Land Use Policy 114 (2022) 105950

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

Evaluations of agri-environmental schemes based on observational farm data: The importance of covariate selection

Reinhard Uehleke [a,*], Martin Petrick [b], Silke Hüttel [a]

[a] Production Economics Group, University of Bonn, Meckenheimer Allee 174, D-53115 Bonn, Germany
[b] Professorship for Agricultural, Food and Environmental Policy, Justus-Liebig-Universität Gießen, Senckenbergerstrasse 3, D-35390 Gießen, Germany

A R T I C L E I N F O

JEL Classification:
Q18
Q24
Q58
D04

Keywords:
Policy impact evaluation
Agri-environmental schemes
DID matching
Kernel matching
Covariate selection

A B S T R A C T

Evaluations of agri-environmental schemes (AES) based on observational farm data generally use a matching algorithm for comparing participating and non-participating farms. To mitigate the potential post-matching covariate imbalances between groups resulting from the use of large covariate sets, this paper proposes a method mix that reduces the covariate set and maximises the utilised number of observations. We test the approach on an evaluation of the European Union's AES in the programming period of 2000–2006, estimating the impacts of AES participation on typical measures of land management, i.e. fertiliser and plant protection expenditures and grassland share. We use Mahalanobis distance matching with exact matching on the entry year of the participating farms and kernel matching with automated bandwidth selection to maximise the utilised sample and increase the estimator's efficiency. Combining cause-and-effect path analysis with statistical covariate selection algorithms reduces the covariate set and improves balance on the characteristics that describe the production environment, farming intensity, productivity, and farmers' preferences. We find that AES generate moderate decreases in plant protection expenditure and moderate increases in grassland shares. We conclude that our proposed method mix ensures an efficient use of information and improves the reliability of AES impact evaluation.

1. Introduction

The European Union's (EU) agri-environmental schemes (AES) were designed to mitigate the adverse environmental effects of farming. Early AES-evaluation studies have noted mixed results for achieving biodiversity goals (Kleijn et al., 2006), while recent studies noted that subsequent schemes were not reversing population declines in farmland birds (Teillard et al., 2015; Jerrentrup et al., 2017; Bellebaum and Koffijberg, 2018; Bowler et al., 2019). Other studies reported positive effects of AES on biodiversity, depending on the landscape context (Scheper et al., 2013; Marja et al., 2019). To date, few environmental impact assessments of AES and similar programs using large quantitative data sets have been undertaken (Baylis et al., 2016; Börner et al., 2017; Yoder et al., 2019), for example, Pufahl and Weiss (2009) for Germany, Chabé-Ferret and Subervie (2013) for France, Arata and Sckokai (2016) for selected EU countries and Cisilino et al. (2019) for Italy.

The AES impacts of interest are identified by comparing the average differences in land management (outcome) over the AES period of participants compared to matched non-participants (Pufahl and Weiss, 2009; Chabé-Ferret and Subervie, 2013; Arata and Sckokai, 2016; Cisilino et al., 2019; Bertoni et al., 2020). Comparative methods depend on collecting large numbers of participating and non-participating farms with several years of observations (panel data), sufficient overlap in the farms' characteristics and parallel outcome trends between control and treated groups. To ensure comparability and mitigate confounding from time-invariant unobserved effects (selection bias), matching algorithms have been combined with difference-in-difference (DID) approaches (Heckman et al., 1997). Conditioning on covariates describing the farm and production characteristics that are relevant for the AES participation decision and the resulting farming outcome, i.e., all confounding characteristics, the unobserved counterfactual is approximated by observations of the non-participants (Imbens and Rubin, 2015; Kainz et al., 2017).

* Corresponding author.
E-mail address: r.uehleke@ilr.uni-bonn.de (R. Uehleke).

https://doi.org/10.1016/j.landusepol.2021.105950
Received 28 October 2020; Received in revised form 16 December 2021; Accepted 17 December 2021
Available online 13 January 2022
0264-8377/

In the EU context, the Farm Accountancy Data Network (FADN) appears most suitable for counterfactual AES evaluations (Castaño et al., 2019). Given the typical AES contract length of five years, a panel length of at least six years is needed to denote the impact of participation in the AES programming period on land management (measures) after participation. But given the rotating panel nature of the FADN (Cisilino et al., 2011), each additional year included in the analysis reduces the number of usable observations at the expense of statistical efficiency. Typically, large sets of covariates capturing regional production, farming, and farm(er) characteristics are used to describe the decision to participate in AES (Pufahl and Weiss, 2009; Pascucci et al., 2013; Lastra-Bravo et al., 2015; Arata and Sckokai, 2016; Zimmermann and Britz, 2016; Bartkowski and Bartke, 2018). However, such a "kitchen sink approach" (Shortreed and Ertefaie, 2017) could require pruning more observations than necessary to achieve covariate balance on the confounding covariates at an additional expense of statistical efficiency. In a worst case, the relevant covariates may remain unbalanced with biased effect estimates, outweighing the advantage of reducing the risk of missing important confounders. Extensive sets of covariates also could inflate standard errors of nonparametric estimates of the desired policy impacts without improving covariate balance and increase bias of the estimate (de Luna et al., 2011; Wooldridge, 2016; Shortreed and Ertefaie, 2017).

Some studies include land management measures prior to participation (pre-treatment outcome) in the set of covariates (e.g. Pufahl and Weiss, 2009; Arata and Sckokai, 2016). DID matching with conditioning on the pre-treatment outcome, however, could introduce bias (Chabé-Ferret, 2015, 2017; Daw and Hatfield, 2018): before entering AES, farms could manage land at lower intensities due to temporary or transitory shocks unrelated to AES participation, such as liquidity constraints. In such cases, conditioning on the pre-treatment outcome violates the parallel trends assumption and also could introduce bias (Chabé-Ferret, 2017).

Motivated by the challenges of AES evaluation, this paper proposes a method mix approach that reduces the full covariate set to the relevant set by using minimal subset identification (de Luna et al., 2011; Persson et al., 2017), while still maximising the utilised number of observations. Our approach increases efficiency in two ways: we maximise the utilised pre-matching sample by including staggered AES entry dates, i.e., the participating farms entering in the first, second, and third years of the respective programming period. Then we combine DID matching on the Mahalanobis distance with exact matching on the entry year to include all relevant entry cohorts and use kernel matching with optimal bandwidth selection (Loader, 1999; Galdo et al., 2008; Huber et al., 2015) which is more efficient than pair matching (Caliendo and Kopeinig, 2008; King and Nielsen, 2019). In contrast to frequently used approaches for lasso-based variable selection in regression contexts (Belloni et al., 2014) which aim at finding the set of covariates that enables an unbiased regression coefficient estimation, we aim for minimal subset identification (de Luna et al., 2011; Persson et al., 2017) to find the relevant set of matching variables. Finding the relevant set of matching variables means gauging between potential bias stemming either from exhaustive sets of covariates for which covariate balance cannot be achieved or from reduced covariates sets that erroneously exclude important confounders. We therefore combine data-driven covariate selection with theory-informed cause-and-effect paths to reduce the risk of missing confounders (Swanson, 2015; Steiner and Kim, 2016).

We estimate the impacts of AES participation on typical measures of land management (intensity and diversity), i.e. fertiliser and plant protection expenditures and grassland share, and apply the approach to data from the FADN for Western Germany and the first AES programming period 2000–2006.[1] Our procedure increased the size of the sample by 97% compared to a sample that only used farms entering in 2000. Covariate selection combined with cause-and-effect paths diagrams reduced the set of covariates by up to 59%. Post-matching covariate balance was achieved with the reduced set of covariates describing the production environment, farming intensity, productivity, and preferences. While only small and statistically insignificant effects on fertiliser expenditures could be attributed to AES, we found moderate reductions in plant protection expenditures and moderate increases in grassland shares.

2. Efficient AES evaluation: data set maximisation and covariate selection

2.1. The Rubin causal model and DID matching

We quantify the average treatment effect of the treated (ATT) of AES on farms' environmental outcome based on Rubin's causal model (Rubin, 1974). The causal effect of AES participation is defined as the difference between participants' outcome under AES and their counterfactual outcome, i.e. the outcome as if they had not participated. Using observational data from the FADN, we rely on the observational approach and apply statistical matching to identify a counterfactual group and the causal effect: By conditioning on observable covariates, capturing farm and farming characteristics, differences in outcomes can attributed to AES participation. Yet, any remaining unobserved (time-constant) determinants for AES participation and outcomes could bias the estimated effects (selection bias), e.g., farms could select into AES because they already operate at low intensities for longer and the programme is an easy source of remuneration (Defrancesco et al., 2008). A direct comparison of the outcomes between groups would then suggest an AES effect on farming intensity, whereas the difference is actually caused by the farming intensity prior to AES participation.

As mentioned, matching algorithms have been combined with difference-in-difference (DID) approaches to account for the observed and unobserved but time constant (permanent and transitory) differences between groups (Heckman et al., 1997). DID matching, however, requires differencing and conditioning on the set of covariates that determine AES participation and respective outcomes, to achieve *conditional ignorability* of the participation decision (Rubin, 1980; Rosenbaum and Rubin, 1983), i.e., that the *conditional mean independence* between the changes in potential land management outcomes and AES participation must hold.

To capture potential transitory confounders and to ensure the parallel trends assumption, conditioning on pre-treatment outcomes, has been proposed (Abadie, 2005). But combining DID matching with conditioning on pre-treatment outcomes potentially introduces bias (Chabé-Ferret, 2015, 2017; Daw and Hatfield, 2018) for two reasons.

[1] In 2000 an overhaul of the Common Agricultural Policy, known as "Agenda 2000", combined existing environmental programmes into a unified framework for the first time, establishing what came to be known as the "second pillar" of the CAP. Prior to the overhaul, the CAP promoted integrated cultivation practices, set-aside land, grassland conservation, organic farming and contracted nature protection measures (Osterburg and Stratmann, 2002). The German federal states used the Agenda 2000 to broaden measures to reduce fertiliser applications in water protection areas, promote environmentally oriented farm management, introduced extend set-aside options and measures to promote biodiversity on pasture land (Osterburg and Stratmann, 2002, p. 264). Some federal states had been offering programmes only for the farm branch-level, e.g., mulch seeding, drill row spacing, or abstaining from mineral fertiliser and chemical plant protection. These farm-level programmes had marginal adoption rates and eventually were replaced with programmes offering a modulated approach (Thomas et al., 2001). Based on the FADN, the average yearly share of AES participants in 2000–2006 increased to 51% compared to 32% in 1993–1999 (own calculations).

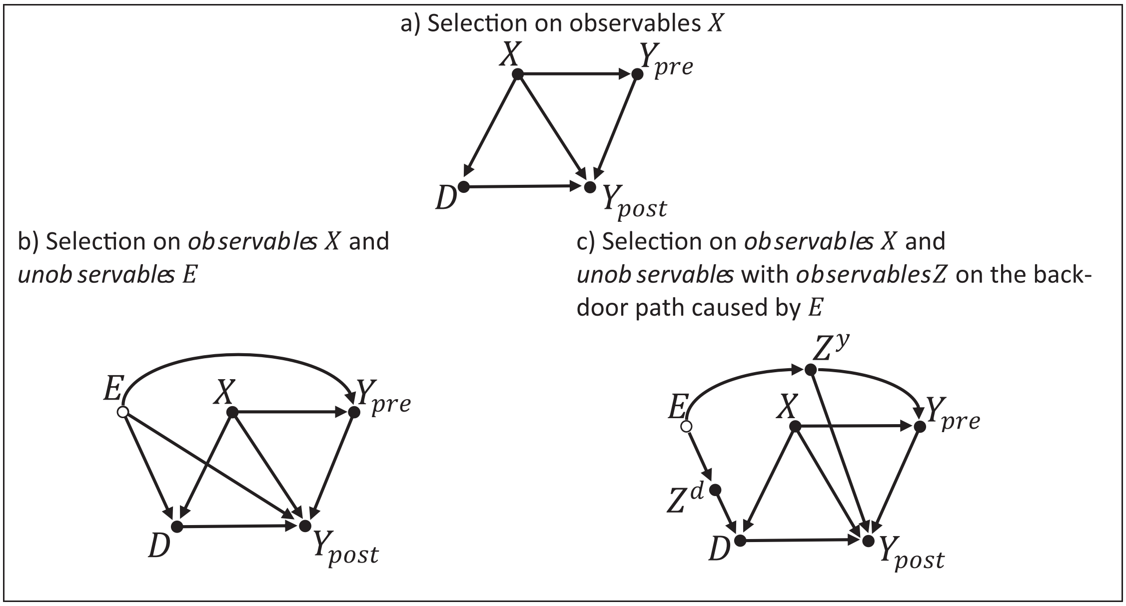

Fig. 1. Potential cause-and-effect paths relevant for AES evaluation. Note: Arrows represent direct causal relationships, nodes show the starting and ending points of the arrows, full nodes show the observable covariates or sets of covariates denoted by X or Z, respectively, hollow nodes show the unobserved ones denoted by E, D denotes the treatment, Y_{pre} and Y_{post} denote outcomes before and after treatment X denotes regional production, farming, farm(er) characteristics, E denotes environmental preferences, $Z^d$and Z^y denote the variables on the back-door path caused by E and related to treatment and outcome, respectively.

First, DID matching with conditioning on the pre-treatment outcome reduces to the matching estimator and is only consistent in the absence of a time-invariant unobserved confounder (Chabé-Ferret, 2017). Second, pre-treatment outcomes such as land management intensity could be affected by temporary or transitory shocks that are unrelated to AES participation. For example, temporary liquidity constraints in the pre-participation year could be a reason for lower fertiliser expenditure. Conditioning on this pre-treatment outcome selects a subgroup with reduced land management intensity into the control group. Due to the transitory nature of such a liquidity shock, these farms would recover and return to their "normal" level of land management intensity. In this case, matching on measures for land management intensity (pre-treatment outcome) violates the parallel trends assumption.

2.2. *Theory-informed covariate selection for matching*

Given the limited panel size in the FADN or comparable data sets for AES evaluation implies that using large sets of matching covariates could fail to obtain a post-matching covariate balance on the most relevant covariates. Theory-informed directed acyclic diagrams (DAGs) (Pearl, 2009) have been used to show the potential cause and effects between observed and unobserved covariates, AES participation (treatment assignment), and land management measures (outcomes). In Fig. 1, the straight arrow from the treatment assignment D to the outcome Y_{post} shows the causal effect of interest. Nodes show the starting and ending points of the arrows, full nodes show the observable covariates or sets of covariates denoted by X or Z, respectively, and hollow nodes show the unobserved ones denoted by E. The trajectories between D and Y_{post} with an arrow pointing to D and ending in Y_{post} show the non-causal associations between D and Y_{post} that denote the back-door paths (Pearl, 1995, definition 3). Conditional mean independence can be achieved if all back-door paths between D and Y_{post} are blocked. A back-door path can be blocked by conditioning on a covariate that lies on the back-door path (Pearl, 1995).

In the case of *selection on observables* (Fig. 1, panel a) back-door paths between D and Y_{post} emerge if the observable variables in X are directly related to Y_{post} and via pre-participation outcome Y_{pre}: $D \leftarrow X \rightarrow Y_{post}$ and $D \leftarrow X \rightarrow Y_{pre} \rightarrow Y_{post}$. Conditioning on the observable covariates in X blocks all back-door paths, and the effect is identified by (simple) matching on X.

Farms could decide for AES participation based on observables such as regional production, and farming and farm(er) characteristics (X) and on unobserved ones such as environmental preferences (Chabé-Ferret and Subervie, 2013; Leonhardt et al., 2021). Unobservables such as environmental preferences denoted by E are related to treatment and outcome separately from X (see Fig. 1 panels b and c). Assuming a time-invariant effect of E on the outcome before and after AES participation creates the back-door path

$$D \leftarrow E \rightarrow Y_{pre} \rightarrow Y_{post}.$$

Alternatively, if farmers with strong pro-environmental preferences increase their efforts for sustainable land management and consequently improve the respective outcome measure over time irrespective of AES participation, it creates the back-door path

$$D \leftarrow E \rightarrow Y_{post}.$$

To capture such effects, we consider that the observed covariates lying on the backdoor path stemming from E and related to the participation, Z^d, or the outcome, Z^y in the set of potential covariates (see Fig. 1, panel c). Conditioning on any of these variables blocks the back-door paths going through E. The covariates in Z^d could include participation in other nature conservation programmes, a previous AES, or in a green political party. The covariates in Z^y could include measures for sustainable land management practices or technologies, e.g., an agro-environmental index (Purvis et al., 2009).

We explicitly consider measures for $Z^y$and Z^d in the set of potential covariates (hereafter, the full set). The data-driven covariate selection procedure described below reduces the full set of available covariates to the set of covariates, jointly associated with treatment assignment and outcome for the specific question and data set (Vansteelandt et al., 2012).

2.3. *Data-driven covariate selection procedure*

Previous AES evaluation studies relied on treatment selection for the matching, where the covariate set was selected based on statistically significant covariates from a logit regression of the treatment indicator on farm characteristics (Pufahl and Weiss, 2009; Arata and Sckokai, 2016). This approach may include covariates that are strongly related to treatment but unrelated to outcome, which increases bias (Brookhart et al., 2006; Pearl, 2009) and inflates standard errors (Shortreed and Ertefaie, 2017). Thus, the latter authors recommend including all covariates associated with the outcome. Adjusting for all outcome related covariates, however, could result in achieving less covariate balance, compared to adjusting only for the confounding set.

To overcome this selection problem, de Luna et al. (2011) suggest finding the confounding set as the minimal subsets of covariates related to treatment and outcomes, such that the treatment and the potential outcomes are independent given these sets. Häggström et al. (2015) use non-parametric regression to find a minimal subset of covariates. In a simulation study, Persson et al. (2017) conclude that lasso regression has the highest success rate to select the subsets that uphold unconfoundedness when using smaller samples. Additionally, lasso offers a reasonable compromise between controlling false positives and discovering true positives (Wang et al., 2020). Therefore, we use two lasso regressions to find a minimal subset by the intersection set of covariates that predict AES participation and the outcome.

Minimal subset selection procedures typically used for matching approaches are distinct from covariate selection procedures used for treatment effect estimation based on (parametric) regression models. To select covariates for a regression that estimates a treatment effect, Belloni et al. (2014) suggest using the union set of covariates from two lasso regressions: one predicting the outcome, and the other the participation status, which is known as a double selection procedure. The rationale is to increase the robustness of the estimation of effects (regression coefficients) by mitigating the problem that lasso-based covariate selection excludes covariates that are weakly correlated with programme participation, or the outcome, or both. The double selection procedure, which tends to include non-confounding covariates, could prohibit achieving post matching covariate balance. To address this shortcoming, we use minimal subset selection, because the selection of fewer covariates improves post matching covariate balance (Abadie and Imbens, 2006; de Luna et al., 2011; Häggström et al., 2015; Persson et al., 2017; Abadie and Cattaneo, 2018). To prevent omitting confounding covariates, we complement the minimum subset selection with cause-and-effect paths using DAGs, and categorise the confounding factors based on theory to ensure including all covariates representing each category in the matching.

2.4. DID matching: efficient data use and matching algorithm

We begin by maximising the sample to include all farms that adopted AES during any year of the first AES programming period. To account for the staggered entry years, we combine Mahalanobis distance matching with exact matching on the respective entry year. Then, we apply kernel matching algorithms combined with optimal bandwidth selection (Loader, 1999; Galdo et al., 2008; Huber et al., 2015) using the R × quant-distance method (Huber et al., 2015) that sets the bandwidth to 1.5 times the 0.90-quantile of the distribution of (non-zero) distances between observations, and two variants based on cross-validation and weighted cross-validation, respectively (Galdo et al., 2008). For the kernel matching, we match on the Mahalanobis distance using the Epanechnikov kernel function. To demonstrate the performance of the kernel estimators we compare them with one-to-one and one-to-five pair matchings based on Mahalanobis distance.

3. Empirical evaluation of AES participation in 2000–2006 for Western Germany

3.1. Treatment, outcome, and covariate definition from the EU FADN

We test our approach on a sample of farms in Western Germany for the first period. Our aim is to identify the effect of AES after the usual five-year contract length, when a policy impact should have been realised. Considering the potential disadvantages of conditioning on the pre-treatment outcomes, we examine the sensitivity of the results when including pre-participation outcomes in the covariate sets.

In Western Germany, AES implementation consists of measures supporting the maintenance of grassland and/or reduced grassland use intensity, sustainable fertiliser application techniques, mulch, and no-till farming, extensive crop rotation, flower strips, and organic farming (Grajewski and Schmidt, 2015). Following previous studies we estimate the effect of AES participation on fertiliser and plant protection expenditures per hectare of farmland (see Arata and Sckokai, 2016 for five EU member states, Chabé-Ferret and Subervie, 2013 for France, and Pufahl and Weiss, 2009 for Germany). We interpret the input intensities as indirect indicators of a farm's sustainability in land management. e.g., a reduction in fertiliser expenditure could imply decreases in soil nutrient input and improvements in groundwater and surface water quality (e.g., Ulén et al., 2007). Reduced fertiliser and plant protection expenditures per hectare of utilised agricultural area (UAA) and increased grassland share could relate to overall lower intensities in land management that contribute to biodiversity (Billeter et al., 2008; Emmerson et al., 2016; Carmona et al., 2020).[2]

We structure the set of covariates related to farms' (dynamic) utility maximisation and the farm optimisation problem under environmental (biological and geophysical) production and farm capacity constraints (e.g., Bowman and Zilberman, 2013; Chabé-Ferret and Subervie, 2013). We characterise farms' *production environment* by farmland (total UAA) and region, because these measures relate to production constraints. To capture potential farm expansion limits, we include share of rented land, e.g., a high share of rented land could imply that a farm expanded and reached other capacity limits (Margarian, 2010) and because ownership affects land management decisions (Leonhardt et al., 2019). Related to input choice as a result of the optimisation problem, we characterise farms by *farming intensity* in several dimensions, including livestock and capital density, and *productivity,* defined as output over input measured by sales per ha, revenue per capital, revenue per labour input, and total output over total input. To acknowledge different *farming, risk* and *time preferences*, we also include type of farming and farmer's age. A diversified farm could be linked to farming preferences or risk management. Diversification could stabilise yields (Rosa-Schleich et al., 2019) or mitigate price risk. Younger farmers could be more likely to adopt AES (Villanueva et al., 2015).

To compare farms according to farmer's *environmental preferences*, we consider participation in environmental programmes prior to 2000 as a measure for Z^d to block the back-door paths caused by environmental preferences E (see DAG in Fig. 1 panel c). Effect identification using matching with conditioning on participation in agri-environmental programmes prior to AES, however, requires farms to change their participation status between programming periods (switchers). If the majority of farms participating in 2000–2006 also participated in a previous programme, effect identification relies on the limited number of farms entering or exiting in 2000–2006. Alternatively, we consider sustainable land management practices as a measure for Z^y. The practices could result from pro-environmental preferences and block the back-door paths going through E (see DAG in Fig. 1 panel b and c). We use the Shannon diversity index as a measure of crop diversity (Areal et al., 2018). For the list of covariates, see Table 1: Farm characteristics.

[2] Studies of other AES impacts include farm income (Udagawa et al., 2014), farm productivity (Mennig and Sauer, 2019), quantity of environmentally relevant input use (Cisilino et al., 2019), and farming structure described by crop diversity and proportion of organically managed cropland (Bertoni et al., 2020).

Table 1
Farm characteristics.

	Starting year of participation in AES					
	2000		2001		2002	
Participation	No	Yes	No	Yes	No	Yes
Production environment						
Land input (ha)	60	54	65	80	64	59
Region						
South	0.06	0.80	0.06	0.25	0.09	0.58
West	0.18	0.17	0.21	0.63	0.21	0.25
North	0.76	0.03	0.73	0.12	0.70	0.17
Proportion of rented land	0.50	0.57	0.51	0.63	0.52	0.58
LFA subsidies (yes/no)	–	–	0.12	0.47	0.13	0.42
Farming intensity						
Proportion of grassland area	0.38	0.39	0.37	0.3	0.34	0.31
Proportion of cereals area	0.32	0.34	0.34	0.43	0.36	0.38
Direct payments crops (€/ha)	161	171	166	214	172	197
Direct payments livestock (€/ha)	30	17	50	33	81	67
LU cattle per ha	1.11	0.94	1.08	0.68	1.07	0.97
LU pigs and poultry per ha	1.01	0.65	1.00	0.76	1.08	0.90
Fertiliser exp. per ha	88	74	104	88	113	108
Plant protection exp. per ha	69	70	73	79	82	83
Fixed capital per ha (1000€)	13.32	13.85	13.41	10.99	13.64	13.36
Productivity						
Sales per ha (1000€)	2.79	2.45	3.04	2.43	3.50	2.85
Revenue per AWU (1000€)	91	64	99	88	102	85
Revenue per capital (€)	0.51	0.29	0.61	0.49	0.47	0.30
Total output over total input	1.17	1.10	1.20	1.09	1.15	1.11
Preferences						
Age of farmer	45	45	45	45	45	47
Farm type						
Crop	0.22	0.19	0.25	0.30	0.27	0.29
Livestock	0.50	0.46	0.48	0.32	0.48	0.44
Livestock crop mixed	0.28	0.34	0.27	0.37	0.26	0.27
Participation in AES before 2000						
Yes	0.05	0.48	0.04	0.25	0.05	0.29
No	0.95	0.52	0.89	0.68	0.83	0.42
Unknown	0.00	0.00	0.07	0.06	0.13	0.29
Shannon diversity index	0.96	1.19	0.99	1.32	1.01	1.23
N	458	740	491	139	482	52

Notes: LU denotes livestock units and ha denotes hectare (total UAA).
Source: FADN.

3.2. Data set

Seventy-one percent of participating farms in the first programming period 2000–2006 adopted AES between 2000 and 2002 (see Appendix Table A.I). We use farms that entered AES in one of these years (three entry cohorts). Using data for the pre-participation years (1999, 2000, or 2001, respectively) and for the five years after entering AES with continuous participation or non-participation yields a sample of 937 treated and 1431 untreated farms.

We exclude farms that switched treatment state from the control group, to avoid comparing early with late adopters, which would not yield insights into the effect of participation.[3] We also exclude organic farms because their participation rate was almost 100%, and horticulture, vineyard, and permanent crop farms because their participation rate was very low.

Table 1 reports that farms show a higher proportion of rented land, but lower livestock density, sales per hectare, lower labour and capital productivity, and fertiliser expenditure, which indicate an overall lower production intensity before participation. Higher values for the Shannon diversity index indicate more crop diversity. Participation rates in Southern Germany are higher than Northern Germany. On average, participating farms receive payments for less favoured areas (LFA) compared to non-participating farms.

Forty-eight percent of farms that started participation in 2000 had participated during the 1990s (minimum two years of participation). Due to panel attrition, the share of farms for which pre-participation status is unknown increased to 29%. Matching on the pre-participation status reduced the available control group to farms participating in 1990s and stopping AES participation later. With only 5% of non-participating farms in the data set, matching on the pre-participation status limits identification of the causal effect,[4] which is why we use the Shannon diversity index to block the back-door paths going through *E*.

The large share of farms already participating in the earlier programmes introduces another challenge for effect identification. While the AES reform in 2000 brought major changes (see footnote 1), some federal states had already implemented similar programme variants. In these states, farms participating before 2000 and continued participating after Agenda 2000 possibly incurred fewer changes in farming operation than the farms participating only after Agenda 2000, i.e., the reform could have had a higher impact on the outcome measures of farms that did not participate before 2000. Therefore, we examine whether impacts of AES participation differ for the farms that did not participate in AES prior to 2000 and estimate the treatment effects for this subgroup of newcomers separately.

3.3. Covariate selection based on lasso

Table 2 lists the results of the covariate selection procedure for the outcome measures of fertiliser and plant protection expenditures per ha UAA and grassland share. The lasso procedure does not select the pre-participation status (participation before 2000) into the set of confounding variables. Determining the relevance of the pre-participation status, however, could suffer from weak identification because only a small number of farms exited the AES program after 2000 ($n = 23$ for year 2000). Again, we use the Shannon diversity index to block the back-door paths going through *E*.

For the outcome fertiliser expenses, covariate selection based on lasso (lasso-selected set) indicates region, proportion of grassland area, pigs and poultry (LU) per ha UAA, fertiliser expenditure, plant protection expenditure, sales per ha, total output over total input, farmer's age, farm type, and interaction between farm size and direct payments. For the outcome plant protection expenditure, the lasso set includes proportion of rented land, reception of LFA subsidies, and interaction effects with farm size, such as grassland share, livestock density, fixed capital, revenue per capital age and the Shannon diversity index. For the outcome grassland share, the lasso set also includes the Shannon diversity index. Other covariates from the lasso-selected set are region, farm size, LFA participation, proportion of grassland and cereal area, cattle density, plant protection expenditure, fixed capital, sales per ha, revenue per capital, and the interaction between farm size and direct payments for livestock.

To evaluate the matching procedures in terms of the post-matching covariate balance, we compare the standardised differences in means before and after matching (Rosenbaum and Rubin, 1985; Imai et al., 2008; Kainz et al., 2017). Guidelines for what constitutes an acceptable standardised difference for a given covariate lie between 0.1 and 0.25,

[3] To represent the counterfactual, Kuhfuss and Subervie (2018) suggested using farms that did not participate in 2000–2006 and participated later, but we would need to extend the minimum panel length to 9 years. Such panel lengths are not widely available for the German FADN. Our data set contains 12 farms which did not participate between 1999 and 2006 and only began participating in 2007.

[4] We thank an anonymous reviewer for this observation.

Table 2
Results of lasso covariate selection.

Outcome	Fertiliser exp.		Plant protection exp.		Grassland share	
	Participation	Outcome	Participation	Outcome	Participation	Outcome
Production environment						
Land input (ha)					**x**	**x**
Region south	x		**x**	**x**	**x**	**x**
Region north	**x**	**x**		x	**x**	**x**
Proportion of rented land	x		**x**	**x**	x	
LFA subsidies (yes/no)	x		**x**	**x**	**x**	**x**
Farming intensity						
Proportion of grassland	**x**	**x**	**x**	**x**	**x**	**x**
Proportion of cereals area	x		**x**	**x**	**x**	**x**
Direct payments crops per ha	x		**x**	**x**		
Direct payments livestock per ha				x		x
LU cattle per ha	x		**x**	**x**	**x**	**x**
LU pigs and poultry per ha	**x**	**x**	**x**	**x**		
Fertiliser exp. per ha	**x**	**x**	**x**	**x**		
Plant protection exp. per ha	**x**	**x**	**x**	**x**	**x**	**x**
Fixed capital per ha (1000€)	x		**x**	**x**	**x**	**x**
Productivity						
Sales per ha (1000€)	**x**	**x**	x		**x**	**x**
Revenue per AWU (1000€)	x		x		x	
Revenue per capital (€)	x		x		**x**	**x**
Total output over total input	**x**	**x**	**x**	**x**	x	
Preferences						
Age of farmer	**x**	**x**	x		x	
Farm type crop		x		x		x
Farm type livestock	x		**x**	**x**	**x**	**x**
Participation before 2000	x		x		x	
Shannon diversity index	x		x		**x**	**x**
Scale effects						
Land input^2	x		x		x	
Land input × Prop. grassland	x		**x**	**x**		x
Land input × Prop. rented land	x		x		x	
Land input × Prop. cereal area	x		x			
Land input × Direct payments livestock	**x**	**x**	x		**x**	**x**
Land input × Cattle (LU/ha)	x		x		x	
Land input × pigs and poultry (LU/ha)			**x**	**x**	x	
Land input × Sales per ha				x		x
Land input × Plant protection exp.	x		x		x	
Land input × Fixed capital			**x**	**x**		
Land input × Revenue per capital			**x**	**x**	x	
Land input × Revenue per working unit				x		x
Land input × Age			**x**	**x**		
Land input × Shannon diversity index	x		**x**	**x**	x	

Notes: LU denotes livestock units and ha denotes hectare (total UAA); boldface denotes common set covariate selections.

depending on the context (Harder et al., 2010). We consider a standardised difference of below 20% as balanced. Table 3 lists the standardised differences for the full covariate set and the lasso-selected set.

For the outcome fertiliser expenditure, standardised differences tend to decrease after matching, but could also increase for single covariates, i.e., grassland share. For the full covariate set, notable differences for region, rented land, cattle density, and productivity remain after matching. For the lasso-selected set, the nearest neighbour matching variants and kernel matching[5] based on cross-validation with respect to the means of covariates X performs best in reducing covariate imbalance. Mean standardised differences (absolute) of all covariates in these matching variants with reduced set could drop below the threshold value of 20%. For the cross-validation with respect to the outcome, the mean standardised difference (absolute) over all covariates higher than for the other variants (0.17), which show mean values from 0.04 to 0.09. Additionally, standardised differences in the two covariates, region and productivity, remain above the threshold level of 0.2.

For the outcome plant protection expenditure, the full set does not achieve covariate balance. For the lasso-selected set (Table 2 columns 4 and 5), covariate balance is achieved for the nearest neighbour matching variants (Table 4). For these variants the mean standardised difference (absolute) reduces to 0.06 (1:1 NN matching) and 0.09 (1:5 NN matching), and no selected covariates have a standardised difference after matching of above 0.2.

For the outcome measure proportion of grassland, we present results with exact matching on region as this improves covariate balance. For the full set, covariate balance could not be achieved and for the reduced set, imbalance remains in the covariates farm size, capital per ha, and cattle density (see Appendix Table A.II). The 1:1 nearest neighbour matching achieves the best covariate balance with a mean difference of 0.09 and two covariates (farm size and fixed capital per ha) above the threshold of 0.2.

3.4. Average treatment effects on the treated

DID relies on parallel trends, but testing this assumption is debatable because the potential outcomes are unobserved (Roth, 2018; Freyaldenhoven et al., 2019) and a careful assessment of the data is required. A

[5] Kernel matching estimators were calculated with the Stata user-written program kmatch (Jann, 2017).

Table 3
Covariate balance in terms of standardised differences using the outcome measure "fertiliser expenditure".

		NN matching		Kernel matching Cross-validation w.r.t.		
	Pre-matching	1:1 NN	1:5 NN	Quantile dist. method	Means of covariate sets X	Means of outcomes Y
Full covariate set						
Production environment						
Land input (ha)	-0.12	0.13	0.06	-0.07	0.14	0.00
Region west	0.09	0.07	0.08	0.13	0.00	0.04
Region north	**-1.93**	**-0.76**	**-1.15**	**-1.9**	**-0.38**	**-1.26**
Proportion of rented land	**0.23**	**0.32**	**0.31**	**0.26**	**0.27**	**0.28**
LFA subsidies (yes/no)	0.03	0.00	0.01	0.15	0.00	0.00
Farming intensity						
Proportion of grassland area	0.02	-0.04	-0.13	**-0.28**	0.00	-0.13
Proportion of cereals area	0.07	0.06	0.12	**0.32**	0.04	0.14
Direct payments crops (€/ha)	0.12	0.05	0.13	**0.31**	0.05	0.15
Direct payments livestock (€/ha)	**-0.36**	-0.05	-0.02	-0.09	-0.08	-0.07
LU cattle per ha	**-0.22**	**-0.24**	**-0.28**	**-0.35**	**-0.24**	**-0.23**
LU pigs and poultry per ha	-0.2	-0.02	-0.01	0.01	-0.06	-0.07
Fertiliser exp. per ha[1]	**-0.27**	-0.09	-0.08	-0.14	-0.15	-0.12
Plant protection exp. per ha	-0.03	0.09	0.15	0.13	0.01	0.05
Fixed capital per ha (1000€)	-0.01	-0.2	-0.12	0.02	**-0.23**	-0.1
Productivity						
Sales per ha (1000€)	-0.20	-0.04	-0.02	-0.05	-0.09	-0.08
Revenue per AWU (1000€)	**-0.57**	-0.11	**-0.21**	**-0.38**	-0.1	**-0.27**
Revenue per capital (€)	-0.15	0.06	0.06	-0.03	0.02	0.00
Total output over total input	**-0.28**	-0.17	**-0.24**	**-0.36**	**-0.23**	**-0.29**
Preferences						
Age of farmer	0.03	0.07	0.03	0.03	0.03	0.02
Farm type livestock	-0.09	-0.07	-0.15	**-0.28**	-0.01	-0.14
Farm type mixed	0.17	0.02	0.08	**0.25**	0.01	0.1
Shannon diversity index	**0.45**	0.13	**0.24**	**0.54**	0.08	**0.3**
Mean standardised difference	0.26	0.13	0.17	0.28	0.10	0.17
Lasso common covariate set						
Production environment						
Region north	**-1.93**	-0.04	-0.09	**-0.27**	0.00	**-1.31**
Farming intensity						
Proportion of grassland area	0.02	-0.08	-0.13	-0.13	-0.11	-0.07
Pigs and poultry (LU) per ha	-0.20	-0.01	0.01	-0.07	-0.04	-0.08
Fertiliser exp. per ha[1]	**-0.27**	-0.08	-0.09	-0.13	-0.11	-0.17
Plant protection exp. per ha	-0.03	0.08	0.14	0.07	0.06	0.01
Productivity						
Sales per ha (1000€)	-0.2	-0.06	-0.08	-0.13	-0.11	-0.1
Total output over total input	**-0.28**	-0.07	-0.11	-0.19	-0.12	**-0.24**
Preferences						
Age of farmer	0.03	0.04	0.04	0.04	0.03	0.06
Scale effects (Farm size × …)						
Direct payments livestock	**-0.32**	0.01	-0.02	-0.02	-0.02	-0.04
Mean standardised difference	0.36	0.04	0.06	0.09	0.05	0.17

Notes: Exact matching on entry year, [1] indicates exclusion of pre-participation outcome from the matching set; boldface denotes standardised differences of larger than 0.2 (absolute).

comparison of trends in fertiliser expenditure between participants in 2000 and non-participants before 2000 reaffirms our use of DID matching (see Appendix Fig. A.I).

Table 5 shows the estimated average treatment effect of the treated (ATT) for all outcome measures. The 1:1 nearest neighbour estimators use only 321 out of 1417 observations available from the non-participants, when including the pre-participation outcome, whereas the kernel matching algorithms use almost all of the observations.

For the outcome fertiliser expenditures, the ATT estimates vary with the matching algorithms and sets of covariates. For example, column 4 shows that when matching on the full set of covariates all of the matching variants except kernel-matching based on bandwidth selection with respect to means of covariates *X* yield statistically significant estimates ($p < 0.1$). For these kernel-matching algorithms, the standardised differences remain higher than the threshold value of 0.2 for several covariates including those such as region and productivity selected into the confounding set by lasso (see Table 3). In contrast to the full set, matching on the reduced set achieves covariate balance for the nearest neighbour matching variants and the cross-validation with respect to the means of covariates *X*. These variants cannot reject a zero impact of AES participation on fertiliser expenditure.

For the outcome plant protection expenditure, ATT estimates are generally lower than the naïve DID estimate of -6.05 ($p < 0.05$). Both 1:1 and 1:5 nearest neighbour matching, which achieve covariate balance on the lasso-selected set (Table 4, columns 3 and 4) suggest an effect of AES participation in reducing plant protection expenditure of 5–6€/ha.

For the outcome grassland share, 1:1 nearest neighbour matching and kernel matching with bandwidth selection based on cross-validation with respect to the means of covariates *X* performs best in terms of the standardised differences (see Appendix Table A.II). The kernel matching estimates suggest an increase of 2.1% points in the grassland share, although the covariates of farm size, capital, and cattle density remain imbalanced.

For the pre-treatment outcome fertiliser expenditure, the imbalance is 0.27 in terms of the standardised differences. Including the pre-treatment outcome in the matching set yields equivalent results for both variants (see Appendix Table A.IV).

Given the discussed limitations of matching on participation in a predecessor programme and the possible differing treatment effects for farms that did not participate in AES prior to 2000, we also estimate ATTs for this group of newcomers (Table 6). For fertiliser expenditure

Table 4
Covariate balance in terms of standardised differences using the outcome measure “plant protection expenditure”.

		NN matching		Kernel matching Cross-validation w.r.t.		
	Pre-matching	1:1 NN	1:5 NN	Quantile dist. method	Means of covariate sets X	Means of outcomes Y
Lasso common covariate set						
Production environment						
Region south	0.09	0.11	0.14	0.15	0.02	0.1
Proportion of rented land	**0.23**	0.14	0.17	**0.24**	0.14	0.2
LFA subsidies (yes/no)	0.03	0.02	0.03	0.14	0.00	0.01
Farming intensity						
Proportion of grassland area	0.02	-0.02	-0.07	**-0.24**	-0.05	**-0.21**
Proportion of cereals area	0.07	0.05	0.07	**0.29**	0.08	**0.24**
Direct payments crops (€/ha)	0.12	0.05	0.09	**0.29**	0.09	**0.24**
LU cattle per ha	**-0.22**	-0.13	-0.17	**-0.35**	-0.13	**-0.28**
LU pigs and poultry per ha	-0.2	-0.03	-0.03	-0.02	-0.08	-0.02
Fertiliser exp. per ha	**-0.27**	-0.07	-0.08	-0.13	-0.13	-0.11
Plant protection exp. per ha[1]	-0.03	0.06	0.08	0.12	0.00	0.08
Fixed capital per ha (1000€)	-0.01	0.08	0.07	0.04	0.08	0.07
Productivity						
Total output over total input	**-0.28**	-0.14	-0.17	**-0.31**	**-0.22**	**-0.28**
Preferences						
Farm type livestock	-0.09	-0.06	-0.09	**-0.24**	0.00	**-0.24**
Farm type mixed	0.17	0.01	0.03	0.2	0.00	0.2
Scale effects (Farm size × …)						
Age	-0.11	0.06	0.08	0.04	0.00	0.04
Grassland share	-0.06	-0.01	-0.03	**-0.22**	-0.09	-0.19
Pigs and poultry (LU/ha)	-0.18	-0.03	-0.03	0.00	-0.07	-0.02
Revenue per capital	-0.16	0.01	0.05	0.00	-0.01	-0.01
Capital per ha	**-0.3**	-0.02	-0.04	-0.14	-0.04	-0.07
Shannon diversity index	0.18	0.12	0.19	**0.29**	0.11	**0.23**
Mean standardised difference	0.14	0.06	0.09	0.17	0.07	0.14

Notes: Exact matching on entry year, [1] indicates exclusion of pre-participation outcome from the matching set; boldface denotes standardised differences of larger than 0.2 (absolute).

the naïve DID estimates slightly increase to − 7.50 ($p < 0.01$), for plant protection expenditure to − 7.20 ($p < 0.05$), and for grassland share to 0.025 ($p < 0.01$). The matching results and covariate balance are equivalent to the results when using the complete sample (see Appendix Table A.III).

4. Discussion

4.1. Benefits of the proposed approach

The challenges of using large covariate sets to satisfy the conditional independence assumption could be overcome by using a method that uses the data efficiently and optimises the covariate selection, to enhance the accuracy and reliability of an impact analysis. To maximise the utilised sample size, we use staggered entry dates and include all farm participants entering in any year of our AES programming period. Compared to a sample that includes only farms entering in the first year (year 2000), our method achieves a 97% increase of the utilised sample. Using kernel matching, we could include up to 1400 non-participating farms and achieve post-matching balances similar or better than using pair matching, which achieves less than 200 matched controls depending on the outcome measure.

To optimise covariate selection, we combine theory-informed cause-and-effect paths with data-driven covariate selection based on lasso. We

Table 5
ATT estimates: changes in outcome measures.

Outcome measure	Fertiliser expenditure			Plant protection expenditure			Grassland share		
Covariate selection/ matching estimator	Matched treated	Matched controls	ATT in €/ha	Matched treated	Matched controls	ATT in €/ha	Matched treated	Matched controls	ATT (Δ share)
Full set of covariates									
NN Mahalanobis matching									
1:1 NN matching	928	321	-7.47* (4.37)	928	326	-6.34* (3.68)	928	190	.017** (0.007)
1:5 NN matching	928	776	-6.37** (3.15)	928	778	-5.51* (2.60)	928	477	.025** (0.009)
Kernel matching									
R × quant-dist. method[a])	908	1369	-3.77* (2.08)	909	1364	-5.10*** (2.08)	905	1375	.023*** (0.008)
Cross-validation w.r.t. X	517	506	-6.82 (5.41)	534	496	-6.78* (3.24)	722	866	.021** (0.009)
Weighted cross-validation w.r.t. Y[b])	826	1178	-4.15* (2.37)	897	1297	-5.43*** (1.86)	928	1413	.019** (0.007)
Lasso covariate selection set									
NN Mahalanobis matching									
1:1 NN matching	928	240	-3.15 (4.17)	928	350	-5.28* (2.70)	931	187	.015(0.009)
1:5 NN matching	928	559	-3.22 (3.50)	928	782	-5.89* (2.40)	931	457	.023*** (0.009)
Kernel matching									
R × quant-dist. method[a]	901	1298	-1.52 (3.01)	908	1353	-5.59*** (1.81)	905	1366	.021** (0.009)
Cross-validation w.r.t. X	848	994	-2.74 (3.15)	796	939	-6.69*** (2.10)	796	974	.021** (0.009)
Weighted cross-validation w.r.t. Y[b]	916	1379	-3.78* (2.22)	908	1353	-5.59*** (1.92)	931	1427	.019** (0.008)
N	928	1417		928	1417		928	1417	

Notes: Exact matching on entry year, excluding pre-treatment outcome from matching set; standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; a) 1.5 × 90% quantile, Huber et al. (2015); b) Galdo et al. (2008); standard errors for NN matching based on Abadie and Imbens (2006) and for kernel matching bootstrapped with 1000 replications; naïve DID: fertiliser expenditure − 6.05 (p < 0.05), plant protection expenditure − 6.00 (p < 0.01), and grassland share 0.019 (p < 0.05).

identify the relevant covariate set by the common set of variables selected by a lasso regression on the treatment and a lasso regression on the outcome. Based on this data-driven covariate selection procedure, we could reduce the full set of covariates by up to 59%, depending on the outcome measure.

Across all outcome measures, the lasso-selected set of covariates includes our measures from each category of covariates: region as a measure for *production environment*; pigs and poultry density, plant protection expenditure, and sales per ha as measures for *farm intensity*; total output over total input as a measure for *productivity*; farm type and farmer's age for the outcome fertiliser expenditure, and the Shannon crop diversity index for the outcome grassland share as measures for *preferences*.

In line with previous studies, we suggest a relation between *farming intensity* measures (prior to participation) and AES adoption (Zimmermann and Britz, 2016) and outcome measures (Billeter et al., 2008). Previous studies, however, have reported mixed results for a relation between *productivity* measures and AES adoption. Pufahl and Weiss (2009) and Salhofer and Streicher (2005) found no correlation between *productivity* and AES adoption for Germany and Austria, respectively. Wąs et al. (2021) reported statistically significant correlations between farm productivity and AES participation for Poland, and based on theories of cultural capital, viewed productivity as an expression of landscape preferences. Such preferences could inhibit participation because AES could imply less desirable landscape changes. Our result supports Chabé-Ferret and Subervie (2013), who found windfall effects resulting from adverse selection, and Pates and Hendricks (2020), who emphasised that AES could offer substantial private benefits. We note that observed productivity also is the result of the farm optimisation problem and depends on technology and input choices under production uncertainty (Chambers and Pieralli, 2020), with possible time-dependencies (Ahn et al., 2000). Thus, current productivity could simultaneously determine farmer's technology choice, such as AES adoption, and relate to environmental outcome measures (Serra et al., 2014; Baldoni et al., 2017).

The lasso regressions also select farmer's age, and the Shannon diversity index in case of the outcome grassland share. Since both variables could reflect farmers' experience and preferences, their relation to outcome measures seems plausible, which is in line with studies that considered behavioural aspects, such as attitudes, perceived norms, and farmer types as predictors for adopting sustainable farming practices and AES (Dessart et al., 2019; Villamayor-Tomas et al., 2019; Leonhardt et al., 2021).

The lasso regressions show a relation between participation in environmental programmes prior to 2000 and participation in AES 2000–2006, but no relation between earlier participation and the outcome measures. Although some studies noted that prior participation is an important predictor for AES adoption (e.g., see overview in Lastra-Bravo et al., 2015), rarely has it been considered in impact evaluation studies. One exception is Chabé-Ferret and Subervie (2013), who discussed experience with prior participation as an important confounder for evaluating AES. We find no difference between the results from separately estimating the effects for newcomers and for farmers with long-standing participation and the results based on the full sample.

For the outcome grassland share, the lasso-selected covariate set includes the Shannon diversity index as a measure for sustainable land management practices and is in line with Fig. 1 mechanism c. This supports our assumption that farm sustainability measures can block the backdoor-path through environmental preferences.

Based on the existence of minimal subsets (de Luna et al., 2011), the applied covariate selection approach of including the relevant covariate set differs from covariate selection approaches proposed for parametric regression that rely on a correct model specification. For example, the

Table 6
ATT estimates: changes in outcome measures for farms that do not participate in environmental programmes prior to AES (1992–1999).

Outcome measure	Fertiliser expenditure			Plant protection expenditure			Grassland share		
Covariate selection/ matching estimator	Matched treated	Matched controls	ATT in €/ha	Matched treated	Matched controls	ATT in €/ha	Matched treated	Matched controls	ATT (Δ share)
Full set of covariates									
NN Mahalanobis matching									
1:1 NN matching	524	220	-5.59 (5.35)	524	229	-5.78(4.22)	524	146	.025** (0.007)
1:5 NN matching	524	653	-5.95 (3.68)	524	658	-5.76* (3.32)	524	391	.032*** (0.011)
Kernel matching									
R × quant-dist. method[a])	513	1306	-5.54* (2.35)	512	1299	-4.53** (2.17)	513	1317	.028*** (0.010)
Cross-validation w.r.t. X	338	556	-4.69 (6.23)	351	572	-3.16(3.02)	399	796	.030*** (0.010)
Weighted cross-validation w.r.t. Y[b])	457	932	-4.22 (3.22)	392	664	-2.65(2.78)	523	1350	.025*** (0.009)
Lasso covariate selection set									
NN Mahalanobis matching									
1:1 NN matching	524	192	-3.33 (4.75)	524	259	-6.59** (3.13)	525	135	.021* (0.013)
1:5 NN matching	524	484	-4.33 (4.01)	524	672	-6.53** (2.40)	525	357	.032*** (0.012)
Kernel matching									
R × quant-dist. method[a])	502	1249	-2.48 (3.24)	519	1284	-5.26** (2.20)	506	1315	.028** (0.011)
Cross-validation w.r.t. X	463	851	-3.73 (3.76)	455	844	-5.64*** (2.16)	403	784	.027** (0.013)
Weighted cross-validation w.r.t. Y[b])	460	823	-4.50 (3.71)	444	819	-5.11** (2.11)	525	1365	.025*** (0.009
N	524	1354		524	1354		525	1365	

Notes: Exact matching on entry year, excluding pre-treatment outcome from matching set; standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; a) 1.5 × 90% quantile, Huber et al. (2015); b) Galdo et al. (2008); standard errors for NN matching based on Abadie and Imbens (2006) and for kernel matching bootstrapped with 1000 replications; naïve DID: fertiliser expenditure − 7.50 ($p < 0.01$), plant protection expenditure − 7.20 ($p < 0.05$), and grassland share 0.025 ($p < 0.01$).

double selection approach by Belloni et al. (2014) uses the union of covariates, because the unbiasedness of the coefficient of interest requires inclusion of covariates with small coefficients, which lasso will discard. Compared to matching on the covariates, the sparsity condition for these treatment effect estimators is relaxed but they are sensitive to poor overlap (D'Amour et al., 2021).

Using the union of covariates, however, was not feasible for our data set because the union excludes only two covariates and six out of 14 scale effects, so achieving post-matching covariate balance is not possible. Based on the assumption that not all covariates in this set are needed to establish conditional mean independence, the lasso-reduced set could be an option for supporting the exclusion of non-confounding covariates. To avoid omitting confounding covariates, we classify the covariate set along the relevant dimensions for the farms' optimisation problem and use theory-informed cause-and-effects paths. Our lasso-reduced set includes covariates from all of these dimensions, which seems reasonable from an economics perspective.

4.2. Estimated average AES effects in the programming period 2000–2006

For the estimated ATT, we show that AES reduces plant protection expenditure by up to 6€ per ha and increases grassland share by 2.1% points. With a baseline mean plant protection expenditure of 72.33€/ha for the treated group, there is an 8.3% reduction in plant protection expenditure. For fertiliser expenditure, we cannot show an effect. We note that kernel matching supports the credibility of the results because the low number of observations used in pair matching could underpower the hypothesis test and falsely suggest no effect of AES participation.

Our results differ from studies of the same AES programming period. Based on farms' bookkeeping data comparable to the FADN for 2000–2005 and DID matching, Pufahl and Weiss (2009) reported an average reduction of 9.4% in fertiliser expenditure, a 4.7% reduction in plant protection expenditure, and an increase in grassland share of almost 9%. Arata and Sckokai (2016), who also used the FADN but constructed a fully balanced panel for 2003–2006, found that AES participants reduce fertiliser expenditure by €33/ha, based on a subsample in which the AES payments comprised more than 5% of the total farm income; the authors also found no statistically significant effect on plant protection expenditure. We attribute the different results to the construction of the treated and control groups. Arata and Sckokai (2016) restricted their data set to farms adopting AES in 2003, whereas Pufahl and Weiss (2009) excluded about 80% of observations to ensure covariate balance. The sample used in Pufahl and Weiss (2009) is also larger and more representative for Northern Germany than the FADN sample.

4.3. Data limitations

While FADN is considered the best available data set for EU agri-environmental policy impact evaluation, it lacked data on farms' participation in specific AES programs. Thus, some programmes in our study period do not necessarily relate to the available intensity measures. Additionally, our data set could include participants for which no change in the outcome could be expected after AES adoption. For instance, participation in the "late mowing of grassland" programme does not necessarily relate to changes in a farm's grassland share, i.e., by including this programme's participants, we could have skewed the effects of other programmes. From a policy perspective, we suggest that an average effect across all programmes on the outcome measure of interest still is a meaningful success indicator for the policy.

The FADN contained only indirect environmental outcome measures, such as input intensity. Therefore, we had to rely on fertiliser expenditure because fertiliser quantities are available only for recent years. Today, more studies are using the Integrated Administrative Control System (IACS) to develop alternative environmental measures (e.g., crop diversity at the farm level (Uthes et al., 2020)) as outcomes indicators because they apply to AES outcomes such as farming intensity as well as landscape diversity and climate change mitigation (Bertoni et al., 2020).

We note that some EU member states collect detailed data on the environmental performance of agricultural land use and combine the data with national farm accountancy data sets (Dabkienė, 2016; Kelly et al., 2018). Buckley et al. (2015) derived farm-gate N and P balances from the Irish National Farm Survey and benchmarked different farming systems against each other. Thus, expanding the FADN could help policy-makers develop alternative outcome measures.

The FADN's rotating panel is a significant limitation. Given a typical AES contract length of five years, the effect of interest will be the impact after five years of participation. In Section 2.4, we stated that a panel length of six years could reduce the sample size considerably. For EU member states with very low or high participation rates (e.g., 77% of all farms in Austria participate in AES (Eurostat, 2017); for a detailed overview, see Zimmermann and Britz, 2016). Rotating panels also reduce the possible number of observations in the treated and control groups.

5. Conclusion

This study addressed the problem of having more covariates that improve the control of confounding factors but reduce both the available data for matching and the precision of the estimates. The proposed approach was designed to increase the efficiency and reduce potential bias of causal impact estimates for AES participation effects on farms' land management using the FADN between 1999 and 2006. The approach maximised the utilised sample while reducing the covariate set based on cause-and-effect paths and automated covariate selection procedures and thus achieves results that are more representative of the complete AES programming period. The results indicated that the approach could increase the robustness of causal impact estimation. Moreover, when combined with improved outcome measures, the approach has potential to improve both the counterfactual analyses within the EU's method mix for AES evaluation and the development of environmental programmes. We suggest that future research could examine other approaches for covariate selection that use parametric (Chernozhukov et al., 2018) and non-parametric regressions (de Luna et al., 2011; Persson et al., 2017).

Acknowledgements

The data used in this study was made available within the EU's 7th Research Framework Program as part of the project "Factor Markets" Grant agreement N°: 245123-FP7-KBBE-2009-3.

Appendix A

See Fig A1.
See Tables AI–AIV.

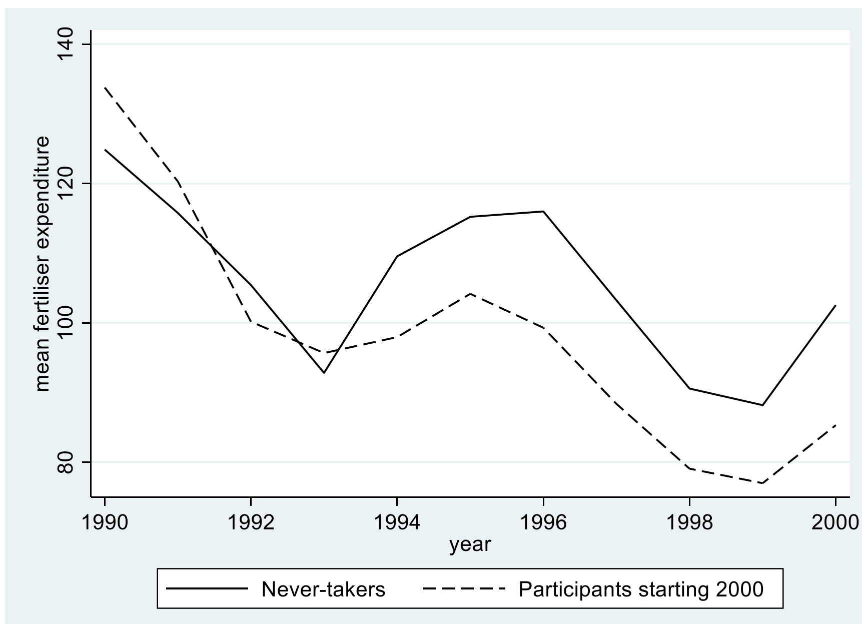

Figure A.I. Pre-treatment outcome trend. Notes: For 2000, 425 observations are denoted as never-takers (5 years of non-participation) and 651 observations as participants starting in 2000 (5 years of continuous participation); sample size reduces with every additional year before 2000; for 1990, only 116 treated and 119 untreated observations are available.

Table A.I
AES adoption by year in Western Germany.

	Number of farms		Number of farms with six years of observations	
Year	Non-participants	Participants entering AES	Non-participants (no switchers)	Participants
1999	4216	–	458	740
2000	2769	1482	491	139
2001	1676	1562	482	52
2002	1643	677	–	–
2003	1651	441	–	–
2004	1567	397	–	–
2005	2502	243	–	–
2006	2636	414	–	–

Source: FADN.

Table A.II
Post-matching balance after DID matching using outcome measure “share of grassland”.

		NN matching		Kernel matching Cross-validation w.r.t.		
	Pre-matching	1:1 NN	1:5 NN	Quantile dist. method	Means of covariate sets X	Means of outcomes Y
Full covariate set						
Production environment						
Land input (ha)	-0.12	**0.39**	**0.47**	**0.53**	**0.37**	**0.51**
Proportion of rented land	**0.23**	**0.49**	**0.57**	**0.69**	**0.46**	**0.7**
LFA subsidies (yes/no)	0.03	0	-0.01	-0.01	0	0.05
Farming intensity						
Proportion of grassland area[1]	0.02	-0.06	**-0.21**	-0.16	-0.04	-0.07
Proportion of cereals area	0.07	0.06	**0.22**	0.15	0.08	0.09
Direct payments crops (€/ha)	0.12	0.11	**0.25**	**0.22**	0.11	0.2
Direct payments livestock (€/ha)	**-0.36**	-0.08	-0.03	-0.04	-0.06	-0.1
LU cattle per ha	**-0.22**	**-0.37**	**-0.59**	**-0.52**	**-0.37**	**-0.5**
LU pigs and poultry per ha	-0.2	0.01	0.06	-0.09	-0.06	-0.13
Fertiliser exp. per ha	**-0.27**	-0.02	-0.04	-0.06	-0.1	-0.06
Plant protection exp. per ha	-0.03	0.16	**0.27**	0.14	0.04	0.04
Fixed capital per ha (1000€)	-0.01	**-0.44**	**-0.5**	**-0.64**	**-0.46**	**-0.65**
Productivity						
Sales per ha (1000€)	-0.2	-0.07	-0.1	-0.2	-0.14	**-0.23**
Revenue per AWU (1000€)	**-0.57**	0.06	0.07	0.02	0.02	-0.07
Revenue per capital (€)	-0.15	0.13	0.14	0.09	0.05	0.07
Total output over total input	**-0.28**	-0.18	**-0.35**	**-0.46**	**-0.31**	**-0.42**
Preferences						
Age of farmer	0.03	0.09	0.03	-0.03	0.04	-0.04
Farm type livestock	-0.09	-0.11	**-0.36**	**-0.25**	-0.14	-0.17
Farm type mixed	0.17	-0.04	0.14	0.11	0.09	0.08
Shannon diversity index	**0.45**	0.05	0.16	**0.28**	0.11	**0.29**
Mean standardised difference	0.18	0.15	0.23	0.23	0.15	0.22
Lasso common covariate set						
Production environment						
Land input (ha)	-0.13	**0.35**	**0.47**	**0.49**	**0.36**	**0.51**
LFA subsidies (yes/no)	0.03	0.00	0.00	0.00	0.00	0.05
Farming intensity						
Proportion of grassland area[1]	0.02	0.06	-0.03	-0.07	0.03	-0.07
Proportion of cereals area	0.07	-0.02	0.03	0.07	0.02	0.08
Direct payments crops (€/ha)	0.11	0.02	0.09	0.14	0.05	0.19
LU cattle per ha	**-0.22**	-0.20	**-0.33**	**-0.37**	**-0.23**	**-0.5**
Plant protection exp.	-0.03	0.03	0.13	0.05	-0.02	0.05
Fixed capital per ha (1000€)	-0.01	**-0.34**	**-0.45**	**-0.58**	**-0.42**	**-0.65**
Productivity						
Sales per ha (1000€)	-0.2	-0.04	-0.1	-0.18	-0.14	**-0.22**
Revenue per capital	-0.12	0.05	0.06	0.05	0.03	0.03
Preferences						
Farm type livestock	-0.09	0.03	-0.12	-0.16	-0.03	-0.17
Shannon diversity index	**0.44**	-0.01	0.04	0.19	0.05	**0.28**
Scale effects (Farm size × …)						
Direct payments livestock (€/ha)	**-0.32**	0.01	0.04	0.01	0.01	-0.05
Mean standardised difference	0.14	0.09	0.15	0.18	0.11	0.22

Notes: Exact matching on entry year and region; boldface denotes standardised differences of larger than 0.2 (absolute); [1] denotes that pre-participation outcome is excluded from the matching set.

Table A.III
Covariate balance in terms of standardised differences, excluding farms that participate in the predecessor programme using the outcome measure “fertiliser expenditure”.

		NN matching		Kernel matching Cross-validation w.r.t.		
	Pre-matching	1:1 NN	1:5 NN	Quantile dist. method	Means of covariates sets X	Means of outcomes Y
Full covariate set						
Production environment						
Land input (ha)	**-0.25**	0.06	-0.01	-0.2	0.08	-0.06
Region west	0.12	0.05	0.11	0.16	0.01	0.09
Region north	**-1.91**	**-0.77**	**-1.08**	**-1.9**	**-0.45**	**-1.22**
Proportion of rented land	0.19	**0.29**	**0.27**	**0.21**	**0.29**	**0.25**
LFA subsidies (yes/no)	0.09	0.00	0.01	0.17	0.00	0.00
Farming intensity						
Proportion of grassland area[1]	0.13	-0.1	-0.16	-0.13	-0.01	-0.09
Proportion of cereals area	-0.02	0.11	0.17	0.19	0.05	0.13
Direct payments crops (€/ha)	0.01	0.13	0.16	0.18	0.06	0.13
Direct payments livestock (€/ha)	**-0.37**	-0.06	-0.03	-0.12	-0.04	-0.1
LU cattle per ha	-0.12	**-0.31**	**-0.33**	**-0.23**	**-0.23**	**-0.24**

(*continued on next page*)

Table A.III (*continued*)

	Pre-matching	NN matching: 1:1 NN	NN matching: 1:5 NN	Kernel matching, Cross-validation w.r.t.: Quantile dist. method	Kernel matching, Cross-validation w.r.t.: Means of covariates sets X	Kernel matching, Cross-validation w.r.t.: Means of outcomes Y
LU pigs and poultry per ha	-0.18	0.04	0.05	-0.02	-0.05	-0.05
Fertiliser exp. per ha	**-0.29**	-0.09	-0.1	-0.16	-0.16	-0.15
Plant protection exp. per ha	-0.05	0.18	0.19	0.07	0.03	0.06
Fixed capital per ha (1000€)	0.06	-0.14	-0.07	0.12	-0.17	-0.01
Productivity						
Sales per ha (1000€)	-0.14	-0.05	0.00	-0.01	-0.11	-0.06
Revenue per AWU (1000€)	**-0.49**	-0.14	-0.19	**-0.36**	-0.12	**-0.23**
Revenue per capital (€)	-0.13	0.04	0.05	-0.04	0.02	0
Total output over total input	-0.19	**-0.32**	**-0.3**	**-0.27**	**-0.35**	**-0.32**
Preferences						
Age of farmer	-0.05	0.00	-0.03	-0.05	-0.05	-0.07
Farm type livestock	0.07	-0.1	-0.18	-0.11	-0.01	-0.12
Farm type mixed	0.05	0.05	0.11	0.14	0.01	0.09
Shannon diversity index	**0.26**	0.13	0.20	**0.35**	0.05	0.20
Mean standardised difference	0.24	0.14	0.17	0.24	0.11	0.17
Lasso common covariate set						
Production environment						
Region north	**-1.91**	-0.1	-0.17	**-0.98**	0.00	0.00
Farming intensity						
Proportion of grassland area[1]	0.13	-0.09	-0.12	-0.03	-0.12	-0.1
Pigs and poultry (LU) per ha	-0.18	0.01	0.02	-0.09	-0.04	-0.04
Fertiliser expenditure	**-0.29**	-0.03	-0.05	-0.14	-0.08	-0.08
Plant protection expenditure	-0.05	0.16	0.16	0.01	0.07	0.06
Productivity						
Sales per ha(1000€)	-0.14	-0.02	-0.02	-0.08	-0.07	-0.07
Total output over total input	-0.19	-0.03	-0.08	-0.15	-0.1	-0.09
Preferences						
Age of farmer	-0.05	0.02	0.04	-0.01	0.02	0.02
Scale effects (Farm size × …)						
Direct payments livestock (€/ha)	**-0.34**	-0.05	-0.04	-0.07	-0.05	-0.04
Mean standardised difference	0.36	0.06	0.08	0.17	0.06	0.06

Notes: Exact matching on entry year; boldface denotes standardised differences of larger than 0.2 (absolute); [1] denotes that pre-participation outcome is excluded from the matching set.

Table A.IV

ATT estimates: changes in outcome measures with matching on the pre-treatment outcome.

Outcome measure / Covariate selection/ matching estimator	Fertiliser expenditure: Matched treated	Fertiliser expenditure: Matched controls	Fertiliser expenditure: ATT in €/ha	Plant protection expenditure: Matched treated	Plant protection expenditure: Matched controls	Plant protection expenditure: ATT in €/ha	Grassland share: Matched treated	Grassland share: Matched controls	Grassland share: ATT (Δ share)
Full set of covariates									
NN Mahalanobis matching									
1:1 NN matching	928	328	-5.69 (4.14)	928	328	-5.68(3.58)	928	193	.017** (0.008)
1:5 NN matching	928	780	-5.96** (2.96)	928	780	-5.43** (2.54)	928	467	.026*** (0.009)
Kernel matching									
R × quant-dist. method[a])	909	1371	-4.47* (1.99)	909	1371	-4.89*** (1.65)	909	1372	.024*** (0.008)
Cross-validation w.r.t. X	474	434	-4.13 (4.72)	474	434	-4.89* (2.95)	691	785	.022** (0.009)
Weighted cross-validation w.r.t. Y[b])	852	1138	-4.51* (2.33)	887	1287	-4.84*** (1.74)	918	1398	.021** (0.008)
Lasso covariate selection set									
NN Mahalanobis matching									
1:1 NN matching	928	252	-5.52 (3.48)	928	356	-5.23* (2.87)	931	185	.016* (0.009)
1:5 NN matching	928	560	-4.00 (2.96)	928	782	-5.37** (2.33)	931	458	.025*** (0.008)
Kernel matching									
R × quant-dist. method[a])	904	1332	-3.74 (2.31)	912	1359	-4.62*** (1.65)	906	1361	.022*** (0.008)
Cross-validation w.r.t. X	851	1063	-2.64 (2.63)	759	868	-6.18*** (2.00)	827	1038	.024** (0.010)
Weighted cross-validation w.r.t. Y[b])	826	956	-3.17 (2.75)	871	1221	-5.66*** (1.78)	912	1381	.022*** (0.008)
N	928	1417		928	1417		928	1417	

Notes: Exact matching on entry year, standard errors in brackets, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; a) 1.5 × 90% quantile, Huber et al. (2015); b) Galdo et al. (2008); standard errors for NN matching based on Abadie and Imbens (2006) and for kernel matching bootstrapped with 1000 replications; naïve DID: fertiliser expenditure − 6.05 (p < 0.05), plant protection expenditure − 6.00 (p < 0.01), and grassland share 0.019 (p < 0.05).

References

Abadie, A., 2005. Semiparametric difference-in-differences estimators. Rev. Econ. Stud. 72 (1), 1–19.
Abadie, A., Cattaneo, M.D., 2018. Econometric methods for program evaluation. Annual Review of Economics 10, 465–503. https://doi.org/10.1146/annurev-economics-080217-053402.
Abadie, A., Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74 (1), 235–267.
Ahn, S.C., Good, D.H., Sickles, R.C., 2000. Estimation of long-run inefficiency levels: a dynamic frontier approach. Econ. Rev. 19 (4), 461–492.
Arata, L., Sckokai, P., 2016. The impact of agri-environmental schemes on farm performance in five EU member states: a did-matching approach. Land Econ. 92 (1), 167–186.
Areal, F.J., Jones, P.J., Mortimer, S.R., Wilson, P., 2018. Measuring sustainable intensification: combining composite indicators and efficiency analysis to account for positive externalities in cereal production. Land Use Policy 75, 314–326.
Baldoni, E., Coderoni, S., Esposti, R., 2017. The productivity and environment nexus with farm-level data. The case of carbon footprint in Lombardy FADN farms. Bio-Based Appl. Econ. J. 6 (2). ⟨https://ideas.repec.org/a/ags/aieabj/276289.html⟩ (Accessed 23 May 2021).
Bartkowski, B., Bartke, S., 2018. Leverage points for governing agricultural soils: a review of empirical studies of European farmers' decision-making. Sustainability 10 (9), 3179.
Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P.J., Lapeyre, R., Persson, U.M., Pfaff, A., Wunder, S., 2016. Mainstreaming impact evaluation in nature conservation. Conserv. Lett. 9 (1), 58–64.
Bellebaum, J., Koffijberg, K., 2018. Present agri-environment measures in Europe are not sufficient for the conservation of a highly sensitive bird species, the corncrake Crex crex. Agric. Ecosyst. Environ. 257, 30–37.
Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. Rev. Econ. Stud. 81 (2), 608–650.
Bertoni, D., Curzi, D., Aletti, G., Olper, A., 2020. Estimating the effects of agri-environmental measures using difference-in-difference coarsened exact matching. Food Policy 90, 101790.
Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., Blust, G., de, Cock, R., de, Diekötter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J.P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W., Zobel, M., Edwards, P.J., 2008. Indicators for biodiversity in agricultural landscapes: a pan-European study. J. Appl. Ecol. 45, 141–150.
Börner, J., Baylis, K., Corbera, E., Ezzine-de-Blas, D., Honey-Rosés, J., Persson, U.M., Wunder, S., 2017. The effectiveness of payments for environmental services. World Dev. 96, 359–374.
Bowler, D.E., Heldbjerg, H., Fox, A.D., Jong, M. de, Böhning-Gaese, K., 2019. Long-term declines of European insectivorous bird populations and potential causes. Conserv. Biol. 33 (5), 1120–1130.
Bowman, M.S., Zilberman, D., 2013. Economic factors affecting diversified farming systems. E&S 18 (1).
Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., Stürmer, T., 2006. Variable selection for propensity score models. Am. J. Epidemiol. 163 (12), 1149–1156.
Buckley, C., Wall, D.P., Moran, B., Murphy, P.N.C., 2015. Developing the EU farm accountancy data network to derive indicators around the sustainable use of nitrogen and phosphorus at farm level. Nutr. Cycl. Agroecosyst. 102 (3), 319–333.
Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. 22 (1), 31–72.
Carmona, C.P., Guerrero, I., Peco, B., Morales, M.B., Oñate, J.J., Pärt, T., Tscharntke, T., Liira, J., Aavik, T., Emmerson, M., et al., 2020. Agriculture intensification reduces plant taxonomic and functional diversity across European arable systems. Funct. Ecol. 34 (7), 1448–1460.
Castaño, J., Blanco, M., Martinez, P., 2019. Reviewing counterfactual analyses to assess impacts of EU rural development programmes: what lessons can be learned from the 2007–2013 ex-post evaluations? Sustainability 11 (4), 1105.
Chabé-Ferret, S., 2015. Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. J. Econ. 185 (1), 110–123.
Chabé-Ferret, S., 2017. Should we combine difference in differences with conditioning on pre-treatment outcomes? Toulouse School of Economics working paper 17-824. Toulouse School of Economics (TSE).
Chabé-Ferret, S., Subervie, J., 2013. How much green for the buck? Estimating additional and windfall effects of french agro-environmental schemes by did-matching. J. Environ. Econ. Manag. 65 (1), 12–27.
Chambers, R.G., Pieralli, S., 2020. The sources of measured US agricultural productivity growth: weather, technological change, and adaptation. Am. J. Agric. Econ. 102 (4), 1198–1226.
Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econ. J. 21 (1), C1–C68.
Cisilino, F., Bodini, A., Zanoli, A., 2019. Rural development programs' impact on environment: an ex-post evaluation of organic farming. Land Use Policy 85, 454–462.
Cisilino, F., Cagliero, R., Scardera, A., 2011. Evaluating rural development programmes using FADN data. Ministry of Agriculture and Forestry. ⟨https://www.researchgate.net/publication/263620520_Evaluating_Rural_Development_Programmes_Using_FADN_Data⟩, (Accessed 17 September 2021).
D'Amour, A., Ding, P., Feller, A., Lei, L., Sekhon, J., 2021. Overlap in observational studies with high-dimensional covariates. J. Econ. 221 (2), 644–654.
Dabkienė, V., 2016. The scope of farms sustainability tools based on FADN data. Sci. Pap. Ser. Manag. Econ. Eng. Agric. Rural Dev. 121–128.
Daw, J.R., Hatfield, L.A., 2018. Matching and regression to the mean in difference-in-differences analysis. Health Serv. Res. 53 (6), 4138–4156.
de Luna, X., Waernbaum, I., Richardson, T.S., 2011. Covariate selection for the nonparametric estimation of an average treatment effect. Biometrika 98 (4), 861–875.
Defrancesco, E., Gatto, P., Runge, F., Trestini, S., 2008. Factors affecting farmers' participation in agri-environmental measures: a northern Italian perspective. J. Agric. Econ. 59 (1), 114–131.
Dessart, F.J., Barreiro-Hurlé, J., van Bavel, R., 2019. Behavioural factors affecting the adoption of sustainable farming practices: a policy-oriented review. Eur. Rev. Agric. Econ. 46 (3), 417–471.
Emmerson, M., Morales, M.B., Oñate, J.J., Batary, P., Berendse, F., Liira, J., Aavik, T., Guerrero, I., Bommarco, R., Eggers, S., et al., 2016. How agricultural intensification affects biodiversity and ecosystem services. Adv. Ecol. Res. 55, 43–97.
Eurostat, 2017. Agri-environmental indicator – commitments. ⟨https://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_commitments#Context⟩, (Accessed 7 March 2020).
Freyaldenhoven, S., Hansen, C., Shapiro, J.M., 2019. Pre-event trends in the panel event-study design. Am. Econ. Rev. 109 (9), 3307–3338.
Galdo, J.C., Smith, J., Black, D., 2008. Bandwidth selection and the estimation of treatment effects with unbalanced data. Ann. Econ. Stat. (91/92), 189.
Grajewski, R., Schmidt, T.G., 2015. Agrarumweltmaßnahmen in Deutschland-Förderung in den ländlichen Entwicklungsprogrammen im Jahr 2013. Thünen Working Paper. ⟨http://hdl.handle.net/10419/117299⟩.
Häggström, J., Persson, E., Waernbaum, I., de Luna, X., 2015. Covsel: An r package for covariate selection when estimating average causal effects. J. Stat. Softw. 68 (1).
Harder, V.S., Stuart, E.A., Anthony, J.C., 2010. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychol. Methods 15 (3), 234–249.
Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Rev. Econ. Stud. 64 (4), 605–654.
Huber, M., Lechner, M., Steinmayr, A., 2015. Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. Empir. Econ. 49 (1), 1–31.
Imai, K., King, G., Stuart, E., 2008. Misunderstandings among experimentalists and observationalists about causal inference. J. R. Stat. Soc. Ser. A 171 (part 2), 481–502.
Imbens, G.W., Rubin, D.B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, Cambridge, p. 625.
Jerrentrup, J.S., Dauber, J., Strohbach, M.W., Mecke, S., Mitschke, A., Ludwig, J., Klimek, S., 2017. Impact of recent changes in agricultural land use on farmland bird trends. Agric. Ecosyst. Environ. 239, 334–341.
Kainz, K., Greifer, N., Givens, A., Swietek, K., Lombardi, B.M., Zietz, S., Kohn, J.L., 2017. Improving causal inference: recommendations for covariate selection and balance in propensity score methods. J. Soc. Work. Res. 8 (2), 279–303.
Kelly, E., Latruffe, L., Desjeux, Y., Ryan, M., Uthes, S., Diazabakana, A., Dillon, E., Finn, J., 2018. Sustainability indicators for improved assessment of the effects of agricultural policy across the EU: is FADN the answer? Ecol. Indic. 89, 903–911.
King, G., Nielsen, R., 2019. Why propensity scores should not be used for matching. Polit. Anal. 27 (4), 435–454.
Kleijn, D., Baquero, R.A., Clough, Y., Díaz, M., Esteban, J., de, Fernández, F., Gabriel, D., Herzog, F., Holzschuh, A., Jöhl, R., Knop, E., Kruess, A., Marshall, E.J.P., Steffan-Dewenter, I., Tscharntke, T., Verhulst, J., West, T.M., Yela, J.L., 2006. Mixed biodiversity benefits of agri-environment schemes in five European countries. Ecol. Lett. 9 (3), 243–257.
Kuhfuss, L., Subervie, J., 2018. Do European agri-environment measures help reduce herbicide use? Evidence from viticulture in France. Ecol. Econ. 149, 202–211.
Lastra-Bravo, X.B., Hubbard, C., Garrod, G., Tolón-Becerra, A., 2015. What drives farmers' participation in EU agri-environmental schemes?: results from a qualitative meta-analysis. Environ. Sci. Policy 54, 1–9.
Leonhardt, H., Braito, M., Uehleke, R., 2021. Combining the best of two methodological worlds? Integrating q methodology-based farmer archetypes in a quantitative model of agri-environmental scheme uptake. Agric. Hum. Values.
Leonhardt, H., Penker, M., Salhofer, K., 2019. Do farmers care about rented land? A multi-method study on land tenure and soil conservation. Land Use Policy 82, 228–239.
Loader, C.R., 1999. Bandwidth selection: classical or plug-in? Ann. Stat. 27 (2), 415–438.
Margarian, A., 2010. Coordination and differentiation of strategies: the impact on farm growth of strategic interaction on the rental market for land. Ger. J. Agric. Econ. 59 (3), 202–216.
Marja, R., Kleijn, D., Tscharntke, T., Klein, A.-M., Frank, T., Batáry, P., 2019. Effectiveness of agri-environmental management on pollinators is moderated more by ecological contrast than by landscape structure or land-use intensity. Ecol. Lett. 22 (9), 1493–1500.
Mennig, P., Sauer, J., 2019. The impact of agri-environment schemes on farm productivity: a did-matching approach. Eur. Rev. Agric. Econ. 104 (6), 1667.
Osterburg, B., Stratmann, U., 2002. Die regionale Agrarumweltpolitik in Deutschland unter dem Einfluss der Förderangebote der Europäischen Union. Ger. J. Agric. Econ. 51 (05), 1–21.

Pascucci, S., de-Magistris, T., Dries, L., Adinolfi, F., Capitanio, F., 2013. Participation of Italian farmers in rural development policy. Eur. Rev. Agric. Econ. 40 (4), 605–631.

Pates, N.J., Hendricks, N.P., 2020. Additionality from payments for environmental services with technology diffusion. Am. J. Agric. Econ. 102 (1), 281–299.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82 (4), 669–688.

Pearl, J., 2009. Causal inference in statistics: an overview. Stat. Surv. 3, 96–146.

Persson, E., Häggström, J., Waernbaum, I., de Luna, X., 2017. Data-driven algorithms for dimension reduction in causal inference. Comput. Stat. Data Anal. 105, 280–292.

Pufahl, A., Weiss, C.R., 2009. Evaluating the effects of farm programmes: results from propensity score matching. Eur. Rev. Agric. Econ. 36 (1), 79–101.

Purvis, G., Louwagie, G., Northey, G., Mortimer, S., Park, J., Mauchline, A., Finn, J., Primdahl, J., Vejre, H., Vesterager, J.P., Knickel, K., Kasperczyk, N., Balázs, K., Vlahos, G., Christopoulos, S., Peltola, J., 2009. Conceptual development of a harmonised method for tracking change and evaluating policy in the agri-environment: the agri-environmental footprint index. Environ. Sci. Policy 12 (3), 321–337.

Rosa-Schleich, J., Loos, J., Mußhoff, O., Tscharntke, T., 2019. Ecological-economic trade-offs of diversified farming systems – a review. Ecol. Econ. 160, 251–263.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Rosenbaum, P.R., Rubin, D.B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am. Stat. 39 (1).

Roth, J., 2018. Pre-test with caution event-study estimates after testing for parallel trends. Working Paper.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psych. 66 (5), 688–701.

Rubin, D.B., 1980. Comment. J. Am. Stat. Assoc. 75 (371), 591–593.

Salhofer, K., Streicher, G., 2005. Self-selection as a problem in evaluating agri-environmental programs. In: Proceedings of the 87th EAAE-Seminar. Assessing rural development of the CAP. 〈http://ecsocman.hse.ru/data/332/691/1219/87eaae_salhofer_streicher.pdf〉.

Scheper, J., Holzschuh, A., Kuussaari, M., Potts, S.G., Rundlöf, M., Smith, H.G., Kleijn, D., 2013. Environmental factors driving the effectiveness of European agri-environmental measures in mitigating pollinator loss-a meta-analysis. Ecol. Lett. 16 (7), 912–920.

Serra, T., Chambers, R.G., Oude Lansink, A., 2014. Measuring technical and environmental efficiency in a state-contingent technology. Eur. J. Oper. Res. 236 (2), 706–717.

Shortreed, S.M., Ertefaie, A., 2017. Outcome-adaptive lasso: variable selection for causal inference. Biometrics 73 (4), 1111–1122.

Steiner, P.M., Kim, Y., 2016. The mechanics of omitted variable bias: bias amplification and cancellation of offsetting biases. J. Causal Inference 4 (2).

Swanson, S.A., 2015. Communicating causality. Eur. J. Epidemiol. 30 (10), 1073–1075.

Teillard, F., Jiguet, F., Tichit, M., 2015. The response of farmland bird communities to agricultural intensity as influenced by its spatial aggregation. PLoS One 10 (3), e0119674.

Thomas, F., Hartmann, E., Luick, R., Poppinga, O., 2001. Analyse von Agrarumweltmaßnahmen: Abschlussbericht des F + E Vorhabens: "Agrarumweltmaßnahmen in der Bundesrepublik Deutschland-Analyse der Umsetzung aus der Sicht des Natur-, Umwelt- und Ressourcenschutzes: Effektivität, Schwachstellen, weitere Entwicklung". 〈https://www.hs-rottenburg.net/fileadmin/user_upload/Forschung/Forschungsprojekte/Management/Evaluierung-AUP/Evaluierung-AUP-Zusammenfassung.pdf〉, (Accessed 29 August 2021).

Udagawa, C., Hodge, I., Reader, M., 2014. Farm level costs of agri-environment measures: the impact of entry level stewardship on cereal farm incomes. J. Agric. Econ. 65 (1), 212–233.

Ulén, B., Bechmann, M., Fölster, J., Jarvie, H.P., Tunney, H., 2007. Agriculture as a phosphorus source for eutrophication in the north-west European countries, Norway, Sweden, United Kingdom and Ireland: a review. Soil Use Manag. 23 (s1), 5–15.

Uthes, S., Kelly, E., König, H.J., 2020. Farm-level indicators for crop and landscape diversity derived from agricultural beneficiaries data. Ecol. Indic. 108, 105725.

Vansteelandt, S., Bekaert, M., Claeskens, G., 2012. On model selection and model misspecification in causal inference. Stat. Methods Med. Res. 21 (1), 7–30.

Villamayor-Tomas, S., Sagebiel, J., Olschewski, R., 2019. Bringing the neighbors in: a choice experiment on the influence of coordination and social norms on farmers' willingness to accept agro-environmental schemes across Europe. Land Use Policy 84, 200–215.

Villanueva, A.J., Gómez-Limón, J.A., Arriaza, M., Rodríguez-Entrena, M., 2015. The design of agri-environmental schemes: farmers' preferences in southern Spain. Land Use Policy 46, 142–154.

Wang, F., Mukherjee, S., Richardson, S., Hill, S.M., 2020. High-dimensional regression in practice: an empirical study of finite-sample prediction, variable selection and ranking. Stat. Comput. 30 (3), 697–719.

Wąs, A., Malak-Rawlikowska, A., Zavalloni, M., Viaggi, D., Kobus, P., Sulewski, P., 2021. In search of factors determining the participation of farmers in agri-environmental schemes – does only money matter in Poland? Land Use Policy 101, 105190.

Wooldridge, J.M., 2016. Should instrumental variables be used as matching variables? Res. Econ. 70 (2), 232–237.

Yoder, L., Ward, A.S., Dalrymple, K., Spak, S., Lave, R., 2019. An analysis of conservation practice adoption studies in agricultural human-natural systems. J. Environ. Manag. 236, 490–498.

Zimmermann, A., Britz, W., 2016. European farms' participation in agri-environmental measures. Land Use Policy 50, 214–228.