Author(s): Amer Ait Sidhoum, Philipp Mennig & Johannes Sauer

Title: Do agri-environment measures help improve environmental and economic efficiency? Evidence from Bavarian dairy farmers

Year: 2023

Version: Published version

Please cite the original version:

Amer Ait Sidhoum, Philipp Mennig, Johannes Sauer, Do agri-environment measures help improve environmental and economic efficiency? Evidence from Bavarian dairy farmers, European Review of Agricultural Economics, 2023;, jbad007, https://doi.org/10.1093/erae/jbad007

European Review of Agricultural Economics Vol **00 (00)** (2023) pp. 1–36
doi:https://doi.org/10.1093/erae/jbad007

Do agri-environment measures help improve environmental and economic efficiency? Evidence from Bavarian dairy farmers

Amer Ait Sidhoum[*,†], **Philipp Mennig**[‡] **and Johannes Sauer**[‡]

[†]*Natural Resources Institute Finland (Luke), Business Economics, Finland;* [‡]*Department of Agricultural Production and Resource Economics, Technische Universität München, Germany*

Received July 2021; final version accepted February 2023

Abstract

This study presents an innovative empirical application to the assessment of agri-environment measures on environmental and economic efficiency. Using a multi-equation representation with desirable technology and its accompanying undesirable by-production technology, we investigate the effects of agri-environment measures on farm-level environmental and economic efficiency. A combination of propensity score matching and a difference-in-difference approach is used to estimate the policy effect. The application focuses on a balanced sample of Bavarian dairy farms surveyed between 2013 and 2018. Results suggest that agri-environment schemes do not alter farms' economic efficiency, whereas environmental efficiency does not seem to be stimulated by schemes participation.

Keywords: agri-environment schemes, policy evaluation, nitrogen pollution, environmental performance, data envelopment analysis

JEL classification: C23, Q12, Q18, Q57

1. Introduction

Nitrogen (N) pollution from agriculture is recognised as one of the most pressing environmental problems humanity faces. The N surplus in the environment is mainly a result of the intensive use of mineral fertilisers and livestock production (Sutton *et al.*, 2011), which has surpassed the planet's boundaries

*Corresponding author: E-mail: amer.ait-sidhoum@luke.fi
Review coordinated by: Nauges, Céline

(Rockström *et al.*, 2009; Steffen *et al.*, 2015). The negative environmental impacts related to this excess as a result of the human-induced disruption of the biochemical nitrogen cycle have only recently begun to overlay the critical role that nitrogen fertiliser has played in boosting agricultural production. Its inefficient application has resulted in losses to the environment in various ways. Via leaching and runoff, nitrogen, mainly in the form of nitrate (NO_3^-), gets into groundwater and surface waters, where it threatens human health when entering drinking water (Davidson *et al.*, 2011) and causes unwanted eutrophication (Howarth *et al.*, 2000). In gaseous forms (ammonia—NH_3, nitric oxide—NO, nitrous oxide—N_2O and N_2), it is emitted to the atmosphere where these types cause considerable harm. NO plays a significant role in tropospheric ozone pollution (Hickman *et al.*, 2017), and NH_3 and NO are precursors to particulate matter air pollution and facilitate the deposition of nitrogen into ecosystems (Galloway *et al.*, 2003). Furthermore, high concentrations of ammonia can be harmful for vegetation in case of direct air contact (Krupa, 2003). N_2O, on the other hand, is a potent greenhouse gas that additionally depletes stratospheric ozone (IPCC, 2014; Kanter *et al.*, 2017). Finally, high nitrogen loadings threaten biodiversity and ecosystem functioning (Rabalais, Turner and Wiseman, 2002; Vitousek *et al.*, 1997).

In order to reduce N pollution, numerous measures have been implemented by countries, associations of states and private and non-profit players all around the globe. The European Union (EU), for example, aims to tackle the problem with a number of environmental directives: the Nitrates Directive, the Habitats Directive, the Water Framework Directive and the National Emissions Ceilings Directive. These directives force member states to act and are accompanied by regional-, national- and EU-level initiatives, one of which are agri-environment schemes (AESs), or agri-environment-climate measures (AECM) as they are lately being referred to, as part of the Common Agricultural Policy (CAP).

AES in the EU agricultural policy can be traced back to the 1985 Agricultural Structures Regulation. They were conceptualised as tools that compensate farmers for income losses and costs associated with environmentally friendly farming practices. After an amendment in 1987, which governed funding sources and cost flows for environmentally sensitive areas, AES became compulsory for all EU member states in 1992 (EU Regulation 2078/92). Since then, they are an important component of the second pillar of the CAP. In the 2014–2020 EU rural development programming period, European, national and regional Pillar II spending on environment protection and resource efficiency (the themes to which AES are mainly associated) amounted to 38 billion Euros, roughly 25 per cent of the total spending on rural development (European Commission, 2021). In the course of the years, AES spending increased and the schemes themselves evolved. Today, top-down action-based schemes still dominate; however, more and more spatially targeted (results-based) AESs created using a multi-actor approach come into play. They are a response to the patchy success of the classical schemes in many environmental categories, including biodiversity (Bisang, Lienhard and Bergamini, 2021; Budka *et al.*, 2019), water quality (Jones *et al.*, 2017; Kay *et al.*, 2012; Slabe-Erker *et al.*, 2017) and greenhouse gas emissions (Coderoni and Esposti, 2018; Stetter, Menning and Sauer, 2020), where diffuse pollution from nitrogen use causes direct harm. The present work attempts to further investigate the impact of AES on nitrogen pollution. In contrast to previous studies, we assess the programme impact in terms of changes in the (nitrogen) environmental and economic efficiency of farms.

Assessing the impact of AES on economic efficiency is critical not only for managers who are driven by profit-maximising objectives but also for other stakeholders and decision-makers such as policymakers and regulators (e.g. environmental agencies). For example, regardless of whether the producer decides to join an environmental programme or not, he or she will be concerned with improving productivity not only in physical terms but also in terms of the economic benefits associated with an optimal allocation of actual quantities of inputs and outputs. For regulators, minimising the level of pollution associated with the production of desirable products is also critical.

In our case, the production process is modelled in the presence of desirable and bad outputs. The desirable output is the value of the dairy products, while the bad output is related to nitrogen pollution as measured by the farm-level nitrogen balance. Few alternatives exist in the technical literature on the proper modelling of undesirable outputs in production processes (see Dakpo, Jeanneaux and Latruffe, 2016). The most recent type of by-production approach relies on the concept that a production technology is a result of an interaction of two separate sub-technologies: one governs the production of good outputs and a second governs the production of bad outputs. Introduced by Murty, Robert Russell and Levkoff (2012) and extended by Serra, Chambers and Oude Lansink (2014) and Førsund (2021, 2018), this approach is one of the most promising methods for modelling pollution-generating technologies. In this study, we rely on the novel by-production approach to estimate our combined economic and environmental[1] efficiency indices.

From a practical point of view, the nitrogen pollution is supposed to be influenced by AES that requires the adoption of specific management practices or environmentally beneficial actions, e.g. a reduction of nitrogen fertiliser use. Clearly, specific agri-environment measures primarily aim at a reduction of absolute nitrogen pollution levels with desirable output losses being offset

1 Although the environmental dimension covers large and complex fields, the environmental impacts of nitrogen pollution are well known. High levels of fertiliser application lead to a large nitrogen surplus and to correspondingly large flows of nitrogen into the soil and into the air. Part of the nitrogen is taken up by crops, but a large portion of these nutrients is emitted to the environment. Nitrogen pollution leads to nitrate contamination of the groundwater aquifers, the most important source of drinking water. Nitrogen also evaporates as ammonia and causes acid rain. Phosphate pollution causes eutrophication of surface water, which endangers plant and fish life. For these reasons, 'nitrogen pollution' is used in this paper as a proxy of the environmental dimension.

by the AES payment. Improving farm efficiency is thus no direct goal of the schemes. Furthermore, a reduced nitrogen balance as a result of being enrolled in an AES might be linked to output losses that keep the output-pressure ratio unaltered. However, studying the effect of AES enrolment on farm efficiency is still important for two reasons. First, authors such as van Grinsven *et al.* (2019), Grovermann *et al.* (2019) and Beltrán-Esteve and Picazo-Tadeo (2017) claim that innovations have a positive impact on farmers' efficiency in minimisation of environmental outcomes given economic outcomes. We hypothesise that AES participation is likely to stimulate eco-innovation, which in turn has a positive impact on economic and environmental efficiency. Second, economic and environmental issues need to be looked at jointly in the light of sustainable agricultural production. If scheme enrolment leads to less environmental pollution in absolute terms and additionally improves economic efficiency, further strengthening the role of second pillar AECM and newly introduced first pillar eco-schemes would be reasonable. As Kumbhakar and Malikov (2018) stated, both economic and environmental efficiencies are crucial from a policy standpoint since they are usually taken as the basis for designing new or assessing existing policy instruments. Furthermore, policymakers may prefer adopting policies aimed at efficiency enhancement rather than implementing restrictive policies (Kuosmanen and Kortelainen, 2005). Economic and environmental efficiency improvements are therefore arguably the most cost-effective way to reduce environmental pollution while maintaining or improving the competitiveness of the agricultural sector.

This paper extends the literature by providing an in-depth analysis of the impact of AES on nitrogen pollution. Specifically, we do not purely focus on the environmental outcome but link it to classical efficiency analysis, thus obtaining measures of environmental and economic efficiency that allow us to assess whether AES can be effective in pursuing economic and ecological goals simultaneously. Our second contribution relates to the combination of the impact evaluation methods, namely, propensity score matching (PSM) and difference-in-difference (DiD) with a multi-equation representation using data envelopment analysis (DEA).

Just as in the rest of Europe, the agricultural sector has undergone rapid structural change in Bavaria in the past decades. While the number of farms amounted to around 440,000 in 1960, it decreased to 88,610 in 2017. This development was mainly driven by differences in working conditions and incomes between economic sectors, technical progress and efficiency gains. Especially the use of mineral fertilisers, pesticides and innovative machinery resulted in increases in yields and lower producer prices, which in turn forced farmers to enlarge their businesses. High fertiliser and pesticide applications, however, created negative externalities in the form of environmental pressure. Persistently, high loads of nitrogen fertilisers in particular negatively affected water quality in Bavaria. Only 15 per cent of all running waters and only 26 out of 50 surface water bodies that were evaluated fulfil the criteria set in the European Water Framework Directive (LfU, 2021a). Furthermore, at 6.4 per cent of all groundwater quality measuring points, the threshold for nitrate of 50 mg/l is exceeded, and at 30 per cent, values higher than 25 mg/l are reported, indicating human-induced pollution (LfU, 2021b). In both cases, runoff and leaching from the application of nitrogen fertilisers play a key role. This pollution needs to be decreased further in order to reach the target of a 'good state' of water bodies by 2027 as stipulated in the Water Framework Directive.

Measures to improve the state of water bodies and groundwater quality, of which AESs are part, need to be effective, while at the same time, they should neither affect a farm's economic performance nor development possibilities. The latter is of particular importance as Bavarian agriculture, which compared to the German average is characterised by rather small-scale family farms and partly unfavourable natural conditions, competes with national and international players. In order to guarantee food supply on the regional level with limited natural resources and a low(er) environmental footprint, the sector needs to follow a path of sustainable intensification, i.e. farms need to adopt production measures that improve either economic or environmental outcomes without compromising each other (in the Bavarian context, more emphasis might be put on the environmental component). From a farm perspective and on the field level, scheme limiting nitrogen fertiliser input, for example, might be such a measure if the payment offsets or is higher than income foregone and costs incurred. The existence of windfall effects is likely to affect actual changes in pollution levels, though.

Given the theoretical promises of such practices with the potential to balance the trade-off between economic and environmental efficiency, surprisingly, little empirical studies evaluating them exist. Few to be found mainly use field trial data (e.g. Townsend, Ramsden and Wilson, 2016) or simulations (e.g. Devkota *et al*., 2016) and put an emphasis on yields. Holistic farm-level assessments, particularly in Central Europe, however, seem underrepresented. Contrary to farmers in the global South, European farmers rather aim at enhancing their ecological output without sacrificing the economic outcome when adopting sustainably intense practices (Charles, Godfray and Garnett, 2014). The undesirable environmental harm as a by-product of agricultural production is thus at the centre of attention. It still interacts with farm-specific technical and social factors, which is why it is important to consider the multidimensionality of outcomes (Ait Sidhoum, Serra and Latruffe, 2020). Among few studies that use a respective approach for measuring the effect of specific farming practices on economic and environmental performance at farm level are the studies by Gadanakis *et al*. (2015) and Pérez Urdiales, Lansink and Wall (2016). Both studies use a radial eco-efficiency measure based on non-parametric DEA. Gadanakis *et al*. (2015) find that arable farms in the UK can improve eco-efficiency by adopting sustainable farming practices, while Pérez Urdiales, Lansink and Wall (2016) identify socio-economic characteristics and attitudes that explain eco-efficiency of Spanish dairy farms. In both cases, however, the approaches used lack explicit causal interpretation. This shortcoming is addressed by Weltin and Hüttel (2019) who investigate the environmental improvement potential of sustainable intensification in the northern German Plain using a directional meta-frontier approach and matching. Their results show that adopters of sustainably intense practices have higher mean eco-efficiency scores compared to non-adopters but still do not fully exploit the potential of ecological improvements.

2. Methodology

Our analytical framework consists of three steps. In the first step, we use PSM to control for potential selection bias arising from observable characteristics. In step 2, we use DEA to estimate farm-level technical and environmental efficiency indices. Third, a DiD specification within a single bootstrap procedure is applied to derive the average treatment effect of AES on efficiency scores. In the following text, we will explain our framework step by step.

2.1. PSM

There are several strong theoretical reasons why AESs might improve farm performance, but how can we be sure that the better farm performance of AES-participating farms compared to non-participating farms is caused by AES adoption (or not)? In most instances, experimental data would have been of great value to obtain information on the counterfactual state that would address the question of causal inference. However, this is not our case, as for each farmer only one situation can be observed, we have a missing data problem (Rubin, 1976). Therefore, to do this, we have to prevent selection bias while seeking to separate the AES impact from other determinants of farm performance.[2] To address the selection bias problem, we use PSM techniques.

The first step of this procedure consists of estimating the propensity scores that are used to match AES-participating with non-participating farms for the pre-treatment period 2013. Because agri-environment measures uptake is a binary choice, we estimate a logit model for the pre-treatment period as follows:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta X_i + \varepsilon_i \quad (1)$$

where $i = 1, \ldots, I$ indexes the number of farms and $\pi_i = P(T_i = 1|X_i)$ indicates the farmers' probability[3] of treatment assignment, conditional on a vector of observed explanatory variables X_i. Successful implementation of the PSM technique requires a set of relevant covariates that could potentially affect AES participation and/or farm performance (technical and environmental efficiency) but otherwise are not affected by the treatment (scheme participation). More specifically, X includes a set of variables representing farm and regional characteristics (see Table A1).

2 Farms' performance is affected by both observed factors (e.g. land size, experience, farmers' education, capital, labour and chemical use) and unobserved factors (e.g. managerial ability, motivation and environmental awareness).

3 Instead of matching on a set of observable covariates, X_i showed that matching on the propensity score (π_i) is sufficient to identify the treatment effect.

Once the propensity scores have been estimated for each of the i farms in the sample, a matching strategy is used to construct counterfactual outcomes for the participating farms. Then, the average treatment effect on the treated (ATT) can be estimated as follows:

$$\text{ATT} = E(Y_i(1)|T_i = 1, \pi(X_i)) - E(Y_i(0)|T_i = 0, \pi(X_i)) \quad (2)$$

where $Y_i(1)$ is the outcome variable (e.g. farm efficiency) when the farm i participates in AESs and $Y_i(0)$ is the outcome variable when the farm i does not participate. T_i is a dummy variable that takes the value of 1 for the participating farms and 0 otherwise, and $\pi(X_i)$ is the propensity score of farm i conditional on observed covariates X_i. The ATT compares the differences in outcomes between the participating farms (first term of the model (2)) and matched non-participating farms (second term of the model (2)). The validity of PSM further relies on two assumptions: the *conditional independence assumption* and the *common support condition* (Khandker, Koolwal and Sammad, 2009). The *conditional independence* assumption states that given a set of observable covariates X_i, the potential outcomes are independent of treatment assignment, which is specified as follows:

$$Y_i(1), Y_i(0) \perp T_i | X_i \quad (3)$$

The common support assumption: $0 < P(T_i = 1|X_i) < 1$ ensures that for each participating farm, there are non-participating farms with similar observed characteristics.

Once the matching has been completed and comparable participant and non-participant farms have been identified, the next step is to derive technical and environmental efficiency measures.

2.2. Efficiency measurement

The traditional firm-specific technical efficiency analysis framework has been extended to allow for externalities. Within the extended framework, both traditional netputs and environmental impacts characterise the production technology. However, there is a debate in the production economics literature on the appropriate method to use when dealing with environmental impacts in production analysis. Two main categories can be distinguished. First, one approach represented by a group of studies (e.g. Chambers, Chung and Färe, 1996; Färe *et al*., 2005) relies on the additive directional distance function that aims at increasing the good outputs, while reducing pollution at the same time. This single-equation approach has been criticised for not being sufficiently suitable to simultaneously capture the trade-off between the desirable output and the undesirable output and thus leading to an overstatement of the overall efficiency. Second, a multi-equation framework that has its foundations by Frisch (1965) that has been followed up in some recent studies (Murty, Robert Russell and Levkoff, 2012) relies on a radial measure of a production technology that combines two independent sub-technologies reflecting a desirable output and undesirable by-production sub-technology. Despite the limiting assumption of two independent sub-technologies,[4] the multi-equation representation remains a very promising approach to tackle the challenges associated with by-production technologies.

Similarly to Murty, Robert Russell and Levkoff (2012), we model firm performance as a composition of two sub-technologies: a desirable production technology and a pollution-generation technology. In their model, they separate inputs into non-polluting and polluting inputs in addition to the output distinction. In the context of our empirical application, the non-polluting inputs (x_n) comprise capital, labour, land and so on, while the polluting input (p) consists of nitrogen input, which generates runoff pollution as by-products, such as nitrogen pollution. In general terms, the 'overall' technology, denoted by T, is represented as the intersection of two sub-technology sets, T^1 and T^2. The overall technology with its sub-components can be represented as follows:

$$T^1 = \{(x_n, p, y, z) : (x_n, p) \text{ can produce } y\} \tag{4}$$

$$T^2 = \{(x_n, p, y, z) : (p) \text{ can produce } z\} \tag{5}$$

where

$$T = T^1 \cap^{T^2} \tag{6}$$

The desirable sub-technology T^1 satisfies classic regularity assumptions such as the free disposability in desirable outputs and inputs, whereas under the undesirable sub-technology T^2, the bad outputs (nitrogen pollution) and the polluting inputs are assumed to be costly disposable. This property implies that both polluting inputs and by-products cannot be disposed without additional cost. In other words and according to Murty, Robert Russell and Levkoff (2012)'s perspective, if a given level of nitrogen generates some minimal level of runoff, then inefficiency in the use of fertilisers may imply that this level of fertilisers application could also generate a higher amount of runoff.

Computing farm efficiency measures that evaluate the farms' performance in each of its sub-technologies requires the estimation of a production frontier. We estimate the production frontier non-parametrically using DEA. First, we construct the frontier of the desirable sub-technology T^1 that only takes into consideration the interaction between conventional inputs (x_n) and desirable output (y) while ignoring the presence of undesirable output. Assuming that the observed inputs and outputs combinations are available for I farms with the superscript $i = 1, \ldots, I$ denoting the individual farms, then the DEA

4 See Dakpo, Jeanneaux and Latruffe (2016) for a good overview.

representation of the desirable sub-technology T^1 is then expressed as:

$$T^1 = \left\{ (x_n, p, y, z) : p \geq \sum_i \lambda^i p^i, x_n \geq \sum_i \lambda^i x_n^i, y \leq \sum_i \lambda^i y^i, \sum_i \lambda^i = 1, \lambda^i \in \mathbb{R}_+^N \right\} \quad (7)$$

While DEA does not require any a priori assumption on the functional form of the production function, some assumptions about how these inputs and outputs interact are needed. Our assumption for the desirable sub-technology T^1follows the direction in which the inequality constraints are expressed in model (7), that is to say the desirable outputs (y) are produced using the polluting (p) and non-polluting (x_n) inputs. Thus, according to this representation, additional units of inputs to an existing set of inputs will positively contribute to the production of the good outputs. The term λ^i represents intensity vector, containing the firms' weights that are used to define the best-practice frontier representing the best-practice technology. The assumption that T^1 maintains variable returns to scale (VRS) is captured by the constraint $\sum_i \lambda^i = 1$, $\lambda^i \in \mathbb{R}_+^N$,

whereas the undesirable sub-technology T^2 under VRS can be given by:

$$T^2 = \left\{ (x_n, p, y, z) : p \leq \sum_i \mu^i p^i, z \geq \sum_i \mu^i z^i, \sum_i \mu^i = 1, \mu^i \in \mathbb{R}_+^N \right\} \quad (8)$$

where μ^i is a vector of intensity variables representing the weight of each farm that are used to construct the efficient frontier for T^2. The undesirable sub-technology refers to the performance of farms in controlling the generation of by-products. The second process (T^2) responds to the assumption that increased use of nitrogen application (p) promotes increased runoff by-products and treats by-products (z) as weakly disposable.

Following Murty, Robert Russell and Levkoff (2012), the global performance that recognises the presence of desirable outputs and by-products can be captured by the following index:

$$\begin{aligned} Eff_{Global}(x_n, p, y, z : T) = {} & \tfrac{1}{2} \min_{\theta_1} \{ \theta_1 | (x_n, p, y \oslash \theta_1, z) \in T^1 \} \\ & + \tfrac{1}{2} \min_{\theta_2} \{ \theta_2 | (x_n, p, y, z \otimes \theta_2) \in T^2 \} \end{aligned} \quad (9)$$

where $y \oslash \theta_1 = y/\theta_1$ and $z \otimes \theta_2 = z\theta_2$ and θ_1 is the efficiency score for desirable output sub-technology, while θ_2 is the efficiency score of the undesirable sub-technology. Basically, the index in model (9) is simply the average[5] efficiency scores of the two sub-technologies T^1 and T^2.

5 See Chambers and Serra (2018) who discussed the possibility to use different weights of the different sub-technologies.

2.3. DiD estimation

In the third step of our analysis, the effect of AES participation on farm efficiency is estimated. While PSM allows controlling for possible selection bias issues arising from observed characteristics, it has been shown however that the unobserved factors could also play a significant role in farmers' decisions to participate in agri-environment programmes (Hynes and Garvey, 2009), and therefore, by failing to control for unobserved heterogeneity, the estimated effect may be biased. One way of dealing with the unobserved heterogeneity and being able to control for AES determinants such as environmental awareness or managerial ability is to combine PSM with DiD method (Heckman, Ichimura and Todd, 1997; Smith and Todd, 2005). More specifically, we employ the DiD[6] approach to estimate the difference in change of farm performance indicators (e.g. efficiency scores) over two periods between the participants and the matched non-participants. For two time periods, the DiD estimator measures the average programme impact as follows (Heckman, Ichimura and Todd, 1997):

$$\mathrm{ATT}^{DiD} = \frac{1}{N}\sum_{i=1}^{I}\left\{\left(Y_{i,t_2}(1) - Y_{i,t_1}(1)\right) - \sum_{j=1}^{J} w_{i,j}\left(Y_{j,t_2}(0) - Y_{j,t_1}(0)\right)\right\} \tag{10}$$

where $t_1$represents the pre-treatment period and t_2 represents the post-treatment period. i and j are, respectively, the treated and the matched non-treated individuals, and N is the number of individuals in the treatment group. $w_{i,j}$ indicates the weights $\left(0 \le w_{i,j} \le 1\right)$ given to the j^{th} non-treated individual matched to i^{th} treated individual.

To derive the ATT, we apply second-stage regression analysis to provide empirical evidence of changes in the efficiency of AES-participating farms relative to changes in the efficiency of non-participating farms. Simar and Wilson (2007) demonstrated that traditional second-stage approaches (e.g. censored Tobit regression) to inference are unreliable due to serial correlation of DEA efficiency measures and show how a truncated regression with bootstrapping procedure provides a reliable technique for inference. Following Simar and Wilson (2007), we use the single bootstrap approach (Algorithm no. 1) which implementation would include the following steps:

1. The use of maximum likelihood (ML) to estimate the parameters of the following truncated regression:

$$\theta_{it} = \alpha + \beta T_i + \gamma t_i + (DD)\, T_i t_i + \delta_j Z_{ij} + \varepsilon_{it} \tag{11}$$

where θ_{it} is the efficiency score for the farm i at time t. T_i is a dummy variable representing scheme participation, and t_i is a time dummy taking the value of 0 for the pre-treatment period and 1 for the post-treatment period.

6 It should be noted that the availability of a panel data structure makes it possible to implement the difference–indifference procedure.

The DiD estimator is represented by the coefficient (Coeff.), DD, which gives the estimate for the impact of the AES on the farm-level efficiency scores. Finally, Z_i represents a matrix of regressors expected to affect farm technical and environmental efficiency, and ε_{it} is the unobserved time-varying error component.

The model (11) generates estimates $\left(\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \overline{DD}, \overline{\delta_j}\right)$ of $(\alpha, \beta, \gamma, DD, \delta_j)$ as well as an estimates of the standard deviation (S.d.) of the error component $(\overline{\sigma_\varepsilon})$.

2. Based on these estimates, a bootstrap procedure is applied over the following three steps 2,000 times:
   (a) For each farm, ε_{it} is drawn from a $N\left(0, \overline{\sigma_\varepsilon^2}\right)$ distribution with left truncation at 1.
   (b) Using the drawn error components ε_{it}, new efficiency scores, $\hat{\theta}_{it} = \bar{\alpha} + \bar{\beta}T_i + \bar{\gamma}t_i + \left(\overline{DD}\right)T_i t_i + \overline{\delta_j}Z_{ij} + \varepsilon_{it}$, are predicted.
   (c) The use of the ML method again to estimate the truncated regression with the predicted efficiency estimates $\hat{\theta}_{it}$ on $(T_i t_i, T_i t_i, Z_{ij})$ to obtain the estimates $\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \widehat{DD}, \widehat{\delta_j}\right)$ and $(\widehat{\sigma_\varepsilon})$
3. Finally, confidence intervals are obtained for $\left(\alpha, \beta, \gamma, DD, \delta_j\right)$ and (σ_ε) by using the bootstrap estimates $\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \widehat{DD}, \widehat{\delta_j}\right)$ and $(\widehat{\sigma_\varepsilon})$ and the original values $\left(\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \overline{DD}, \overline{\delta_j}\right)$ and $(\overline{\sigma_\varepsilon})$.

3. Data and empirical model

In this study, we use balanced panel data of 1,626 Bavarian dairy farms,[7] covering the period 2013–2018 drawn from Farm Accountancy Data Network in Bavaria. This database has been complemented with the official agricultural support data (InVeKoS) that provide further details on farm characteristics as well as on Pillar I and Pillar II payments received. We also used publicly available data from the Bavarian State Statistical Office containing information about socio-economic characteristics (e.g. gross domestic product, unemployment and labour force) at municipality level to improve our matching accuracy. The timeframe of concern is from 2014 to 2018 as the last rural development programme in the region of Bavaria was initiated in 2014, in accordance with the 7-year Multi-annual Financial Framework of the CAP. The Bavarian rural development programme offered a large number of AESs targeting various environmental categories. Most schemes have a commitment period of 5 years and scheme participation is voluntary, which means that farmers self-select into treatment based on utility considerations. For this reason, selection bias has to be accounted for. A widely used and well-accepted method to address selection bias is PSM (see Ali and Abdulai (2010); Mayen, Balagtas and Alexander (2010) or Pufahl and Weiss (2009) for examples of empirical studies

7 The milk revenue of our specialised dairy farms represents at least 70 per cent of their total revenue coming from animal and crop production. Because of their markedly different technology, organic farming systems were not taken into account.

in the field of agricultural economics). For the purpose of performing PSM, the year 2013 was considered to identify comparable treated farms and untreated farms based on observable variables before the scheme's implementation.

The choice of the period 2013–2018 and the pre-treatment year 2013 was motivated by two factors. First, the CAP's programming periods follow a 7-year framework, which means that for all 7 years, the first and second pillar measures are refined. The period before the 2014–2020 phase ended in 2013, so did the 2007–2013 AES, which differed to a certain extent from those of the period 2014–2020 and thus represent a 'new' treatment. The year 2013 can consequently be considered the pre-treatment year. The time range that was chosen for the DiD analysis including the pre-treatment year is linked to the 5-year commitment period of most AESs. Second, nitrate pollution of groundwater in Bavaria has not improved during the 2007–2013 programming period, despite 2007–2013 AES targeting water quality. This resulted in additional efforts in the 2014–2020 period.[8]

In our empirical analysis, we are interested in estimating the impact of environmental subsidies on farm-level combined technical and environmental efficiency over the period 2014–2018. With the year 2013 being the pre-treatment phase, the year 2018 is considered to be the post-treatment phase. We, therefore, focus on those farms that did not participate in any AESs in the initial time period to assess the impact of environmental subsidies. As a result, farms that did not receive environmental subsidies in 2013 but did receive them during the 2014–2018 period were assigned to the treated group, while farms that did not receive any subsidy during the whole period (2013–2018) were assigned to the control group.

Then, PSM was performed to balance farm characteristics between farms that participated/did not participate in AESs. A central challenge when implementing PSM is which variables to include within the PSM model. While the literature has thoroughly discussed the issues associated with which variables to include in the propensity score model, yet still there is no scientific consensus regarding the choice of variables in the PSM model (Austin, Grootendorst and Anderson, 2007).

Heckman *et al.* (1997, 1998) demonstrate that data quality is an important component of any accurate estimation technique. They showed that estimators are only consistent when they are applied on data from a comparison group that meets the following requirements: (i) the same data sources are used for participants and non-participants, allowing characteristics to be quantified in a similar way, (ii) participants and non-participants are from the same area and (iii) the data include a diverse set of indicators that influence programme participation as well as outcomes. These requirements are obviously met by our data.

8 Twenty out of a total of 34 individual schemes (almost 60 per cent) of the most important Bavarian agri-environment programme KULAP can be expected to have either a direct or an indirect effect on nitrogen reduction. Among these, 20 schemes are the most popular ones with large agricultural areas under programme (e.g. extensive use of grassland, catch crops, organic farming).

While there is no scientific consensus on how many variables to include in a propensity score binary model, we are aware of the fact that the inclusion of non-significant variables can increase the variance of the propensity score estimates; however, they do not bias the estimates nor make them inconsistent. Augurzky and Schmidt (2001) discuss the pros and cons of including a lot of variables that are not significant. They show that a smaller set of covariates can lead to a better estimation of the treatment effect. However, in two of their three testing sets, they explicitly include variables that do either not or only weakly influence the outcome and variables that are relevant for the outcome but irrelevant to the treatment decision. The smallest and best-performing set only includes covariates that influence both, which is a necessary condition for a successful matching procedure. In their conclusion, Augurzky and Schmidt (2001, p. 27) emphasise that 'the main criterion of success for matching remains the balance of the relevant covariates and not the proper estimation of the selection equation'. According to Rubin and Thomas (1996), even if the variables are not statistically significant, they should be included. Our variable selection is based on theoretical economic arguments and empirical evidence (Arata and Sckokai, 2016; Chabé-Ferret and Subervie, 2013; Defrancesco *et al.*, 2007; Matzdorf and Lorenz, 2010; Mennig and Sauer, 2020). Table A1 presents descriptive statistics for the variables included in the PSM estimation.

After having defined the treated and untreated farms and the potential relevant covariates for the matching procedure, the propensity score[9] is calculated using a logit regression as a measure of the probability that a farm will be classified as a programme participant. Logit model results for the PSM are presented in Table A2. The likelihood ratio test is statistically significant at 1 per cent level, indicating that all farm characteristics considered are jointly significant in explaining programme participation.

Propensity scores were calculated for each observation based on the parameter estimates of the logit model, which were then used to match participant and non-participant farms. Different matching algorithms[10] were tested prior to selecting the nearest neighbour estimator (1:1) without replacement. Before matching, significant differences are assumed to be found between the treated and control groups, and therefore, the resultant balance of the relevant covariates assesses the success of propensity score estimation. Table 1 shows

9 The propensity score represents the conditional probability of participation for farm given a set of $i\ X = xi$ observed characteristics $p\ (X) = \Pr\ (P = 1|X = x\ i)$. The propensity score is estimated from a logit model in which the binary treatment variable (AES) serves as the dependent variable conditional upon the observed variables (covariates).

10 We tested the most common matching algorithms: kernel matching, radius matching and nearest neighbour matching without and with replacement from 1 to 10 neighbours. We compared the different matching algorithms and found that 1:1 nearest neighbour matching without replacement using a calliper width of 0.3 performed best. Rosenbaum and Rubin (1985) propose the use of standardised bias (SB) to compare treated unit means and untreated unit means before and after matching as a measure of covariate balance. As noted by Caliendo and Kopeinig (2008), an SB below 5 after matching would be seen as sufficient. Our findings indicate that the overall SB was reduced from 38.5 to 3.2 per cent by the matching procedure. By comparison, the overall SB was reduced to 5.4 per cent, 3.6 per cent and 4.5 per cent with nearest neighbour matching with replacement, radius matching and kernel matching, respectively.

Table 1. Mean and median bias reduction of relevant covariates before and after matching the pre-treatment year 2013

	Before matching		After matching		SB	
Variables	Control mean	Treated mean	Control mean	Treated mean	Before matching	After matching
Livestock units per ha	1.050***	0.840	0.949	0.947	−59.8	−0.7
Labour per ha	0.036**	0.028	0.032	0.032	−60.1	6
Capital depreciation per ha	628.810**	572.210	592.270	602.590	−16.5	3
Sales per ha	3,913.70**	3,275.60	3,564.100	3,574.100	−51.7	0.8
Fertilisers per ha	178.080**	158.750	163.830	166.070	−24.6	2.9
Pesticide per ha	62.950	63.460	60.068	59.446	1.3	2.6
Feed per ha	606.980***	501.440	564.640	553.360	−32.4	−3.5
Share of arable land	0.571	0.589	0.571	0.573	9	1.1
Share of grassland	0.427	0.411	0.425	0.426	−8.2	0.5
Yield index per ha	77.833***	49.025	56.518	58.738	−51.4	4
GDP	28,240.000***	27,454.000	28,031.000	28,033.000	−16.6	0.1
Number of dairy farms	124	147	69	69		
Total number of farms	271		138			

** and *** indicate statistical significance at 1 per cent and 0.1 per cent, respectively, of a t-test on the equality of mean differences between observations from the treated and the control groups.
Note 1: When using a 1:1 nearest neighbour matching algorithm, observations that fall outside the region of common support as well as observations for which there is no matching partner for a treated individual that is closest in terms of the propensity score have to be discarded from the analysis. Consequently, the sample size is reduced after matching.
Note 2: PSM can eliminate the self-selection bias issues arising from observed characteristics. Although we have included as many covariates as possible for matching purposes, some unobservable factors may cause estimation biases. Therefore, we use the Rosenbaum sensitivity analysis to check whether there are other factors that affect the quality of matching. The analysis indicates that the results are quite robust to unobserved factors. These results are available from the authors upon request.

covariates' mean values before and after matching between the two groups. These results suggest that no significant differences between participating and non-participating farms remain after matching. We can therefore conclude that the applied matching algorithm worked well, as the existing observable differences have been controlled for.[11] The matched treatment and control groups are used to estimate the average effect of AES participation on the combined farm-specific technical and environmental efficiency using a DEA-DiD model.

In our empirical efficiency analysis, we rely on balanced panel data of 138 Bavarian dairy farms covering the period 2013–2018 (828 observations in total). Seven input variables were defined and used in the analysis: livestock units expressed in number of cows (x_1); total labour (x_2 measured in man-work units); utilised land (x_3 in hectares); capital depreciation (x_4 in Euros); crop-specific inputs, namely expenses for chemicals (pesticides and fertilisers) (x_5 in Euros); expenses for feed (x_6 in Euros) and quantities of nitrogen input (q in kilograms). The desirable output is total farm sales (y in Euros). Finally, the undesirable output is represented by the nitrogen balance[12] (z in kilograms), computed as the difference between the nitrogen contained in the inputs and the nitrogen contained in the outputs. Table 2 shows the average values of the main variables. On average, and for the period under consideration (2013–2018), firms in our sample generate total sales of around 223,000 Euros. On average, our sample farms manage 58 ha of land, have average livestock units of 58 cows, devote 1.75 man-work units of labour per year, have a capital depreciation value of 35,575 Euros and spend around 13,000 and 32,200 Euros on chemicals (fertilisers and pesticides) and feed, respectively.

The average nitrogen balance per farm in our sample is slightly less than 6,285 kg per year. The nitrogen balance shows substantial heterogeneity between farms, with a Coeff. of variation of 0.68. On a per hectare basis, average nitrogen balances declined from 109 kg/ha in 2014 to 83 kg/ha in 2018. Eulenstein *et al.* (2008) calculate the average nitrogen surplus to be on the order of 91 kg/ha, while OECD (2013) estimates the annual average nitrogen balance in the period 2007–2009 for the EU 15 agricultural sectors to be 65 kg/ha, with substantial variation from 204 kg/ha in the Netherlands to 25 kg/ha in Greece. OECD (2013) places this value around 85 kg/ha for Germany. The difference in the results reported in the literature usually depends on the method used and the type of farms under investigation.

11 The non-participating farms were chosen to be similar to the participating ones across several observable characteristics including, among others, labour, land, livestock density and capital depreciation. It is, therefore, reasonable to assume that all producers share the same technology.

12 In order to estimate the nitrogen balance, we follow Gamer and Bahrs (2010). We use Wendland *et al.* (2018)'s Coeff. to approximate the quantity of nitrogen contained in milk and meat outputs as well as the nitrogen content in feed input and the LFL (2013) Coeff. to approximate the quantities of nitrogen that have been fixed by legumes, while for mineral fertilisers, the quantities of nitrogen can be computed from the information provided in STATBA (2018).

Table 2. Summary statistics (average and S.d.—in parenthesis) for the main variables in the sample (828 farms)

Variable	Symbol	Dimension	2013	2014	2015	2016	2017	2018	Full period (2013–2018)
Total sales	y	Euros	210,665.51 (104,027.02)	237,141.27 (116,735.25)	219,212.20 (108,614.00)	201,702.05 (100,998.89)	214,421.12 (112,800.45)	257,445.99 (142,765.69)	223,431.36 (112,602.11)
Livestock units	x_1	Number	55.49 (26.39)	57.69 (28.25)	58.28 (29.06)	59.05 (30.02)	60.14 (32.06)	60.60 (33.57)	58.54 (29.50)
Labour	x_2	Man-work units	1.70 (0.54)	1.73 (0.57)	1.75 (0.58)	1.76 (0.55)	1.79 (0.58)	1.79 (0.58)	1.75 (0.54)
Land	x_3	hectares	57.26 (24.56)	57.88 (24.67)	58.03 (24.55)	59.18 (25.95)	59.68 (26.62)	60.31 (26.82)	58.72 (25.35)
Capital depreciation	x_4	Euros	35,737.05 (22,372.23)	36,383.51 (23,241.15)	36,261.95 (23,414.60)	34,497.50 (23,314.33)	34,543.50 (25,116.38)	36,028.37 (27,437.59)	35,575.31 (23,338.10)
Chemicals	x_5	Euros	13,764.52 (10,303.63)	14,334.55 (9,619.82)	13,857.73 (10,208.55)	13,406.20 (10,525.50)	11,546.54 (8,401.37)	11,112.76 (7,792.39)	13,003.72 (9,183.28)
Feed	x_6	Euros	31,994.36 (20,020.74)	33,225.19 (20,315.27)	31,012.28 (19,996.46)	32,042.08 (21,409.53)	31,710.19 (21,654.38)	33,238.65 (22,950.68)	32,203.79 (20,333.09)
Nitrogen input	q	kg	7,657.81 (4,795.30)	9,121.79 (5,221.79)	8,470.47 (5,271.83)	8,849.60 (5,844.01)	8,584.21 (5,713.23)	8,029.11 (5,051.09)	8,452.16 (5,155.12)
Nitrogen balance	z	kg	5,285.20 (3,956.37)	6,460.38 (4,148.09)	5,960.46 (4,353.46)	6,188.52 (4,829.82)	5,941.28 (4,681.52)	5,230.56 (3,907.13)	6,284.89 (4,288.59)

Note: Monetary variables were deflated to 2015 values using the appropriate price indices.

4. Results and discussion

4.1. Technical and environmental efficiencies

It is well known that non-parametric efficiency calculations are very sensitive to the presence of outliers. We have used the super-efficiency method proposed by Banker and Chang (2006) to detect and remove potential outliers in the dataset. Thus, observations with extreme super-efficiency ratings were eliminated from the sample. By using the tolerance level of 1.2, 18 farms were found to be dominating observations (outliers) and thus have been removed.[13] As a result, we have decided to trim 13 per cent (18 farms out of 138) of the observations and work with this trimmed sample going forward.[14]

The frequency distributions and the mean values of the overall efficiency and its technical and environmental components are summarised in Table 3. Non-parametric kernel density distributions of the overall efficiency and both technical and environmental efficiencies are presented in Figure 1. The environmental efficiency distribution (in blue) shows a highly right-skewed density distribution, with around 50 per cent of the sample having an efficiency score below 0.65. These results point to a significant difference between calculated mean values for technical efficiency (0.920) and nitrogen balance efficiency (0.649). This suggests that the sample dairy farms could on average decrease their nitrogen pollution levels by around 35 per cent, while maintaining their desirable output efficiency unaltered. These differences between technical and environmental performance are in line with earlier research that has shown that a high level of economic efficiency is associated with a lower level of environmental performance (Ait Sidhoum, Hervé Dakpo and Latruffe, 2022; Mayen, Balagtas and Alexander, 2010). On average, only one farm was found to be fully efficient in controlling nitrogen pollution over the analysed period, while the most inefficient farm had a nitrogen efficiency rating of 0.260, implying that it could maintain its level of desirable output while reducing nitrogen pollution by 74 per cent. For instance, using the summary statistics of the nitrogen balance (Table 2), the calculated nitrogen pollution efficiency ratings indicate that the representative farm could reduce its surplus of nitrogen by $0.35 \times 6284.89 = 2199.71$ Kg N.

Our findings indicate a relatively high degree of environmental inefficiency in Bavarian dairy farms. This could be due to the fact that fertiliser control regulations are not implemented effectively (Kirschke *et al.*, 2019). This finding is not specific to the region; in fact, various related studies in the literature reported low environmental efficiency scores, where some studies explicitly focused on dairy farms. Examples of such research include the study by

13 Note that the removed outliers have been identified from both participants and non-participants farms (nine observations in each group), and comparison between treated and untreated units has been repeated without these outliers, but results did not change significantly.

14 Despite the fact that such a sample size might appear to be small, it is in line with other studies in the efficiency and productivity measurement literatures (Alvarez and Del Corral, 2010; Fraser and Cordina, 1999; Pérez Urdiales, Lansink and Wall, 2016; Skevas and Serra, 2016).

Table 3. Frequency distribution of efficiency scores over the period 2013–2018

Year	Technical	Environmental	Overall
$0 < \lambda < 0.1$	0	0	0
$0.1 \leq \lambda < 0.2$	0	0	0
$0.2 \leq \lambda < 0.3$	0	1	0
$0.3 \leq \lambda < 0.4$	0	3	0
$0.4 \leq \lambda < 0.5$	0	14	0
$0.5 \leq \lambda < 0.6$	0	24	0
$0.6 \leq \lambda < 0.7$	0	39	24
$0.7 \leq \lambda < 0.8$	17	21	42
$0.8 \leq \lambda < 0.9$	21	13	41
$0.9 \leq \lambda < 1$	61	4	12
$\lambda = 1$	21	1	1
Average score	0.920	0.649	0.785

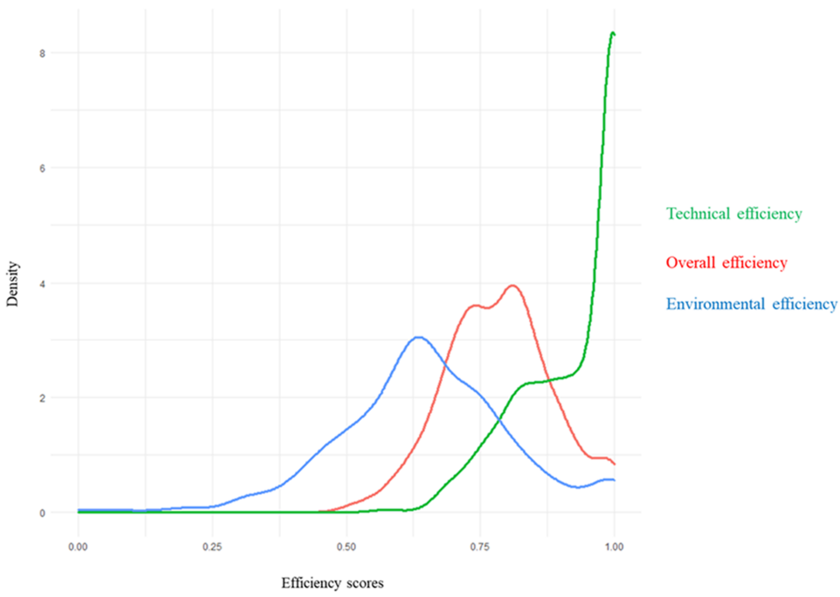

Fig. 1. Estimated kernel density distributions of the overall, environmental and technical efficiency scores (2013–2018).

Reinhard and Thijssen (2000), who reported nitrogen pollution efficiency levels of 0.56 for Dutch dairy farms, while Adenuga *et al*. (2019), who estimated environmental efficiency of a sample of Irish dairy farms, found relatively high-efficiency ratings of the order of 0.92. The use of different methods precludes a direct comparison of the efficiency results; however, Guesmi and Serra (2015) and Ait Sidhoum, Serra and Latruffe (2020) have used a comparable approach to estimate the nitrogen efficiency of a sample of Spanish arable

Table 4. Average efficiency scores over the period 2013–2018

Year	Technical	Environmental	Overall
2013	0.926 (0.084)	0.617 (0.159)	0.772 (0.096)
2014	0.917 (0.082)	0.723 (0.128)	0.820 (0.086)
2015	0.912 (0.104)	0.672 (0.133)	0.792 (0.095)
2016	0.925 (0.091)	0.633 (0.159)	0.779 (0.095)
2017	0.919 (0.098)	0.631 (0.161)	0.775 (0.104)
2018	0.921 (0.099)	0.619 (0.200)	0.770 (0.117)

Note: S.d. is given in parentheses.

crops and found similar patterns of nitrogen efficiency ratings, 0.69 and 0.57, respectively.

From Table 4, we observe that the technical efficiency scores show only a small difference over the period 2013–2018.[15] The absence of fluctuating technical efficiency levels in the milk sector can be explained by the abolition of the milk quotas in 2015, which prevents farmers from adjusting their production if the established quotas are exceeded and thus farmers are more likely to follow a well-structured and well-designed strategy that aims to expand milk production if enough land is available (Samson, Gardebroek and Jongeneel, 2016). The left-skewed distribution for technical efficiency (in green) shows evidence that many farms are performing very well in terms of desirable outputs. This implies that most of the farms are highly efficient, while only a few of them have an efficiency score below 0.800.

As mentioned, our results point to a relatively high level of technical efficiency in Bavarian dairy farms, but the fact that no prior studies have been carried out to estimate the technical efficiency of Bavarian dairy farms using the approach we use precludes a direct comparison of our findings with existing results reported in the literature. However, an extensive empirical literature exists on estimating dairy farm technical efficiency using DEA. A short and non-exhaustive list includes the studies by Fraser and Cordina (1999); Hansson and Öhlmér (2008); Jaforullah and Whiteman (1999); Latruffe, Fogarasi and Desjeux (2012); Reinhard, Knox Lovell and Thijssen (2000); Shortall and Barnes (2013) and Stokes, Tozer and Hyde (2007). For Bavarian dairy farms, Wimmer and Sauer (2020) applied a stochastic frontier analysis using an input distance function to estimate technical efficiency for a sample of dairy farms (2000–2014) and found an average technical efficiency of around 0.750, whereas in a study of 1,530 dairy farmers covering the period 2007–2011, Mennig and Sauer (2020) found an average technical efficiency for dairy farms of 0.902.

15 Since the technological level varies from year to year because of technological advances and because of different weather conditions, it is common to run efficiency analyses for each year separately (Blancard *et al.*, 2006).

4.2. Effectiveness of agri-environment measures

European citizens are expecting more effective spending of EU budgets. It is therefore of interest to investigate the effectiveness of agri-environment measures for delivering environmental benefits and improving farms' economic efficiency. This is done by calculating the difference between the farms participating in AES and non-participating farms in terms of average efficiency levels.

There is a growing body of research reporting on the impact of AES on farm performance (Arata and Sckokai, 2016; Baráth, Fertő and Bojnec, 2020; Mary, 2013; Mennig and Sauer, 2020). However, our reading of the environmental economics literature reveals that the impact of AES on the combined environmental and technical efficiency measures has not been investigated so far. Our analysis clearly brings a new perspective on the link between AESs and farm performance. Our results are shown in Table 5. The average treatment effect of AESs on efficiency scores is represented by the DiD estimator (obtained from model (11)), which compares the changes in efficiency scores between participants and the matched non-participants over the period 2013–2018.[16] A positive DiD estimator represents an increase in the mean efficiency scores of the participants that is greater than the increase of their matched counterparts or that the decrease in the mean efficiency scores of the participants is lower than the decrease of the non-participants.

Although agri-environment measures have been launched initially to prevent the negative impacts of intensive agricultural systems on environmental aspects, several studies have highlighted the significant role that economic factors can play in farmers' willingness to adopt AESs (Defrancesco *et al.*, 2007; Mozzato *et al.*, 2018). Thus, the environmental effectiveness of the agri-environment policies cannot be assessed in isolation but should always involve the economic efficiency of the schemes. The corresponding DiD estimator that captures the impact of the AES on economic efficiency is not significant. This finding is consistent with the notion in which the loss of economic resources associated with the adoption of environmentally friendly practices has been adequately compensated by the AES. Our results are partially consistent with those recently reported in the literature. The difference between AES-participating and non-participating farms in terms of productivity growth was found to be positive for arable farms and not significant for dairy farms (Mennig and Sauer, 2020). Arata and Sckokai (2016) found that the economic impact of AES varies across countries.

16 Since we are relying on an aggregated variable to captures the impact of AES participation on efficiency scores and in an attempt to disaggregate the different individual schemes and their potential effects at least partly, we formed four agri-environment categories and assigned the individual schemes to these categories and included the categories as dummy variables in the regression models. This approach was chosen as the number of individual schemes amounts to 34. The agri-environment categories that were defined followed the main programme goals set by the Bavarian State Ministry of Food, Agriculture and Forestry: climate protection, soil and water conservation, biodiversity conservation and cultural landscape preservation. Detailed results are reported in Tables A5–A7.

Table 5. Average treatment effect of AES on efficiency scores, 2013–2018

	Technical efficiency		Environmental efficiency		Overall efficiency	
	Treated mean	Control mean	Treated mean	Control mean	Treated mean	Control mean
Pre-treatment	0.925	0.927	0.629	0.606	0.777	0.766
Post-treatment	0.924	0.914	0.666	0.645	0.795	0.779
DiD estimator	0.015		–0.006		0.005	
	(0.035)		(0.028)		(0.019)	
Share of grassland	–0.035*		–0.004		–0.003	
	(0.018)		(0.013)		(0.009)	
Insurance	–0.031		0.034**		0.004	
	(0.023)		(0.016)		(0.011)	
Capital/labour	–0.012		0.067***		0.021***	
	(0.015)		(0.01)		(0.007)	
Age	–0.01		–0.105***		–0.082***	
	(0.039)		(0.029)		(0.021)	
Soil type	–0.031**		0.017		0.012	
	(0.014)		(0.011)		(0.008)	

Notes: Bootstrapped standard errors are shown in parenthesis.
Significance at the 10 per cent, 5 per cent and 1 per cent level is indicated by *, ** and ***, respectively.

AES participation is projected to affect environmental effectiveness, as they offer incentives to participant farmers to adopt environmentally sustainable farming practices. However, our findings for the Bavarian dairy farms suggest that AES payments do not significantly affect the environmental efficiency, indicating that the mean change in the efficiency in controlling nitrogen pollution from 2013 to 2018 does not significantly differ between the subsidised and the non-subsidised farms. The absence of the effect of AES payments on farm-level environmental efficiency is surprising insofar as participation in these schemes[17] requires the adoption of sustainable practices that, among others, impose restrictions on the use of mineral fertilisers which could be expected to reduce nitrogen pollution and thus improves the environmental farm performance, *ceteris paribus*. In fact, practice-based schemes continue to prevail among agri-environment measures because they are relatively easy to comply with and do not require a real change of agricultural practices (Burton and Schwarz, 2013; Kleijn *et al.*, 2011). This might explain why we do not find a significant impact of AES on environmental efficiency. Moreover, the literature might provide insights on possible reasons for this non-effectiveness. Batáry *et al.* (2015) perform a meta-analysis that covers a wide swath of literature over the last 20 years and conclude that agri-environment measures have been

17 The most widespread AESs in Bavaria were related to measures for arable land and grassland measures. Dairy farms are located mainly in the alpine region and Bavarian forest where grassland is one of the most dominating agricultural land uses and most of the adopted AES were under the category of grassland measures (mainly aimed at an input reduction and production extensification).

generally successful in improving farmland biodiversity. However, in their meta-analysis, Batáry *et al*. (2015) asserted that the environmental effectiveness of AES depends on the compatibility of the scheme's design with respect to the specific region in which they are implemented. Kleijn and Sutherland (2003) have reached a similar conclusion on the ambiguous pattern of environmental effectiveness of AES aimed at improving farmland biodiversity. Our findings are in agreement with the above-mentioned results and found no clear support that participating farms differ significantly from non-participants with regard to the environmental efficiency. Furthermore, selection bias could not be excluded either in this study or in the agri-environment programmes in general, as farmers operating in areas experiencing high levels of environmental degradation are unlikely to join a programme that limits their decision-making options (Brady *et al*., 2009; Schmit and Rounsevell, 2006), which in turn will affect the study findings. When it comes to the effect of AES on the overall efficiency scores, the results are similar to those obtained from the above-discussed AES effects. This result is expected since the overall effectiveness (economic and environmental) of the schemes is particularly penalised by the lack of the combined economic and environmental effectiveness.

In sum, our findings show that there is no significant connection between AES and farm performance, and some possible explanations have been discussed earlier. Still, it is important to mention that there is a range of studies that highlighted the link between subsidies and the deterioration of farmers' motivation to produce efficiently (Ahovi, Schneider and Lansink, 2021; Martinez Cillero *et al*., 2018; Skevas, Lansink and Stefanou, 2012). This could lead to increased use of inputs and poor farming practices (Latruffe and Fogarasi, 2009). Therefore, in order to test for the robustness of our results, we replicated our estimations considering short post-treatment periods. While participating farmers are usually tied to management plans for at least 5 years, these robustness checks allow us to examine whether participating farmers were able to improve their performance within a short time frame. Results of the average treatment effect of AES on efficiency scores over three different periods are presented in Table A3. Results remain unchanged.

Table 5 also summarises the results of the regression of the efficiency scores against potential influencing variables. Table A4 provides descriptive statistics of these variables. The contextual variables we choose are farmers' age, share of grassland, insurance, the ratio of capital to labour which is used as a proxy for technology and soil type. Following Mennig and Sauer (2020), the soil type dummy variable is measured on the basis of the yield index unit. Our farms have been divided into two subsamples, one containing farms with a yield index above the median (1 = high-quality soil type) and farms below that level (0 = low-quality soil type). Positive and negative signs of the related variables represent the increase and decrease of efficiency scores, respectively. Full regression results are available in Tables A5–A7.

The results from the regression of technical efficiency scores against these potential influencing variables show that farmers' age does not appear to have a significant impact on technical efficiency. This result is not surprising given the fact that the effect of age on economic performance is widely contested in the literature; older farmers who may have more experience are suggested to be more efficient (Coelli and Battese, 1996). On the other hand, other studies suggest that younger farmers are expected to be more efficient since they are more knowledgeable about technical developments (Weersink, Turvey and Godah, 1990). A higher share of grass on total farm land area is associated with efficiency levels. This result is in line with previous analyses of German farms within a single output framework (Lakner and Brümmer, 2008; Wimmer and Sauer, 2020). Results show that insurance, capital-to-labour ratio and soil type do not exert a statistically significant effect on technical efficiency scores. The lack of significant effect of the soil type Coeff. is surprising, as the theory and common sense suggest that better soils are likely to increase yields, leading to increased farmers' technical efficiency. However, a similar result was reported by Mennig and Sauer (2020), who concluded that better soil conditions do not lead to productivity improvements.

Turning now to the outcome of the regression of the environmental efficiency scores against the same set of explanatory variables, farms operating under better soil conditions do not seem to perform well in terms of environmental efficiency. However, the effect of such factor is debatable because unfavourable conditions may promote better management practices or may have the opposite effect. Indeed, environmental impacts are observed to be higher on fertile soils that favour the development of weeds. Farmers with higher insurance coverage have higher environmental efficiency. This result is in line with Mishra, Wesley Nimon and El-Osta (2005), who explained that the coverage provided by the insurance may reduce the use of agrochemicals inputs and thus improve nitrogen pollution control. Results show that the share of grassland does not exert a statistically significant effect on environmental efficiency scores. Although a higher capital-to-labour ratio may increase economic outcomes through labour-enhancing investments, the impact of investments on environmental performance would depend on whether the investments are categorised as green technologies or not. In our case, we find a positive effect. Finally, we find that younger farmers have higher environmental efficiency since they may be more aware of the external effects of nitrogen pollution. This finding is consistent with several other studies. Concerning the results from the regression of the overall efficiency scores, we find similar patterns as those observed for the environmental efficiency measures.

5. Conclusion

Understanding the effects of AESs on farm-level performance is important for policymakers in making sound decisions when designing schemes that meet economic and environmental goals. For a sample of Bavarian dairy farms over the period 2013–2018, we propose a theoretical framework to analyse the impact of agri-environment measures on farm-level technical and environmental efficiency indices. In our empirical application, environmental efficiency is proxied by nitrogen pollution efficiency as nitrogen surplus from livestock farming is recognised as one of the most pressing environmental concerns affecting water quality in Bavaria. To control for potential selection bias arising from observable and unobservable characteristics, we combine matching with DiD estimators to identify comparable participating and non-participating farms. Our farm performance measures are based on a multi-equation representation of the production frontier (Murty, Robert Russell and Levkoff, 2012).

Our results show that the sample farms have a technical efficiency of 0.920 on average. Our environmental performance measures focusing on nitrogen pollution show an average score of 0.649, implying that there is a considerable reduction potential in terms of nitrogen pollution. The lack of effective environmental regulation to control dairy farmers' nitrogen pollution in Bavaria in the observation period may explain the low environmental efficiency estimates. It was only in 2017 and 2020 that through major reforms of the German fertilisation ordinance, more stringent measures aiming at a reduction of nitrogen losses to the environment were introduced. AESs do not seem to have been successful in reducing nitrogen pollution. Our empirical findings show a non-significant impact of AES on environmental efficiency scores, as the average change over the period 2013–2018 does not significantly differ between participants and their matched non-participants. The same impact was found on technical efficiency levels. This finding does not violate the well-known standards of the World Trade Organization that agri-environment programmes should not distort trade or production. In general, however, much more research is needed on the economic and environmental effectiveness of AES, especially with respect to how scheme payments may influence economic performance (Ansell *et al.*, 2016; Hasler *et al.*, 2022).

This study illustrates the relevance of accounting for runoff by-products when modelling the overall production process of a farm. The inability to adequately handle the existence of unintended outputs not only precludes researchers from assessing the *ceteris paribus* reduction in unintended outputs but also distorts the reliability of the performance measures that have been derived for the desirable production technology. From a policy point of view, these aspects are very important, since both economic and environmental performance, as well as other technological indicators, are frequently taken as the basis for the implementation of new or assessment of already existing policy measures (Kumbhakar and Malikov, 2018). Furthermore, a critical point of departure for evaluating the capacity for policies to build synergies or trade-offs between environmental and economic performance in agriculture is to consider how these types of performance are related before any policy action takes place. Although the widespread idea that the contribution of agricultural activities to the depletion of natural resources may suggest a trade-off relationship between environmental and economic performance, complementarities and synergies are likely to occur and create opportunities for potential win–win situations where the pursuit of economic objectives might generate corollary benefits in terms of environmental aspects. To address this debate, we calculated the Spearman's rank correlation Coeff. to study the association between technical and environmental efficiency scores. Our result indicates a significant positive rank correlation (of the order of 0.420) between technical efficiency and nitrogen pollution efficiency. This positive association does not necessarily invalidate the need for policy intervention. In such cases of synergies, it is recommended from a policy perspective to support those farmers lying behind in terms of overall efficiency rather than trying to enhance environmental performance without considerably compromising economic performance (DeBoe, 2020). For instance, policymakers should consider implementing complementary interventions that enforce linkages between agri-environment policies and insurance policies that sustain economic growth and enable farmers to adopt innovative environmental technologies.

Finally, it is worth mentioning that there are other potential avenues for future research. First, our empirical application is limited by data availability and only nitrogen pollution is considered; thus, a potential extension of the current research could be the incorporation of other environmental indicators such as pesticides pollution (Ait Sidhoum, Serra and Latruffe, 2020) and biodiversity index (Sipiläinen and Huhtala, 2013). Second, while our results, naturally, cannot be extended beyond the specific geographic context of Bavaria, it would be interesting to test the robustness of our findings with those derived from other datasets or alternative methods, such as the one described by Centorrino *et al*. (2021), which involve the use of stochastic frontier procedures. Finally, a promising avenue for future research would be more context-specific research to investigate the effects of specific AES measures on technical and environmental performance.

Acknowledgements

This research is part of the EFFECT project. The authors gratefully acknowledge financial support from the EU Horizon 2020 Programme for Research, Technological Development, and Demonstration under grant agreement no. 817903.

References

Adenuga, A. H., Davis, J., Hutchinson, G., Donnellan, T. and Patton, M. (2019). Environmental efficiency and pollution costs of nitrogen surplus in dairy farms: a parametric hyperbolic technology distance function approach. *Environmental and Resource Economics* 74: 1273–1298.

Ahovi, E., Schneider, K. and Lansink, A. O. (2021). Technical inefficiency of Dutch vegetable farms: specific-input analyses. *PLoS One* 16: e0250494.

Ait Sidhoum, A., Serra, T. and Latruffe, L. (2020). Measuring sustainability efficiency at farm level: a data envelopment analysis approach. *European Review of Agricultural Economics* 47: 200–225.

Ait Sidhoum, A., Hervé Dakpo, K. and Latruffe, L. (2022). Trade-offs between economic, environmental and social sustainability on farms using a latent class frontier efficiency model: evidence for Spanish crop farms. *PLoS One* 17: e0261190.

Ali, A. and Abdulai, A. (2010). The adoption of genetically modified cotton and poverty reduction in Pakistan. *Journal of Agricultural Economics* 61: 175–192.

Alvarez, A. and Del Corral, J. (2010). Identifying different technologies using a latent class model: extensive versus intensive dairy farms. *European Review of Agricultural Economics* 37: 231–250.

Ansell, D., Freudenberger, D., Munro, N. and Gibbons, P. (2016). The cost-effectiveness of agri-environment schemes for biodiversity conservation: a quantitative review. *Agriculture, Ecosystems & Environment* 225: 184–191.

Arata, L. and Sckokai, P. (2016). The impact of agri-environmental schemes on farm performance in five E.U. member States: a DID-matching approach. *Land Economics* 92: 167–186.

Augurzky, B. and Schmidt, C. M. (2001). The propensity score: a means to an end. IZA Discussion Papers, No. 271.

Austin, P. C., Grootendorst, P. and Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine* 26: 734–753.

Banker, R. D. and Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research* 175: 1311–1320.

Baráth, L., Fertő, I. and 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: 853–876.

Batáry, P., Dicks, L. V., Kleijn, D. and Sutherland, W. J. (2015). The role of agri-environment schemes in conservation and environmental management. *Conservation Biology* 29: 1006–1016.

Beltrán-Esteve, M. and Picazo-Tadeo, A. J. (2017). Assessing environmental performance in the European Union: eco-innovation versus catching-up. *Energy Policy* 104: 240–252.

Bisang, I., Lienhard, L. and Bergamini, A. (2021). Three decades of field surveys reveal a decline of arable bryophytes in the Swiss lowlands despite agri-environment schemes. *Agriculture, Ecosystems & Environment* 313: 107325.

Blancard, S., Boussemart, J. P., Briec, W. and Kerstens, K. (2006). Short- and long-run credit constraints in French agriculture: a directional distance function framework using expenditure-constrained profit functions. *American Journal of Agricultural Economics* 88: 351–364.

Brady, M., Kellermann, K., Sahrbacher, C. and Jelinek, L. (2009). Impacts of decoupled agricultural support on farm structure, biodiversity and landscape mosaic: some EU results. *Journal of Agricultural Economics* 60: 563–585.

Budka, M., Jobda, M., Szałański, P. and Piórkowski, H. (2019). Effect of agri-environment measure for the aquatic warbler on bird biodiversity in the extensively managed landscape of Biebrza Marshes (Poland). *Biological Conservation* 239: 108279.

Burton, R. J. F. and Schwarz, G. (2013). Result-oriented agri-environmental schemes in Europe and their potential for promoting behavioural change. *Land Use Policy* 30: 628–641.

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

Centorrino, S., Pérez-Urdiales, M., Bravo-Ureta, B. and Wall, A. (2021). Binary endogenous treatment in stochastic frontier models with an application to soil conservation in El Salvador. In *95th Annual Conference*, March 29-30, 2021, Warwick, UK (Hybrid) 312058, Agricultural Economics Society - AES.

Chabé-Ferret, S. and Subervie, J. (2013). How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. *Journal of Environmental Economics and Management* 65: 12–27.

Chambers, R. G., Chung, Y. and Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory* 70: 407–419.

Chambers, R. G. and Serra, T. (2018). The social dimension of firm performance: a data envelopment approach. *Empirical Economics* 54: 189–206.

Charles, H., Godfray, H. and Garnett, T. (2014). Food security and sustainable intensification. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20120273.

Coderoni, S. and Esposti, R. (2018). CAP payments and agricultural GHG emissions in Italy. A farm-level assessment. *Science of the Total Environment* 627: 427–437.

Coelli, T. and Battese, G. (1996). Identification of factors which influence the technical inefficiency of Indian farmers. *Australian Journal of Agricultural and Resource Economics* 40: 103–128.

Dakpo, K. H., Jeanneaux, P. and Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research* 250: 347–359.

Davidson, E. A., David, M. B., Galloway, J. N., Goodale, C. L., Haeuber, R., Harrison, J. A., Howarth, R. W., Jaynes, D. B., Lowrance, R. R., Thomas Nolan, B., Peel, J. L., Pinder, R. W., Porter, E., Snyder, C. S., Townsend, A. R. and Ward, M. H. (2011). Excess nitrogen in the U.S. environment: trends, risks, and solutions. *Issues in Ecology* 15.

DeBoe, G. (2020). Economic and environmental sustainability performance of environmental policies in agriculture. Paris: OECD Publishing.

Defrancesco, E., Gatto, P., Runge, F. and Trestini, S. (2007). Factors affecting farmers? Participation in agri-environmental measures: a Northern Italian perspective. *Journal of Agricultural Economics* 59: 114–131.

Devkota, K. P., McDonald, A. J., Khadka, L., Khadka, A., Paudel, G. and Devkota, M. (2016). Fertilizers, hybrids, and the sustainable intensification of maize systems in the rainfed mid-hills of Nepal. *European Journal of Agronomy* 80: 154–167.

Eulenstein, F., Werner, A., Willms, M., Juszczak, R., Schlindwein, S. L., Chojnicki, B. H. and Olejnik, J. (2008). Model based scenario studies to optimize the regional nitrogen balance and reduce leaching of nitrate and sulfate of an agriculturally used water catchment. *Nutrient Cycling in Agroecosystems* 82: 33–49.

European Commission. (2021). European agricultural fund for rural development. https://cohesiondata.ec.europa.eu/funds/14-20.

Färe, R., Grosskopf, S., Noh, D. W. and Weber, W. (2005). Characteristics of a polluting technology: theory and practice. *Journal of Econometrics* 126: 469–492.

Førsund, F. R. (2018). Multi-equation modelling of desirable and undesirable outputs satisfying the materials balance. *Empirical Economics* 54: 67–99.

Førsund, F. R. (2021). Performance measurement and joint production of intended and unintended outputs. *Journal of Productivity Analysis* 55: 157–175.

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

Frisch, R. (1965). *Theory of Production*. Dordrecht: D. Reidel Publishing.

Gadanakis, Y., Bennett, R., Park, J. and Areal, F. J. (2015). Evaluating the Sustainable Intensification of arable farms. *Journal of Environmental Management* 150: 288–298.

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B. and Cosby, B. J. (2003). The nitrogen cascade. *Bioscience* 53: 341–356.

Gamer, W. and Bahrs, E. (2010). Bilanzen von potenziell umweltbelastenden Nährstoffen (N, P, K und S) der Landwirtschaft in Baden-Württemberg. Forschungsauftrag des Ministeriums für Ernährung und Ländlichen Raum Baden-Württemberg, Stuttgart.

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

Guesmi, B. and Serra, T. (2015). Can we improve farm performance? The determinants of farm technical and environmental efficiency. *Applied Economic Perspectives and Policy* 37: 692–717.

Hansson, H. and Öhlmér, B. (2008). The effect of operational managerial practices on economic, technical and allocative efficiency at Swedish dairy farms. *Livestock Science* 118: 34–43.

Hasler, B., Termansen, M., Nielsen, H. Ø., Daugbjerg, C., Wunder, S. and Latacz-Lohmann, U. (2022). European agri-environmental policy: evolution, effectiveness, and challenges. *Review of Environmental Economics and Policy* 16: 105–125.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica* 66: 1017–1098.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* 64: 605–654.

Hickman, J. E., Huang, Y., Wu, S., Diru, W., Groffman, P. M., Tully, K. L. and Palm, C. A. (2017). Nonlinear response of nitric oxide fluxes to fertilizer inputs and the impacts of agricultural intensification on tropospheric ozone pollution in Kenya. *Global Change Biology* 23: 3193–3204.

Howarth, R. W., Swaney, D. P., Butler, T. J. and Marino, R. (2000). Climatic control on eutrophication of the Hudson River estuary. *Ecosystems* 3: 210–215.

Hynes, S. and Garvey, E. (2009). Modelling farmers' participation in an agri-environmental scheme using panel data: an application to the rural environment protection scheme in Ireland. *Journal of Agricultural Economics* 60: 546–562.

IPCC. (2014). Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva.

Jaforullah, M. and Whiteman, J. (1999). Scale efficiency in the New Zealand dairy industry: A non-parametric approach. *Australian Journal of Agricultural and Resource Economics* 43: 523–541.

Jones, J. I., Murphy, J. F., Anthony, S. G., Arnold, A., Blackburn, J. H., Duerdoth, C. P., Hawczak, A., Hughes, G. O., Pretty, J. L., Scarlett, P. M., Gooday, R. D., Zhang, Y. S., Fawcett, L. E., Simpson, D., Turner, A. W. B., Naden, P. S. and Skates, J. (2017). Do agri-environment schemes result in improved water quality? *Journal of Applied Ecology* 54: 537–546.

Kanter, D. R., Wentz, J. A., Galloway, J. N., Moomaw, W. R. and Winiwarter, W. (2017). Managing a forgotten greenhouse gas under existing U.S. law: an interdisciplinary analysis. *Environmental Science & Policy* 67: 44–51.

Kay, P., Grayson, R., Phillips, M., Stanley, K., Dodsworth, A., Hanson, A., Walker, A., Foulger, M., McDonnell, I. and Taylor, S. (2012). The effectiveness of agricultural stewardship for improving water quality at the catchment scale: experiences from an NVZ and ECSFDI watershed. *Journal of Hydrology* 422–423: 10–16.

Khandker, S. R., Koolwal, G. B. and Sammad, H. A. (2009). *Handbook on Impact, Quantitative Methods and Practices*. Washington, DC: The World Bank.

Kirschke, S., Häger, A., Kirschke, D. and Völker, J. (2019). Agricultural nitrogen pollution of freshwater in Germany. The governance of sustaining a complex problem. *Water (Switzerland)* 11: 2450.

Kleijn, D., Rundlöf, M., Scheper, J., Smith, H. G. and Tscharntke, T. (2011). Does conservation on farmland contribute to halting the biodiversity decline? *Trends in Ecology & Evolution* 26: 474–481.

Kleijn, D. and Sutherland, W. J. (2003). How effective are European agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology* 40: 947–969.

Krupa, S. V. (2003). Effects of atmospheric ammonia (NH_3) on terrestrial vegetation: a review. *Environmental Pollution* 124: 179–221.

Kumbhakar, S. C. and Malikov, E. (2018). Good modeling of bad outputs: editors' introduction. *Empirical Economics* 54: 1–6.

Kuosmanen, T. and Kortelainen, M. (2005). Measuring eco-efficiency of production with data envelopment analysis. *Journal of Industrial Ecology* 9: 59–72.

Lakner, S. and Brümmer, B. (2008). Determinants of technical efficiency of organic grassland farming in Germany. *Konferenzbeitrag Auf Dem IAMO-Forum* 25: 27.

Latruffe, L., Fogarasi, J. and Desjeux, Y. (2012). Efficiency, productivity and technology comparison for farms in Central and Western Europe: the case of field crop and dairy farming in Hungary and France. *Economic Systems* 36: 264–278.

Latruffe, L. and Fogarasi, J. (2009). Farm performance and support in central and western Europe: a comparison of Hungary and France. Working Paper SMART-LERECO.

LfL. (2013). Basisdaten (Düngeberatung/Düngeverordnung). Bayerische Landesanstalt für Landwirtschaft.

LfU. (2021a). Ökologischer Zustand der Oberflächengewässer - LfU Bayern. https://www.lfu.bayern.de/umweltdaten/indikatoren/natur_landschaft/oekozustand_oberflaechengewaesser/index.htm.

LfU. (2021b). Nitrat im Grundwasser - LfU Bayern. https://www.lfu.bayern.de/umweltdaten/indikatoren/umwelt_gesundheit/nitrat_im_grundwasser/index.htm.

Martinez Cillero, M., Thorne, F., Wallace, M., Breen, J. and Hennessy, T. (2018). The effects of direct payments on technical efficiency of Irish beef farms: a stochastic frontier analysis. *Journal of Agricultural Economics* 69: 669–687.

Mary, S. (2013). Assessing the impacts of pillar 1 and 2 subsidies on TFP in French crop farms. *Journal of Agricultural Economics* 64: 133–144.

Matzdorf, B. and Lorenz, J. (2010). How cost-effective are result-oriented agri-environmental measures?-An empirical analysis in Germany. *Land Use Policy* 27: 535–544.

Mayen, C. D., Balagtas, J. V. and 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.

Mennig, P. and Sauer, J. (2020). The impact of agri-environment schemes on farm productivity: A DID-matching approach. *European Review of Agricultural Economics* 47: 1045–1053.

Mishra, A. K., Wesley Nimon, R. and El-Osta, H. S. (2005). Is moral hazard good for the environment? Revenue insurance and chemical input use. *Journal of Environmental Management* 74: 11–20.

Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E. and Sartori, L. (2018). The role of factors affecting the adoption of environmentally friendly farming practices: Can geographical context and time explain the differences emerging from literature? *Sustainability (Switzerland)* 10: 3101.

Murty, S., Robert Russell, R. and Levkoff, S. B. (2012). On modeling pollution-generating technologies. *Journal of Environmental Economics and Management* 64: 117–135.

OECD. (2013). *OECD Compendium of Agri-environmental Indicators*. Paris: OECD.

Pérez Urdiales, M., Lansink, A. O. and Wall, A. (2016). Eco-efficiency among dairy farmers: the importance of socio-economic characteristics and farmer attitudes. *Environmental and Resource Economics* 64: 559–574.

Pufahl, A. and Weiss, C. R. (2009). Evaluating the effects of farm programmes: results from propensity score matching. *European Review of Agricultural Economics* 36: 79–101.

Rabalais, N. N., Turner, R. E. and Wiseman, W. J. (2002). Gulf of Mexico hypoxia, a.k.a. "The dead zone. *Annual Review of Ecology and Systematics* 33: 235–263.

Reinhard, S., Knox Lovell, C. A. and Thijssen, G. J. (2000). Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA. *European Journal of Operational Research* 121: 287–303.

Reinhard, S. and Thijssen, G. (2000). Nitrogen efficiency of Dutch dairy farms: A shadow cost system approach. *European Review of Agricultural Economics* 27: 167–186.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., De Wit, C. A., Hughes, T., Van Der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P. and Foley, J. A. (2009). A safe operating space for humanity. *Nature* 461: 472–475.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39: 33–38.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63: 581–592.

Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52: 249–264.

Samson, G. S., Gardebroek, C. and Jongeneel, R. A. (2016). Explaining production expansion decisions of Dutch dairy farmers. *NJAS - Wageningen Journal of Life Sciences* 76: 87–98.

Schmit, C. and Rounsevell, M. D. A. (2006). Are agricultural land use patterns influenced by farmer imitation? *Agriculture, Ecosystems & Environment* 115: 113–127.

Serra, T., Chambers, R. G. and Oude Lansink, A. (2014). Measuring technical and environmental efficiency in a state-contingent technology. *European Journal of Operational Research* 236: 706–717.

Shortall, O. K. and Barnes, A. P. (2013). Greenhouse gas emissions and the technical efficiency of dairy farmers. *Ecological Indicators* 29: 478–488.

Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136: 31–64.

Sipiläinen, T. and Huhtala, A. (2013). Opportunity costs of providing crop diversity in organic and conventional farming: would targeted environmental policies make economic sense? *European Review of Agricultural Economics* 40: 441–462.

Skevas, T., Lansink, A. O. and Stefanou, S. E. (2012). Measuring technical efficiency in the presence of pesticide spillovers and production uncertainty: the case of Dutch arable farms. *European Journal of Operational Research* 223: 550–559.

Skevas, T. and Serra, T. (2016). The role of pest pressure in technical and environmental inefficiency analysis of Dutch arable farms: an event-specific data envelopment approach. *Journal of Productivity Analysis* 46: 139–153.

Slabe-Erker, R., Bartolj, T., Ogorevc, M., Kavaš, D. and Koman, K. (2017). The impacts of agricultural payments on groundwater quality: spatial analysis on the case of Slovenia. *Ecological Indicators* 73: 338–344.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125: 305–353.

STATBA. (2018). Statistisches Bundesamt (annual reports). Fachserie 4 Reihe 8.2 Düngemittelversorgung. Statistisches Bundesamt, Wiesbaden.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., De Vries, W., De Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B. and Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science* 347: 1259855.

Stetter, C., Menning, P. and Sauer, J. (2020). Going beyond average—using machine learning to evaluate the effectiveness of environmental subsidies at micro-level. Working Paper to be presented at the 94th Annual Conference of the Agricultural Economics Society.

Stokes, J. R., Tozer, P. R. and Hyde, J. (2007). Identifying efficient dairy producers using data envelopment analysis. *Journal of Dairy Science* 90: 2555–2562.

Sutton, M. A., Oenema, O., Erisman, J. W., Leip, A., Van Grinsven, H. and Winiwarter, W. (2011). Too much of a good thing. *Nature* 472: 159–161.

Townsend, T. J., Ramsden, S. J. and Wilson, P. (2016). How do we cultivate in England? Tillage practices in crop production systems. *Soil Use and Management* 32: 106–117.

van Grinsven, H. J. M., van Eerdt, M. M., Westhoek, H. and Kruitwagen, S. (2019). Benchmarking eco-efficiency and footprints of Dutch agriculture in European context and implications for policies for climate and environment. *Frontiers in Sustainable Food Systems* 3: 13.

Vitousek, P. M., Mooney, H. A., Lubchenco, J. and Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science (1979)* 277: 494–499.

Weersink, A., Turvey, C. G. and Godah, A. (1990). Decomposition measures of technical efficiency for Ontario dairy farms. *Canadian Journal of Agricultural Economics/Revue Canadienne D'agroeconomie* 38: 439–456.

Weltin, M. and Hüttel, S. (2019). Farm eco-efficiency: can sustainable intensification make the difference? FORLand-Working Paper 10. Berlin: Humboldt-Universität zu Berlin.

Wendland, M., Diepolder, M., Offenberger, K. and Raschbacher, S. (2018). Leitfaden für die Düngung von Acker-und Grünland. Bayerische Landesanst. für Bodenkultur und Pflanzenbau.

Wimmer, S. and Sauer, J. (2020). Diversification economies in dairy farming - Empirical evidence from Germany. *European Review of Agricultural Economics* 47: 1338–1365.

Appendix

Table A1. Summary statistics for the covariates used in the PSM in the pre-treatment year 2013

	Dimension	Average	S.d.	Min	Max
AES	1 if yes, 0 if no	0.54	0.50	0	1.00
Livestock unit	number	57.11	27.43	8.00	162.00
Labour	Man-work units	1.79	0.62	0.40	5.00
Land	Hectares	66.48	35.55	12.75	290.05
Capital depreciation	Euros/ha	597.55	339.56	29.81	3,145.65
Total sales	Euros/ha	3,567.60	1,256.20	1,390.59	10,294.10
Fertilisers	Euros/ha	167.49	78.84	0	487.14
Pesticides	Euros/ha	63.22	39.48	0	231.47
Feed	Euros/ha	6.13	0.64	2.69	7.73
Farmer's age	Number	56.96	9.98	33.00	91.00
Share of arable land	%	0.58	0.19	0	0.97
Share of grassland	%	0.42	0.19	0.03	1.00
Share of rented land	%	0.61	0.36	0.02	2.89
Yield index	Number/ha	62.07	56.58	5.31	317.65
Agricultural income	Euros/ha	1,091.56	602.35	−572.70	4,059.40
Dummy variable 'Swabia'	1 if yes, 0 if no	0.15	0.36	0	1.00
Dummy variable 'Lower Franconia'	1 if yes, 0 if no	0.10	0.30	0	1.00
Dummy variable 'Middle Franconia'	1 if yes, 0 if no	0.23	0.42	0	1.00
Dummy variable 'Upper Franconia'	1 if yes, 0 if no	0.20	0.40	0	1.00
Dummy variable 'Upper Palatinate'	1 if yes, 0 if no	0.17	0.38	0	1.00
Dummy variable 'Lower Bavaria'	1 if yes, 0 if no	0.03	0.17	0	1.00
Dummy variable 'Upper Bavaria'	1 if yes, 0 if no	0.12	0.33	0	1.00
Dummy variable 'no agric. education'	1 if yes, 0 if no	0.04	0.21	0	1.00
Dummy variable 'skilled worker'	1 if yes, 0 if no	0.54	0.50	0	1.00
Dummy variable 'University education'	1 if yes, 0 if no	0.41	0.49	0	1.00
Gross value added in agriculture, forestry, fishing	Euros (million)	72.46	32.34	6.00	144.00

(continued)

Table A1. (Continued)

	Dimension	Average	S.d.	Min	Max
Gross domestic product per capita	Euros	27,818.60	4,782.03	18,470.00	55,265.00
Unemployment rate	%	0.03	0.01	0.01	0.07
Workforce	Number	36,464.44	12,961.64	21,672.00	76,017.00
Farmland rental price	Euros/ha	227.78	74.47	108.00	412.00
Number of observations			271		

Table A2. Estimation of the propensity score

Logistic regression
LR chi^2(27) = 101.26
Prob > chi^2 = 0.0000
Log likelihood = −135.624
Pseudo R^2 = 0.272
Number of observations = 271
Dependent variable: AES

Regressors	Coef.	*z*-stat	*P*-value
Livestock per ha	–0.984	−1.13	0.26
Labour per ha	–2.441	–0.14	0.889
Land	0.040	3.88	0.000
Capital depreciation per ha	–0.040	–0.11	0.909
Total sales per ha	0.829	0.62	0.538
Fertilisers per ha	0.001	0.22	0.83
Pesticides per ha	–0.286	−1.36	0.174
Feed per ha	–0.326	–0.89	0.376
Ln farmers' age	–0.917	–0.97	0.334
Share arable land	–7.580	–2.69	0.007
Share grassland	–2.576	–2.5	0.013
Share rented land	0.155	0.62	0.537
Ln yield index per ha	0.674	1.71	0.087
Agricultural income per ha	0.475	0.63	0.528
Dummy variable 'master's certificate or university degree'	1.012	1.31	0.19
Dummy variable 'in education or skilled worker'	0.621	0.81	0.418
Dummy variable 'Swabia'	−17.295	–0.01	0.993
Dummy variable 'Lower Franconia'	−16.946	–0.01	0.993
Dummy variable 'Middle Franconia'	−15.657	–0.01	0.993
Dummy variable 'Upper Franconia'	−14.582	–0.01	0.994
Dummy variable 'Upper Palatinate'	−15.675	–0.01	0.993
Dummy variable 'Upper Bavaria'	−17.383	–0.01	0.993
Ln gross domestic product per capita	1.752	1.51	0.132
Unemployment rate	0.446	0.42	0.674
Gross value added in agriculture, forestry, fishing	0.818	2.25	0.024
Intercept	– 9.502	–0.01	0.996

Table A3. Average treatment effect of AES on efficiency scores with varying length of post-treatment period

	Technical efficiency	Environmental efficiency	Overall efficiency
2014 as post-treatment			
DiD estimator	–0.002	0.019	0,008
Bootstrapped Std. Err.	0.029	0.030	0.020
z-value	–0.14	0.63	0.40
P>|z|	0.889	0.528	0.692
2015 as post-treatment			
DiD estimator	0.018	0.012	0.013
Bootstrapped Std. Err.	0.036	0.025	0.019
z-value	0.52	0.47	0.70
P>|z|	0.602	0.636	0.481
2016 as post-treatment			
DiD estimator	0.018	0.006	0.011
Bootstrapped Std. Err.	0.035	0.024	0.018
z-value	0.54	0.24	0.62
P>|z|	0.589	0.807	0.533

The results for the other explanatory variables are not shown to conserve space but are available from the authors.

Table A4. Descriptive statistics of the contextual variables (720 observations)

Variable	Dimension	Average	S.D.
Share of grassland	(%)	41	0.17
Insurance	Euros	6018.76	2294.10
Capital/labour	Ratio	19,689.07	10,400.47
Age	Years	54	9.45
	Category	(%)	
Soil type	1	50	
	0	50	
Climate AES	1	23	
	0	77	
Soil water AES	1	31	
	0	69	
Biodiversity AES	1	25	
	0	75	
Cultural landscape AES	1	8	
	0	92	

Table A5. Detailed results of the truncated regression analysis with bootstrapped confidence intervals (technical efficiency)

Variables	Coef.	Bootstrap Std. Err.	*z*	*P*>|z|	95% Confidence interval Lower	Upper
Technical efficiency						
Constant	1.334	0.255	5.220	0.000	0.844	1.828
AES	–0.013	0.035	–0.360	0.717	–0.081	0.058
Time	–0.035	0.030	–1.150	0.250	–0.095	0.023
DID	0.015	0.036	0.410	0.684	–0.058	0.084
Share of grassland	–0.035	0.018	–1.930	0.054	–0.071	–0.001
Insurance	–0.031	0.023	–1.340	0.179	–0.077	0.013
Capital/labour	–0.012	0.015	–0.790	0.428	–0.040	0.017
Age	–0.010	0.039	–0.260	0.792	–0.085	0.070
Soil type	–0.031	0.014	–2.280	0.023	–0.058	–0.003
D2014	–0.006	0.029	–0.210	0.831	–0.066	0.051
D2015	–0.034	0.029	–1.160	0.247	–0.092	0.024
D2016	–0.013	0.030	–0.430	0.667	–0.072	0.047
D2017	–0.041	0.029	–1.430	0.154	–0.099	0.014
D2018	–	–	–	–	–	–
Climate AES	0.002	0.018	0.130	0.896	–0.033	0.038
Soil water AES	–0.034	0.017	–1.980	0.048	–0.068	–0.001
Biodiversity AES	–0.016	0.015	–1.100	0.273	–0.044	0.014
Cultural landscape AES	0.078	0.033	2.390	0.017	0.021	0.148

Note: The Coeff. of D2018 has been omitted because of collinearity.

Table A6. Detailed results of the truncated regression analysis with bootstrapped confidence intervals (environmental efficiency)

Variables	Coef.	Bootstrap Std. Err.	*z*	*P*>|z|	95% Confidence interval Lower	Upper
Environmental efficiency						
Constant	0.066	0.186	0.350	0.724	–0.307	0.431
AES	0.049	0.027	1.820	0.068	–0.004	0.103
Time	–0.006	0.023	–0.270	0.790	–0.050	0.042
DID	–0.006	0.028	–0.230	0.820	–0.062	0.049
Share of grassland	–0.004	0.013	–0.320	0.750	–0.030	0.023
Insurance	0.034	0.016	2.080	0.038	0.003	0.066
Capital/labour	0.067	0.010	6.890	0.000	0.048	0.086
Age	–0.105	0.029	–3.610	0.000	–0.162	–0.047

(continued)

Table A6. (Continued)

Variables	Coef.	Bootstrap Std. Err.	z	$P>\|z\|$	95% Confidence interval Lower	Upper
Soil type	0.017	0.011	1.570	0.116	–0.004	0.039
D2014	0.112	0.022	5.020	0.000	0.068	0.156
D2015	0.058	0.022	2.590	0.010	0.015	0.103
D2016	0.021	0.022	0.960	0.339	–0.020	0.066
D2017	0.010	0.022	0.470	0.638	–0.032	0.055
D2018	–	–	–	–	–	–
Climate AES	–0.019	0.013	−1.370	0.169	–0.045	0.006
Soil water AES	–0.026	0.013	–2.010	0.044	–0.051	0.000
Biodiversity AES	–0.056	0.012	–4.590	0.000	–0.079	–0.032
Cultural landscape AES	0.047	0.021	2.250	0.024	0.004	0.088

Note: The Coeff. of D2018 has been omitted because of collinearity.

Table A7. Detailed results of the truncated regression analysis with bootstrapped confidence intervals (overall efficiency)

Variables	Coef.	Bootstrap Std. Err.	z	$P>\|z\|$	95% Confidence interval Lower	Upper
Overall efficiency						
Constant	0.849	0.130	6.520	0.000	0.601	1.101
AES	0.022	0.018	1.180	0.237	–0.014	0.058
Time	–0.011	0.016	–0.690	0.491	–0.042	0.020
DID	0.005	0.019	0.290	0.776	–0.032	0.042
Share of grassland	–0.003	0.009	–0.370	0.710	–0.022	0.015
Insurance	0.004	0.011	0.370	0.708	–0.018	0.027
Capital/labour	0.021	0.007	3.100	0.002	0.007	0.034
Age	–0.082	0.021	–3.970	0.000	–0.123	–0.042
Soil type	0.012	0.008	1.620	0.105	–0.002	0.027
D2014	0.049	0.016	3.140	0.002	0.018	0.079
D2015	0.020	0.015	1.350	0.177	–0.009	0.049
D2016	0.003	0.016	0.210	0.837	–0.028	0.034
D2017	–0.004	0.015	–0.270	0.787	–0.034	0.025
D2018	–	–	–	–	–	–
Climate AES	–0.009	0.009	–0.940	0.347	–0.028	0.009
Soil water AES	–0.017	0.009	–1.900	0.058	–0.035	0.000
Biodiversity AES	–0.036	0.009	–4.210	0.000	–0.053	–0.019
Cultural landscape AES	0.035	0.014	2.460	0.014	0.007	0.063

Note: The Coeff. of D2018 has been omitted because of collinearity.