production economics such a type of classification is generally considered arbitrary (Orea & Kumbhakar, 2004; Mekonnen et al., 2015). Producers within the same region or country may operate under different production circumstances, whereas producers in different countries may be more similar and share a production frontier. For instance extensive grazing systems for dairy production can be found in Poland as well as in France or Austria. Latent differences are thus considered more important than a strict geographic categorization when estimating production frontiers.

We employed a latent class model to allocate dairy farms to groups that are characterized by a higher degree of homogeneity. Subsequently, gross margin and efficiency were analyzed for certified and non-certified dairy enterprises. A similar approach had for example been used to examine efficiency of dairy farming in Spain (Alvarez & Corral, 2010) or to study the effects of innovation systems on eco-efficiency in agriculture (Mekonnen et al., 2015; Grovermann et al., 2019). To account for technological choice and key farm and agro-climatic characteristics, the class splitting model was specified using the following variables: Costs per dairy cow (feed, forage, machinery, other direct costs), labor per dairy cow, farm area, stocking density, forage, and fodder areas as well as average yearly minimum and maximum temperatures, rainfall, the number of hot days, and the number of dry days. Farms that possess similar attributes in the above variables are more likely to be in the same class[1]. Optimal class size was determined by applying the Schwarz Bayesian Information Criterion (SBIC) and the rule of no small classes (<than 5% of total observations). This rule has long been used in practice as a part of the idea of obtaining more useful results, but has

[1] For a more detailed explanation of the method, including the equations required for the assignment of observations to classes, see Mekonnen et al. (2015) or Grovermann et al. (2019).

TABLE 1 Class-specific average production and farm characteristics (standard deviations in brackets)

	Class 1		***Class 2***		***Class 3***		***Class 4***	
Description	*Dairy farming under cool conditions*		*Dairy farming under temperate conditions More intensive Less intensive*				*Dairy farming under warm conditions*	
Organic certified	***NO***	***YES***	***NO***	***YES***	***NO***	***YES***	***NO***	***YES***
Farms (#)	4815	1066	20162	1029	9161	439	5116	115
Production data								
Production (kg/cow)[1]	6869.97	6282.22	7247.26	6285.68	5247.69	4601.19	6667.26	6039.61
	(2064.61)	(1644.07)	(1656.14)	(1544.46)	(1552.40)	(1360.86)	(2105.79)	(2221.70)
Land (ha/cow)	2.13	2.70	1.09	1.31	1.01	1.59	0.65	1.11
	(1.95)	(2.37)	(0.66)	(0.69)	(0.72)	(0.91)	(0.70)	(1.30)
Labor (days/cow)	0.06	0.06	0.02	0.03	0.07	0.09	0.04	0.04
	(0.04)	(0.04)	(0.01)	(0.02)	(0.04)	(0.05)	(0.03)	(0.02)
Feed costs (€/cow)	1104.73	951.83	770.33	790.32	447.08	307.69	1199.07	1187.01
	(579.57)	(674.48)	(383.54)	(577.77)	(214.11)	(205.62)	(560.71)	(727.02)
Forage costs (€/cow)[2]	88.82	37.28	159.56	58.41	95.61	26.37	85.10	62.40
	(94.34)	(52.66)	(86.14)	(54.07)	(62.49)	(34.88)	(85.57)	(79.70)
Mach. costs (€/cow)	202.86	237.87	168.61	204.55	83.37	131.73	72.53	62.59
	(163.14)	(194.22)	(101.47)	(120.95)	(66.85)	(114.63)	(75.68)	(73.05)
Other costs (€/cow)	323.84	351.91	363.39	391.71	110.37	131.24	183.17	147.14
	(253.88)	(283.21)	(181.08)	(203.12)	(82.86)	(103.18)	(152.69)	(138.14)
Min. temperature (C)	2.30	2.45	6.37	5.95	4.61	4.54	9.43	10.40
	(1.93)	(1.79)	(1.33)	(1.13)	(0.68)	(0.99)	(2.50)	(2.00)
Max. temperature (C)	9.88	10.47	13.86	13.44	12.59	12.92	19.10	19.70
	(2.22)	(2.23)	(1.54)	(1.37)	(1.03)	(1.29)	(1.91)	(1.72)
Farm characteristics data								
Farm size (ha)	163.82	99.77	166.16	144.77	46.74	41.07	51.95	53.39
	(369.82)	(151.09)	(359.01)	(244.58)	(149.01)	(59.79)	(186.30)	(50.98)
Econ. Size (ESU)	196.54	126.08	374.90	323.73	65.07	57.54	205.00	179.72
	(372.01)	(160.49)	(604.28)	(381.42)	(133.67)	(68.32)	(324.86)	(158.36)
Stocking d. (cows/ha)	1.17	0.90	2.13	1.44	2.12	1.23	8.30	3.03
	(1.50)	(0.44)	(1.72)	(0.45)	(1.82)	(0.52)	(55.34)	(2.61)
Subsidies (€/ha)	318.94	254.07	346.52	302.88	205.84	252.90	1119.49	432.06
	(362.61)	(165.52)	(153.42)	(95.21)	(174.10)	(103.06)	(7,227.97)	(468.69)
Assets (€)	1782.69	2162.21	1288.69	1700.11	1542.28	1885.41	808.71	679.03
	(1460.61)	(1562.04)	(926.35)	(1149.22)	(1,169.87)	(1602.62)	(877.94)	(741.99)
Forage area (ha)	116.68	85.16	103.43	112.46	27.46	31.34	35.50	48.09
	(235.26)	(121.89)	(175.18)	(171.01)	(83.71)	(43.85)	(78.35)	(48.84)
Specialization (prop.)	0.65	0.61	0.68	0.69	0.62	0.59	0.72	0.67
	(0.14)	(0.14)	(0.16)	(0.14)	(0.16)	(0.14)	(0.15)	(0.15)
Family labor (prop.)	0.81	0.89	0.86	0.82	0.96	0.95	0.90	0.84
	(0.31)	(0.22)	(0.26)	(0.25)	(0.14)	(0.17)	(0.21)	(0.26)
Renting land (Y/N)	0.88	0.85	0.94	0.95	0.72	0.71	0.75	0.64
	(0.32)	(0.36)	(0.24)	(0.22)	(0.45)	(0.45)	(0.44)	(0.48)
Less favored (Y/N)	163.82	99.77	166.16	144.77	46.74	41.07	51.95	53.39
	(369.82)	(151.09)	(359.01)	(244.58)	(149.01)	(59.79)	(186.30)	(50.98)
Fallow (Y/N)	0.03	0.05	0.08	0.11	0.03	0.08	0.05	0.08
	(0.19)	(0.21)	(0.29)	(0.33)	(0.25)	(0.27)	(0.22)	(0.43)

[1]Milk sale revenues do not include subsidies.

[2]Forage costs were estimated based on seed, fertilizer, and pesticide expenditures.

recently been discovered to also have some theoretical justifications based on the posterior distribution of the class proportions (Nasserinejad et al., 2017).

2.3 | Performance measurement

The first performance measurement variable, gross margins, was calculated using per cow revenue (without subsidies) and per cow variable cost figures. Cost items included in the calculation were labor, feed, forage, machinery maintenance, and other costs, as specified in the Data section. The gross margin was then derived by simply subtracting variable costs from sales revenues.

The efficiency scores, as second performance measurement, were estimated for individual farms by applying the stochastic production frontier framework. This is a standard approach for efficiency analysis of agricultural production systems with applications ranging from animal to crop production, from high-income to medium- and low-income economies and from farm-level to country-level analysis (for some recent examples see Houssain et al., 2012; Mekonnen et al., 2015 or Finger et al., 2018). A meta-study by Bravo-Ureta et al. (2007) comprised 167 farm level efficiency estimations of which the majority are based on stochastic frontier models, with nonparametric deterministic and parametric deterministic frontier models being other key estimation techniques found in the literature.

In this study we computed observation-specific efficiency scores using an output-oriented measure. This means that a farm is inefficient if a higher level of output is attainable for the given input use. Key specification decisions relate to the choice of the functional form of the production frontier and the distribution of the inefficiency term in the model. Stochastic frontier analysis distinguishes between a term that captures statistical noise and a term that accounts for inefficiency (Alvarez & Arias, 2014). As explained by Kumbhakar et al. (2015) assuming a half-normal distribution for the inefficiency term implies that the majority of the producers are operating at rather efficient levels. For the highly competitive dairy sector, this is considered an appropriate specification. For the estimation of the frontier model a Cobb-Douglas functional form was selected, following a constant returns to scale assumption. We also tested a translog specification as robustness check. Apart from some slight divergence, mostly in class four, both specifications resulted in similar efficiency estimates and regression results (see Tables A4 and A7 in the appendix for details). Therefore, output is considered to increase in proportion with an increase in all production factors. Estimations were not only separated by class, but also performed separately for conventional farms and for organic farms, with all parts then merged again into one dataset for further analysis. To start with, the eight subclasses were created and then in a second step the empirical stochastic frontier model was estimated, based on the following parameterization:

$$\ln(\mathrm{OUT_i})|_{c,o}=\beta_0+\beta_i \ln(\mathrm{X_i})\,|_{c,s}+\beta \mathrm{T}+\mathrm{v_i}|_{c,s}-\mathrm{u_i}|_{c,s} \quad (1)$$

where the vertical bars with index c and index o mean that different models were separately estimated for each class c and for organic and conventional systems s. The dependent variable OUT_i measures milk output in kg per dairy cow for observation i in class c and system s. X_i constitutes an observation-, class and system-specific vector of output-enhancing inputs, these being per cow land requirements, labor, feed costs, forage costs, machinery costs, and other costs. T are dummies for the years 2012 and 2013 (2011 being the reference year). The systematic error component v_i is assumed to be an independently and identically distributed random error term with a normal distribution. The inefficiency term u_i is measured as the ratio of observed output to the corresponding class-specific stochastic frontier and follows in this application, as explained above, a half-normal distribution.

2.4 | Impact analysis

The focus of our interest is the impact of certification on the gross margin and efficiency scores of dairy farm enterprises in each of the four classes. To obtain a valid measure of impact from the certification intervention some pre-processing of the data was needed to avoid the comparison being confounded with other factors (White & Raitzer, 2017). For a valid comparison, producers in the participating and non-participating groups should not significantly differ in characteristics that are not related to certification, but rather possess similar production conditions (such as agro-climatic circumstances) or farm traits (such as size or specialization). Due to non-random assignment, a particular challenge when comparing certified to non-certified producers is self-selection bias, that is, the fact that those producers that choose to participate in a standard often significantly differ in a number of characteristics from those producers that do not participate (Meemken & Qaim, 2018; Ssebunya et al., 2019). Several techniques to control for such a bias are available, of which propensity score matching, reweighting, or instrumental variable (IV) approaches, are among the most widely used (see reviews by Lopez-Avila et al, 2017 or Knook et al., 2018). While the latter can control for selection on properties that are often unobserved, such as entrepreneurship or risk behavior for example, matching and reweighting methods rely on observable information, that is, a dataset that captures all important variables that directly or indirectly determine selection. In the present analysis endogeneity due to unobserved properties can occur in some or all instances of the class-outcome combinations. Therefore, IV was selected as primary estimation strategy, backed up by a reweighting approach.

2.5 | Endogenous treatment model

To exploit the advantage of full maximum likelihood estimation for IV, we follow Cerulli (2015) and employ an endogenous treatment model, using the *etregress* routine in Stata. The model tests selection on unobservable characteristics and, where appropriate, corrects for such a bias (Fischer & Qaim, 2012; Tambo & Wünscher, 2014). Contrary to standard IV regression, the endogenous treatment model can estimate an average treatment effect, which is of higher policy relevance than a local average treatment effect (White & Raitzer, 2017). The core concept of the selected model is that the variation between predicted probabilities of treatment and actual treatment can be captured by adding terms in the outcome regression, which absorb the effects of unobservable determinants of treatment (Nichols, 2007; White & Raitzer, 2017). First a binary participation variable is regressed on observable characteristics using a probit model. From this, two ancillary terms, the selection hazard rate and the inverse mills ratio, are predicted, which are then inserted in a linear regression determining the effect of participation on the outcome of interest. In that way, unbiased treatment effects can be estimated. Results of the endogenous treatment model are based on the premise that at least one valid instrument has been identified, that is, an independent variable included in the probit model that influences participation, but not the outcome of interest.

Based on observed differences in the data, we postulate that dairy farm enterprises with fallow land are more likely to adopt organic certification, but that this is not associated with their performance. As information on fallow land was available in the dataset, it was decided to use this variable as an instrument in the estimation of the endogenous treatment model for classes one to three. Following Fischer and Qaim (2012) and Tambo and Wünscher (2014), the exogeneity of the instrument was tested by including it as an additional regressor in a "placebo" regression model with explanatory variables X, using only data from non-certified farm enterprises. Fallows were not significantly associated with either gross margins or efficiency (see Tables A8 and A9 in the appendix for details).

The outcome equation of the endogenous treatment model was defined for each class c (as indicated by the vertical bars with index c) as follows:

$$\begin{aligned}\mathrm{PERF_i}|_c = \beta_0+\beta_1\,(\mathrm{FARM1_i})\,|_c+\beta_2\mathrm{ORG_i}\lambda_{1_i}|_c \\ +\,\beta_3\,(1-\mathrm{ORG_i})\,\lambda_{0_i}|_c+\beta_4\mathrm{L}+\beta_5\mathrm{T}+\mathrm{v_i}|_c \end{aligned} \quad (2)$$

where $PERF_i$ represents the two performance outcome variables, per cow gross margins and efficiency for each observation i. A range of farm characteristics are captured by the vector $FARM1_i$: Farm size, economic size, stocking density, subsidy payments, farm assets, forage area, degree of specialization, share of family labor, rented land, position in a less favored area. Details on each variable are displayed in Table 1. Similar control variables have been used in other impact evaluation studies (Mayen et al., 2010; Läpple et al., 2013; Tambo & Wünscher, 2014). However, data on certain characteristics of the farm manager, such as age or education, was unavailable in our case. However, these properties are assumed to be implicit in the other variables or are captured by the added term that accounts for unobservables. The variable ORG is equal to one if a dairy enterprise possesses organic certification and zero otherwise. λ_{1i} (inverse mills ratio) and λ_{0i} (hazard rate) are parameters produced by the joint estimation of Equations (2) and (3) in order to absorb the selection bias. Besides controlling for time lapse, heterogeneity between countries, for example, due to differences in regulatory or support schemes, was accounted for in the model. Country and time fixed effects were included through location dummies L and T. Lastly, v_i is the random error term.

To remedy the potential endogeneity problem, the selected model estimates jointly with Equation (2) an equation for the certification decision using the variable representing fallow land as instrument. The participation equation takes the following form:

$$\mathrm{ORG_i}|_c=\delta_0+\delta_1\ln\,(\mathrm{FARM2_i})\,|_c+\delta_2\mathrm{FAL_i}|_c+\mathrm{v_i}|_c \quad (3)$$

where the binary dependent variable ORG_i captures the organic certification decision. The vector $FARM2_i$ contains the same variables as the vector *FARM1i* in the outcome equation, apart from stocking density and actual subsidies, which are considered to be somewhat influenced by organic farming rather than the other way around. The covariate $\mathrm{FAL_i}$ stands for the instrumental variable.

The Wald test for independent equations is used to support interpretation of the output of the endogenous treatment model. Where the test statistic is insignificant, indications are that there is no need to control for unobservable characteristics, thus matching or reweighting methods are sufficient. For this case, for estimating effects in class four and for better understanding the results overall, we therefore complemented the analysis with a balancing through entropy weights, a data pre-processing technique proposed by Hainmüller (2012) and implemented in Stata through the *ebalance* routine. A very similar approach was used by Meemken and Qaim (2018) to analyze the impact of food standards. Due to exact adjustment of covariate moments, it is considered an appealing alternative to standard matching or reweighting methods when estimating causal effects from observational studies (Zhao & Percival, 2016). Using entropy balancing, covariate balance for mean and variance moments could be directly incorporated in the estimation. Observation-specific weights were thus generated using the covariates in the vector $FARM2_i$. Including these weights in further analysis, differences correlated with selection and existing prior to or independent of treatment can be controlled or eliminated. Certification effects were estimated using a weighted regression for the gross margin and technical efficiency outcomes, based on the same outcome equation as specified for the endogenous treatment model (Equation 2). We thus used a doubly robust approach as described by Hainmueller (2012). Following the approach developed by Oster (2019), we also conducted robustness checks on the model choice by calculating the relative degree of selection. This method is increasingly applied to understand the role of unobservable factors that affect the outcomes of interest (e.g., Gunes & Tsaneva, 2020; Wuepper et al., 2021). Lastly, all estimations were adjusted for clustered standard errors to account for regional clusters at the level of nuts2 regions. This specifies that observations are independent across clusters, but not necessarily within clusters. Farmers in one region are likely to implement more similar approaches and often share similar values.

3 | RESULTS

3.1 | Classification

The overall class splitting approach resulted in four distinct groupings, as shown in Table 1. This was considered suitable, with the smallest class including 12% of observations, whereas a five-class model would have produced a distinct minor class with less than 4% of observations. Class one represents dairy farming under cool conditions (including e.g., Scandinavian and Baltic farms), class two characterizes more intensive dairy farming under temperate conditions (including for example many Dutch, German or Irish Farms), while class three characterizes more extensive dairy farming under temperate conditions (including for example many Polish or Austrian farms), and class four relates to dairy farming under warm conditions (including for example many Italian or Spanish farms). While there is no strict geographic split, there are geographical tendencies, for example, 71% of Italian farms are found in class four and almost all Dutch farms were assigned to class two. However, farms that exhibit distinct features, such as mountain dairy enterprises in Italy for example, are not allocated to class four, but rather to classes one or three (for more information on the distributions of farms by country and class see Table A1 in the Appendix). To illustrate differences among farms that are organic certified or not, data in Table 1 are further subdivided according to certification status.

3.2 | Efficiency scores

Efficiency scores take on values between zero and one. Our results show that across all four classes and across certification status efficiency levels are high, peaking among the dairy farm enterprises in class two.

From the first to the fourth class, average scores for non-certified dairy enterprises are .86, .87, .86, and .85 respectively. Those for certified enterprises lie at .87, .87, .86, .90 for each of the four classes. Distribution of efficiency scores are presented in Figure 1, revealing no major differences due to certification status, apart from class four. Also distributions across classes are not highly distinct. Detailed regression outputs of the class-specific stochastic frontier estimations are provided in Tables A2 and A3 in the appendix. All coefficients had the expected positive sign across the four classes, apart from those for land in class four, which is however not significant in this class.

3.3 | Certification impacts

The effects of organic certification vary among classes in terms of their magnitude, but are positive across all four classes, as regards gross margins. In classes two and four the Wald test suggests that an IV estimation approach is justified. For classes one and three, entropy balancing results apply. Certification impacts on variable profit are highest among the predominantly Mediterranean farms in class four, with the estimated effect amounting to 182%. Impacts are 50% in class three, regrouping mostly extensive central European farms. Gross margin differences are less pronounced among the farm enterprises in cool northern European and intensive central European environments, being 43% in class one and 38% in class two (see Table 2). The results from the respective bias correcting techniques differ somewhat in terms of magnitude, but both show a similar trend for classes one to three. Contrary to the other classes, the findings for class four diverge substantially. This could be due to an important role of unobservable characteristics, which we further analyzed (Oster, 2019).

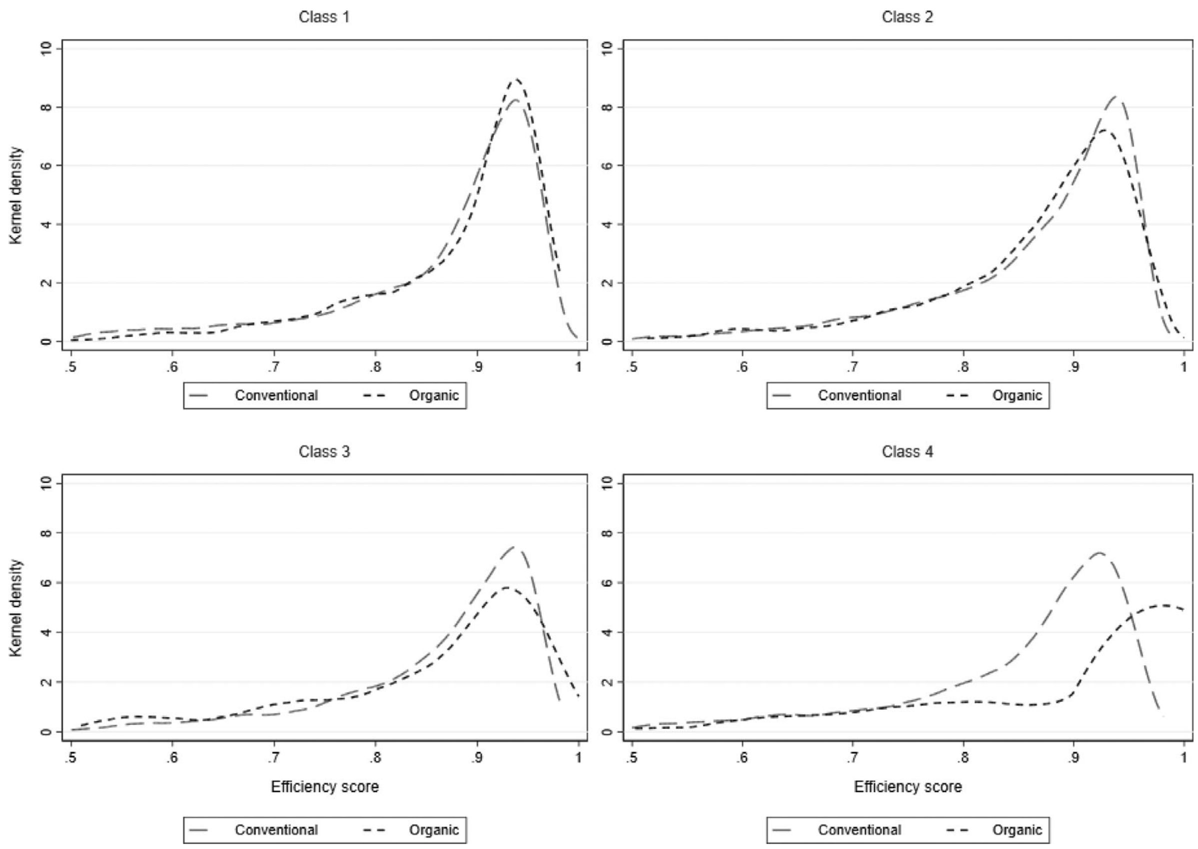

FIGURE 1 Distribution of efficiency scores for conventional and organic dairy enterprises in classes 1 to 4

TABLE 2 Certification impacts for gross margin and efficiency outcomes across four classes

	Class 1		*Class 2*		*Class 3*		*Class 4*	
	GM(€/cow)	*EFF(0-1)*	*GM(€/cow)*	*EFF(0-1)*	*GM(€/cow)*	*EFF(0-1)*	*GM(€/cow)*	*EFF(0-1)*
(1) Endogenous treatment model								
ATT	101	0.037	***158***	0.018	119	0.001	***376***	0.075
Sig.	Ns	***	***	Ns	***	***	***	
Comparison	−220	0.851	***413***	0.863	290	0.859	***207***	0.837
% change	46%	4%	***38%***	2%	41%	0%	***182%***	9%
Wald test	Ns	Ns	***	**	Ns	ns	***	ns
(2) Entropy balancing model								
ATT	***104***	***0.014***	102	−0.007	***66***	−0.004	1	***0.062***
Sig.	**	***	***	Ns	**	ns	Ns	***
Comparison	***−242***	***0.835***	399	0.879	***132***	0.829	211	***0.865***
% change	***43%***	***2%***	26%	−1%	***50%***	−1%	0%	***7%***
Delta (Rmax = 1.3[*R])	***−17.5***	***−8.7***	1.2	−2.8	***−4.4***	−1.2	0.1	−4.6
N	5881		21,191		9600		5231	

[*]*Notes*: Overall significant estimations underlined and in bold.

ATT = Average Treatment Effect on the Treated; Comparison = Mean reference value in the comparison group; GM = Gross Margin; EFF = Efficiency; Wald test = Wald test of independent equations to test for selection on unobservable characteristics; Delta = Coefficient of proportionality that describes how large the effect of unobservables needs to be in proportion to the effect of observables for the treatment effect to be equal to 0, given a maximum value of the R-squared[***] = 1% significance level; [**] = 5% significance level; [*] = 10% significance level.

[^]Full regression outputs for the estimations of the endogenous treatment model are shown in tables A5 and A6 in the appendix.

TABLE 3 Simple comparison (without any pre-processing)^ of certified and non-certified farms

Organic certified	*Class 1* NO	*Class 1* YES	*Class 2* NO	*Class 2* YES	*Class 3* NO	*Class 3* YES	*Class 4* NO	*Class 4* YES
Gross margin (€/cow)								
Mean	−211.61	−157.77	419.53	371.64	301.25	152.84	205.75	316.60
Standard dev.	(855.05)	(998.72)	(584.80)	(651.51)	(557.97)	(650.27)	(742.41)	(650.77)
Sign. diff.	**		***		***		*	
Efficiency								
Mean	0.85	0.88	0.86	0.86	0.86	0.85	0.84	0.89
Standard dev.	(0.14)	(0.10)	(0.10)	(0.10)	(0.10)	(0.13)	(0.12)	(0.17)
Sign. diff.	***		ns		***		***	

***1% significance level.

**5% significance level.

*10% significance level.

^The simple comparison shown here does not involve the careful construction of a counterfactual and the balancing of properties among certified and uncertified dairy enterprises.

The estimates of the proportional degree of selection reveal that selection on unobservables in the gross margin models is an issue particularly in class four, and to some extent in class two. As Table 2 exhibits, the delta values in these two cases are positive and either below unity or close to it (i.e., the bias from unobservables is higher or similar to the bias from observables).The gross margin coefficients of proportionality (*delta*) for classes one and three are negative. Negative values suggest that including additional controls in the model increases the effect of organic certification and that unobservables are negatively correlated with the controls. The threat of upward bias comes mainly from positive selection only. Therefore, negative values also confirm robustness of the results (Gunes & Tsaneva, 2020).

Significant efficiency effects were found only in classes one and four. These amount to 2% and 7% respectively. These relatively small estimated effects are predominantly due to the fact that the majority of all farms are operating at rather high efficiency levels. Coefficients of proportionality are all negative in the efficiency estimations.

According to our results, the Wald test of independent equations proved highly significant in three out of eight estimations. Therefore, the endogenous treatment model is a reasonable choice in these three instances (for full model estimations, see Tables A5 and A6 in the appendix). For the remaining cases, where selection on unobservable characteristics appears to be no issue, results from the entropy balancing approach can be considered more appropriate.

The per cow gross margin calculations indicate that several dairy enterprises are making a loss, as becomes also evident from the estimates in class one. In this context it is important to clarify that subsidies as well as fixed costs have not been taken into account in the computation of the gross margin variable, as we intended to focus on a variable profit measure that reflects the farm economics without government support. While in the counterfactual analysis dairy farm enterprises with organic certification appear slightly more efficient and considerably more lucrative, in a simple comparison, as shown in Table 3, certified farms perform worse in terms of gross margin for classes one to three. This indicates that an impact evaluation approach produces results that differ substantially from a simple comparative approach.

4 | DISCUSSION AND CONCLUSIONS

Using a latent class counterfactual approach we were able to systematically quantify the variable profit and efficiency impact of the organic standard for dairy farms across Europe. The study shows that certification pays off for those dairy farmers that converted to organic farming. This supports findings from previous more location-specific studies about the positive economic effects of organic certification (Bolwig et al., 2009; Hoop et al., 2017; Tran & Goto, 2019). In addition, findings indicate that not only were higher profits achieved on average, but small average efficiency gains also resulted from organic certification in two instances. The latter outcome is contrary to findings from previous studies that compared efficiency among organic and conventional dairy farms (Oude Lansink et al., 2002; Kumbhakar et al., 2009; Nehring et al., 2009; Mayen et al., 2010). Mayen et al. (2010) point out that efficiency estimates for certified farms are lower if a common frontier is assumed for organic and conventional farms. In the context of the present analysis, we estimated separate functions for organic and conventional dairy enterprises. Foremost however, we apply a class splitting model. The latent differences among farms are considered more central for addressing heterogeneity among farms than a using a priori defined geographic areas. The resulting groups are entirely based on the statistical characteristics of the data with the aim to identify most distinct groups within the data, while farms within a group transform economic inputs in a similar way—as far as reflected in the data—into economic output. The opposite to this data-driven approach is to operate with predefined groups based on administrative units or agronomically motivated cut-offs, such as production regions (e.g., plain, hills, mountains). The advantage of such groups is that they are mutually exclusive along all single dimensions and make sense from an agronomic classification point of view—but may not reflect the "reality" regarding production technologies as captured in the input-output relations provided by the data analyzed. The data-driven grouping on the other hand reflect economic input-output relations but may not be easily interpreted from an agronomic point of view. Furthermore, they are not clear-cut and show overlaps along all dimensions.

Many organic dairy enterprises, just like conventional ones, can define and be very close to the specific frontiers, as our analysis revealed. The findings indicate that, given the respective production environments, they can achieve slightly greater technical efficiency than conventional enterprises in the same class, but not across all classes. In this regard it should be noted that technical efficiencies refer to the "own" frontiers of organic and conventional farms, which means that we can compare the efficiency scores, but cannot compare relative productivity of specific inputs between organic and conventional production. Another consideration is that our analysis focuses on efficiency per cow. While organic farms achieve greater economic performance per cow, due to the lower stocking rates within organic systems, the results per hectare of land may differ. Overall, the fact that high levels of efficiency can be found across all classes, suggests that the scope for efficiency improvements in Europe under current conditions is rather restricted.

As in similar impact studies (Läpple et al., 2013; Tambo & Wünscher, 2014), the endogenous treatment model proved valuable in addressing selection bias by taking into account both observable and unobservable characteristics. The appropriate statistical test suggested that the model was appropriate in three out of eight estimations. For the remaining estimation and for checking the general direction of the effects, estimates from entropy balancing are provided. The overall approach, testing for selection on unobservables and otherwise using entropy balancing as data pre-processing method, follows the logic outlined by Meemken and Qaim (2018). The results of the endogenous treatment model are based on the premise that at least one valid instrument can be identified. Using fallow land as instrumental variable relies on the assumption that it is associated with deciding on organic certification, but not with the performance of dairy production. This is considered a valid assumption and was tested. Besides IV and reweighting approaches, further analysis of the impacts of organic certification could seek to exploit certification differences across time applying a regression discontinuity approach (Cattaneo, Idrobo, & Titiunik, 2020). Following Wuepper et al. (2020) geographic boarders–here delineating clear thresholds in terms of organic farming adoption–could provide a case for employing a fuzzy spatial regression discontinuity design. Such spatial features can be the result of distinct regional or national policies for promoting organic farming.

Overall, the results point out that organic dairy production appears to be an economically sensible strategy for certified dairy farmers in Europe. It is however important to note that organic dairy farming has been found to be more risky than conventional production (e.g., in a study for the Netherlands by Berentsen et al., 2012), with organic farmers appearing to be generally less risk averse than conventional farmers (Iver et al., 2020). While certification might thus not be a viable option for several of the existing conventional farms, the results point out that non-certified farms with characteristics that are similar to those of the certified farms may benefit from conversion. The findings also show that higher dairy enterprise gross margins are partially associated with higher efficiency.

An important factor to take into consideration when deriving from the present analysis any broader recommendation on transitioning to organic dairy farming is the consumer side. There have been signs that despite substantial recent growth, the potential of future organic dairy sales growth might be limited due to market saturation. This implies a need for policy interventions to target the demand side, for example, through attempting to include some external costs in the price of conventional milk. Following our results and economic reasoning, supply of sufficient organic dairy products seems less problematic, as financial incentives for conversion exist. Following the findings of Ramankutty et al. (2019), an additional policy measure lies in the greening of conventional dairy production including fairer pricing and a blend of best practices from different systems. This might be a good alternative also for those farmers who do not consider organic certification as an option for their dairy enterprise. Seufert et al. (2017) point out that environmental principles are inadequately represented by organic regulations. The concept of eco-efficiency might offer a promising indicator for studies that integrate the assessment of economic and environmental performance aspects. For future impact research on dairy farming, it would be of interest to extend the analysis beyond technical efficiency and gross margin outcomes to include aspects of fair price, animal welfare, and ecosystem services. Flexibility is another important potential indicator, that is, the capability to adapt production output without major shifts in the average cost structure. Flexibility can contribute to maintaining profitability during economic shocks. Recent studies have shown this as well as a tradeoff between flexibility and technical efficiency (Hirsch et al., 2020; Renner et al., 2014). More research on this issue in organic farming is required.

ACKNOWLEDGMENT

This research received funding from GenTORE, a Horizon 2020 project running from 1 June 2017 to 31 May 2022, as part of the European Union's H2020 Research and Innovation Program under agreement No. 727213.

DECLARATION OF INTEREST

The authors have no competing interests to declare.

REFERENCES

Alvarez, A., & Corral, J. (2010). Identifying different technologies using a latent class model. Intensive vs. extensive dairy farms. *European Review of Agricultural Economics*, *37*, 231–250.

Alvarez, A., & Arias, C. (2014). A selection of relevant issues in applied stochastic frontier analysis. *Economics and Business Letters*, *3*, 3–11.

Berentsen, P. B., Kovacs, K., & Van Asseldonk, M. A. (2012). Comparing risk in conventional and organic dairy farming in the Netherlands: An empirical analysis. *Journal of dairy science*, *95*, 3803–3811.

Bolwig, S., Gibbon, P., & Jones, S. (2009). The economics of smallholder organic contract farming in tropical Africa. *World Development*, *37*, 1094–1104.

Bouttes, M., Bize, N., Maréchal, G., Michel, G., San Cristobal, M., & Martin, G. (2019). Conversion to organic farming decreases the vulnerability of dairy farms. *Agronomy for Sustainable Development*, *39*, 19

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

Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge, UK: Cambridge University Press.

Cerulli, G. (2015). *Econometric evaluation of socio-economic programs: theory and applications*. Berlin; Germany: Springer.

Crowder, D. W., & Reganold, J. P. (2015). Financial competitiveness of organic agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 7611–7616.

EPRS. (2018). *The EU dairy sector: Main features, challenges and prospects. Service Briefing PE 630.345*. Brussels, Belgium: European Parliament Research (EPRS).

EUROSTAT. (2018). HICP- inflation rate 2018, Retrieved from https://ec.europa.eu/eurostat/databrowser/view/tec00118/default/table?lang=en

EUROSTAT. (2019). Price indices of the means of agricultural production - annual data, Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apri_pi05_ina&lang=en

FADN. (2018). *European Dairy Farms Report 2018*. Brussels, Belgium: DG-AGRI European Commission, Farm Accountancy Data Network (FADN).

FAO. (2018). *10 Elements of Agroecology*. Rome, Italy: Food and Agriculture Organization of the United Nations.

Finger, R., Dalhaus, T., Allendorf, J., & Hirsch, S. (2018). Determinants of downside risk exposure of dairy farms. *European Review of Agricultural Economics*, *45*, 641–674.

Fischer, E., & Qaim, M. (2012). Linking smallholders to markets: Determinants and impacts of farmer collective action in Kenya. *World Development*, *40*, 1255–1268.

Grovermann, C., Wossen, T., Muller, A., & Nichterlein, K. (2019). Eco-efficiency and agricultural innovation systems in developing countries: Evidence from macro-level analysis. *Plos One*, *14*, e0214115.

Gunes, P., & Tsaneva, M. (2020). The effects of teenage childbearing on education, physical health, and mental distress: Evidence from Mexico. *Journal of Demographic Economics*, *86*, 183–206.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*, 25–46.

Hirsch, S., Mishra, A., Möhring, N., & Finger, R. (2020). Revisiting firm flexibility and efficiency: Evidence from the EU dairy processing industry. *European Review of Agricultural Economics*, *47*, 971–1008.

HLPE. (2019). *Agroecological and other innovative approaches for sustainable agriculture and food systems that enhance food security and nutrition*. Rome, Italy: High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security (HLPE).

Hoop, D., Dux, D., Jan, P., Renner, S., & Schmid, D. (2017). *Rapport 2016, Echantillon situation du revenu: Dépouillement centralisé des données comptables*, Taenikon, Switzerland: Agroscope.

Hossain, M. K., Kamil, A., Baten, M. A., & Mustafa, A. (2012). Stochastic frontier approach and data envelopment analysis to total factor productivity and efficiency measurement of Bangladeshi rice. *Plos One*, *7*, e46081.

Ibanez, M., & Blackman, A. (2016). Is eco-certification a win–win for developing country agriculture? Organic coffee certification in Colombia. *World Development*, *82*, 14–27.

Iyer, P., Bozzola, M., Hirsch, S., Meraner, M., & Finger, R. (2020). Measuring farmer risk preferences in Europe: A systematic review. *Journal of Agricultural Economics*, *71*, 3–26.

Knook, J., Eory, V., Brander, M., & Moran, D. (2018). Evaluation of farmer participatory extension programmes. *The Journal of Agricultural Education and Extension*, *24*, 309–325.

Kumbhakar, S. C., Tsionas, E. G., & Sipilainen, T. J. (2009). Joint estimation of technology choice and technical efficiency: An application to organic and conventional dairy farming. *Journal of Productivity Analysis*, *31*, 151–161.

Kumbhakar, S. C., Wang, H., & Alan, P. (2015). *A Practitioner's guide to stochastic frontier analysis using stata*. Cambridge, UK: Cambridge University Press.

Läpple, D., Hennessy, T., & Newman, C. (2013). Quantifying the economic return to participatory extension programmes in Ireland: An endogenous switching regression analysis. *Journal of Agricultural Economics*, *64*, 467–482.

Lopez-Avila, D., Husain, S., Bhatia, R., Nath, M., & Vinaygyam, R. (2017). *Agricultural innovation: An evidence gap map. 3ie Evidence*

Gap Map Report 12. New Delhi, India: International Initiative for Impact Evaluation (3ie).

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.

Meemken, E., & Qaim, M. (2018). Can private food standards promote gender equality in the small farm sector? *Journal of Rural Studies*, *58*, 39–51.

Mekonnen, D. K., Spielman, D. J., Fonsah, E. G., & Dorfman, J. H. (2015). Innovation systems and technical efficiency in developing-country agriculture. *Agricultural Economics*, *46*, 689–702.

Mendoza, T. C. (2004). Evaluating the benefits of organic farming in rice agroecosystems in the Philippines. *Journal of Sustainable Agriculture*, *24*, 93–115.

Moakes, S., Lampkin, N., & Gerrard, C. (2016). *Organic farm incomes in England and Wales, 2014/15*. Newbury, UK: Organic Research Centre.

Mockshell, J., & Villarino, E. J. (2019). Agroecological intensification: Potential and limitations to achieving food security and sustainability. In (P. Ferranti, E. M. Berry, & J.R. Anderson Eds.), *Encyclopedia of Food Security and Sustainability* (pp. 64–70). Amsterdam, Netherlands: Elsevier.

Nasserinejad, K., van Rosmalen, J., de Kort, W., & Lesaffre, E. (2017). Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. *PlosOne*, *12*, e0168838.

Nehring, R., Gillespie, J., Sandretto, C., & Hallahan, C. (2009). Small U.S. dairy farms: Can they compete?. *Agricultural Economics*, *40*, 817–825.

Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, *7*, 507–541.

Orea, L., & Kumbhakar, S. C. (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, *29*, 169–183.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*, 187–204.

Oude Lansink, A., Pietola, K., & Backman, S. (2002). Efficiency and productivity of conventional and organic farms in Finland 1994–1997. *European Review of Agricultural Economics*, *29*, 51–65.

Ramankutty, N., Ricciardi, V., Mehrabi, Z., & Seufert, V. (2019). Trade-offs in the performance of alternative farming systems. *Agricultural Economics*, *50*, 97–105.

Renner, S., Glauben, T., & Hockmann, H. (2014). Measurement and decomposition of flexibility of multi-output firms. *European Review of Agricultural Economics*, *41*, 745–773.

Schleifer, P., & Sun, Y. (2020). Reviewing the impact of sustainability certification on food security in developing countries. *Global Food Security*, *24*, 100337.

Schreinemachers, P., Schad, I., Tipraqsa, P., Williams, P., Neef, A., Riwthong, S., … Grovermann, C. (2012). Can public GAP standards reduce agricultural pesticide use? The case of fruit and vegetable farming in northern Thailand. *Agriculture and Human Values*, *29*, 519–529.

Serra, T., Zilberman, D., & Gil, J. M. (2008). Differential uncertainties and risk attitudes between conventional and organic producers: The case of Spanish arable crop farmers. *Agricultural Economics*, *39*, 219–229.

Seufert, V., Ramankutty, N., & Mayerhofer, T. (2017). What is this thing called organic? – How organic farming is codified in regulations. *Food Policy*, *68*, 10–20.

Ssebunya, B., Morawetz, U., Schader, C., Stolze, M., & Schmid, E. (2019). Group membership and certification effects on incomes of coffee farmers in Uganda. *European Review of Agricultural Economics*, *46*, 109–132.

Tambo, J. A., & Wünscher, T. (2014). Building Farmers' Capacity For Innovation Generation: What Are The Determining Factors? Selected Paper at the International Congress of the European Association of Agricultural Economists, August 26–29, 2014, Ljubljana, Slovenia.

Tran, D., & Goto, D. (2019). Impacts of sustainability certification on farm income: Evidence from small-scale specialty green tea farmers in Vietnam. *Food Policy*, *83*, 70–82.

van der Ploeg, J. D., Barjolle, D., Bruil, J., Brunori, G., Costa Madureira, L. M., Dessein, J., … Wezel, A. (2019). The economic potential of agroecology: Empirical evidence from Europe. *Journal of Rural Studies*, *71*, 46–61.

Weber, S. (2010). bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *The Stata Journal*, *10*, 331–338.

Willer, H., & Lernoud, J. (2019). *The World of Organic Agriculture. Statistics and emerging trends*. Frick, Switzerland: Research Institute of Organic Agriculture (FiBL); Bonn, Germany: IFOAM – Organics International.

White, H., & Raitzer, D. (2017). *Impact evaluation of development interventions - A practical guide*. Manila, Philippines, Asian Development Bank (ADB).

Wuepper, D., Wimmer, S., & Sauer, J. (2020).,Is small family farming more environmentally sustainable? Evidence from a spatial regression discontinuity design in Germany. *Land Use Policy*, *90*, 104360.

Wuepper, D., Roleff, N., & Finger, R. (2021). Does it matter who advises farmers? Pest management choices with public and private extension. *Food Policy*, *99*, 101995.

Zhao, Q., & Percival, D. (2016). Entropy Balancing is Doubly Robust. *Journal of Causal Inference*, *5*, 101515.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Grovermann C, Quiédeville S, Muller A, Leiber F, Stolze M, & Moakes S. Does organic certification make economic sense for dairy farmers in Europe? - A latent class counterfactual analysis. *Agricultural Economics*. 2021;*52*:1001–1012.
https://doi.org/10.1111/agec.12662