LIFT

Low-Input Farming and Territories – Integrating knowledge for improving ecosystem based farming

Research and Innovation action: H2020 – 770747
Call: H2020-SFS-2016-2017
Type of action: Research and Innovation Action (RIA)
Work programme topic: SFS-29-2017
Duration of the project: 01 May 2018 – 30 April 2022

Farm, farm-group and territorial level impact of policies on the adoption of ecological approaches and the performance and sustainability of ecological agriculture

Lionel Védrine[1]*, Sophie Legras[1], Vincent Larmet[1], Legrand Saint Cyr[1], Valentin Bellassen[1], Andrew Barnes[2], Thierno Bocar Diop[1], Stéphane Blancard[1], Philippe Jeanneaux[3], Julie Le Gallo[1], Sébastien Marchand[3], Kato Van Ruymbeke[4], Katarzyna Zawalińska[5]

[1] INRAE (France), [2] SRUC (Scotland), [3] VetAgroSup (France), [4] KU Leuven (Belgium), [5] IRWiR PAN (Poland)

* Deliverable leader – Contact: lionel.vedrine@inrae.fr

DELIVERABLE D6.2

Workpackage N°6

Due date: M42

Actual delivery date: 31/10/2021

Dissemination level: Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770747

About the LIFT research project

Ecological approaches to farming practices are gaining interest across Europe. As this interest grows there is a pressing need to assess the potential contributions these practices may make, the contexts in which they function and their attractiveness to farmers as potential adopters. In particular, ecological agriculture must be assessed against the aim of promoting the improved performance and sustainability of farms, rural environment, rural societies and economies, together.

The overall goal of LIFT is to identify the potential benefits of the adoption of ecological farming in the European Union (EU) and to understand how socio-economic and policy factors impact the adoption, performance and sustainability of ecological farming at various scales, from the level of the single farm to that of a territory.

To meet this goal, LIFT will assess the determinants of adoption of ecological approaches, and evaluate the performance and overall sustainability of these approaches in comparison to more conventional agriculture across a range of farm systems and geographic scales. LIFT will also develop new private arrangements and policy instruments that could improve the adoption and subsequent performance and sustainability of the rural nexus. For this, LIFT will suggest an innovative framework for multi-scale sustainability assessment aimed at identifying critical paths toward the adoption of ecological approaches to enhance public goods and ecosystem services delivery. This will be achieved through the integration of transdisciplinary scientific knowledge and stakeholder expertise to co-develop innovative decision-support tools.

The project will inform and support EU priorities relating to agriculture and the environment in order to promote the performance and sustainability of the combined rural system. At least 30 case studies will be performed in order to reflect the enormous variety in the socio-economic and bio-physical conditions for agriculture across the EU.

Project consortium

No.	Participant organisation name	Country
1	**INRAE - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement**	**FR**
2	VetAgro Sup – Institut d'enseignement supérieur et de recherche en alimentation, santé animale, sciences agronomiques et de l'environnement	FR
3	SRUC – Scotland's Rural College	UK
4	Teagasc – Agriculture and Food Development Authority	IE
5	KU Leuven – Katholieke Universiteit Leuven	BE
6	SLU – Sveriges Lantbruksuniversitet	SE
7	UNIBO – Alma Mater Studiorum – Universita di Bologna	IT
8	BOKU – Universitaet fuer Bodenkultur Wien	AT
9	UBO – Rheinische Friedrich-Wilhelms – Universitat Bonn	DE
10	JRC – Joint Research Centre – European Commission	BE
11	IAE-AR – Institute of Agricultural Economics	RO
12	MTA KRTK – Magyar Tudományos Akadémia Közgazdaság – és Regionális Tudományi Kutatóközpont	HU
13	IRWiR PAN – Instytut Rozwoju Wsi i Rolnictwa Polskiej Akademii Nauk	PL
14	DEMETER – Hellinikos Georgikos Organismos – DIMITRA	GR
15	UNIKENT – University of Kent	UK
16	IT – INRAE Transfert S.A.	FR
17	ECOZEPT Deutschland	DE

Table of contents

List of acronyms and abbreviations

AECM: agri-environmental and climate measures

AEP: agri-environmental payments

AES: agri-environmental schemes

ASP: Agence des Services et Paiements

ATT: Average Treatment Effect on the Treated

AWU: annual working units

CAP: Common Agricultural Policy

CDI: crop diversity index

CF: conventional farming

CGE: Computable General Equilibrium

COP: cereal, oilseed and protein

CUMA: Coopérative d'Utilisation du Matériel Agricole (formal machinery sharing cooperative)

DID: Difference-In-Difference

ECO-E: Eco-efficiency

ES: Ecosystem Services

FADN: Farm Accountancy Data Network

FAT: Funnel plot asymmetry test

FPP: Färe-Primont productivity

GHG: greenhouse gas

LCSFM: Latent Class Stochastic Frontier Model

LI: low-input farming

MRA: meta-regression analysis

ODR: Observatoire du Développement Rural (Observatory of Rural Development)

OF: organic farming

PES: Payment for Environmental Services

PET: Precision effect test

PSM: Propensity Score Matching

RE: Random Effect

SGP: Standard Gross Product

SUTVA: Stable Unit Treatment Value Assumption

TFP: Total factor Productivity

TE: Technical efficiency

TFI: Treatment Frequency Index

UAA: Utilised Agricultural Area

VIF: Variance inflation factors

WLS: Weighted Least-Squares

1 Summary

This deliverable presents the results of the research carried out in WP6 task 6.2 of the LIFT project, on the impact of policies on the adoption of ecological approaches and on the performance and sustainability of ecological agriculture. We first provide a short synthesis of the policy implications of the studies carried out in WP2, WP3 and WP4 of the LIFT project. These studies highlight some drawbacks of currently implemented schemes, such as the current Common Agricultural Policy (CAP) first and second pillar subsidies that may not be adequate for extensive technologies. In addition, these studies advocate policy compensation schemes that take into consideration the income forgone, given the regional potential, both in terms of agricultural production and environmental endowments. We then focus on the effect of Payments for Environmental Services (PES) using meta-analysis and quasi-experimental methods for about 150 PES-schemes implemented worldwide. We find that the effect of PES largely depends on their characteristics. Among others, eligibility of Ecosystem Services (ES) providers, contract length, reference design, payment constraint, monitoring system and the implementation zone of the PES schemes appear to be correlated with the probability of achieving positive environmental results. Nevertheless, the effectiveness of the PES-schemes investigated in this meta-analysis is shown to especially depend on the monitoring system implemented to ensure compliance and on the eligibility of ES providers. Using econometric analysis on French farm data, we also find that farmers' incomes are not affected by their ecological practices, once the extra cost of these practices has been covered by the Agri-Environmental Schemes (AES) payment or promote some efficiency gains. The real cost of the transition is therefore on average well compensated by these payments. It does not imply that farms earn extra profit, and thus appears to respect World Trade Organisation (WTO) rules.

2 Introduction

This document presents the results of the research carried out in WP6 task 6.2 of the LIFT project, on the impact of policies on the adoption of ecological approaches and on the performance and sustainability of ecological agriculture. In the context of the CAP reform, including the Green Deal and the FARM2FORK strategies, it seems essential to highlight the impact of the previous and current CAP programmes on the adoption of ecological practices, the labour demand and the economic performance of farms which are three key issues of the European agricultural sector.

The deliverable is structured in three parts. The first part (Section 3) is a synthesis of the policy implications of the studies carried out in the LIFT project, in terms of adoption of ecological practices, employment and economic performance. We focus only on work with direct implications for public policy in promoting ecological practices. These studies are briefly summarised and we highlight their main implications for public policy. They address the impacts at the farm level, the group of farms' level and the territorial level. This part highlights that (i) current CAP subsidies received by farmers have no impact on intensive farms, but they reduce the technical efficiency of extensive farms and (ii) technical efficiency might be increased with special agricultural training for low input farm managers. Scale efficiency is sensitive to policies regarding taxes. Reconsideration of agricultural tax policies might improve scale efficiency of low input farms. Despite their effectiveness, agricultural tax policies can be complicated to implement given a lack of social acceptability. Therefore, the other two parts

(Sections 4 and 5) focus on exploring the scope for improvement in the implementation of PES, which is considered to be a second-best solution.

The second part (Section 4) is a meta-analysis of the additionality of PES (Engel, 2016). More precisely, our meta-analysis is the first to analyse the impact of the characteristics of PES-schemes on both their effectiveness, in terms of the probability to increase ES provision, and their efficiency, considering the level of additionality. The latter is defined as the “direct changes in land or natural resource-use among participants induced by the PES scheme, compared to a baseline without the PES” (Börner et al., 2017). We use meta-regression analysis on a sample of 114 individual studies that investigate the determinants of the performance of about 150 PES-schemes implemented worldwide. We find that increased effectiveness of PES schemes is strongly associated with periodical third-party monitoring, as well as generic reference design and targeted eligibility to a lesser extent. Cash payments and private buyers are associated with higher additionality whereas periodical monitoring is associated with lower additionality. Result-based payments however are found to be associated neither with higher effectiveness nor with higher efficiency. These results are relevant for policy makers seeking to improve the design of PES schemes to promote ecological practices.

While the second part relies on an international comparison of PES implementation on its effectiveness, the third part (Section 5) focuses on the last generation of AES implemented in France. More precisely, we conduct a counterfactual analysis of the impact of agri-environmental and climate measures (AECM) and organic farming on the adoption of ecological practices, labour demand and economic performance of farms adopting these ecological practices. Our study applies a Difference-In-Difference propensity score matching to identify the causal effect of AECM both on farm practices, labour demand and economic performance. We use a sample of French farms over the 2010-2020 period. Our results reveal a large heterogeneity over the type of AECM and farm specialisation.

3 A synthesis of public policy implication from LIFT studies

This section presents the results of the LIFT project with a direct implication of the effect of public policies on the adoption of ecological practices and economic performance, at the individual level (Section 3.1), group of farms (Section 3.2) and territorial level (Section 3.3).

3.1 Impact at farm level

3.1.1 Technical efficiency of intensive and extensive technology in dairy farming in the European Union – LIFT project, Deliverable 3.1, Latruffe et al. (2021) and Jin et al., (2021)

Latruffe et al. (2021) analyse and compare the technical efficiency of intensive and extensive dairy farms in Ireland, France and Austria, with Farm Accountancy Data Network (FADN) data from the period 2014–2015. The latent class stochastic frontier model (LCSFM) is used to classify farms into extensive and intensive technologies. The LCSFM allows an endogenous categorisation of farms into classes of more or less intensive farms, represented by different production technologies, while simultaneously estimating the technical efficiency of each class. The results show that extensive farms performed worse economically than intensive farms. Jin et al. (2021) found the same results in their study on Ireland. In addition, current CAP subsidies received by farmers have no impact on intensive farms, but they reduce the technical efficiency of extensive farms. Public policies could design the CAP greening payment using the LCSFM method to adequately compensate farmers for losses due to more extensive production practices. Policy design should nonetheless also account for the results from the inefficiency drivers: while current CAP subsidies received by farmers have no impact on intensive farms, they reduce the technical efficiency of extensive farms, suggesting that the current type of subsidies may not be adequate for extensive technologies.

3.1.2 Differences in efficiency and productivity between conventional and organic farms: the case of Hungarian cereal oilseed and protein (COP) crop producing farms (2010-2015) – LIFT project, Deliverable 3.1, Baráth et al. (2021).

Baráth et al. (2021) compare total factor productivity (TFP) and its components, computed with stochastic frontier models using data from FADN, of low input and conventional Cereals, Oilseed and Protein (COP) crop producing farms in Hungary. They identify low input farms based on the typology protocol that was developed within the LIFT project in WP1 (Rega et al., 2021). Statistical tests allow them to compare low input and conventional farms. Results show that TFP scores of low input farms are smaller than TFP scores of conventional farms, and the difference is statistically significant. As the difference in technical efficiency (TE) and scale efficiency between low input and conventional farms was significant, the results suggest that the performance of these farms can be increased with appropriate policy measures. TE might be increased with special agricultural training for low input farm managers. Scale efficiency is sensitive to policies relying on taxes. Reconsideration of agricultural tax policies might improve scale efficiency of low input farms.

3.1.3 Dynamics of productivity and efficiency performance in Poland's dairy farms: comparative analysis by different degrees of ecological approaches – LIFT project, Deliverable 3.1, Zawalińska and Krupin (2021)

Zawalińska and Krupin (2021) aim at comparing the efficiency of dairy farms in Poland according to their degree of ecological approaches with data from FADN in 2015. The LIFT FADN protocol developed in WP1 (Rega et al., 2021) was used to distinguish different ecological types of dairy farms. The authors computed efficiency with Färe-Primont productivity change index (FPP) with its decomposition into technological change and efficiency changes, and meta frontier FPP. Over the period of 2006-2015, standard, integrated and changeable farms experienced a TFP growth while low input-integrated and mixed farms experienced a decline. The highest growth was in case of integrated farms, followed by standard and changeable. The TFP decline was higher for mixed ecological types compared to low input-integrated farms. All ecological types experienced a TFP decline in 2009 revealing that they were affected by the global crises in 2008. As far as the meta-frontier analysis is concerned, the dairy sector in Poland over the decade 2006-2015 experienced a slight TFP growth thanks to progress in technological change, but the efficiency decreased in the meantime. The decline was mainly due to deterioration in TE, while scale efficiency slightly improved, with the residual mix efficiency almost not changing. The findings suggest that the measures aimed at increasing the uptake of ecological approaches in Polish agriculture within the previous (2007-2013) and ongoing (2014-2020) CAP have not been efficient enough. Therefore, this needs to be considered in the next CAP programming period.

3.1.4 Pesticide efficiency of French wheat producers under a stochastic frontier framework – LIFT project, Deliverable 3.1, Dakpo and Féménia (2021)

Dakpo and Féménia (2021) develop an input-specific measure using stochastic frontier analysis (SFA) to derive the potential reduction of pesticide while maintaining the level of farm's profitability. They derive pesticide demand and then adjust this function to account for inefficiency. The obtained model is estimated, using a maximum likelihood approach, on a large sample of about 2,000 French wheat producers over the period 1998 to 2014. The results show a potential reduction of pesticides by about 15%. This figure is much lower than the figures reported in previous studies (Lechenet et al., 2017). The authors also find crop diversification impacts positively damage abatement only for low levels of pesticide use. Two types of recommendations for pesticide reduction policies are drawn from these results. First, a public intervention seems necessary to encourage farmers to reduce their pesticide use, either by compensating them for the loss of profit due to the decrease in their productivity or by making pesticides more costly through a taxation policy. Second, the results on the effects of crop diversification on damage abatement suggest that farmers could fully benefit from the positive impacts of crop diversification, and therefore the reduction of productivity loss, and thus would benefit if a public policy was implemented to discourage high use of pesticides.

3.1.5 Integrating a crop diversity index to eco-efficiency measurement for cropland farms in Sweden – LIFT project, Deliverable 3.1, Huang et al. (2021)

Huang et al. (2021) measure eco-efficiency (ECO-E) in cropland farms by incorporating the dynamic effects of crop diversity into the production function when assessing ECO-E. ECO-E is defined as ecologically adjusted production efficiency by incorporating the dynamic crop diversity index (CDI) into the production function using directional distance function. The study uses an unbalanced panel of data for 209 farms covering 937 observations from the Swedish FADN database for the period 2009-2016. Their results show that the average ECO-E is estimated to be 0.873. The average ECO-E in Northern Sweden is slightly higher than that in Southern Sweden, and the minimum ECO-E in Northern Sweden is much higher than the minimum ECO-E in Southern Sweden. CDI is estimated to positively contribute to ECO-E of crop production significantly. In other terms, crop diversification is a burden on the farm economy, and is related to the regional economic and environmental characteristics. As the new CAP 2021-2027 is aiming at better targeting for securing stable economic incomes and intensifying the environmental and climate actions, the authors recommend policy compensation schemes that take into consideration the income forgone, given the regional potential, both in terms of agricultural production and environmental endowments.

3.1.6 Analysis of the effect of ecological farming practices on the intensity of labour use – LIFT project, Deliverable 3.4 (Davidova et al., 2021)

Davidova et al. (2021) study the effect of ecological farming approaches on labour use. They compare situations in five different countries (Greece, France, Hungary, Poland and the United Kingdom) and use FADN data from 2004 to 2015. Their results show that agri-environmental payments (AEP) increase labour use across all different countries under study. Moreover, low input and capital intensity increase the intensity of labour use as well. These results imply that public policies promoting ecological agriculture may increase farm and rural employment, besides environmental benefits. The authors recommend public authorities to set up policies that aim to reduce transaction costs in the job market. This will allow more flexible adjustments of hired labour. This is important as farmers can fear labour costs that will increase with adoption of ecological practices and be reluctant to partake in.

3.1.7 Agri-environmental schemes (AES) adoption, technical efficiency and environmental indicator: evidence from France – LIFT project, Diop et al. (2021)

AES are voluntary payment–based measures that intend to incentivise farmers to adopt environmental practices on the farm while giving farmers a premium to cover the cost of adoption or the loss of income related to the change in practices. Many studies have analysed the role of farms and farmers' characteristics on AES participation, but none of them has examined the effects of past performance and environmental indicators on multiple AES adoption by farmers. Diop et al. (2021) intend to fill this gap by analysing the role of past four-year average technical efficiency and environmental indicators of farmers on the probability of adopting an AES. This work differs from existing literature and contributes to it in two different ways. Firstly, this study is the first to examine the effect of TE and environmental indicators on multiple AES in France, as Latruffe and Nauges (2014) study the effect of TE only on organic farming. By contrast, the inclusion of environmental indicators allows to bring some clarity in the debate about the windfall effect of AES, i.e. the situation where the most efficient farmers in terms of environment will be the one adopting the AES, which, in turn, will reduce its real environmental effect. Second, the authors adopt a novel methodology proposed by Puhr et al. (2017) based on Firth's penalisation, which is more convenient for small samples, rare events and complete separation problems. Diop et al.'s (2021) study is the first application of this methodology in the agricultural sector.

The results show heterogeneous effects of TE on AES adoption, possible presence of windfall effects. For AES "AES02 – Rotation", the environmental performance plays a positive role in partaking in the AES only if farms reach a certain level of efficiency. Less inefficient farmers might think that adopting this AES will not decrease their efficiency. They consider that they are able to comply with the AES requirements without reducing their efficiency. Therefore, the AES will be more appealing for those less inefficient farmers as they have high environmental performances. Based on the results, windfall effects may exist only for the more efficient farmers. For AES "AES08 – Reduction of chemical crop protection use", TE has a negative and significant effect on partaking in the AES. The more technically efficient farmers are, the less likely they join this AES. This result may be explained by the fact that farmers can see in this AES a threat to their efficiency. It is mainly the case if the requirements of the scheme are seen to be difficult. If farmers depend heavily on this to be efficient, they will not be willing to risk joining the AES and reducing their efficiency. As far as the environmental indicators are concerned, they appear to negatively affect the probability of adopting this AES. It means that more intensive farmers in fertiliser usage, i.e. the ones who have greater negative impact on the environment, tend to not partake in this AES. This result also highlights the possible presence of windfall effect for "AES – Reduction of chemical crop protection use" as in "AES – Rotation". Finally, for AES “AES09 – Fertilisation”, the TE has a negative but not significant effect on the probability of adopting the scheme. Similarly, the environmental indicator does not significantly affect the probability to join the AES.

These results highlight the potential for currently implemented AES to induce windfall effects depending on which types of farmers actually adopt them. They call for a better targeting of farmers to recruit, in particular those putting the most pressure on the environment. The link with TE is less clear-cut. In addition, farmers putting more pressure on the environment are less likely to participate in AES. These results highlight a problem of targeting. If policy makers want the AES to have a greater impact on the environment, much effort should be made to better target farmers and incentivise them to join the schemes. It can be accompanied by an increase in subsidies but also more information about the schemes through farmer unions or cooperatives, for instance.

3.1.8 The impact of European policies on the uptake of ecological approaches – legume production on dairy farms challenged by European policy interaction – LIFT project, Heinrichs et al. (2021)

Heinrichs et al. (2021) evaluate agronomic, economic and environmental impacts of increasing legume production (peas, fava beans and alfalfa) in France and in Germany with the bioeconomic programming farm-scale model FarmDyn. More precisely, the study investigates the interaction of policy measures that, on the one hand, aim at promoting legume production (with voluntary coupled support (VCS) in the frame of the CAP greening) and, on the other hand, potentially constrain their production by regulating nitrogen (N) supply (with Nitrate Directive). Their results show that large legume shares induced by high VCS do not lead to substantial environmental benefits in the analysed dairy farms. In other words, the impact of crop diversification on environmental sustainability of livestock farms is limited. Public policies that can be derived from their results should be to implement stricter regulations in terms of livestock density or manure handling. These policies are more ambitious agro-environmental measures that target animal production directly. The authors also recommend implementing policies that facilitate the processing of legumes at farm level, such as investments in specific storage and improved sorting.

3.2 Impact at farm group level

3.2.1 Collaborative networks and adoption of ecological practices – LIFT project, Task 2.3

When it comes to adoption of ecological farm management practices by farmers, the cognitive barriers farmers face when switching to management practices with specialised equipment is often overlooked. Nonetheless, evidence demonstrates that a lack of know-how and sense of cohesion amongst farmers regarding management practices is equally as restrictive as a lack of economic means (Liu et al., 2018; Mozzato et al., 2018). Collaborative networks - formal and informal networks designed to share, manage, and/or exchange equipment, labour and/or immaterial resources between farmers (Lucas et al., 2018) - offer an opportunity for farmers to overcome economic barriers by sharing mechanisation and labour costs, as well as overcoming cognitive barriers by sharing experiences and know-how regarding (the application of) ecological management practices (Groupe de Bruges, 2014; Lucas et al., 2018). In a discrete choice experiment (DCE) carried out across three case study areas in Belgium and France, preferences amongst farmers to engage in collaborative networks aimed at overcoming cognitive and economic barriers through the sharing of knowledge, labour and machinery were assessed. In France, such collaboration amongst farmers is led by the CUMA (Coopérative d’Utilisation du Matériel Agricole, formal machinery sharing cooperative), which allows its members (farmers) to share their resources such as machinery, sheds, or workshops. In Belgium, however, collaborative behaviour amongst farmers is far less engrained in the farming system. From the results it was found that preferences to join collaborative networks varied between respondents based on their previous experience with collaborative behaviour. Respondents who had previous and/or current experience with collaborative behaviour were found to have a stronger preference to join collaborative networks. Likewise, those respondents without any previous experience demonstrated an aversion to joining such collaborative networks. Unfortunately, the small sample size meant that interactions between characteristics of collaborative networks and ecological farm management practices could not be estimated, such that no evidence could be found in support of collaborative networks increasing adoption rates of ecological farm management practices. Nonetheless, the preferences observed for knowledge and labour sharing illustrate that there is a demand for such collaborative behaviour amongst farmers in Belgium and France, and highlight the need for policy makers to invest in fostering collaboration amongst farmers, whether it be formal through collaborations such as the CUMA, or informal farmer-led collaborations. While the impact of collaborative networks on adoption of ecological farm management practices remains to be assessed, results from this DCE seem to imply there is at least a demand amongst farmers to overcome certain cognitive barriers through collaboration with fellow farmers.

3.2.2 Using rodent outbreak as a model for experimenting low level of pesticides systems for reducing their impact at a farm and local territory scales – LIFT project, Task 6.2

Since the 1970s, grass-based breeding systems in mid-mountain regions in France (Auvergne and Franche-Comté) have been significantly affected by vole outbreaks. Every 5 to 10 years, depending on local landscape, vole outbreaks cause major damages to meadows breeding systems, because they disturb the good economic operation of agricultural systems, especially by weakening their fodder self-sufficiency (Couval et al., 2014). Several strategies have been developed over the years. Poisons have been used to kill voles, but these chemical treatments come to be inefficient and dangerous for voles' predators and auxiliary fauna when voles are in a high-density phase. Actually, regulation of rodent population must go through an early, accurate, and exhaustive monitoring that should be collectively managed. Dureau (2020) has analysed different experiences over a long time. Long-term monitoring of collective action in Franche-Comté showed that it was efficient to control vole outbreaks and to protect fodder production systems (Giraudoux et al., 2017). Therefore, the regulation regarding vole outbreaks (as a pest classified in the second category of sanitary dangers) is rather clear and is supposed to oblige farmers to gather and to plan collective action (there are "mandatory action decrees" at sub-regional scales). But there is neither control nor sanction planned for farmers who do not get involved in collective action. There are main specificities of this risk requiring a high level of coordination: space-related and time-related cycles, high intensity and high frequency, link with landscape and agricultural practices, systemic character at a local scale, difficulty to find out a stable and sufficient premium base. The second part of the analysis has shown the limits of the current risk management strategy, focused on "hazard" and ignoring the vulnerability factors. Risk perception is a main limiting factor to collective action because it causes mistrust between stakeholders (lack of vertical coordination) and between farmers (lack of horizontal coordination). Olson's paradigm regarding collective action allows to better understand this lack of coordination and opens up some new perspectives for a better policy, especially regarding the design of contracts (reducing information asymmetries and organisation costs). These issues need specific tools to get and share information to monitor vole populations (webmapping, mobile phone application…) and a social organisation at the local and regional levels (forecast network, advising system) to coordinate a collective and preventive action against outbreaks. The authors recommend policy compensation schemes that support these specific tools and social organisation.

3.3 Impact at territorial level

3.3.1 Socio-economic impact of ecological agriculture at the territorial level – LIFT project, Deliverable 4.2 (Bailey et al., 2021)

Bailey et al. (2021) study the socio-economics effects of ecological approaches on territories with the help of qualitative approaches to stakeholders. These results show that an increase in skill level and quality of life in on-farm employment is expected. Respondents also predict an increase in the trade of inputs such as compost and manure, but also more sharing of capital and labour. Adoption of ecological farming can have a mixed effect on contracting, machinery purchasers and machinery traders and dealers. In addition, ecological practices are expected to decrease intermediaries, leading to shorter supply chains. Ecological farming can also contribute to higher local consumer demand, as the rural environment can be more attractive. Furthermore, it may increase the cooperation between farmers, sharing best practices and trading input. A key policy implication is that economic returns from ecological agriculture not only contribute to quality of life and the health of rural communities and local environment, but they are also a key factor to farm survival.

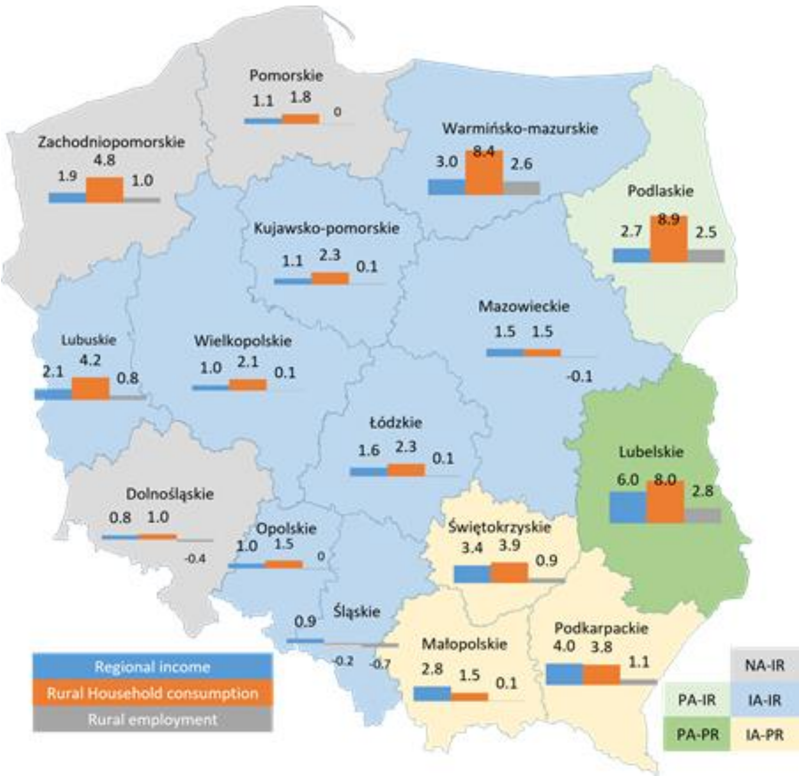
3.3.2 Territorial impact of an organic production increase in Poland using spatial CGE – LIFT project, Task 4.2

Based on territorial scenarios developed in the LIFT project, the potential increase in organic production by 10% over 10 years is analysed for Poland using spatial Computable General Equilibrium (CGE) model developed by Zawalińska et al. (2013). The investigated scenario assumes achieving this goal by shifting more Pillar 2 expenditures to agri-environmental measures, and distributing them to NUTS2 regions with respect to their initial level of agri-environmental expenditures. The socio-economic results of such a scenario show a positive impact on regional income which is in line with predictions of the Polish stakeholders reported in Bailey et al. (2021). The CGE results show that the highest increase in regional income is predicted in South-East regions: Lubelskie, Podkarpackie, Świętokrzyskie, Małopolskie and Podlaskie – out of which the first four are predominantly rural (PR) (indicated in light yellow and light green colour) and the last one is intermediate rural (IR) (light blue and light green), but at the same time also predominantly agricultural (PA) (green backgrounds) (see Map 1)[1].

There is also an overall positive impact on rural households' consumption among the regions except in Śląskie which is one of the most urban regions (with only 23% of the population living in rural areas and 22% of the rural population working in agriculture). The highest estimated increase in rural household consumption is in Podlaskie, Warmińsko-Mazurskie and Lubelskie where the first two are intermediate rural while the first and the third ones are predominantly agricultural regions (see Map 1).

[1] The types of regions are defined according to their rurality and dependence on agriculture as follows: Predominantly rural regions (PR) have above 50% of rural population, Intermediate rural (IR) between 15%-50% and Predominantly urban (PU) less than 15%; and analogically the Predominantly agricultural regions (PA) have above 50% of rural population working in agriculture, Intermediate agricultural (IA) between 15%-50% and Non-agricultural (NA) less than 15%.

Map 1: Regional impact on income, rural household consumption and rural employment in case of 10% increase in organic production, at the Polish NUTS2 regions



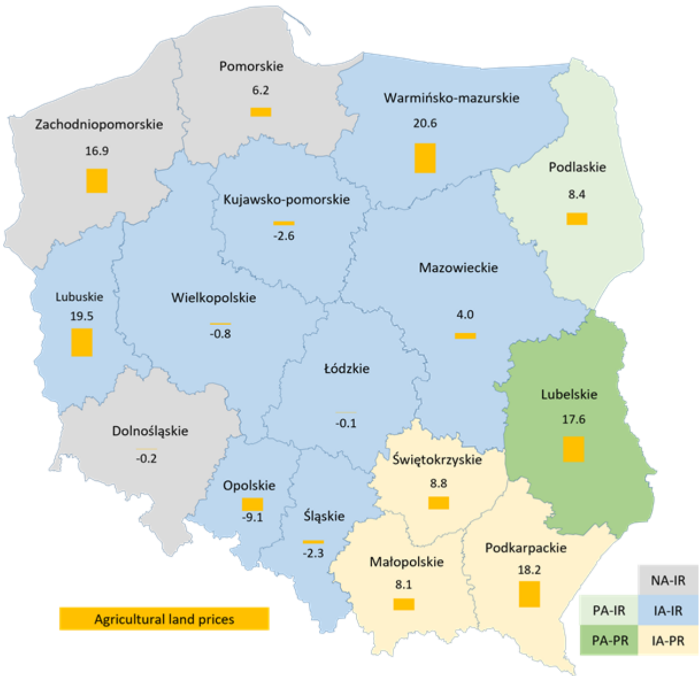
Source: Author's own calculations. Legend for the map's background: PA-IR: predominantly agricultural intermediate rural region; PA-PR: predominantly agricultural predominantly rural regions; NA-IR: non-agricultural intermediate rural region; IA-IR: intermediate agricultural intermediate rural; IA-PR: intermediate agricultural predominantly rural.

The impact on rural employment is diverse, it increases in all predominantly agricultural and all predominantly rural regions but declines in most urban and industrial regions (Śląskie, Dolnośląskie and Mazowieckie). In the case study regions the stakeholders predicted an increase in wages, and hence in the rural employment due to the introduction of ecological production, as reported in Bailey et al. (2021). This is in line with the CGE results showing that, in Lubelskie and Podlaskie, the rural employment increases by 2.8% and 2.5% respectively in the analysed scenario.

The effect on agricultural land prices is mixed – see Map 2. The highest increase is expected primarily in predominantly rural regions (Lubelskie, Podkarpackie, Świętokrzyskie, Małopolskie) but also in border regions (Warmińsko-Mazurskie, Lubuskie, Zachodniopomorskie, Pomorskie). Decline in agricultural land prices is expected in most urban regions (e.g. Opolskie, Śląskie).

All in all, the analysed scenario of 10% increase in organic production would be the most beneficial to predominantly agricultural regions and both intermediate and predominantly rural ones.

Map 2: Regional impact on agricultural land prices in case of 10% increase in organic production, at the Polish NUTS2 regions



Source: Author's own calculations. Legend for the map's background: PA-PR: predominantly agricultural predominantly rural regions; PA-IR: predominantly agricultural intermediate rural region; NA-IR: non-agricultural intermediate rural region; IA-IR: intermediate agricultural intermediate rural; IA-PR: intermediate agricultural predominantly rural (see the definitions in footnote 1).

4 What gives the PES additionality? Some evidence from a quantitative meta-analysis

4.1 Literature review on PES, PES definition and issues

PES schemes have become a popular tool to address environmental degradation and to promote sustainable management of ecosystem services (Engel, 2016; Ezzine-de-Blas et al., 2016; Börner et al., 2017). While several programs to reward the provision of ES are implemented worldwide, there still exists an academic debate on what PES cover (Derissen and Latacz-Lohmann, 2013). Economic theory defines PES as a transaction between ES users and providers (Wunder, 2015; Muradian et al., 2010) but this definition gives room for many interpretations. A narrow definition, which we will call Coasean PES[2], restricts PES to transactions between private agents, where both suppliers and buyers are free to trade or not. A broader definition, which we will call Pigouvian PES[3], includes schemes where buyers are forced to purchase a certain amount of ES. These mandatory buyers can either be private firms subject to a cap-and-trade system or the State itself in the case of result-based subsidies. Ranging from a narrow to a broad definition, PES schemes have been implemented in various geographical and socio-economic contexts. As PES schemes may have different objectives, they may also achieve different levels of performance, which depends on design characteristics as well as on the context of implementation (Wunder, 2015).

The performance of PES typically rests on two synergetic pillars: enrolment and efficiency. To achieve significant environmental impacts, a PES must indeed enrol a reasonably large number of potential ES providers, often farmers or community members (Wunder et al., 2008). However, efficiency is also paramount: if a faulty design results in payments for what most ES providers would have done anyway, a PES will have little effectiveness even if it manages to enrol a large number of potential ES providers.

Various studies investigate the performance of several PES schemes and show mixed evidence on their actual performance (Pattanayak et al., 2010). Besides qualitative studies (e.g., Engel et al., 2008; Alston et al., 2013; Börner et al., 2017; Kaiser et al., 2021), previous quantitative meta-analyses most often focus on PES that target specific ES or specific zones or countries (e.g., Keijn and Sutherland, 2003; Brouwer et al., 2011; Scheper et al., 2013). Ezzine-de-Blas et al.'s (2016) work is the first to summarise the effect of PES schemes implemented around the world based on a broad quantitative meta-analysis.

The contribution of this study is fourfold. Firstly, we perform a meta-analysis on the impact of PES on the provision of ES (effectiveness) considering a wide range of schemes implemented worldwide. Note that there is no consensus in the literature on what PES actually cover. We do not enter the debate on the ideal definition of PES and therefore, our meta-analysis includes empirical studies on both Cosean and Pigouvian PES, expanding from the Ezzine-de-Blas et al. (2016) meta-analysis that focuses on Coasean PES.

Secondly, we assess the additionality of PES as a continuous variable - by how much did the PES increase the provision of the environmental service? - whenever possible. In these cases, we also re-assess effectiveness in order to distinguish characteristics that improve enrolment from those which improve efficiency.

Thirdly, our sample of PES allows us to pay particular attention to five key design variables untested by Ezzine-de-Blas et al. (2016): Should payment be based on practices or results? Should payment be

[2] In which there is a direct market arrangement between the 'buyer' and 'provider' of services (Coase, 1960).

[3] In which governments provide the payments to reach specific policy objectives (Pigou, 1920).

in cash or in kind? Who should the payment recipient be: individuals/firms or communities? How often should the monitoring take place? How should the counterfactual reference be estimated?

Fourthly, we combine different estimation techniques to obtain robust estimations of the average level of PES additionality and to study the impact of PES-schemes characteristics on their effectiveness and efficiency. To estimate the impact of PES-schemes characteristics on their effectiveness, we implement an original tobit specification within a meta-regression analysis (MRA) in order to control for possible censoring in the data. To estimate the average additionality of PES, we combine a precision effect test (PET) with a Funnel plot asymmetry test (FAT) to control for potential publication bias (Doucouliagos and Stanley, 2009).

These specifications are used to test our hypotheses through a number of PES-scheme characteristics identified in a sample of 114 individual studies that investigate the performance of about 150 PES-schemes of different types implemented worldwide. We find evidence that a number of characteristics of the PES schemes included in this meta-analysis indeed play a significant role in their effectiveness and their efficiency. In addition, the main characteristics of PES schemes that influence the probability of increasing ES provision are different from those that affect the level of additionality. Increased effectiveness of PES schemes is found to be associated with periodical monitoring, generic reference design and targeted eligibility of potential ES providers. Conversely, the voluntary nature of demand in the PES-schemes and the type of payments received by the participants seem to be the main factors that influence their efficiency.

The rest of this section is structured as follows. The next section (Section 4.2) presents the main hypotheses to be tested (Section 4.2.1), the selection process of the documents included in the meta-analysis (Section 4.2.2) and the methods used to test the hypotheses (Section 4.2.3). The main results are presented in Section 4.3.

4.2 Material and method

4.2.1 Main hypotheses tested in the meta-analysis

Various studies investigate the performance of several PES schemes and show mixed evidence on their actual performance (Pattanayak et al., 2010). Besides qualitative studies (e.g., Engel et al., 2008; Alston et al., 2013; Börner et al., 2017; Kaiser et al., 2021), previous quantitative meta-analyses most often focus on PES that target specific ES or specific zones or countries (e.g., Keijn and Sutherland, 2003; Brouwer et al., 2011; Scheper et al., 2013). Ezzine-de-Blas et al.'s (2016) work is the first to summarise the effect of PES schemes implemented around the world based on a broad quantitative meta-analysis.

We go further as the main objective of our work is to determine, based on a meta-analysis, whether PES schemes achieve significant environmental impacts (effectiveness), measured as the probability to increase ES provision. On a restricted smaller sample (including only studies with sufficient material to compute additionality), we also assess their level of additionality (efficiency). Following Börner et al. (2017), we define additionality as the "direct changes in land or natural resource-use among participants induced by the PES scheme, compared to a baseline (i.e., without the PES)". Our estimate of additionality is the difference between the level of ES with the payment and an estimate of what this level would have been without the payment, expressed as a percentage of the latter. We also identify the role of the main PES characteristics on their additionality, and, by re-assessing effectiveness on the same, smaller, sample, we are able to decompose the drivers of effectiveness between enrolment and efficiency. Because both effectiveness and additionality assessment depend on an ex-post evaluation of the scheme (Wunder, 2015), we focus only on ex-post analyses of PES impacts to test the following hypotheses. These assumptions correspond to the observable variations in the characteristics of PES that we suspect influence both additionality, enrolment and efficiency.

H1: Coasean PES > Pigouvian PES.

Our first hypothesis is Coasean PES are more effective than Pigouvian PES. This outcome is predicted by contract theory, due to the principal-agent problem (Laffont and Tirole, 1993; Engel et al., 2008; Ferraro, 2008). In Coasean PES, the end-consumer of the ES is directly involved in the contract and is therefore interested in closely scrutinising that the expected outcome is achieved. In Pigouvian PES, the end-consumer - which can be very large groups of people, up to the global population for climate mitigation - is represented by an agent – e.g., a Non-Government Organisation, a regulator, an auditor, etc. - whose objectives may not be perfectly aligned with those of the end-consumer. As a result, the level of scrutiny of the expected outcome may be suboptimal.

H2: single objective > multiple objectives.

PES with multiple objectives including rural development and recipient welfare are expected to be less effective as some environmental goals may be dominated by the other objectives and some degree of non-compliance may be tolerated (Wunder, 2015; Wunder et al, 2008; Engel, 2016).

H3: spatial targeting +.

Spatial targeting is expected to increase the efficiency of PES. The rationale is that spatial targeting is a way to capture part of the information rent of ES suppliers (e.g., Antle et al., 2003; Canton et al., 2009). This has long been identified empirically (Muñoz-Piña et al., 2008; Ezzine-de-Blas et al., 2016). Moreover, as budgets are usually too scarce to enrol all potential ES suppliers in a scheme, spatially targeting high-ES density and high-threat areas may also lead to increased PES efficiency (Wunder, 2015; Wünscher et al., 2008; Sims et al., 2014).

H4: result-based payments = practice-based payments.

Result-based payments are intuitively expected to be more efficient than practice-based payments, because in principle they allow for better targeting. This is why the European Commission is currently pushing towards more result-based subsidies in the CAP (European Commission, 2018). In practice however, very few schemes directly monitor results (e.g., through soil samples to measure soil carbon storage) and actual monitoring requirements span along a continuum of more or less accurate estimates of targeted services (e.g., area with cover crops multiplied by the average carbon storage per hectare of cover crops to estimate soil carbon storage) (Burton and Schwarz, 2013; Herzon et al., 2018).

H5: individual reference > generic reference.

Implicitly or explicitly, PES value environmental services provided in addition to a baseline level, which may be a generic zero. When this baseline is set stringently and on a case-by-case basis, based on the individual characteristics of the recipient, the PES is expected to be more efficient (Bento et al., 2015; Cormier and Bellassen, 2013).

H6: individual cash payments > community in-kind rewards.

Individual cash payments tend to be preferred to community-level rewards by potential ES providers (Costedoat et al., 2016). Because individual cash payments are more targeted, they are also expected to be more efficient than community-level payments hampered by free-riding.

H7: long-term contracts > short-term contracts.

PES offering long-term contract to ES providers may be more effective than those implemented with other types of contracts for two reasons. Firstly, many ecosystem services may take time to be regenerated and secondly, long-term contract may be more secure for the beneficiaries (Engel, 2016; Meyer et al., 2015; Moxey et al., 1999).

H8: regular third-party monitoring > one-off internal monitoring.

Monitoring systems that constrain compliance are expected to increase PES efficiency (Honey-Rosés et al., 2009). For example, most climate-related PES request regular third-party verifications of the emissions reductions claimed by ES providers (Bellassen et al., 2015). A trade-off between the costs and benefits of increasing monitoring stringency and frequency should however be considered (Bellassen and Shishlov, 2017), especially as monitoring/verification costs make up the bulk of the overall transaction costs of PES (Bellassen et al., 2015).

4.2.2 Material

An intensive search of studies has been conducted in a large range of databases in order to select the highest possible number of relevant studies for the meta-analysis. The search algorithm included a large number of keywords related to the ES targeted such as biodiversity, deforestation, water quality, etc. In order to restrict our selection to the studies that explicitly analyse the impact of PES, this main list of keywords was combined to two other lists of keywords that qualify the compensation for the ES targeted and the PES effect. We screened ScienceDirect (ScD), Web of Sciences (WoS), EconLit, AGRICOLA (USDA), JSTOR, AgEcon, Pascal and Francis (P&F) and OAIster during the period April-May 2019. Extensive internet searches on Google Scholar were also performed to select documents that were not identified in the databases consulted. We kept publications in English only as a majority of studies are in English language and to avoid a time-consuming translation procedure. After removing duplicate items, 4,898 documents were found from the literature search.

From the initial screening results, we identified 449 studies that dealt with the analysis of PES effectiveness. From this sample of eligible studies, we removed (i) literature reviews and meta-analyses, as these studies are syntheses of results from other articles that are included in our analysis, (ii) Ph.D. thesis and books, as parts of these works may be included in published articles, and (iii) ex-ante analyses and evaluations using simulation methods on fictive PES. We analysed the remaining 233 documents and excluded additional studies from the meta-analysis for various reasons: (i) the measure did not qualify as a PES even based on the broadest definition; (ii) lack of information on PES characteristics and/or on the study characteristics needed for the analysis; and (iii) inconsistency in the analysis (some studies assessing the impact of a PES on an environmental outcome not targeted by the scheme). A total of 114 documents that (i) present an ex-post analysis of the impact of (ii) a well-defined and implemented PES on (iii) the provision of a target ES, are included in this meta-analysis. The selection process is presented in Figure S1 in the Appendix.

The selected studies date from 1983 to 2019 and about 88% are published articles. The studies deal with a wide range of topics and were conducted in different locations (see Figure 1). A majority of studies worldwide concern PES schemes on biodiversity (32%) and deforestation (24%). The studies conducted in Europe mostly focus on biodiversity (67%), while a majority of studies in Central America, South America and Africa address deforestation. In the United States of America (USA), PES schemes targeting water quality (27%) and greenhouse gas (GHG) emission (27%) are those that are the most evaluated.

Each study in the sample may investigate more than one PES scheme and/or may provide results from different estimation methods or for different locations. In the database, we thus recorded the results obtained for the impact of a PES scheme and all of this information if they are available from the study. Our final sample includes 511 individual observations. On average, there are eight observations per study with a minimum of one (about 30% of the studies) and a maximum of 55 observations from a single study. Overall, the number of individual observations reporting a positive effect of PES on ES provision is greater than that reporting negative or null effects (Figure 2 panel (a), blue bars). Nevertheless, it should be noted that not all the studies report a level of significance for the results. Furthermore, the effect-size reported in some studies is not significantly different from zero at a 10% level. About 67% of observations reporting a negative effect and 63% of those reporting a positive effect have a p-value larger than 0.1 (as shown in Figure 2, panel (a), orange bars).

As stated before, our meta-analysis aims to investigate the impact of PES characteristics on their effectiveness considering both their enrolment and efficiency. However, not all the necessary information was always explicitly present in the articles selected for this analysis. We thus completed the database in two ways.

First, we added data on PES characteristics by running internet search on the PES under consideration or by consulting other documents that contain the necessary information on the PES. We coded a characteristic as missing when it was not possible to find the information. A study was excluded when information about the PES-schemes characteristics were not sufficient to test the hypotheses formulated in Section 4.2.1. The studies in the sample investigate the effectiveness of 150 PES schemes. Even if some of the PES schemes evaluated have the same name, we considered them different as long as some of the main characteristics were different. This can happen for PES schemes implemented in different locations. The distribution of the main characteristics of the PES schemes in the sample is presented in Section 4.2.4.

Figure 1: PES study topics by location

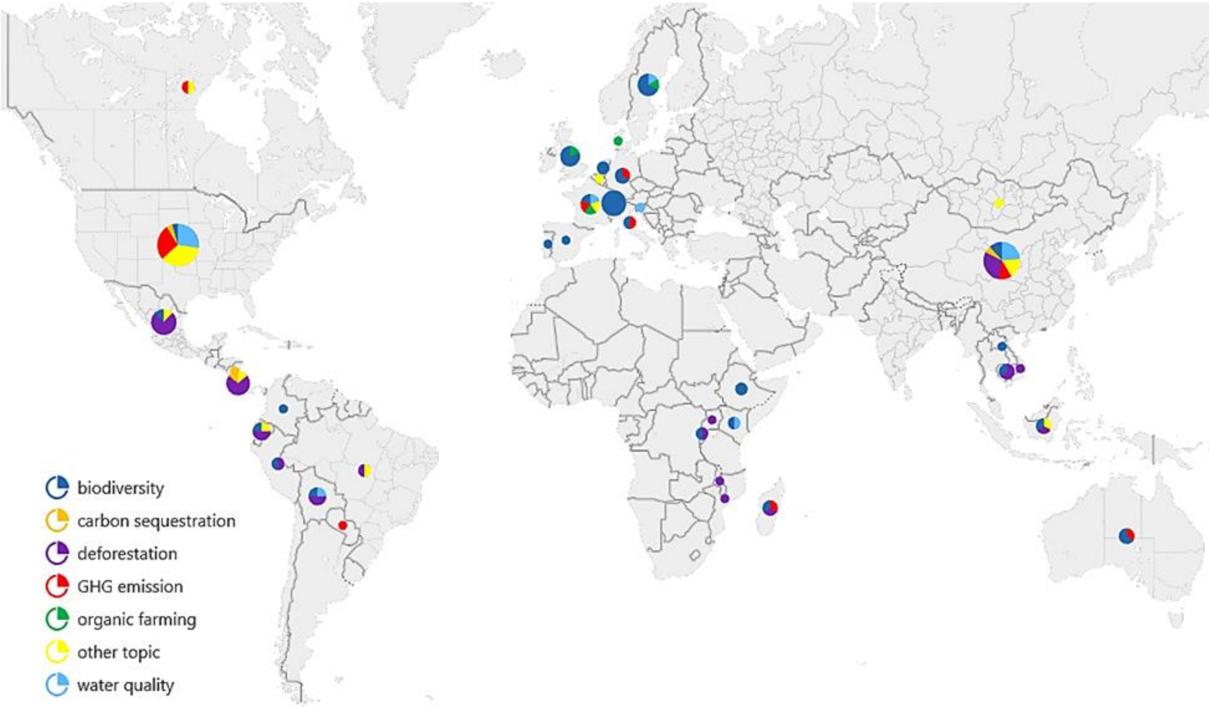
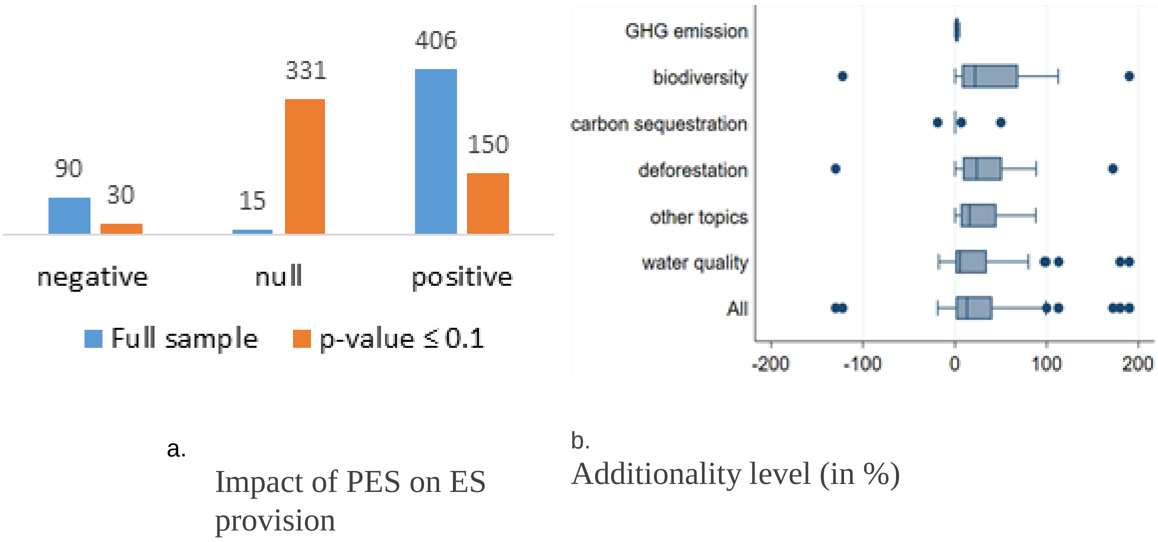


Figure 2: Reported relationship between PES schemes and ES provision (panel a), and additionality level by topics (panel b) from the sample of studies



Note: Statistics on panel (a) and (b) are for the full sample (N = 511) and for the sample with available information on additionality (N = 178), respectively. Organic farming is removed from the list of topics in panel (b) because there is only one observation. On panel (a), results are shown for the effect reported in the whole sample (blue) and after recoding status based on reported p-value (orange).

Second, we computed the additionality level of the PES dividing the treatment effect on the treated group by the observed level of the ES for this group when these two statistics were available but when additionality was not explicitly calculated. We also derived its standard errors using the delta method (Oehlert, 1992) and the p-value statistic for a normal distribution of the error terms. The level of additionality ranges from -130% to 172% with less than 8% of negative values (see Figure 2 panel b). Negative values mean that the PES actually decreased ES provision, an unintended outcome which rarely happens. This suggests a left censoring because it is uncommon to have a negative level of additionality. We control for this particular kind of censoring in our meta-regression. Among the 178 observations for which the additionality level was recorded in the database, we were able to derive standard errors for 105 observations which represent 26 individual studies in the sample. Values higher than 100% - happening only twice in our sample - mean that ES provision was not only efficient in the PES but decreased in the control group.

Additionality is very heterogeneous among the study topics. The average level of additionality for the PES schemes in the sample is 26%, with PES schemes targeting biodiversity (35%) and GHG emission (2.3%) exhibiting the highest and lowest levels, respectively.

4.2.3 Estimation procedure

We use meta-regression analysis (MRA) to investigate determinants of the effectiveness of the PES in our sample of studies - evaluated as the direction of the effect - and of their level of additionality. Because of the specificities in the structure of our data, we use different identification strategies and apply some adaptations of MRA. The general formulation of the model we estimate is:

$$Effect = f(P, X) \tag{1}$$

The effect of PES schemes is thus expressed as a function of the main characteristics of these PES schemes (*P*) presented in Section 4.2.1 and the study characteristics (*X*) that allow between-study heterogeneity to be controlled for. The data collected for the analysis contains studies using various designs, estimation techniques and types of data. This implies large between-study heterogeneity that should be controlled for in order for a single overall estimate of treatment effect to be obtained (Thompson, 1994; Hardy and Thompson, 1998). Our meta-regressions relate to two different outcome variables (the direction of the effect, and the level of additionality) each requiring specific modelling.

4.2.3.1 PES effectiveness

In order to keep all kinds of assessment methods, our study selection procedure keeps studies which provide a Boolean assessment of effectiveness (it worked/it did not work), without quantifying the level of ES provision. Therefore, we use a logistic regression in order to analyse the direction of the effect. Assuming that only a significant and positive effect of the PES schemes is desirable for policy makers, we group the studies reporting significantly negative effects and insignificant effects together. The outcome then takes the value 1 if the reported effect is significant at 5% and positive, and zero otherwise. The result from the logit model may be interpreted as the probability that a PES scheme achieves a significant and positive effect (i.e., increase ES provision) given its characteristics and controlling for between-study heterogeneity. Under the logistic distribution assumption, the model thus writes as:

$$y_i = Pr(P_i, X_i) = \frac{1}{1+e^{-A}} \tag{2}$$

where $A = exp(\sum_1^J \quad \beta_j p_{ij} + \sum_1^K \quad \gamma_k x_{ik})$; y_i is the effect of the PES scheme obtained by study i; P_i and $X_i$are the vectors of the PES characteristics and the study characteristics, respectively; β_j and γ_k are unknown parameters to be estimated. These parameters represent the contribution of each of the characteristics included in the model specification to the probability of observing a positive effect of PES schemes on ES provision.

We include results from unpublished studies to directly control for publication bias (Cook et al., 1993; MacLean et al., 2003; Rothstein et al., 2005). As a robustness check, we also investigate different levels of significance for a "yes" answer to the effectiveness question. It should be noted that only the confidence interval is reported for the estimated treatment effect in some studies. When this happens, we use these confidence intervals to recover a p-value.

In the model specification, we control for heterogeneity in the study topics and other study characteristics such as the study design, the type of data used and the estimation techniques. As a study may report more than one result, standard errors on the estimates were derived by clustering at the level of the individual studies.

4.2.3.2 PES additionality

Following Doucouliagos and Stanley (2009), we combine the FAT-PET with the MRA to investigate the impact of PES schemes characteristics on their levels of additionality. This method thus allows to control for both publication bias and between-study heterogeneity while estimating the impact of PES characteristics on additionality. From Equation (3), the FAT-PET-MRA model thus writes:

$$\widehat{\theta_i} = \theta_0 + \alpha SE_i + \sum_1^J \beta_j p_{ij} + \sum_1^K \gamma_k x_{ik} + \epsilon_i \qquad (3)$$

where $\widehat{\theta_i}$ is the estimated effect in study i, *SE* is the reported standard error; $p_{ij}$and x_{ik} are PES characteristics, respectively; θ_0, α , γ_k and $\beta_j$are parameters to be estimated; and ϵ_i is an *i.i.d*. error term drawings from a normal distribution.

As the data recorded on the additionality level of PES schemes is likely to be left-censored (Figure 2 panel b), a tobit specification is used. The evaluated schemes are considered to increase ES provision only if the level of additionality is higher than 0.005. We first estimate the level of additionality excluding the PES characteristics from Equation (3). A test for a significant effect of PES schemes is H0: $\theta_0 = 0$ which is the PET, and a valid test for publication bias is H0: $\alpha = 0$ which is the FAT. We then investigate the effect of PES characteristics on the level of additionality using the full specification of Equation (3).

The error term in Equation (3) can be divided into two components: $u_i \sim N(0, \tau^2)$ where τ^2 is the between-study or heterogeneity variance; and $e_i \sim N(0, {\sigma_i}^2)$ where ${\sigma_i}^2$ is the within-study variance (Stanley et al., 2017). The parameters of the model can be estimated in the tobit regression using as weights $\frac{1}{({\sigma_i}^2+\tau^2)}$, where ${\sigma_i}^2$ is the standard error of the estimates of the estimated effect in the study and $\tau^2$is the between-study heterogeneity parameter estimated from the sample of studies controlling for the study characteristics. Once again, the standard errors on the estimates are clustered at the study-level as multiple additionality levels may be reported in one study.

For robustness checks, we estimate the parameters from Equation (3) using two different specifications of the model. Firstly, we estimate the parameters with the commonly used random effects (RE) model also called the mixed model (Berkey et al., 1995). This model uses the same weighting system as in the tobit model by allowing the residual heterogeneity to be incorporated also assuming that individual variances are additive. The RE model is well suited in the presence of large between-study heterogeneity (Stanley and Doucouliagos, 2017). Secondly, we estimate the parameters by weighted least-squares (WLS). Conversely to the RE, the WLS uses the precisions of the estimates in the studies (i.e. $\frac{1}{\sigma^2}$) as weights.

4.2.3.3 Decomposing effectiveness into enrolment and efficiency

In an attempt to distinguish between enrolment drivers and efficiency drivers on the same sample, the meta-regression specification used for the assessment of effectiveness on the entire sample (Section 4.2.3.1) is reiterated on the studies for which an estimate of additionality is available (Section 4.2.3.2). Indeed, the enrolment effect of a given independent variable can be deduced from the estimates of its effects on effectiveness and efficiency. For example, if a variable is associated with higher efficiency but not with higher effectiveness, one can deduct that it is likely associated with lower enrolment, explaining why higher efficiency does not translate into higher effectiveness.

4.2.4 Explanatory variables and identification strategy

We use a set of explanatory variables to test the hypotheses stated in Section 4.2.1. We test hypotheses H1 to H7 using, respectively, the variables describing PES type (Coasean or not), PES objective (ES targeted or not), eligibility of ES provider (spatial targeted or not), payment mode (input-based, output-based or both), reference design (individual or not), payment type (cash, in-kind or both) and contract length (short-term, medium-term or long-term). We use variables that describe monitoring systems (self, third-party or none) and monitoring frequency (one-off or periodical) to test hypothesis H8. In addition to these characteristics of interest, we include in the specification of the models a number of control variables that capture other PES-schemes characteristics (see Table 1).

To deal with multicollinearity problems, we select the explanatory variables to be included in the specification of the models based on the variance inflation factors (VIF). We use a three-step selection approach by iteratively eliminating the variables that present the highest level of VIF. In step 1, we select study characteristics that allow between-study heterogeneity to be controlled for. In step 2, we eliminate the characteristics of PES schemes that show a high level of multicollinearity between them (VIF > 5, starting with the highest values). In step 3, we use the selected variables in step 1 and step 2 and discriminate only on PES characteristics based on the level of VIF. When this procedure contributes to eliminating specific indicators in categorical variables with more than two categories, we force the algorithm to group multinomial variables (e.g., contract length and monitoring frequency) in order to have more interpretable results. The criterion for a variable to be included in the models is that its VIF must be less than 5, which is the highest value generally accepted for the resulting coefficients to be interpretable (Akinwande et al., 2015).

Table 1: Variable description included in the full specification of the models and summary statistics

Variables	Modalities	Description	Mean	Std. Dev.
PES characteristics (hypotheses)				
PES type (H1)	*Coasean_pes*	if the PES is private-financed and ES are voluntarily provided	0.037	0.189
PES objective (H2)	*es_targeted*	if specific ES categories are targeted	0.769	0.422
Eligibility (H3)	*spatial_targeted*	if specific ES providers or specific locations are eligible	0.485	0.500
Payment mode (H4)	*output-based*	if the payment is directly linked to the ES provision	0.327	0.470
	input&output-based	if both input and output-based payments are used	0.010	0.099
Reference design (H5)	*individual_reference*	if the payment is estimated on the basis of individual reference	0.306	0.461
Payment type (H6)	*cash*	if the payment is made in cash	0.883	0.322
	cash&in-kind	if both cash and in-kind payments are provided	0.078	0.268
Contract length (H7)	*medium-term*	if the contract length is from 5 to 10 years	0.646	0.479
	long-term	if the contract length is more than 10 years	0.112	0.316
Monitor (H8)	*self*	if PES implementation is monitored by ES providers or private funders	0.084	0.277
	third-party	if PES implementation is monitored by public or independent auditor	0.869	0.338
Monitoring frequency (H8)	*one-off*	if the PES is monitored only at the end of the contract term	0.088	0.283
	periodical	if the PES is monitored at different steps of the implementation	0.864	0.343
PES characteristics (controls)				
Payment time	*after-delivery*	if payment is provided after the ES delivery	0.738	0.440
	upfront&after-delivery	if both upfront and after-delivery payments are considered	0.173	0.379
Payment constraint	*none*	if no payment constraint is imposed	0.725	0.447
	limited	if the maximum payment that the ES provider can receive is capped	0.241	0.428
Payment source	*discount*	if only a share of the total ES is paid for	0.030	0.170
	public	if the PES is linked to governmental or international programs	0.910	0.285
ES provider	*private-for-profit*	if PES is a private investment by local or non-local business	0.022	0.146
	community	if ES are provided by a local community	0.103	0.305
	firm	if ES are provided by a private firm	0.698	0.460

Note: For each variable, one modality is omitted and used as the reference for each variable; each modality can be considered as an indicator variable that take a value 1 if the condition is fulfilled, and zero otherwise; all variable modalities are presented on Figure S2 in the Appendix.

4.3 Results and discussion

4.3.1 PES effectiveness

The average predicted probability of a positive effect of PES schemes on the provision of ES varies across study topics (Figure 3): it is highest for organic farming (0.95) and lowest for carbon sequestration and water quality (0.72 and 0.70 respectively). These probabilities are all different from zero at a 10% level of significance. However, only the predicted probability for PES schemes targeting organic farming is different from the others. Pairwise comparisons with Wald tests for all the categories of topics are reported in Table S1 in the Appendix.

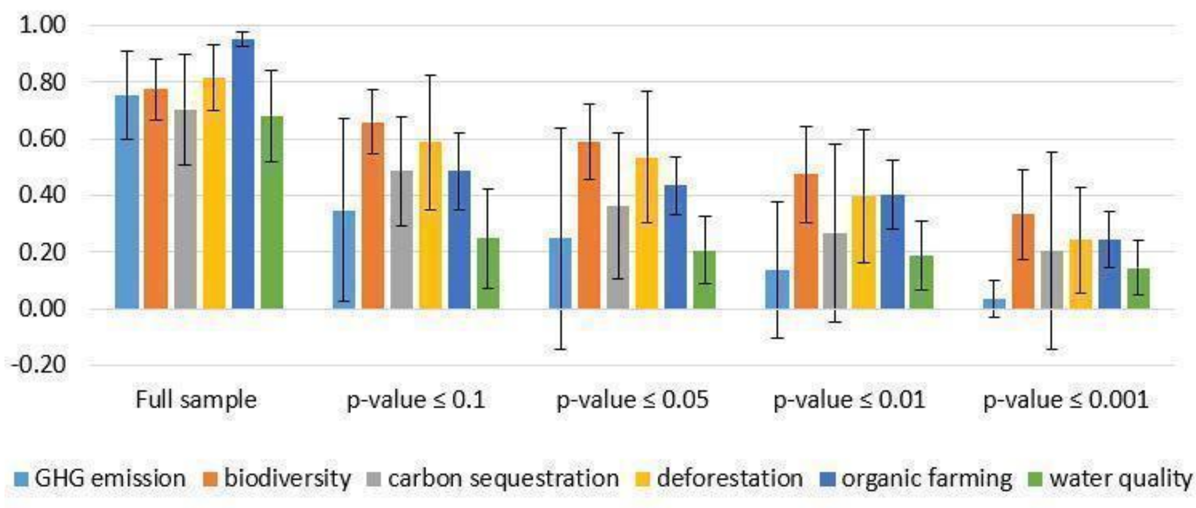
While the average probability of a positive effect of PES schemes logically decreases with the p-value threshold to consider an effect to be significant and positive for all the topics (Figure 3), it remains significant whatever the level of significance used as threshold for biodiversity, deforestation, organic farming and water quality. Carbon sequestration and water quality are consistently at the lower end of effectiveness, but so are PES aimed at reducing GHG emissions when a p-value threshold is considered. At the higher end, organic farming is surpassed by PES targeting deforestation and biodiversity (H2: PES objective).

Considering only the reported effects that are significant at a 1% level, the predicted probabilities of a positive effect are 0.47, 0.40, 0.40 and 0.19 for PES schemes targeting biodiversity, deforestation, organic farming and water quality, respectively. There again, these predicted probability levels are not different from each other at a 5% significance level except for water quality that shows a lower probability level than other topics.

The marginal effects of all the explanatory variables and for various levels of significance for the reported effects are provided in Table S2 in the Appendix. The reported directions of the effects of PES schemes on the provision of ES do not differ across study type. The estimated coefficient of the indicator variable for peer-reviewed studies is not significantly different from zero at a 5% level in four of our five settings. Moreover, this indicator variable is never significantly different from zero when the PES characteristics are included in the model specification which confirms the absence of publication bias. Best designed studies (randomised control trials and natural experiments) and studies using individual data (as opposed to aggregated data) are also more likely to conclude that a PES is effective. This likely results from the suppression of attenuation bias: when the outcome is measured more precisely, the estimated effect increases.

Figure 4 presents the marginal effect of the characteristics of PES schemes that allow our hypotheses to be tested on the probability of increasing ES provision (effectiveness). The results from the full sample and when selecting a demanding 0.001 p-value threshold for a positive effect of PES schemes on ES provision should be interpreted with caution. Indeed, the full sample contains some observations that do not have a high level of significance while 0.001 may be too restrictive a threshold. The conclusions are thus mostly based on the settings with p-value thresholds between 0.1 and 0.01.

Figure 3: Average probability of a positive effect of PES schemes on ES provision with a 95% confidence interval by topic from a logistic regression using the full sample and the subsample with a p-value statistic, using different p-value thresholds for the definition of an effective PES

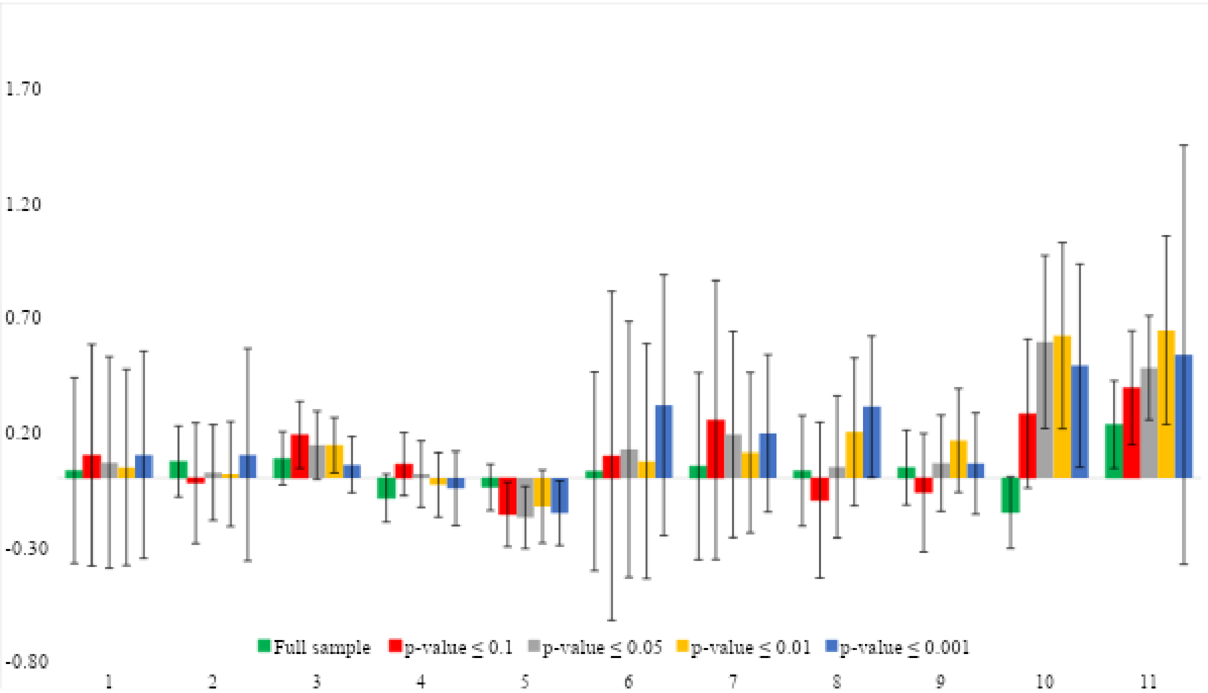


As expected, the PES schemes that target ES providers, either within predefined intervention areas or based on individual characteristics, seem to be more likely to increase ES provision in the zones of implementation (H3: spatial targeting +). However, the estimated effect is significantly different from zero only with a p-value threshold between 0.1 and 0.01. In these settings, the probability of achieving positive outcomes increases from 14% to 19% when the PES-schemes target specific intervention areas. This confirms that spatially targeted PES schemes are more effective than PES schemes that use a horizontal eligibility strategy.

Also, as expected, output-based payments do not seem to be more effective than input-based payments (H4: result-based payments = practice-based payments). The coefficient is not significantly different from zero at a 10% level whatever the p-value threshold considered for the estimation. This result would confirm that other design features seem to dominate the expected effect of result-based payments, or that the definition of "result-based" may not be precise enough to allow for a proper impact to be detected (Bonvillain et al., 2020).

Type of monitoring and monitoring frequency are also found to be positively correlated with PES effectiveness (H8: regular third-party monitoring > one-off internal monitoring). The results show that this probability increases up to 62% when the monitoring of ES provision is undertaken by a third-party (as opposed to the ES provider itself), and by 65% when schemes provision is monitored periodically (as opposed to once and for all at a given stage of project implementation). This confirms that more constraining monitoring systems help to ensure compliance with the objectives of the PES-schemes and thereby increase ES provision. Moreover, the monitoring system seems to be the main characteristic that influences the performance of the PES schemes in our sample of studies. These indicator variables present the highest marginal effects with the highest level of significance in almost all the settings.

Figure 4: Marginal effects of PES-schemes characteristics on the probability of a positive effect on ES provision from a logistic regression with a 95% confidence interval. Results are shown for the full sample and for each subsample with specific p-value statistics used as the threshold for a significant effect-size



However, the type (H1: Coasean PES > Pigouvian PES) and specificity of the PES objective do not seem to play a role on their probability of achieving positive outcomes. Neither does the type of payments (H6: cash vs in-kind) and the presence of a payment constraint (e.g., maximum amount or stringent reference level). One explanation for these non-intuitive results may be that the expected effect is captured by other PES characteristics such as the monitoring system used to ensure compliance with the objectives of the PES schemes. Indeed, both non-compliance and free-riding problems may be avoided with good monitoring systems. Furthermore, free-riding problems could be captured through the variable ES provider which does not show significant effect on the probability of providing positive outcomes. Even if these variables present low to moderate level of correlation – except for PES type and payment source that show high correlation – the chi-square tests are statistically significant (see Table S3 in the Appendix).

Conversely to our expectation, we find that individual reference is negatively correlated to the probability of PES schemes to increase ES provision (H5: reference design). Indeed, the results show that payments based on individual reference may decrease the probability of a positive effect by about 17% compared to PES schemes that use a generic reference. This unexpected result may be caused by the much discussed possibility for project developers to game regulators where an individual reference has to be estimated (Dechezleprêtre et al., 2014; Shishlov and Bellassen, 2012). However, this result must be interpreted with caution because different types of references were grouped in one indicator to avoid multicollinearity problems: customised references based on historical provision were grouped with customised references based on a projected scenario, such as carbon offset schemes where the reference emissions are the emissions of the most profitable alternative to the project.

Contract length may be positively correlated with the probability to increase the effectiveness of the PES schemes (H7: contract length). However, the coefficient for the indicator of long-term contract is significantly different from zero only when the p-value threshold selected to identify effective PES is 0.001. The non-significance of the effect of contract length may be because short-term and even medium-term PES contracts may be subject to multiple renewals as for some of the PES implemented in the EU. The participants may internalise this by anticipation and multiple renewals of the short- and medium-term contracts may also contribute to increasing the probability of achieving positive outcomes.

Finally, we also find that PES-schemes implemented in Asia are likely to be more effective than those in the other continents. However, this effect is significantly different from zero only with a p-value threshold at 10% level. This positive effect may be explained by the fact that a large number of these evaluated PES schemes are implemented in China, where the participants to the schemes may be constrained to compliance due to the possibility of sanctioning (Hou et al., 2021). Another possible explanation is that in Africa and Asia the participants to the PES schemes may be poorer than the beneficiaries from other continents such as Europe and North America. As long as payment is constrained by compliance, these poorer ES providers may have a higher incentive to participate.

4.3.2 PES additionality

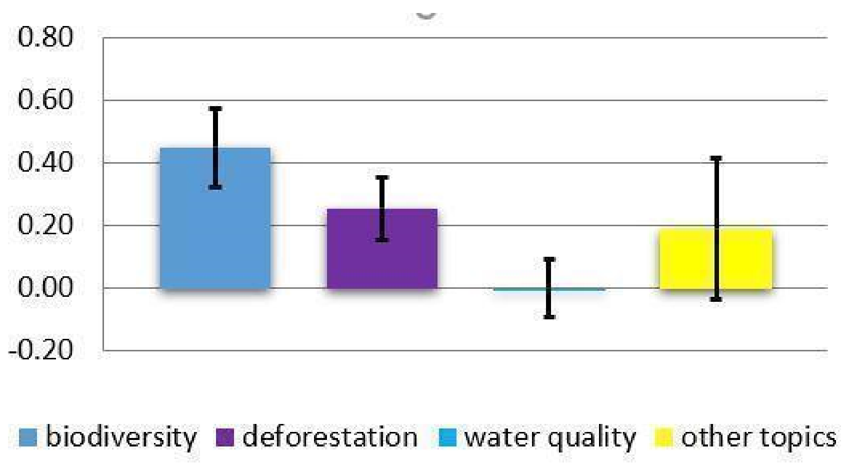
As presented in Section 4.2.3.2 for PES additionality, we estimate the FAT-PET-MRA using a tobit specification to analyse PES additionality in the subsample of studies (N = 90) where it is estimated or can be derived from provided estimates. Given large between-study heterogeneity and high level of correlation between study topics and PES characteristics, we group the seven topics in few categories. We focus only on biodiversity, deforestation and water quality, as they account for more than 90% of the sample, and we group the other topics in the category “other topics”.

We do not find evidence of publication bias in our sample of studies when the between-studies heterogeneity is controlled for using study topics. The estimated coefficient for the standard error included in the model specification is not significant at a 10% level. Both the RE and the WLS specification confirm this result (see Table S3 in the Appendix). This suggests that the asymmetry observed in the Funnel plot is due to the high between-study heterogeneity (see Figure S3 in the Appendix). However, the parameter measuring the between-studies heterogeneity (I2 statistic) is close to 100%, suggesting very high between-studies heterogeneity even when we control for the observed study characteristics. This justifies our strategy to include the resulting heterogeneity parameter (τ) in the weighting scheme to control for excess of heterogeneity in our tobit estimation. These results are confirmed by the RE specification but not by the WLS. The results from the WLS specification should however be interpreted with caution as the VIF values for some of the variables are very large given that the weighting factor (1/2) used in the specification is different from the tobit model.

PES schemes achieve different levels of additionality according to the type of ES targeted (Figure 5 and Figure S4 in the Appendix). PES schemes that target biodiversity provide the most additionality with a level of 45%. This means that ES provision is expected to increase by almost a half the nominal amount paid to ES providers. The PES schemes that target deforestation present the second highest level of additionality with about 25%. While the estimated level of additionality for biodiversity and deforestation are highly significant, the values are not significantly different from each other. Pairwise comparisons with Wald tests confirm this result at a 5% significance level (see Table S2 in the Appendix).

Conversely, the level of additionality is not significantly different from zero at a 10% level for both water quality and the other topics taken all together. One possible explanation for the non-significant effect for PES schemes targeting water quality is that the reduction of inorganic inputs imposed by certain schemes such as organic farming encourages the use of manure that may cause an increased runoff (Torstensson et al., 2006).

Figure 5: Average additionality level of PES schemes with a 95% confidence interval by topic from a tobit regression



Given our small sample of studies (N = 90), few variables other than those used to control for between-studies heterogeneity are selected based on the VIF statistics because of a persistence of multicollinearity problems. The data allows only six out of our eight hypotheses to be tested (H1, H3, H4, H5, H6 and partly H8 since only monitoring frequency is selected). The results are reported in Table 2. Among the variables retained in the model specification, the type of PES schemes (H1), the type of payment (H6), the monitoring frequency (H8) and the zone of implementation seem to play a role on the level of additionality. Furthermore, payment mode does not have an effect on the additionality level, not rejecting H4. These results are confirmed by the RE specification but not by the WLS. There again, the results from the WLS specification should be interpreted with caution.

The results confirm our expectation that output-based payments do increase neither PES efficiency, nor effectiveness (H4). The coefficient is not significantly different from zero at a 10% level. Also, as expected, both PES and payment types are positively correlated to the level of additionality. Coasean PES are about 31% more additional than Pigouvian ones (H1) which points the role of voluntariness in achieving additionality. However, the Coasean vs Pigouvian distinction may also capture the effects of other characteristics of the PES schemes, such as monitoring mode, funding source as these variables show a high level of correlation (see Table S5 in the Appendix). Indeed, Coasean PES are usually applied on a smaller scale, targeting fewer providers, with more local knowledge, which may also explain this higher additionality than Pigouvian schemes. Likewise, providing payments only in cash for ES provision is expected to increase the level of additionality by about 43% (H6). An explanation for this result is that payments in cash may be considered as more enticing than in-kind rewards. Then, it may increase participation in the PES schemes and the level of additionality all other things being equal. Alternatively, it could be explained by unobserved components of the level of scrutiny, as ES buyers may be warier about efficiency when they are doling out cash than when they are paying in-kind. Moreover, results indicate that PES schemes implemented in Africa are more additional than those implemented in other geographical contexts, which may be related to the point raised in the case of PES effectiveness about the financial status of ES providers.

Finally, we find that PES schemes are less additional when the implementation is monitored periodically, a reverse result compared to what we find for ES provision (H8). Let us note that since PES effectiveness depends on both enrolment and additionality level, a different sign on effectiveness and efficiency is not impossible. To better assess this potential trade-off between effectiveness and efficiency, we run again our effectiveness analysis on the restricted sample use to assess efficiency. In this restricted sample, monitoring frequency is no longer associated with higher effectiveness.

4.3.3 Enrolment/Efficiency trade-offs

In an attempt to distinguish between enrolment drivers and efficiency drivers on the same sample, the meta-regression specification used for the assessment of effectiveness on the entire sample in Section 4.4.1 is reiterated on the studies used in Section 4.4.2, that is where an estimate of additionality is available (Tables 2 and 3).

The only design feature that has a clear impact on PES effectiveness in this smaller sample is the provision of the ES by communities rather than firms or individuals. It is likely also influential in the full sample, but it may be superseded there by funding source, a variable which is highly correlated with community provision (Table S5 in the Appendix).

In principle, putting these restricted results on effectiveness in perspective with the results on additionality (Table 1) is a more interesting purpose. For example, Coasean PES which are clearly associated with higher efficiency are not so clearly associated with higher effectiveness, indicating a potential trade-off where higher efficiency is obtained at the cost of lower enrolment. In practice however, the small sample size combined with the Boolean nature of the dependent variables reduces the power of the statistical tests we perform and we cannot test the impact of monitoring frequency and cash payments because of too little variability in the restricted sample. Indeed, all but two PES are implementing a periodical monitoring, and only one is offering in-kind payments. Our results should therefore be interpreted with precaution until a larger sample of studies is available.

Table 2: Meta-regression of PES-schemes additionality level on PES characteristics

Variables		Tobit	RE	WLS
	intercept	0.600**	0.460*	0.654**
		(-0.263)	(-0.247)	(-0.258)
	standard_error	0.069	0.258	-0.108
		(-0.240)	(-0.185)	(-0.612)
PES characteristics				
PES type (H1)	*Coasean*	0.312**	0.272*	0.450
		(-0.126)	(-0.133)	(-0.266)
Eligibility (H3)	*spatial_targeted*	-0.071	-0.041	-0.001
		(-0.066)	(-0.058)	(-0.002)
Payment mode (H4)	*output-based*	0.008	0.068	0.000
		(-0.059)	(-0.051)	(-0.002)
Payment type (H6)	*cash*	0.432**	0.466**	0.319
		(-0.188)	(-0.214)	(-0.247)
Reference design (H5)	*individual*	-0.175	-0.121	-0.013
		(-0.114)	(-0.101)	(-0.011)
Provider	*community*	0.133	0.098	0.103
		(-0.115)	(-0.111)	(-0.219)
Monitoring frequency (H8)	*periodical*	-0.757***	-0.732***	-0.925***
		(-0.137)	(-0.159)	(-0.006)
Study zone	*Africa*	0.341***	0.234**	0.138
		(-0.105)	(-0.083)	(-0.246)
Study characteristics				
Study topic	*biodiversity*	0.038	0.088	0.190***
		(-0.141)	(-0.124)	(-0.042)
	deforestation	0.039	0.061	0.091**
		(-0.095)	(-0.092)	(-0.042)
	water_quality	-0.455**	-0.356**	-0.507**
		(-0.185)	(-0.162)	(-0.226)
Study design	*best_design*	-0.092	-0.103	-0.278
		(-0.060)	(-0.077)	(-0.205)
Estimation method	*robust_est*	0.073	0.084	0.245
		(-0.062)	(-0.077)	(-0.219)
Data type	*aggregate*	0.197*	0.159	0.464*
		(-0.103)	(-0.094)	(-0.230)
	var(e.additionality)	0.026***		
		(-0.009)		
Observations		90	90	90
Pseudo/adjusted R-sq		0.795	0.569	0.992

Note: all the dependent variables were selected based on the VIF using the tobit model; var(e.additionality) is the estimated variance of the regression; standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 3: Effects of PES characteristics on PES effectiveness (restricted sample)

Variables		p-value ≤ 0.1	p-value ≤ 0.05	p-value ≤ 0.01	p-value ≤ 0.001
PES characteristics					
PES type (H1)	*Coasean*	-0.068	0.019	-0.074	0.434**
		(0.141)	(0.177)	(0.318)	(0.215)
Eligibility (H3)	*spatial_targeted*	0.203*	0.117	-0.127	-0.071
		(0.116)	(0.161)	(0.184)	(0.140)
Contract length (H7)	*long_term*	-0.036	-0.024	0.150	0.271*
		(0.181)	(0.183)	(0.224)	(0.153)
Payment mode (H4)	*output-based*	-0.013	-0.140	-0.157	-0.342**
		(0.092)	(0.220)	(0.318)	(0.164)
Reference design (H5)	*individual*	0.021	0.257	0.254	0.092
		(0.118)	(0.178)	(0.195)	(0.103)
Provider	*community*	2.697***	0.519**	0.401	0.315**
		(0.397)	(0.257)	(0.408)	(0.129)
Study zone	*Asia*	-2.559***	-0.334	-0.014	-0.545**
		(0.399)	(0.294)	(0.451)	(0.233)
Study characteristics					
Study topic	*biodiversity*	0.229	0.222	0.263	0.541**
		(0.189)	(0.175)	(0.308)	(0.236)
	deforestation	0.330*	0.035	-0.016	0.098
		(0.181)	(0.205)	(0.357)	(0.303)
	water_quality	0.173	-0.020	-0.092	0.277
		(0.244)	(0.283)	(0.320)	(0.289)
Study design	*best_design*	-0.217	-0.193	-0.193	0.168
		(0.158)	(0.171)	(0.225)	(0.162)
Estimation method	*robust_est*	0.141	-0.061	-0.058	0.149
		(0.139)	(0.126)	(0.226)	(0.139)
Data type	*aggregate*	0.075	0.123	-0.040	0.119
		(0.091)	(0.149)	(0.211)	(0.154)
Observations		85	85	85	85

Note: The dependent variable (ES increase) takes the value 1 if the PES effect is positive and zero otherwise; estimation in columns use samples with the dependent variable set to 1 if the p-value statistic is ≤ 0.1, ≤ 0.05, ≤ 0.01 and ≤ 0.001, respectively; five observations are removed from the initial sample by the logit model; standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

5 Does European AES affect ecological practices' adoption and farm performance? Evidence from quasi-experimental analysis in France

This section aims to estimate the additionality of ecological policies put in place in Europe (organic farming; AECM) on the adoption of ecological practices, labour demand and farm economic performance.

5.1 Context and issues

Since the 1990s, the CAP has evolved its objectives to respond to environmental and climate issues of the agricultural sector. The consideration of these issues was mainly conducted by the introduction of agri-environmental and climate measures (AECM). AECM are financial incentives for farmers who are committed to adopting or maintaining ecological practices. The logic that underpins these measures is that during the adoption or maintaining such practices that will benefit the entire society, farmers support costs individually (Kuhfuss and Subervie, 2018). The amounts of these AECM are calculated in such a way as to offset the average costs generated by the implementation of the obligations of the specifications and the losses of agricultural income related to the adoption of ecological practices.

There are four types of AECM (Duval et al., 2016): entry-level measurements, localised measures, genetic resource preservation measures, and farm-system measures. The entry-level measures and localised measures have been preferred in previous programming. They consist mainly in maintaining the viability of extensive grassland systems for the former, and to respond to a specific environmental issue in an explicitly defined geographical area for the latter.

Since entry-level measures are experiencing significant participation, their impact on the maintenance of practices (e.g. maintenance of grasslands) is quite low (Chabé-Ferret and Voia, 2021). The results of the evaluation of localised measures highlight a fairly low participation of farmers in these measures (Lastra-Bravo et al., 2015). In addition, estimates of their impacts on changes in practices are very heterogeneous and relatively low (Chabé-Ferret and Subervie, 2013).

Environmental gains can also be compromised by unexpected behaviours, for example when localised measures increase the profitability of agricultural production compared to other land uses and can then induce transfers to the detriment of forests or semi-natural areas (Claasen et al., 2008). The causes of these "substitution" effects can be exacerbated by the targeting of localised AECM, which are engaged on part of the farmer's plots and target areas where the environmental challenges are greatest.

By offering farm-level commitments, system AECM help contain substitution effects, at least at the farm level. The "production system" logic of these AECM seeks to support a transition of the entire operation to a more sustainable system. Unlike localised measures, they are designed to limit the pressure of agricultural activities on the environment. This mainly involves modifying farming practices and changes in rotations. In addition, the main innovation in terms of AECM compared to the former programming period 2007-2013 is the introduction of a set of AECM systems to cover the majority of production systems (crops, grassland and pastoral systems, mixed cropping and livestock farming).

If the localised AECM have already been the subject of several evaluations around the world (*cf.* Uthes and Matzdorf, 2013; for a review of the literature) and in France during previous programmes, the work of evaluation of AECM based on system logic is not widespread. This study aims to identify the effect of adoption of an AECM and organic farming on the adoption of ecological practices (corresponding to an estimation of additionality), labour demand, economic performance and farm income (corresponding to an estimation of the real private cost of adoption).

But assessing this is far from being trivial since the participation of farms in an AECM is voluntary. Farmers with the lowest cost to meet the program requirements are generally more likely to participate in the AECM. Consequently, the AECM could generate windfall effects and compensate certain farmers for a change/maintenance of practice that they would have made in the absence of participation in the scheme. Thus, it is important to take into account the potential presence of adverse self-selection. We use a difference-in-difference (DID) propensity score matching to control for the selection bias. DID-matching combines a matching procedure with first-differencing with respect to a pre-treatment period (Blundell and Costa-Dias, 2000).

The rest of this study is organised as follows. First, we present the empirical strategy used to identify these effects. Then we present our data and our sample of farms. Section 5.3 is dedicated to the analysis of the results. Finally, Section 5.4 discusses the results.

5.2 Empirical strategy

The propensity score matching methods are generally applied to assess the impact of a policy in a non-experimental setting, i.e. when the assignment of the policy evaluated is not random. Their use is very common in econometric work (Angrist and Krueger, 1999) and becomes increasingly developed in environmental (List et al., 2003) and agricultural economics (Pufhal and Weiss, 2009; Behaghel et al., 2019). We first present the evaluation problem, then the propensity score matching methods used in this study.

To assess the effect of an AECM on a farm i (Δ_i), it would be necessary to observe, at the same time, the level of practice[4] of the same farm i with and without the AECM (Smith and Todd, 2005). In this context, the effect of the scheme would be the difference between the two levels of practice:

$$\Delta_i = Y^1{}_i - Y^0{}_i \tag{4}$$

However, we only observe one of the practice levels for each farm. For a farm i adopting an AECM, we observe $Y^1{}_i$ but not $Y^0{}_i$, while the observed level of practice for a non-beneficiary farm i is $Y^0{}_i$. The aim of the propensity score matching is to estimate the level of unobserved practice (defined as the counterfactual result), i.e. which would have been observed if the AECM had not been implemented. To this end, it is necessary to be able to identify non-beneficiary farms which would differ from beneficiary farms in only one dimension: their choice to participate in an AECM, to control for adverse self-selection.

We want to estimate the average effect of the AECM on the level of practice of the beneficiaries (effect of the scheme on beneficiaries, average treatment effect on the treated - ATT), which can be written as:

$$\Delta_{ATT} = E(Y^1{}_i - Y^0{}_i | D_i = 1) \tag{5}$$

D_i is a binary variable defining the participation of the farm i in AECM. $D_i = 1$ if the farm participates in the AECM evaluated, and 0 otherwise.

The average effect of the AECM on beneficiaries is the difference between the average level of practice with the AECM $E(Y^1{}_i | D_i = 1)$ and the result that beneficiaries would have received if they had not benefited from the AECM $E(Y^0{}_i | D_i = 1)$. While the first term is observed, the second is not. A credible estimator of this counterfactual level of practice must therefore be found.

[4] Defined as potential levels of practice.

A first assumption in this framework is the absence of externalities from the AECM (Stable Unit Treatment Value Assumption - SUTVA), which makes it possible to reduce the number of potential levels of practice to two in the case of an evaluation of participation in an AECM as presented above. The participation of the farm i only affects its own level of practice, not those of other farms. This hypothesis implies the absence of diffusion effects of practices between beneficiary and non-beneficiary farms, directly caused by the AECM (mimetism or diffusion effects). In the context of our study, this hypothesis seems credible in the context of the evaluation of the majority of the AECMs evaluated in this study, as it is likely that farmers who decide to change their practices will adopt an AECM in order to be paid for this transition.

The SUTVA hypothesis also implies that AECM do not influence the equilibrium on markets, especially the land market. In this case, we can envisage that the evaluation of AECM based on grassland conservation might not respect this hypothesis if it generates land transactions (rent, sale) between beneficiary and non-beneficiary farms - e.g. a transfer of a grassland plot from a non-beneficiary farm to a beneficiary farm.

In a non-experimental setting, the use of all non-beneficiary farms as an estimate of the counterfactual practice level of beneficiary farms poses a problem of selection bias because the voluntary participation of farms implies that the scheme is not randomly assigned. The beneficiary and non-beneficiary groups differ not only in the status of the scheme (participation in AECM) but also in other characteristics. The presence of variations (between groups) in the characteristics simultaneously influencing the decision to adopt an AECM and the counterfactual practices creates a source of confusion, making it impossible to interpret the difference in practices observed between the two groups as the effect of the AECM on the level of practice.

When the assignment mechanism is based on the characteristics of the farms that we can observe, matching on observable characteristics (Rubin, 1977) is the most frequently used method. Matching consists in associating each farm benefiting from an AECM with a farm that is not and whose observable characteristics X_i are similar. The aim is to approximate a controlled experiment by ensuring that the control group is as similar as possible to the beneficiary group in terms of the distribution of variables affecting the probability of benefiting from the policy. The identification of the causal effect of AECM on beneficiaries is based on the assumption that their selection is independent of potential outcomes, conditional on their characteristics (conditional independence assumption):

$$Y^0{}_i \perp D_i | X_i \tag{6}$$

In our case, this is equivalent to assuming that the participation of farm i is based on its observable characteristics only. Our choice of a counterfactual for each beneficiary is based on the estimation of a propensity score $p(X_i)$ (Rosenbaum and Rubin, 1983). This is a two-step method where we first estimate (e.g. using logistic regression) the probability of benefiting from the AECM for the whole sample, before matching the farms on the basis of this probability (propensity score). We thus reduce the matching to the most relevant dimension to deal with selection bias, i.e. the dimension of participation in the evaluated AECM. If the assumption of conditional independence is verified for the control variables, then the potential levels of practices are also independent of participation in the AECM, conditional on the propensity score (Rosenbaum and Rubin, 1983):

$$Y^0{}_i \perp D_i | p(X_i) \tag{7}$$

On the other hand, the estimated propensity score must effectively capture the differences in observable characteristics between the beneficiary and non-beneficiary farms (propensity score balance assumption). To ensure a good quality of match, a majority of the beneficiary farms must also have similar characteristics to those farms in the non-beneficiary group (common support assumption):

$$0 < P(D_i = 1|X_i) < 1 \tag{8}$$

To restrict our sample to the common support, we use the 'min/max' method which consists in discarding beneficiary farms whose propensity score is higher than the maximum score observed for non-beneficiary farms (Dehejia and Whaba, 1999).

The general form of the propensity score matching estimator of the average effect on beneficiaries is the average of the differences between the observed level of practice of each beneficiary farm i and a weighted average of the observed levels of practice of non-beneficiaries j matched to each beneficiary farm i:

$$\hat{\Delta}_{ATT} = \frac{1}{n_I}\sum_{i\in I\cap CS} \left(Y_i^1 - \sum_{j\in J} w_{ij}Y_j^0\right) \tag{9}$$

Y_i^1 is the observed level of practice for group I of beneficiaries, Y_j^0 is the level of practice in the absence of AECM, CS is the common support, n_I is the number of farms *i* in group I of beneficiaries, j the farms in the J group of non-beneficiaries and w_{ij} a weighting (function of the distance between the propensity scores of i and j) characterised by the selected matching algorithm.

There are several algorithms to conduct the matching. We will use the most common one which consists of matching each beneficiary farm to its k nearest neighbours (on the propensity score) from the non-beneficiary group. The number of neighbours is fixed for all beneficiary farms. Our estimates were made for $k = \{1,5,10,20\}$.

We also implemented a “caliper” matching consisting in selecting, for each beneficiary farm i , all the non-beneficiary farms located in a fixed $+/-h$ interval around i. In other words, the difference between the score of the treated and untreated farms must not exceed a given threshold[5].

The advantage of the caliper algorithm is that it limits the frequency of poor quality matches compared to matching to the k nearest neighbours (the distance between the beneficiary farm i and its k^{th} neighbour can be large). Studies comparing different matching algorithms have tended to find that calibration matching provides a better reduction in selection bias (Baser, 2006; Caliendo and Kopeinig, 2008).

Despite the richness of the control variables we use, some unobservable determinants can simultaneously influence the decision to adopt an AECM and farm practices (or performances); therefore a selection bias can remain. These determinants could lead to a selection bias on the non-observables. To limit this problem, we combine the propensity score matching approach with a Difference-In-Difference (DID), as suggested by Blundell and Costa Dias (2000). This approach consists of subtracting ATTs before and after the start of the evaluated scheme in order to remove time invariant unobserved bias. Propensity score matching eliminates selection bias due to observed covariates by comparing recipient farmers to similar non-recipients while DID control for selection bias due to time-invariant unobservable factors (Abadie, 2005).

5.3 Data and sample

The estimation of the additionality of the AECM requires to build a representative database, including beneficiaries and non-beneficiaries, before and after the implementation of the scheme studied. It is

[5] The value of *h* has been defined as half the standard deviation of the propensity score.

essential to observe the characteristics of each farm (whether or not it is a beneficiary of the AECM being evaluated) before the system is put in place. As a single database containing all this information does not exist, we have to build it by merging several databases.

Our study requires two kinds of data. First, we need data on the participation of farms in the different schemes (participation variables). Second, we need to be able to describe the participation of farms in the schemes on the basis of their individual characteristics (control variables).

Finally, the evaluation requires the construction of outcome variables on which we will measure the additionality of AECM. For each farm, it is necessary to know the practices before and after the start of the programme, as well as the observable characteristics of the farm before the start of the programme, which requires a matching between the different data sources. This matching requires a common variable in each database.

Our database is made up of data from the French FADN and administrative data on AECM contracts. The FADN database, implemented in France since 1968, is based on an annual survey carried out in the Member States of the European Union (EU) according to common rules and principles. In particular, it makes it possible to collect accounting and technical-economic data with a view to analysing the diversity of incomes and their formation on agricultural holdings, to draw up economic and financial diagnoses and to simulate the impact of measures taken under the CAP. The main subjects dealt with the structures (surface area and livestock) of farms, the workforce (number of employees, working hours, type of contract, etc.), and the socio-demographic characteristics of farm managers (age, education, gender, etc.). The FADN also provides data on the economic results of production, on intermediate consumption, on value added, etc. Finally, it contains elements of the balance sheet of the fixed assets of the farms (land, equipment, breeding animals), current assets, capital, debts, income and expenses. The data are collected through the agricultural accounting offices by direct input into a national computer application, and collection requires the consent of the farmer. The scope of the FADN includes agricultural holdings operated by an accountant and having a certain economic size. These holdings must cover more than 95% of the country’s agricultural potential. In practice, these are farms whose standard gross production (SGP) is at least 25,000 Euros in metropolitan France and 15,000 Euros in the overseas administrative regions. The variables used to characterise the farms before the start of the program will be built from the 2010 French agricultural census (RA2010). Made every 10 years by the French Ministry of Agriculture, the agricultural census exhaustively describes the French agricultural landscape. In 2010, the agricultural census concerned nearly 475,000 farms. The data collected relates to land use, livestock, labour and other information on the management of the operation. Thus, the majority of control variables come from this database.

The AECM data relate to the commitments contracted by farmers per campaign within the framework of the AECM and aid for the conversion and maintenance of organic farming under the 2014-2020 third programming of the European Rural Development Regulation (RDR3). We also use data on the beneficiaries and payments of these aids between 2007 and 2014, which corresponds to the period of application of the second programming of the European Rural Development Regulation (RDR2). Made available by the “Observatoire de Développement Rural” (ODR, Observatory of Rural Development) (US-ODR INRAE), the contracting and payment data come from the “Agence des Services et Paiements” (ASP). Our analysis focuses on the 2010-2020 programming of the RDR2 and RDR3, the data includes the contractual payments and areas of each beneficiary for the maintenance and conversion to organic farming aid, agri-environmental measures dedicated to farm systems (crop, mixed, livestock/dairy), and AECM with localised issues.

More specifically, we study the following AECM:

- AECM “SHP” consists in maintaining practices designed to preserve the sustainability and agro-ecological balance of permanent grasslands with diversified flora and certain pastoral areas,
- AECM “SPE” and “SPM” consist in making farms evolve towards a better interaction between the livestock and crop activities, including a better feed autonomy; favouring the production of grass; a diversification of crop rotation and a lengthening of rotations; an optimised management of N fertilisation,
- AECM “HERBE” consists in managing grasslands in a more extensive manner (delayed mowing, maintenance of grasslands with high biodiversity, wetlands, etc.),
- AECM “PHYTO” consists in reducing crop protection products or implementing direct seeding on cultivated land,
- AECM “COUVER” consists in promoting soil cover (inter-row cover, grassing of crops, creation of floral cover and networking of ecological regulation zones in the plots),
- AECM “LINEA” consists in maintaining landscape elements (hedges, trees, ponds).

We construct two samples to distinguish specialisations between crop and livestock (or mixed) farms. For each specialisation and each AECM organic or mixed organic-conventional farming, Tables 4 and 5 describe the size of the control group, size of the beneficiary group of the evaluated scheme, the number of control farms matched to the beneficiaries, and the number of unmatched treated farms.

Table 4: Sample size for selected AECM (crop farms)

AECM	Control group	Matched group	Treated group	Unmatched treated
Organic	8,337	855	171	0
Mixed organic-conventional	8,337	695	139	0
COUVER	14,959	405	81	0
HERBE	14,917	320	64	0
LINEA	15,084	54	28	0
PHYTO	14,978	510	102	0

We observe that the size of the different treatment groups is smaller than the control group, a consequence of the relatively low adoption of these schemes. Due to the small number of farms in the AECM “LINEA” for crop farms and “PHYTO” for livestock farms, the results of our estimations should be taken with caution.

Table 5: Sample size for selected AECM (livestock farms)

AECM	Control group	Matched group	Treated group	Unmatched treated
Organic	9,532	815	163	0
Mixed organic-conventional	9,532	305	80	19
COUVER	16,717	325	67	2
HERBE	15,292	1,765	358	5
LINEA	16,650	400	84	4
PHYTO	16,812	150	30	0
SHP	16,858	675	135	0
SPE	16,924	345	69	0
SPM	16,843	750	150	0

5.4 Results

We present our results by distinguishing the estimated effects on the sample of crop farms (Tables 6, 8 and 10) from the estimated effects on livestock farms (Tables 7, 9 and 11). Each table is composed as follows: The columns correspond to the AECM schemes evaluated. For each outcome (row), we have the estimate of the ATT, its standard error, and the corresponding additionality. We will focus our attention on the additionality that informs us about the average variation between the beneficiary farms and their matched farms over the counterfactual value of the outcome.

Changes in crop diversity, fertiliser and crop expenditure alongside share of grassland in utilised agricultural area (UAA) are considered as proxies to indicate the effectiveness of AECM in enhancing environmentally friendly farm practices. Labour demand is measured by family labour, permanent hired labour and total hired labour. Changes in the other outcomes depict the AECM effect on farm economic performance. Moreover, the effect of AECM on farm income may be considered as a way to evaluate the capacity of AECM payment to offset the real farm cost of AECM adoption.

5.4.1 Effect on ecological practices adoption

Crop farms

We observe that the AECM contracted by crop farms strongly influence their expenditure per hectare on fertilisers and crop protection. Only the AECM dedicated to the preservation of grassland does not significantly influence these expenditures (Table 6). We observe a decrease in expenditure per hectare on fertilisers by 404 Euros/ha for AECM “PHYTO” (+/- 81 Euros/ha), 352 Euros/ha (+/-100) for AECM “COUVER”, and -294 euros/ha (+/- 105) for organic farms.

These estimates correspond to significant additionalities. For example, we identify a 73% decrease in fertiliser expenditure and a 23% decrease in crop protection expenditure for farms under AECM

"PHYTO", compared to matched farms. The additionality of the AECM "COUVER" is comparable to the AECM "PHYTO" for fertiliser expenditure (-63%) but significantly higher for crop protection expenditure (-49%).

The AECM "PHYTO", which requires a reduction in the Treatment Frequency Index (TFI) of the beneficiary farms (in relation to a regional baseline), is close to a design with an obligation of results, and does not seem to be more effective than the cover measures (AECM "COUVER"), which are based on an obligation of means (grassing under the crop, inter-rows, creation and maintenance of floral and faunal target zones). Finally, we find that organic farms have 50% less expenditure on fertilisers and 36% less expenditure on crop protection than their matched conventional farms (Table 6).

Although the AECM "HERBE" (for grassland maintenance) do not significantly influence the reduction of chemical inputs, we observe that they achieve their objectives in terms of UAA share devoted to grassland (Table 6). Indeed, we observe an increase by nearly 6 pp. of grassland share for the beneficiary farms, corresponding to an additionality close to 49%. This additionality is relatively high, probably because the average share of grassland is typically small for crop farms. Finally, we do not find any significant effect on crop diversity in our crop farm sample. On the contrary, crop diversity is lower for organic farms and those entering the AECM "PHYTO", compared to their matched farms.

Table 6: Average treatment effect on treated (ATT) of organic and AECM adoption on ecological practices adoption (crop farms)

Outcome		Organic	Mixed organic-conventional	COUVER	HERBE	LINEA	PHYTO
Fertiliser expenditure (€/ha)	**ATT**	-294.4***	-55.39*	-352.8***	4.2	22.337	-404***
	Standard Error	(105.58)	(32.96)	(100.01)	(12.083)	(21.272)	(81.867)
	Additionality	-0.487	-0.217	-0.635	0.026	0.181	-0.732
Crop protection expenditure (€/ha)	**ATT**	-196.8***	-3.21	-219.6***	-24.11*	0.546	-112.4**
	Standard Error	(41.02)	(41.577)	(36.331)	(12.777)	(16.159)	(53.882)
	Additionality	-0.365	-0.008	-0.49	-0.129	0.004	-0.23
Share of grassland (% of UAA)	**ATT**	-0.026***	0.059***	-0.003	0.057***	0.053	-0.008
	Standard Error	(0.005)	(0.02)	(0.01)	(0.019)	(0.033)	(0.011)
	Additionality	-0.738	1.149	-0.118	0.485	0.722	-0.237
Crop diversity	**ATT**	-0.301***	0.027	0.037	-0.051	-0.1	-0.172***
	Standard Error	(0.03)	(0.043)	(0.047)	(0.031)	(0.073)	(0.037)
	Additionality	-0.609	0.059	0.08	-0.073	-0.157	-0.41

Note: *** p<0.01, ** p<0.05, * p<0.1.

Livestock farms

Compared to their matched farms, we observe a significant decrease in fertiliser expenditure (per hectare) for the AECM "SPE" (-20%), "SPM" (-40%) and organic farms (-71%) (Table 7). The results are similar for expenditure on pesticides (Table 6). A strong decrease in expenditure is identified for the organic (-80%), "SPE" (-39%) and "SPM" (-54%) farms (Table 7). In contrast to crop farms, AECM "PHYTO" does not seem to significantly influence the expenditure on chemical inputs on livestock farms (Table 7).

Table 7: Average treatment effect on treated (ATT) of organic and AECM adoption on ecological practices adoption (livestock farms)

Outcome		Organic	Mixed organic-conventional	HERBE	LINEA	PHYTO	SHP	SPE	SPM
Fertiliser expenditure (€/ha)	**ATT**	-63.43***	-13.25	1.633	2.41	2.538	0.93	-14.87**	-25.41***
	Standard Error	(4.253)	(8.969)	(2.905)	(6.114)	(13.71)	(2.476)	(6.718)	(8.002)
	Additionality	-0.706	-0.158	0.027	0.042	0.031	0.028	-0.192	-0.404
Crop protection expenditure (€/ha)	**ATT**	-50.04***	5.104	0.84	8.415	-5.672	-0.239	-18.67***	-14.51***
	Standard Error	(3.105)	(17.787)	(2.75)	(5.35)	(11.482)	(1.172)	(4.021)	(3.742)
	Additionality	-0.834	0.078	0.022	0.253	-0.102	-0.02	-0.385	-0.536
Share of grassland (% of UAA)	**ATT**	0.088***	0.004	-0.012	0.01	0.014	0.043**	-0.152***	-0.013
	Standard Error	(0.027)	(0.036)	(0.017)	(0.028)	(0.054)	(0.018)	(0.036)	(0.03)
	Additionality	0.243	0.012	-0.024	0.021	0.053	0.058	-0.381	-0.026
Crop diversity	**ATT**	-0.268***	-0.028	0.029*	0.01	-0.025	0.05**	0.058	0.062***
	Standard Error	(0.025)	(0.035)	(0.016)	(0.033)	(0.061)	(0.021)	(0.041)	(0.022)
	Additionality	0.572	-0.063	0.085	0.028	-0.051	0.31	0.138	0.249

Note: *** p<0.01, ** p<0.05, * p<0.1.

Furthermore, we observe that farms involved in AECM "SHP" have a larger share of grassland than their matched farms (6%), confirming the additionality of this measure. Nevertheless, territorial AECM dedicated to grassland conservation ("HERBE") alongside AECM "SPM" does not significantly influence the share of grassland of beneficiary farms (Table 7).

5.4.2 Effect on labour demand

The results of the previous section confirm that the adoption of AECM and organic production encourage a shift towards more environmentally friendly practices, notably by limiting the use of chemical inputs. We therefore expect that these changes in farming practice lead to modification of farmers' choices in labour demand.

Crop farms

The effect of AECM and organic farming on employment is assessed using three indicators: family labour, permanent hired labour (excluding seasonal employment) and total hired labour. All of these variables are expressed in full-time equivalent, namely annual work units (AWU).

Organic farms do not seem to use more family labour than their conventional matched farms (Table 8). On the other hand, we do observe a higher demand for salaried labour on organic farms of 30% for permanent salaried workers and 33% for total salaried workers. For partially organic farms, the pattern of labour demand is reversed: we observe an increase in the demand for family labour (12.3%), while no significant effect is observed on the demand for salaried labour (Table 8).

Regarding the AECM, we observe very different patterns, which confirms our methodological choice of distinguishing them. First of all, it is observed that grassland maintenance measures AECM “HERBE” induce a negative effect on the family labour demand (8%) compared to matched farms, but no significant effect on salaried labour demand (Table 8).

Family labour is less used in AECM “PHYTO” farms (16%), as well as total salaried labour (55%) and permanent salaried labour (56%). AECM “COUVER” also uses fewer permanent workers (24%) and total salaried labour (64%). Finally the AECM “LINEA” positively influences family work (20%) and permanent salaried workers (17%). These results are confirmed when we measure these indicators reported to the UAA: “PHYTO”, “HERBE” and “COUVER” operations are strongly less intensive working than their matched farms.

Table 8: Average treatment effect on treated (ATT) of organic and AECM adoption on labour demand (crop farms)

Outcome		Organic	Mixed organic-conventional	COUVER	HERBE	LINEA	PHYTO
Family labour	**ATT**	-0.003	0.182**	-0.076	-0.109*	0.268**	-0.225***
	Standard Error	(0.052)	(0.083)	(0.095)	(0.06)	(0.125)	(0.056)
	Additionality	-0.002	0.123	-0.056	-0.079	0.205	-0.16
Permanent hired labour	**ATT**	1.49***	-0.112	-0.534***	-0.047	0.24*	-0.373**
	Standard Error	(0.296)	(0.219)	(0.166)	(0.112)	(0.128)	(0.186)
	Additionality	0.667	-0.049	-0.242	-0.029	0.173	-0.162
Total hired labour	**ATT**	1.56***	-0.041	-1.293***	-0.136	-0.004	1.107**
	Standard Error	(0.335)	(0.347)	(0.254)	(0.136)	(0.176)	(0.522)
	Additionality	0.87	-0.017	-0.638	-0.182	-0.015	0.561

Note: *** p<0.01, ** p<0.05, * p<0.1.

Livestock farms

The schemes studied here do not appear to strongly influence labour demand for the sample of livestock farms (Table 9). Indeed, only farms with "SPM" and "SPE" AECM respectively experienced a decrease in the demand for non-salaried labour (-14%) and salaried labour (-54%). Our results therefore reveal a strong heterogeneity in responses, in terms of labour demand, to different measures and for different types of agriculture.

Table 9: Average treatment effect on treated (ATT) of organic and AECM adoption on labour demand (livestock farms)

Outcome		Organic	Mixed organic-conventional	HERBE	LINEA	PHYTO	SHP	SPE	SPM
Family labour	**ATT**	0.043	-0.017	0.019	0.064	-0.057	0.051	0.11	-0.221***
	Standard Error	(0.071)	(0.084)	(0.044)	(0.082)	(0.103)	(0.053)	(0.094)	(0.062)
	Additionality	0.025	-0.011	0.011	0.04	-0.035	0.037	0.068	-0.141
Permanent hired labour	**ATT**	0.042	0.031	0.048	-0.034	-0.057	0.052	-0.098	0.09
	Standard Error	(0.087)	(0.193)	(0.061)	(0.186)	(0.158)	(0.06)	(0.112)	(0.087)
	Additionality	0.021	0.017	0.024	-0.017	-0.029	0.034	-0.051	0.06
Total hired labour	**ATT**	0.034	0.126	0.021	-0.156	-0.069	0.015	-0.182***	0.017
	Standard Error	(0.06)	(0.216)	(0.039)	(0.161)	(0.143)	(0.046)	(0.052)	(0.08)
	Additionality	0.089	0.2	0.056	-0.307	-0.15	0.079	-0.541	0.054

Note: *** p<0.01, ** p<0.05, * p<0.1.

5.4.3 Effect on economic performance

The results of the previous two sections indicate that the adoption of AECM mainly influences farm practices, and to a lesser extent their labour demand. This section is devoted to presenting the effect of AECM adoption on farm economic performance. We use several indicators of economic performance. Productive intensity will be measured through the gross product per AWU (productivity per work unit), the farm operating result per hectare and the farm current income per hectare. The latter is used to assess the impact of the AECM on the income of the farm manager.

Crop farms

We observe that crop farms that have adopted an organic production system, or AECM "HERBE" or "PHYTO" have a significantly lower total gross production per AWU than their matched farms: the estimated difference is respectively -31%, -11% and -32% (Table 10). However, the organic farms seem to be detached from the AECM, as the impact of the production method does not seem to significantly influence either farm income or the operating result. Indeed, we observe a significant decrease in (economic) productivity, as their operating result per hectare is lower than their matched farms, of the farms adopting an AECM “COUVER” (-137%), “HERBE” (-10%), “LINEA” (-16%) and “PHYTO” (-107%). This result leads to an equivalent decrease in farm income per hectare, which means that the total gross output is large enough not to be counterbalanced by a decrease in inputs expenditure (Table 10).

Table 10: Average treatment effect on treated (ATT) of organic and AECM adoption on farm economic performance (crop farms)

Outcome		Organic	Mixed organic-conventional	COUVER	HERBE	LINEA	PHYTO
Farm income (€) (log)	**ATT**	-0.002	0.009	-0.013*	-0.001	0.004	-0.033*
	Standard Error	(0.01)	(0.01)	(0.007)	(0.005)	(0.008)	(0.017)
Farm income (€/ha) (log)	**ATT**	-0.006	-0.117	-1.373***	-0.097***	-0.157*	-1.069***
	Standard Error	(0.226)	(0.083)	(0.285)	(0.03)	(0.083)	(0.199)
Farm operating result (€) (log)	**ATT**	-0.001	0.01	-0.013*	0.001	0.005	-0.031*
	Standard Error	(0.01)	(0.011)	(0.007)	(0.005)	(0.008)	(0.016)
Farm operating result (€/ha) (log)	**ATT**	-0.006	-0.125	-1.37***	-0.096***	-0.156*	-1.067***
	Standard Error	(0.225)	(0.083)	(0.285)	(0.03)	(0.083)	(0.198)
Total gross product (€/AWU)	**ATT**	-41700***	-6720	23813	15369*	-14190	-40690***
	Standard Error	(7902.3)	(6659.9)	-17222	(8249.3)	-19225	(7801.6)
	Additionality	-0.315	-0.064	0.196	0.109	-0.092	-0.324

Note: *** p<0.01, ** p<0.05, * p<0.1.

Livestock farms

Except for AECM “LINEA”, “PHYTO” and “SHP”, we observe a significant decrease in total gross product per AWU for organic (-10%), partially organic (-14%), as well as AECM “HERBE” (-7%), “SPE” (-9%) and “SPM” (-9%) farms (Table 11). However, these average differences from the matched farms are smaller than those estimated for the sample of crop farms. Therefore, our results reveal that productivity per worker is more strongly impacted by the introduction of AECM or an organic production mode for the crop farm sector (Table 10).

Then, we observe that the other economic results of the organic farms are not significantly different from the matched conventional farms. On the opposite, we observe that the adoption of the AECM “HERBE”, “PHYTO”, “SHP” and “SPM” results in a decrease of income (farm operating result and farm current income) per ha of -4.9%, -3.7%, -6.5% and -12%, respectively (Table 11).

Finally, none of these transitions to ecological practices seem to significantly affect either the farmers’ income or the operating result of the farm, apart from a slight decrease for the farms adopting the AECM “LINEA” (-2%).

Table 11: Average treatment effect on treated (ATT) of organic and AECM adoption on farm economic performance (livestock farms)

Outcome		Organic	Mixed organic-conventional	HERBE	LINEA	PHYTO	SHP	SPE	SPM
Farm income (€) (log)	**ATT**	0.001	-0.007	-0.002	-0.02**	-0.003	-0.001	0.007	0.001
	Standard Error	(0.004)	(0.006)	(0.003)	(0.008)	(0.014)	(0.002)	(0.005)	(0.003)
Farm income (€/ha) (log)	**ATT**	-0.323	-0.09	-0.049***	-0.014	-0.037**	-0.066***	-4.277	-0.122***
	Standard Error	(0.198)	(0.092)	(0.016)	(0.015)	(0.017)	(0.016)	(3.986)	(0.041)
Farm operating result (€) (log)	**ATT**	0.002	-0.006	-0.001	-0.019**	-0.004	-0.001	0.008	0.002
	Standard Error	(0.004)	(0.004)	(0.003)	(0.008)	(0.014)	(0.002)	(0.005)	(0.003)
Farm operating result (€/ha) (log)	**ATT**	-0.322	-0.089	-0.049***	-0.014	-0.037**	-0.065***	-4.268	-0.122***
	Standard Error	(0.198)	(0.092)	(0.016)	(0.015)	(0.017)	(0.016)	(3.978)	(0.041)
Total gross product (€/AWU)	**ATT**	-14900**	-17830***	-9893***	-4645	3561.8	-651.7	-13060*	-10890**
	Standard Error	(6085.8)	(6666.4)	(3286.7)	(5438.9)	-12185	(3605.9)	(7798.9)	(4585.7)
	Additionality	-0.104	-0.145	-0.072	-0.033	0.025	-0.006	-0.089	-0.087

Note: *** p<0.01, ** p<0.05, * p<0.1.

5.5 Discussion

Our main results suggest that the impact on farms of transitions to ecological practices is different depending on the obligations induced by the different AECM implemented. These AECM also have a different impact on crop farms than on livestock farms. These aspects have been rarely studied in the previous literature (Arata and Sckokai, 2016). Our study fills this gap by combining DID and propensity score matching for the same panel sample of French farms. Our study covers three major dimensions to investigate the implications of AES adoption for farms: the impact on the adoption of ecological practices (additionality of measures), on labour demand, and on economic performance (productivity, compensation of the private cost of adoption).

First, our study shows the additional effect of AECM and organic conversion on the adoption of ecological practices. For crop farms, we identified an additional effect of measures dedicated to the reduction of chemical inputs on the expenditure per ha on pesticides and fertilisers. For livestock farms, while we cannot conclude that there is a significant effect on farms adopting an AECM "PHYTO", we observe a reasonable additionality of the "SPE" and "SPM" AECM (dedicated to mixed crop-livestock systems) on these same chemical input expenditures. These results are consistent with those of Arata and Sckokai (2016) obtained on a sample of French farms. However, we observe that by decomposing by type of AECM, we obtain a higher additionality than those identified by these authors for all AECM combined. We explain this difference in results by the fact that the majority of AECM adopted in France concern the conservation and maintenance of extensive grasslands which do not include obligations to reduce the use of chemical inputs.

Moreover, on both crop and livestock farms, we observe a strong additionality of the adoption of an organic production mode, confirming previous results (Chabé-Ferret and Subervie, 2013; among others). The latter found that organic farming support measures had the highest additionality of the five AES analysed in their study in the period 2000-2006. Due to the lack of more precise data on chemical input use practices, our study was limited to measuring additionality on expenditure. For pesticide use, a more precise estimate of additionality would require measuring the effects on indicators of agricultural practices such as the TFI or the potential risks associated with pesticide use (pesticide load index). Using survey data on farming practices, Védrine and Larmet (2021) calculated these indicators and assessed the additionality of a set of AES implemented between 2014 and 2017 using propensity score matching.

Védrine and Larmet (2021) highlighted the significant effect of the adoption of AECM “PHYTO”, “SPE” and “SPM” on the reduction of the use of phytosanitary products. For the AECM “PHYTO”, the effect is more marked on the reduction in the use of herbicides than on the use of insecticides and fungicides. The AECM dedicated to mixed crop-livestock systems (“SPE” and “SPM”) provide a greater incentive for beneficiary farms to significantly reduce their use of herbicides and fungicides. An important contribution of their work compared to previous studies is the estimation of the additionality of AECM on the reduction of potential risks related to the use of pesticides. The results of their study suggest that the adoption of AES reduced the risks associated with the use of pesticides by 25-35%. Nevertheless, they were unable to demonstrate a significant effect of the AECM “PHYTO” on the evolution of the potential risks associated with the use of these products. However, the AECM “PHYTO” seems to allow a reduction in the potential risks to human health associated with the use of pesticides.

The measures to reduce N inputs (“HERBE”, “SPE”, “SPM”) had limited effects. The study could not identify any significant effect of the AECM “HERBE” on total N input and mineral N input. As for the AECM “SPE” and “SPM”, they certainly allowed a significant reduction in N inputs, but the additionality did not exceed 10%.

We also identified a statistically significant, but modest, additional effect of measures for crop farms and measures for livestock farms on permanent grassland conservation. This result is consistent with the findings of Pufahl and Weiss (2009) which reveal a positive and significant treatment effect of AES programmes on grassland areas for a sample of German farms, alongside the results of Arata and Sckockai (2016). However, the analysis carried out by Védrine and Larmet (2021) based on a larger sample than the one used in our study suggests that, with the exception of a positive effect of the AECM “SHP” on the rate of grassland specialisation, the additionality of AECM on the maintenance of grassland and the extensification of practices is very small, thereby hampering to conclude that these measures are cost-effective. In addition, the specific features of the AES, whose objective is the maintenance of grasslands and extensive practices, make the interpretation of our estimates more complex, as does the causal effect of these measures. Indeed, the implementation of these AES is likely to modify the equilibrium on the land market. By increasing the demand for permanent grassland by beneficiaries, relative to other crops, the AES for grassland would imply an increase in the price of grassland plots. A complementary analysis, on a territorial scale, would certainly make it possible to estimate the effect of a variation in the number of beneficiaries (i.e. contracting rate, sum of payments per hectare on a territorial scale) on changes (on a territorial scale) in the area under permanent grassland. Yet, the estimated effect would correspond to the effect of these AES, net of interactions on the land market, under the strong assumption that the territorial scale selected internalises all the transactions between farms induced by the grassland AECM.

Then, we observed a contrasting situation between the additionality of AECM on crop diversity depending on the type of farm (crop vs livestock). For livestock farms, we observe a significant effect of “HERBE”, “SHP” and “SPM” AECM on crop diversity, whereas no measure seems to influence this trend in the case of crop farms. Thus, our results only partially confirm the additional effects highlighted by Chabé-Ferret and Subervie (2013), Bertoni et al. (2020) and Védrine and Larmet (2021). Bertoni et al. (2020) show a non-negligible effect of AES dedicated to crop diversity, notably on the number of crops in the crop rotation. Without being able to distinguish between the different AES, Arata and Sckokai (2016) show an average effect of AES on the number of crops per beneficiary farm in Great Britain and Italy, while AES do not seem to significantly influence crop diversity on German, Spanish and French farms. Moreover, this effect is more marked when the payments associated with the AES represent more than 5% of the farm income. The effects estimated by Chabé-ferret and Subervie (2013) are more modest and their cost-effectiveness analysis suggests that these measures are costly. Védrine and Larmet (2021) also show that system measures are interesting for achieving greater crop diversity in the last AES programme. Indeed, they find that the average difference between farms benefiting from the AECM “SPE” and their matched farms is 11 to 18 percentage points, corresponding to an increase of 3 to 4% in our crop diversity indicator. This increase is probably a combination of the increase in the number of crops and the decrease in the share of the dominant crop on these farms, compared to the situation they would have had if they had not adopted the AECM “SPE”. This result is confirmed by our DID estimates, for which we obtain an additionality of around 2%. They also show that the beneficiaries of the AECM “SHP” have, on average, an indicator of crop diversity that is 4 to 6% higher than that of matched non-beneficiaries. Finally, they show that localised grassland and cover crop measures. Based on the estimation of spatial econometric models at the LAU1 regional level for France, Desjeux et al. (2015) show a negative relationship between the evolution of the crop diversity index (2007-2010) and the AES oriented to water and biodiversity preservation. Nevertheless, this study highlights the delayed effects of the first programming of the Rural Development Regulation (RDR1) (2000-2006), since the 2007-2010 evolution of the crop diversity index is positively influenced by the contracting rates of the Less favoured Areas (LFA) payments and grassland measures, and negatively influenced by diversification measures (Desjeux et al., 2015).

The effect of organic production on crop diversity may seem strange, as we observe a negative difference between organic and matched farms. The literature is generally not conclusive about the effect of organic farming on crop diversity, without highlighting a negative effect: Bertoni et al. (2020) do not show a significant effect on the number of crops or the share of the dominant crop, while Chabé-ferret and Subervie (2013) show a small but statistically significant increase in crop diversity on converted organic farms. We interpret our result by the presence of strong heterogeneity in costs to implement crop diversification within our treatment group (organic farms) as identified in the literature (Sipiläinen and Huhtala, 2013; Ang et al., 2018). In this case, the ATT is not informative and this particular situation would deserve to be addressed by exploring this heterogeneity.

Second, the effects of the adoption of organic farming and the different AECM on labour demand are not uniform. Indeed, we find a significant positive effect of organic farming on the demand for hired labour, but only for crop farms. These results are consistent with the literature, which concludes that organic farming is more labour intensive than conventional farming. Although the intensity of the relationship differs between studies, we observe that our results are consistent with those of Vérot (1998) or Finley et al. (2017). This result can be explained a substitution from chemical inputs uses to labour-intensive crop protection practices (soil maintenance, mechanical or manual weeding, etc.). As far as the AECM are concerned, the schemes aimed at the maintenance and management of grassland and mixed farming systems have a negative effect on the labour demand of beneficiary farms. In livestock systems, the decrease in labour demand is significant, whether it is family or salaried labour. As far as crop farms are concerned, we can note an ambiguous effect of the adoption of the AECM “PHYTO” on the demand for labour: this measure reduces the use of family labour, to the benefit of an increase in non-permanent salaried labour. Finally, the AECM “LINEA” have a clearly positive effect on the demand for labour on crop farms.

Overall, the effect of these measures is therefore negative, according to Arata and Sckokai (2016), we warn about the effects on rural employment if the practices induced by AECM adoption are generalised. We think that it is essential that future CAP retains strong support mechanisms for rural employment, particularly for diversification into non-agricultural activities, in order to ensure the development and greater economic resilience of rural territories.

Third, productive intensity is negatively influenced by the change in practice associated with the adoption of organic farming or AES. This is particularly the case for the “PHYTO, “HERBE” AECM, and organic farming for crop farms, and the “HERBE”, “SPE”, “SPM” AECM and organic farming for livestock farms. Although this reduction in production intensity results in a reduction in farm income related to UAA, the amounts of the various AES payments cover on average the cost of adopting ecological practices. In fact, only crop farms having adopted a “PHYTO” or “COUVER” AECM, and livestock farms having adopted a “LINEA” AECM have a slightly lower farm income than the matched farms.

In conclusion, our study suggests that farmers’ incomes are not affected by the farm’s ecological practices, once the extra cost of these practices has been covered by the AES payment or promote some efficiency gains. In France, the real cost of the transition is therefore on average well compensated by these payments. It does not imply that farms earn extra profit, and thus appears to respect WTO rules.

6 General conclusion

Against the background of the ambitious goal of the EU to achieve an increasing uptake of ecological approaches in its farming sector, it is crucial to assess how past and current public policies and governance systems implemented in the EU have affected the adoption of ecological approaches and the performance of ecological agriculture at the farm, farm-group and territorial levels.

The studies presented here adopt various approaches to accomplish this goal (econometric analyses, meta-analysis, treatment effect analysis, bio-economic model, regional CGE model). Section 3 of this deliverable focussed on insights from works undertaken in other LIFT WP (WP2, WP3 and WP4) regarding policy designs and impacts. They inform various pieces of EU legislation including CAP, Nitrates Directive, Water Framework Directive and Directive on the sustainable use of pesticides. These results highlight some drawbacks of currently implemented schemes, namely that current CAP subsidies received by farmers reduce the technical efficiency of extensive farms, suggesting that the current type of subsidies may not be adequate for extensive technologies; or that currently implemented AES have the potential to induce windfall effects depending on the technical efficiency of farmers who actually adopt them. In terms of future policy recommendations, they call for more ambitious measures to fulfil the obligations under the Nitrates and Water Framework Directives, by targeting animal production directly and facilitating legume processing at the farm level. In terms of pesticides management, the results put in perspective that the secondary benefits from crop diversification, besides pesticides reduction, could be important for farmers by way of decreased productivity loss. These studies also advocate policy compensation schemes that take into consideration the income forgone, given the regional potential, both in terms of agricultural production and environmental endowments. To facilitate the adoption of ecological practices that are intensive in labour, transaction costs on the labour market should be reduced to allow for more flexibility of hired labour. Finally, both formal and informal arrangements that foster collaborative efforts should be facilitated by policy-makers.

While PES are major instruments available to governmental or private organisations to support the conservation of ecosystems that provide environmental services, their implementation is very heterogeneous. In this context, it is necessary to identify how the design of a PES affects its success, which can be assessed differently. In this respect, the meta-analysis performed in Section 4 is the first to analyse the impact of PES-schemes design on both their effectiveness, measured as the probability to increase ES provision, and their efficiency, assessed based on the level of additionality. We show that these two possible measures of PES performance are driven by different characteristics of PES schemes. However, our results on the impact of PES design on their effectiveness should be considered more robust than those on efficiency, given the smaller available sample with additionality measurements and the associated econometric limitations.

Among others, eligibility of ES providers, contract length, reference design, payment constraint, monitoring system and the implementation zone of the PES schemes appear to be correlated to the probability of achieving positive outcomes. Nevertheless, the effectiveness of the PES-schemes investigated in this meta-analysis is shown to especially depend on the monitoring system implemented to ensure compliance and on the eligibility of ES providers. Conversely, the type of agents involved in the PES-schemes and the type of payments received by the participants seem to be the main factors that influence their efficiency given that they have been found to be positively correlated to the level of additionality.

The monitoring system in place to ensure compliance, both in terms of by whom and how often it is performed, appears as a key driver of performance of PES. We show that regular third-party monitoring is more conducive to increased ES provision compared to one-off internal monitoring. However, the frequency of monitoring could be associated with a lower level of additionality. The trade-off between the costs and benefits of monitoring (Bellassen and Shishlov, 2017) in the sample of schemes reviewed in our meta-analysis is hence not so clear-cut, although one must bear in mind that the efficiency results are less robust than those on effectiveness. Since PES direction depends both on individual additionality and the capacity of the PES to enrol ES providers, our results call for more studies on delineating the impact of different features of the monitoring system on both capacities to enrol and additionality. They also highlight the need for careful consideration of the monitoring system during the design phase of the PES schemes to ensure their performance.

Our meta-analysis also shows that contrary to common expectations, result-based payments are not more efficient than practice-based payments. However, assessments of result-based approaches hinge on the definition of what result is retributed. In their analysis of payment-by-results schemes for biodiversity in Europe, Herzon et al. (2018) identify only five "pure result-based payment" schemes, in which solely biodiversity results are measured and no specific management actions are specified or required. Besides, they find a number of hybrid schemes, with baseline management requirements, or in which the result-based payment is optional above a baseline practice-based payment. Then, other design features seem to dominate the expected effect of result-based payments (Burton and Schwarz, 2013; Herzon et al., 2018). Given the increasing interest of policy-makers for result-based approaches, as illustrated by the latest orientations of the CAP (European Commission, 2018), it is thus crucial to undertake more ex-post studies of actual result-based approaches to provide more robust assessments of their performance.

The few numbers of studies from which we could derive additionality estimates is a limitation of this meta-analysis. In this respect, it is important to mention that policy-makers are best positioned to improve policy evaluation, for instance by anticipating the data requirements of ex-post evaluation during the very first stages of policy design or by using random experimentation at early stages of policy implementation (Behaghel et al., 2019). In this respect, interactions between researchers and policy makers are paramount at early stages of policy design and implementation. Thus, we advocate that the governance structure should be smarter, e.g. by including researchers who can propose and test (through Random Control trial) innovative policies.

Finally, the results of Section 5 suggest that farmers' incomes are not affected by their ecological practices, once the extra cost of these practices has been covered by the AES payment or promote some efficiency gains. The real cost of the transition is therefore on average well compensated by these payments. According to WTO standards, AES payments should distort neither trade nor production but instead only compensate for income forgone and costs incurred (Arata and Sckokai, 2016). It does not imply that farms earn extra profit, and thus appears to respect WTO rules.

This work has made it possible to construct the first elements of an econometric analysis of the additionality and the cost-effectiveness of AECM which could be reproduced in the framework of ex-post evaluations in other Member States. It also allowed us to identify several limitations. Indeed, the specificities of the AECM whose objective is the maintenance of grasslands and extensive practices make the interpretation of our estimates, such as the causal effect of these measures, more complex. Indeed, the implementation of these AECM is likely to modify the balance on the land market. By increasing the demand for grassland by beneficiaries, relative to other crops, the AECM for grassland would imply an increase in the price of grassland. A complementary analysis, on a territorial scale, would certainly make it possible to estimate the effect of a variation in the number of beneficiaries (i.e. contracting rate, sum of payments per hectare on a territorial scale) on changes (on a territorial scale) in the area under grassland. However, the estimated effect would correspond to the effect of these AECM, net of interactions on the land market, under the strong assumption that the selected territorial scale internalises all the transactions between farms induced by the grassland AECM. A second approach can be envisaged, with the main advantage of maintaining an analysis at the farm level, while relaxing the hypothesis of the absence of interference between individuals (SUTVA). We could consider using the "constant response to treatment" model proposed by Manski (2013), in which potential outcomes depend both on the farm's commitment to an AECM, but also on the proportion of farms in the neighbourhood committed to an AECM.

A first sensitivity analysis of the results obtained by matching to the proportion of plots under AECM in the neighbourhood could be considered. If the hypothesis of no interaction effect is valid, then the counterfactual results estimated by this method will not vary with the proportion of neighbouring farms engaged in an AECM. Otherwise, we could identify both the individual (net additionality of transfers on the land market) and interaction effects of AECM using a two-step procedure proposed by Ferracci et al. (2014) consisting in, in a first step, the distribution of the treatment within each neighbourhood and, in a second step, the distribution of the treatment between farms.

Moreover, our study does not show a clear relationship (in terms of trade-off or synergy) between the average economic and environmental performance of farms adopting an AECM or an organic production method. However, limitations related to the size of our samples, notably the low number of AECM beneficiaries in the FADN databases, do not allow us to go further into the study of this relationship.

The small size of our beneficiary group is particularly affected by a low adoption of some of the AECM. For this reason, we cannot explore the heterogeneity of transition costs to greener practices induced by the adoption of AES. We consider that identifying heterogeneous costs according to farm characteristics and/or spatial variations in these costs is a natural extension of our study. This would allow us to determine whether the low adoption of some AECM is explained by high heterogeneity in adoption costs and would inform us on the interest, to increase participation, of setting up differentiated payments according to farm characteristics. Using Machine Learning algorithms, the introduction of such payments could improve the cost-effectiveness of AES, as shown for public schemes in other areas (Andini et al., 2018).

7 Deviations or delays

None

8 References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1-19.

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, 5, 754.

Alston, L. J., Andersson, K., & Smith, S. M. (2013). Payment for environmental services: Hypotheses and evidence. *Annual Reviewof Resource Economics*, 5, 139–159.

Andini, M., Ciani, E., De Blasio, G., D'Ignazio, A., & Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156, 86-102.

Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics* 3, 1277-1366, Elsevier.

Ang, F., Mortimer, S. M., Areal, F. J., & Tiffin, R. (2018). On the opportunity cost of crop diversification. *Journal of Agricultural Economics*, 69(3), 794-814.

Antle, J., Capalbo, S., Mooney, S., Elliott, E., & Paustian, K. (2003). Spatial heterogeneity, contract design, and the efficiency of carbon sequestration policies for agriculture. *Journal of Environmental Economics and Management*, 46, 231–250.

Arata, L., & Sckokai, P. (2016). The impact of agri-environmental schemes on farm performance in five EU member states: a DID-matching approach. *Land Economics*, 92(1), 167-186.

Bailey, A., Davidova S., Henderson, S., Ayouba, K., Bakucs, Z., Benedek, Z., Billaude, L., Brumă, L., Chițea, M., Doboș, S., Eckart, L., Gerner, L., Fereira, J., Florian, V., Gouta, P., Hansson, H., Jeanneaux, P., Jendrzejewski, B., Kantelhardt, J., Konstantidelli, V., Lascano Galarza, M.X., Latruffe, L., Legras, S., Lepicier, D., Manevska-Tasevska, G., Niedermayr, A., Polge, E., Rusu, M., Schaller, L., Simion, G., Tanasa, L., Tzouramani, I., Codrin Dinu Vasiliu, Walder, P., Zavalloni, M., & Zawalinska, K. (2021). Socio-economic impact of ecological agriculture at the territorial level. LIFT Deliverable 4.3.

Barath, L., Bakucs Z., & Fertő I. (2021). Differences in Efficiency and Productivity between conventional and organic farms: the case of Hungarian Cereal Oilseed and Protein (COP) crop producing farms (2010-2015). Unpublished LIFT working paper.

Baser, O. (2006). Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health*, 9(6), 377-385.

Behaghel, L., Macours, K., & Subervie, J. (2019). How can randomised controlled trials help improve the design of the common agricultural policy? *European Review of Agricultural Economics*, 46, 473–493.

Bellassen, V., & Shishlov, I. (2017). Pricing monitoring uncertainty in climate policy. *Environmental and Resource Economics*, 68, 949–974.

Bellassen, V., Stephan, N., Afriat, M., Alberola, E., Barker, A., & Chang, J.-P. (2015). Monitoring, reporting and verifying emissions in the climate economy. *Nature Climate Change*, 5, 319–328.

Bento, A. M., Kanbur, R., & Leard, B. (2015). Designing efficient markets for carbon offsets with distributional constraints. *Journal of Environmental Economics and Management*, 70, 51–71.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, 14, 395–411.

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.

Blundell, R., & Costa Dias, M. (2000). Evaluation methods for non-experimental data. *Fiscal Studies*, 21(4), 427-468.

Bonvillain, T., Foucherot, C., & Bellassen, V. (2020). Will the obligation of environmental results green the CAP? A comparison of the costs and effectiveness of six instruments for the transition to sustainable agriculture. I4CE.

Börner, J., Baylis, K., Corbera, E., de-Blas, D. E., Honey-Rosés, J., Persson, U. M., & Wunder, S. (2017). The effectiveness of payments for environmental services. *World Development*, 96, 359-374.

Brouwer, R., Tesfaye, A., & Pauw, P. (2011). Meta-analysis of institutional-economic factors explaining the environmental performance of payments for watershed services. *Environmental Conservation*, 380–392.

Burton, R. J., & 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., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

Canton, J., De Cara, S., & Jayet, P.-A. (2009). Agri-environmental schemes: Adverse selection, information structure and delegation. *Ecological Economics*, 68, 2114–2121.

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. *Journal of Environmental Economics and Management*, 65(1), 12-27.

Chabé-Ferret, S., & Voia, A. (2021). Are Grassland Conservation Programs a Cost-Effective Way to Fight Climate Change? Evidence from France. TSE Working Paper, n. 21-1248.

Claassen, R., Cattaneo, A., & Johansson, R. (2008). Cost-effective design of agri-environmental payment programs: US experience in theory and practice. *Ecological Economics*, 65(4), 737-752.

Coase, R. H. (1960). The problem of social cost. Classic papers in natural resource economics. Palgrave Macmillan, London, 1960, 87–137.

Cook, D. J., Guyatt, G. H., Ryan, G., Clifton, J., Buckingham, L., Willan, A., & Oxman, A. D. (1993). Should unpublished data be included in meta-analyses? Current convictions and controversies. *Jama*, 269, 2749–2753.

Cormier, A., & Bellassen, V. (2013). The risks of CDM projects: How did only 30% of expected credits come through? *Energy Policy*, 54, 173–183.

Costedoat, S., Koetse, M., Corbera, E., & Ezzine-de-Blas, D. (2016). Cash only? Unveiling preferences for a PES contract through a choice experiment in Chiapas, Mexico. *Land Use Policy*, 58, 302–317.

Dakpo, H., & Féménia, F. (2021). Pesticides efficiency of French wheat producers under a stochastic frontier framework. Unpublished LIFT working paper.

Davidova, S., Bailey, Alastair, Henderson, S., Bakucs, Z., Benedek, Z., Desjeux, Y., Gouta, P., Jendrzejewski, B., Konstantidelli, V., Latruffe, L., Tzouramani, I., Védrine & L., Zawalińska, K. (2021) Employment effects of ecological farming at the farm level. LIFT Deliverable 3.4.

Diop, T.B, Védrine, L., Blancard, S., Legras, S., & Marchand, S., 2021. Agri-Environmental Schemes (AESs) Adoption, Technical Efficiency and Environmental Indicator: Evidence from France. Unpublished LIFT working paper.

Dechezleprêtre, A., Colmer, J., Gennaioli, C., Glachant, M., & Schröder, A. (2014). Assessing the Additionality of the Clean Development Mechanism: Quasi-Experimental Evidence from India. Mimeo.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.

Derissen, S., & Latacz-Lohmann, U. (2013). What are PES? A review of definitions and an extension. *Ecosystem Services*, 6, 12-15.

Desjeux, Y., Dupraz, P., Kuhlman, T., Paracchini, M. L., Michels, R., Maigne, E., & Reinhard, S. (2015). Evaluating the impact of rural development measures on nature value indicators at different spatial levels: Application to France and The Netherlands. *Ecological Indicators*, 59, 41-61.

Doucouliagos, H., & Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47, 406–428.

Duval L., Binet T., Dupraz P., Leplay S., Etrillard C., Pech M., Deniel E., & Laustriat M. (2016). Paiements pour services environnementaux et méthodes d'évaluation économique. Enseignements pour les mesures agro-environnementales de la politique agricole commune. Etude réalisée pour le ministère en charge de l'agriculture. Rapport final.

Engel, S. (2016). The devil in the detail: A practical guide on designing payments for environmental services. *International Review of Environmental and Resource Economics*, 9, 131-177.

Engel, S., Pagiola, S., & Wunder, S. (2008). Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics*, 65, 663-674.

European Commission (2018). Proposal for a regulation of the European Parliament and of the Council establishing rules on support for strategic plans to be drawn up by Member States under the Common agricultural policy (CAP Strategic Plans) and financed by the European Agricultural Guarantee Fund (EAGF) and by the European Agricultural Fund for Rural Development (EAFRD) and repealing Regulation (EU) No 1305/2013 of the European Parliament and of the Council and Regulation (EU) No 1307/2013 of the European Parliament and of the Council.

Ezzine-de-Blas, D., Wunder, S., Ruiz-Pérez, M., & del Pilar Moreno-Sanchez, R. (2016). Global patterns in the implementation of payments for environmental services. *PloS one*, 11.

Ferracci, M., Jolivet, G., & Van Den Berg, G. J. (2014). Evidence of treatment spillovers within markets. *Review of Economics and Statistics*, 96(5), 812-823.

Ferraro, P. J. (2008). Asymmetric information and contract design for payments for environmental services. *Ecological Economics*, 65, 810–821.

Finley, L., Chappell, M., Thiers, P., & Moore, J. R. (2017). Does organic farming present greater opportunities for employment and community development than conventional farming? A survey-based investigation in California and Washington. *Agroecology and Sustainable Food Systems,* 42, 552–572.

Groupe de Bruges, 2014. Enhancing territorial cooperation for the provision of public goods in the context of the CAP Reform, in: Territorial Cooperation for the Provision of Public Goods in the Context of the CAP Reform. Wageningen.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in medicine*, 17, 841–856.

Heinrichs, J., & Britz, W. (2021). The impact of agri-environmental and climate measures on sustainable farm performance – a German case study analysis. Unpublished LIFT working paper.

Heinrichs, J., Jouan, J., Pahmeyer, C., & Britz, W. (2021). Integrated assessment of legume production challenged by European policy interaction: A case-study approach from French and German dairy farms. *Q Open,* 1.

Herzon, I., Birge, T., Allen, B., Povellato, A., Vanni, F., & Hart, K. (2018). Time to look for evidence: Results-based approach to biodiversity conservation on farmland in Europe. *Land Use Policy*, 71, 347–354.

Honey-Roses, J., Lopez-Garcia, J., Rendon-Salinas, E., Peralta-Higuera, A., & Galindo-Leal, C. (2009). To pay or not to pay? Monitoring performance and enforcing conditionality when paying for forest conservation in Mexico. *Environmental Conservation*, 36, 120–128.

Hou, L., Xia, F., Chen, Q., Huang, J., He, Y., Rose, N., & Rozelle, S. (2021). Grassland ecological compensation policy in China improves grassland quality and increases herders' income. Nature Communications, 12(1), 1-12.

Huang, W., Manevska-Tasevska, G., Hansson, H. (2021). Technical and economic performance of arable farms in Sweden: does the degree of ecological approaches matter? Unpublished LIFT working paper.

Jin, Y., Kilcline, K., Ryan, M., & O'Donoghue, C. (2021). Identifying heterogeneous technologies in Irish dairy farming. Unpublished LIFT working paper.

Kaiser, J., Haase, D., & Krueger, T. (2021). Payments for ecosystem services: a review of definitions, the role of spatial scales, and critique. *Ecology and Society*, 26.

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

Laffont, J.-J., & Tirole, J. (1993). A Theory of Incentives in Procurement and Regulation. MIT Press.

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. *Environmental Science & Policy*, 54, 1-9.

Latruffe, L., & Nauges, C. (2014). Technical efficiency and conversion to organic farming: the case of France. *European Review of Agricultural Economics*, 41(2), 227–253.

Latruffe, L., Niedermayr, A., Desjeux, Y., Dakpo, H.K., Ayouba, K., Schaller, L., Kantelhardt, J., Jin, Y., Kilcline, K. Ryan, M., & O'Donoghue, C. (2021). Technical efficiency of intensive and extensive technology in dairy farming in the European Union. Unpublished LIFT working paper.

Lechenet, M., F. Dessaint, G. Py, Makowski, & Munier-Jolain, N. (2017). Reducing pesticide use while preserving crop productivity and profitability on arable farms. *Nature Plants*, 3(3), 17008.

List, J. A., Millimet, D. L., Fredriksson, P. G., & McHone, W. W. (2003). Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. *Review of Economics and Statistics*, 85(4), 944-952.

Liu, T., Bruins, R.J.F., & Heberling, M.T. (2018). Factors influencing farmers' adoption of best management practices: A review and synthesis. Sustainability, 10, 1–26.

Lucas, V., Gasselin, P., Douwe, J., & Der Ploeg, V. (2018). Local inter-farm cooperation: A hidden potential for the agroecological transition in northern agricultures. Agroecology and Sustainable Food Systems, 43, 145–179.

MacLean, C. H., Morton, S. C., Ofman, J. J., Roth, E. A., Shekelle, P. G., & others. (2003). How useful are unpublished data from the Food and Drug Administration in meta-analysis? *Journal of Clinical Epidemiology*, 56, 44–51.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1), S1-S23.

Meyer, C., Reutter, M., Matzdorf, B., Sattler, C., & Schomers, S. (2015). Design rules for successful governmental payments for ecosystem services: Taking agri-environmental measures in Germany as an example. *Journal of Environmental Management*, 157, 146–159.

Moxey, A., White, B., & Ozanne, A. (1999). Efficient contract design for agri-environment policy. *Journal of Agricultural Economics*, 50, 187-202.

Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E., & 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*, 10, 1–23.

Muñoz-Piña, C., Guevara, A., Torres, J. M., & Braña, J. (2008). Paying for the hydrological services of Mexico's forests: Analysis, negotiations and results. *Ecological Economics*, 65, 725–736.

Muradian, R., Corbera, E., Pascual, U., Kosoy, N., & May, P. H. (2010). Reconciling theory and practice: An alternative conceptual framework for understanding payments for environmental services. *Ecological Economics*, 69, 1202-1208.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46, 27–29.

Pagiola, S., & Platais, G. (2006). Payments for Environmental Services: From Theory to Practice. Tech. rep., World Bank, Washington DC.

Pattanayak, S. K., Wunder, S., & Ferraro, P. J. (2010). Show me the money: do payments supply environmental services in developing countries? *Review of Environmental Economics and Policy*, 4, 254–274.

Pigou, A. C. (1920). Some problems of foreign exchange. *The Economic Journal*, 30(120), 460-472.

Pufahl, A., & Weiss, C. R. (2009). Evaluating the effects of farm programmes: results from propensity score matching. *European Review of Agricultural Economics*, 36(1), 79-101.

Puhr, R., G. Heinze, M. Nold, L. Lusa, and A. Geroldinger (2017). Firth's logistic regression with rare events: accurate effect estimates and predictions? *Statistics in Medicine*, 36(14), 2302–2317.

Rega, C., Thompson, B., D'Alberto, R., Niedermayr, A., Kantelhardt, J., Gouta, P., Konstantidelli, V., Tzouramani, I., Desjeux, Y., Latruffe, L., Billaudet, L., Paracchini, M.L. (2021). LIFT Deliverable 1.4.

Reinhard, S., & Linderhof, V. (2015). Convergence of EU nitrogen surplus, the RDP indicator of water quality. *Ecological Indicators*, 59, 19-26.

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.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. John Wiley & Sons.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2(1), 1-26.

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. *Ecology Letters*, 16(7), 912-920.

Shishlov, I., & Bellassen, V. (2012). 10 lessons from 10 years of the CDM, Climate Report. CDC Climat Research, Paris.

Sims, K. R., Alix-Garcia, J. M., Shapiro-Garza, E. L., Fine, L. R., Radeloff, V. C., Aronson, G., & Yanez-Pagans, P. A. (2014). Improving environmental and social targeting through adaptive management in Mexico's payments for hydrological services program. *Conservation Biology*, 28, 1151–1159.

Sipiläinen, T., & 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 Economic*s, 40(3), 441-462.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2), 305-353.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, 70, 103–127.

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, 8, 19–42.

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, 36, 1580–1598.

Thompson, S. G. (1994). Systematic review: Why sources of heterogeneity in meta-analysis should be investigated. *Bmj*, 309, 1351–1355.

Torstensson, G., Aronsson, H., & Bergström, L. (2006). Nutrient use efficiencies and leaching of organic and conventional cropping systems in Sweden. *Agronomy Journal*, 98, 603-615.

Uthes, S., & Matzdorf, B. (2013). Studies on agri-environmental measures: a survey of the literature. *Environmental management*, 51(1), 251-266.

Van Ruymbeke, K., Dakpo, K.H., Latruffe, L., Vranken, L. 2021. Farm environmental performance depending on the degree of ecological approaches. LIFT Low-Input Farming and Territories - Integrating knowledge for improving ecosystem based farming. LIFT Deliverable D3.3.

Védrine, L., & Larmet, V. (2021). Additionnalité des Mesures Agro-Environnementales et Climatiques: évaluation contrefactuelle de l'efficacité environnementale. Rapport d'étude, INRAE, 83p.

Vérot, D. (1998). Agriculture biologique: évaluation d'un gisement d'emplois. FNAB, Ministère de l'Aménagement du territoire et de l'environnement.

Wunder, S. (2015). Revisiting the concept of payments for environmental services. *Ecological Economics*, 117, 234-243.

Wunder, S., Brouwer, R., Engel, S., Ezzine-de-Blas, D., Muradian, R., Pascual, U., & Pinto, R. (2018). From principles to practice in paying for nature's services. *Nature Sustainability*, 1, 145–150.

Wunder, S., Engel, S., & Pagiola, S. (2008). Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries. *Ecological Economics*, 65(4), 834-852.

Wünscher, T., Engel, S., & Wunder, S. (2008). Spatial targeting of payments for environmental services: a tool for boosting conservation benefits. *Ecological Economics*, 65, 822–833.

Zawalińska, K., Giesecke, J., & Horridge, M. (2013). The consequences of Less Favoured Area support: a multi-regional CGE analysis for Poland. *Agricultural and Food Science*, 22(2).

Zawalińska, K., & Krupin, V. (2021). Dynamics of productivity and efficiency performance in Poland's dairy farms: comparative analysis by different degrees of ecological approaches. Unpublished LIFT working paper.

9 Appendix: Supplementary material of the meta-analysis in Section 4

9.1 Document search algorithm

Given the objective of the meta-analysis to include studies that evaluate the impact of PES of different types, a large number of keywords are used in the searching algorithm in order to increase the chance of selecting all relevant studies. The keywords are divided into three groups. The first group consists in the four following lists:

a) biodiversity, conservation, agrobiodiversity, ecological goods, habitat

b) ecosystem services, environmental services, agri-environmental services, organic farming

c) inorganic input reduction, pesticides reduction, fertilisers reduction, water quality

d) deforestation, carbon sequestration, carbon offset, greenhouse gas, climate change, climate mitigation

To minimise the number of non-target documents, the two following groups of keywords are used:

i) payment, reward, compensation, investment, incentive, subsidy

ii) evaluation, assessment, additionality, effectiveness, impacts, effects

The boolean "OR" is used between two keywords and the boolean "AND" is used between two different groups of keywords. For example, using the list of keywords in a) combined to the lists i) and ii), the following algorithm is run:

(biodiversity OR conservation OR agrobiodiversity OR "ecological goods" OR "natural habitat")

AND (payment OR reward OR compensation OR investment OR incentive OR subsidy)

AND (evaluation OR assessment OR additionality OR effectiveness OR impacts OR effects).

9.2 Figures and tables

Figure S1: Selection procedure of the documents

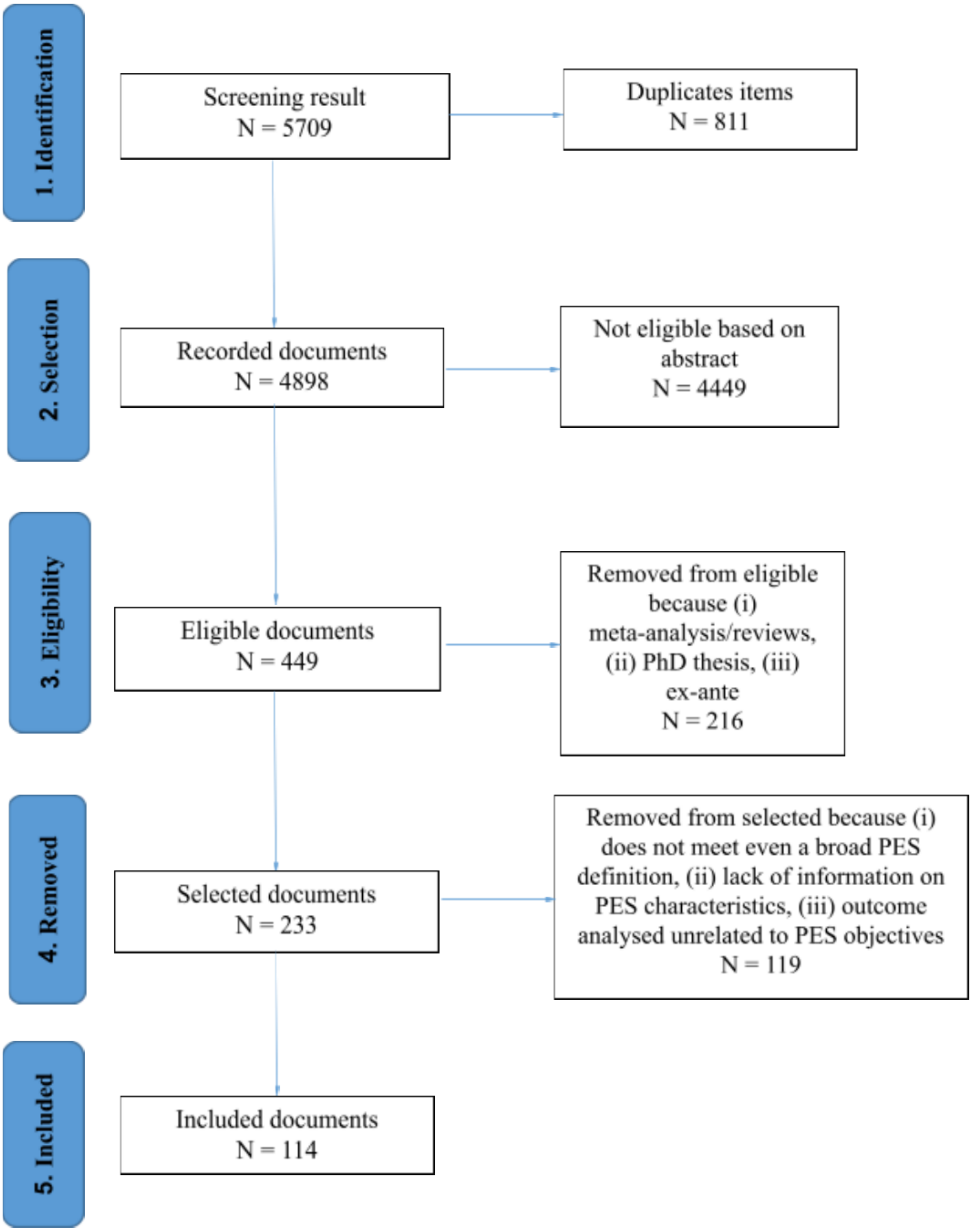


Figure S2: Distribution of PES characteristics in the sample

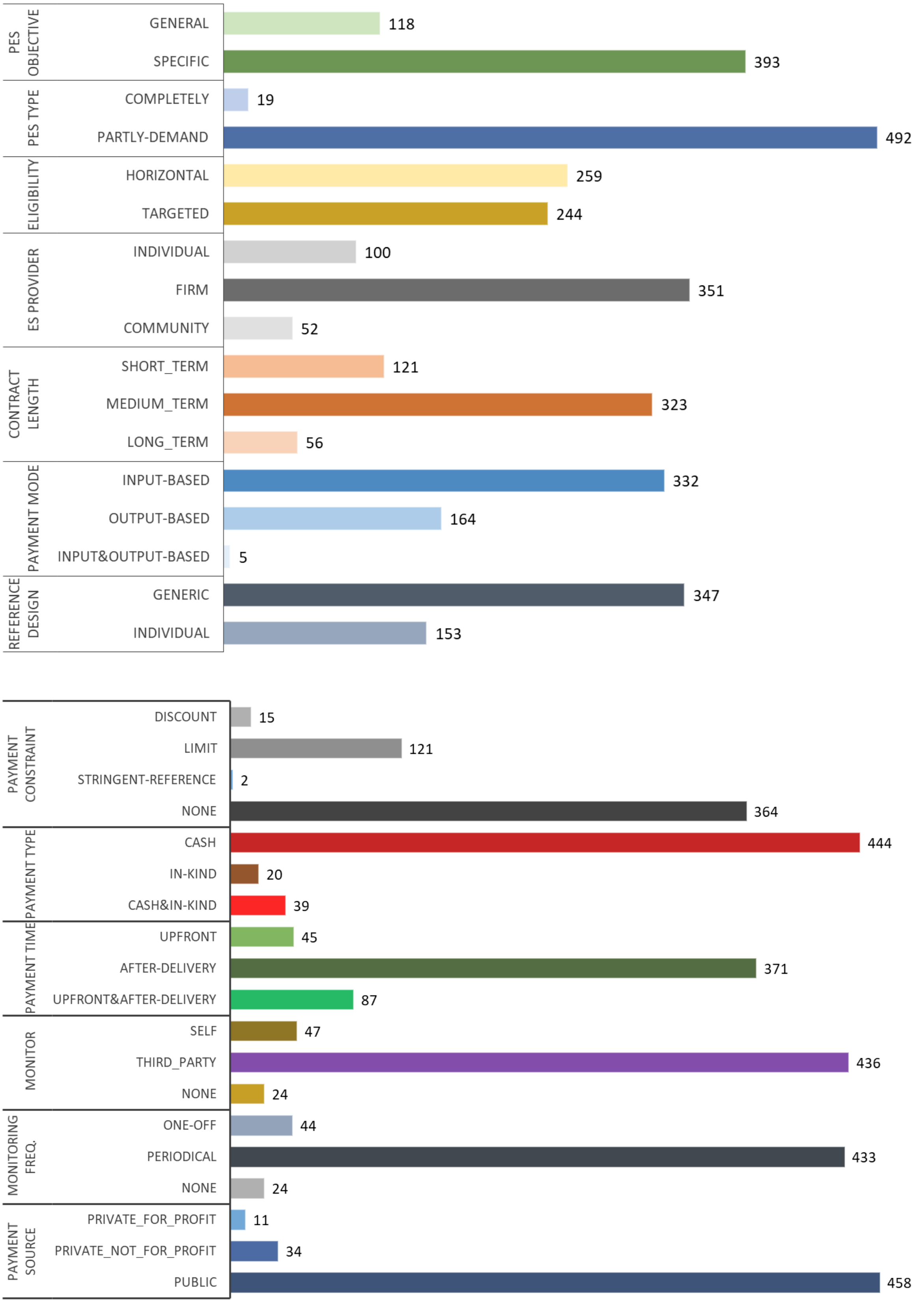


Figure S3: Funnel plot of the additionality level in the selected studies in the sample

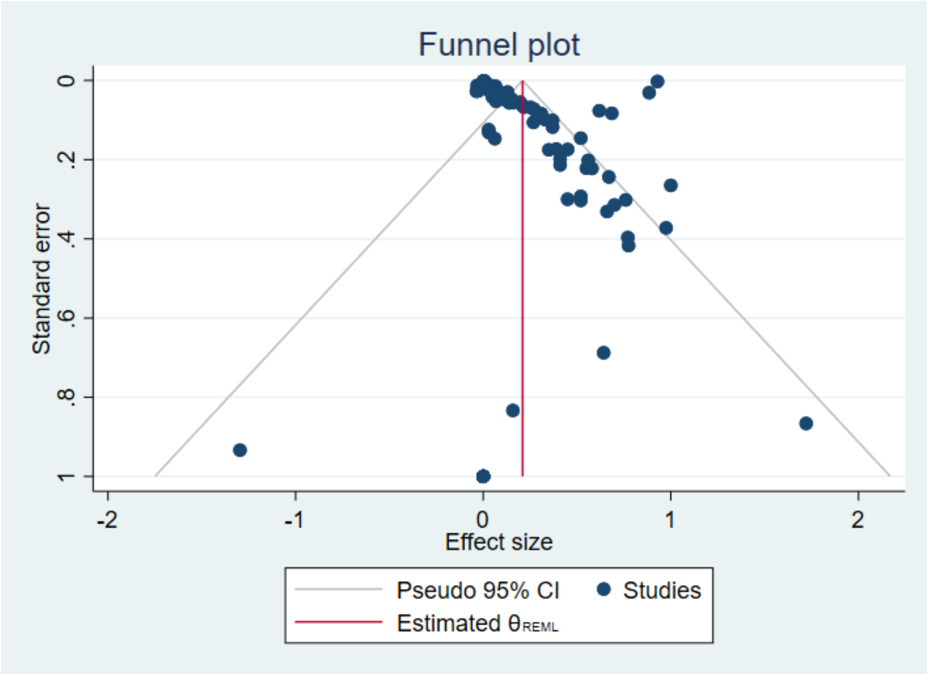


Figure S4: Forest plot for the additionality level in the selected studies in the sample

Study	TE	seTE		95%-CI	Weight
Woodward 2016	0.00	1.0000	0.00	[−1.96; 1.96]	0.1%
Lu 2014	0.98	0.3724	0.98	[0.25; 1.70]	0.5%
Lu 2014	0.64	0.6877	0.64	[−0.71; 1.99]	0.2%
Ito 2015	0.00	0.0020	0.00	[0.00; 0.00]	1.8%
Ito 2015	0.04	0.0130	0.04	[0.02; 0.07]	1.8%
Zheng 2013	0.03	0.0270	0.03	[−0.02; 0.09]	1.8%
Zheng 2013	0.05	0.0425	0.05	[−0.03; 0.13]	1.7%
Zheng 2013	0.03	0.1313	0.03	[−0.23; 0.29]	1.3%
Zheng 2013	0.03	0.1244	0.03	[−0.21; 0.27]	1.4%
Coderoni 2018	0.00	0.0010	0.00	[0.00; 0.00]	1.8%
Coderoni 2018	0.00	0.0010	0.00	[0.00; 0.00]	1.8%
Jayachandran 2017	0.41	0.2134	0.41	[−0.01; 0.83]	0.9%
Jayachandran 2017	0.77	0.3967	0.77	[−0.01; 1.55]	0.4%
Jayachandran 2017	0.76	0.3017	0.76	[0.17; 1.35]	0.6%
Jayachandran 2017	0.55	0.2217	0.55	[0.12; 0.98]	0.9%
Jayachandran 2017	0.70	0.3143	0.70	[0.08; 1.32]	0.6%
Jayachandran 2017	0.41	0.1985	0.41	[0.02; 0.80]	1.0%
Sanchez−Azofeifa 2006	0.00	1.0000	0.00	[−1.96; 1.96]	0.1%
Eshoo 2018	1.00	0.2650	1.00	[0.48; 1.52]	0.8%
Arriagada 2012	0.13	0.0290	0.13	[0.07; 0.19]	1.8%
Arriagada 2012	0.15	0.0467	0.15	[0.06; 0.24]	1.7%
Arriagada 2012	0.11	0.0336	0.11	[0.04; 0.18]	1.8%
Arriagada 2012	0.05	0.0152	0.05	[0.02; 0.08]	1.8%
Arriagada 2012	0.11	0.0348	0.11	[0.04; 0.18]	1.8%
Arriagada 2012	0.14	0.0566	0.14	[0.03; 0.25]	1.7%
Arriagada 2012	0.16	0.0564	0.16	[0.05; 0.27]	1.7%
Arriagada 2012	0.10	0.0330	0.10	[0.04; 0.16]	1.8%
Arriagada 2012	0.13	0.0376	0.13	[0.06; 0.20]	1.8%
Arriagada 2012	0.09	0.0307	0.09	[0.03; 0.15]	1.8%
Grenestam 2018	0.00	1.0000	0.00	[−1.96; 1.96]	0.1%
Grenestam 2018	−0.04	0.0274	−0.04	[−0.09; 0.02]	1.8%
Grenestam 2018	0.01	0.0050	0.01	[0.00; 0.01]	1.8%
Grenestam 2018	0.02	0.0066	0.02	[0.00; 0.03]	1.8%
Grenestam 2018	0.05	0.0364	0.05	[−0.02; 0.12]	1.8%
Grenestam 2018	0.00	0.0064	0.00	[−0.01; 0.02]	1.8%
Grenestam 2018	0.01	0.0012	0.01	[0.00; 0.01]	1.8%
Grenestam 2018	0.07	0.0147	0.07	[0.04; 0.09]	1.8%
Grenestam 2018	−0.03	0.0265	−0.03	[−0.08; 0.03]	1.8%
Grenestam 2018	0.01	0.0036	0.01	[0.00; 0.01]	1.8%
Grenestam 2018	0.00	0.0019	0.00	[0.00; 0.01]	1.8%
Grenestam 2018	−0.03	0.0127	−0.03	[−0.06; −0.01]	1.8%
Grenestam 2018	0.01	0.0030	0.01	[0.00; 0.01]	1.8%
Grenestam 2018	−0.00	0.0042	−0.00	[−0.01; 0.01]	1.8%
Grenestam 2018	−0.01	0.0080	−0.01	[−0.03; 0.00]	1.8%
Grenestam 2018	−0.00	0.0025	−0.00	[−0.01; 0.00]	1.8%
Alix−Garcia 2012	0.37	0.1009	0.37	[0.17; 0.57]	1.5%
Alix−Garcia 2012	0.35	0.1750	0.35	[0.01; 0.69]	1.1%
Alix−Garcia 2012	0.58	0.2222	0.58	[0.14; 1.02]	0.9%
Alix−Garcia 2012	0.66	0.3310	0.66	[0.01; 1.31]	0.6%
Alix−Garcia 2012	0.52	0.3029	0.52	[−0.07; 1.11]	0.6%
Alix−Garcia 2012	0.33	0.0990	0.33	[0.14; 0.52]	1.5%
Alix−Garcia 2012	0.56	0.2018	0.56	[0.16; 0.96]	1.0%
Alix−Garcia 2012	0.37	0.1177	0.37	[0.14; 0.60]	1.4%
Alix−Garcia 2012	0.52	0.2928	0.52	[−0.05; 1.09]	0.7%
Alix−Garcia 2012	0.39	0.1733	0.39	[0.05; 0.73]	1.1%
Alix−Garcia 2012	0.52	0.1457	0.52	[0.23; 0.81]	1.3%
Alix−Garcia 2012	0.67	0.2438	0.67	[0.19; 1.15]	0.8%
Chabe−Ferret 2013	0.62	0.0766	0.62	[0.47; 0.77]	1.6%
Chabe−Ferret 2013	0.93	0.0026	0.93	[0.92; 0.93]	1.8%
Schall 2015	0.05	0.0224	0.05	[0.00; 0.09]	1.8%
Schall 2015	0.03	0.0216	0.03	[−0.01; 0.08]	1.8%
Alix−Garcia 2015	0.45	0.1742	0.45	[0.11; 0.79]	1.1%
Sims 2017	0.29	0.0800	0.29	[0.13; 0.44]	1.6%
Sims 2017	0.29	0.0840	0.29	[0.13; 0.46]	1.6%
Sims 2017	0.31	0.0840	0.31	[0.15; 0.47]	1.6%
Sims 2017	0.29	0.0840	0.29	[0.13; 0.46]	1.6%
Sims 2017	0.25	0.0680	0.25	[0.12; 0.38]	1.6%
Sims 2017	0.06	0.1470	0.06	[−0.23; 0.35]	1.3%
Ruggiero 2019	1.72	0.8662	1.72	[0.02; 3.42]	0.1%
Ruggiero 2019	0.88	0.0309	0.88	[0.82; 0.95]	1.8%
Ruggiero 2019	−1.30	0.9336	−1.30	[−3.13; 0.53]	0.1%
Ruggiero 2019	0.21	0.0659	0.21	[0.08; 0.34]	1.7%
Ruggiero 2019	0.16	0.8333	0.16	[−1.47; 1.79]	0.1%
Ruggiero 2019	0.78	0.4170	0.78	[−0.04; 1.59]	0.4%
Wu 2013	0.00	1.0000	0.00	[−1.96; 1.96]	0.1%
Wu 2013	0.00	1.0000	0.00	[−1.96; 1.96]	0.1%
Hayes 2017	0.12	0.0490	0.12	[0.02; 0.21]	1.7%
Random effects model			**0.20**	**[0.14; 0.26]**	**100.0%**

Heterogeneity: $I^2 = 100\%$, $\tau^2 = 0.0507$, $p = 0$

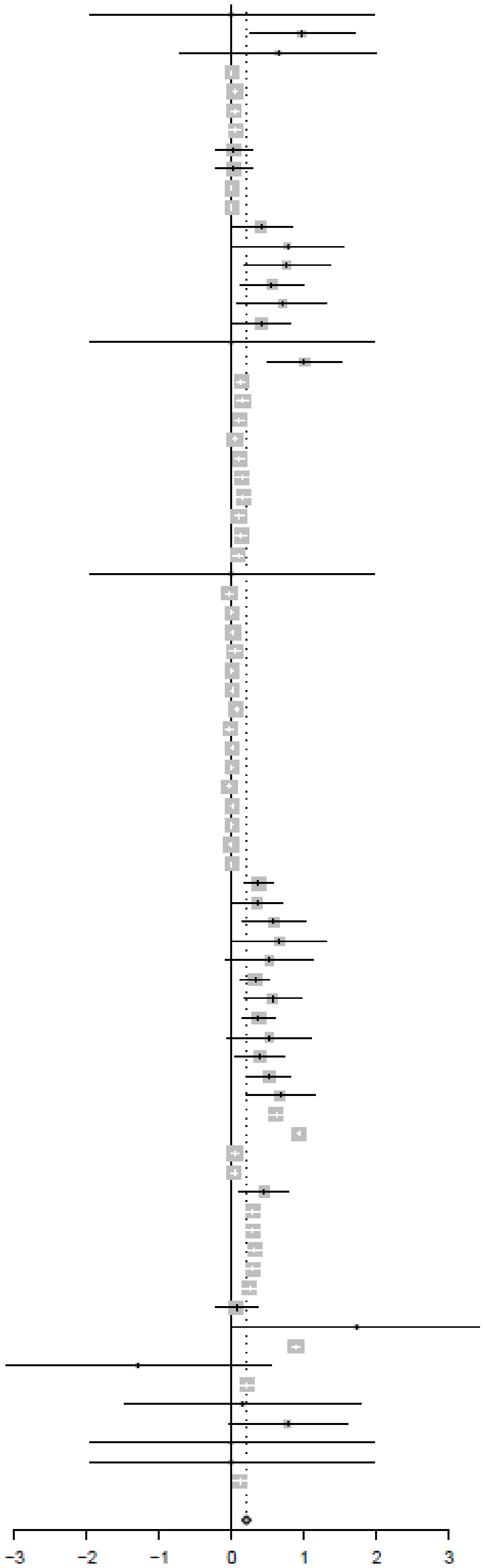


Table S1: Pairwise comparison of the probability of PES to increase ES from the logistic regression

Pairwise comparisons	Full sample	p-value ≤ 0.1	p-value ≤ 0.05	p-value ≤ 0.01	p-value ≤ 0.001
biodiversity vs GHG_emission	0.022	0.312*	0.341*	0.339**	0.299***
carbon_sequestration vs GHG_emission	-0.050	0.138	0.112	0.133	0.171
deforestation vs GHG_emission	0.064	0.239	0.287	0.263	0.208**
organic_farming vs GHG_emission	0.198***	0.138	0.185	0.267**	0.209***
other_topics vs GHG_emission	0.191*	0.292	0.346	0.291	0.334***
water_quality vs GHG_emission	-0.071	-0.100	-0.042	0.051	0.110*
carbon_sequestration vs biodiversity	-0.072	-0.174	-0.229	-0.206	-0.128
deforestation vs biodiversity	0.042	-0.073	-0.055	-0.076	-0.091
organic_farming vs biodiversity	0.176***	-0.175*	-0.156*	-0.072	-0.090
other_topics vs biodiversity	0.169**	-0.020	0.005	-0.047	0.035
water_quality vs biodiversity	-0.094	-0.412***	-0.384***	-0.287***	-0.188*
deforestation vs carbon_sequestration	0.114	0.102	0.175	0.130	0.038
organic_farming vs carbon_sequestration	0.248**	0.000	0.073	0.134	0.038
other_topics vs carbon_sequestration	0.241**	0.154	0.234	0.159	0.163
water_quality vs carbon_sequestration	-0.022	-0.238*	-0.154	-0.081	-0.060
organic_farming vs deforestation	0.134**	-0.102	-0.101	0.004	0.001
other_topics vs deforestation	0.127*	0.053	0.059	0.028	0.126
water_quality vs deforestation	-0.136	-0.339**	-0.329**	-0.212	-0.098
other_topics vs organic_farming	-0.007	0.155	0.161	0.025	0.125
water_quality vs organic_farming	-0.270***	-0.238**	-0.228***	-0.215***	-0.098*
water_quality vs other_topics	-0.263***	-0.392***	-0.389***	-0.240*	-0.223

Note: The dependent variable (ES increase) takes the value 1 if the PES effect is positive and zero otherwise; estimation in columns use the full sample and subsamples with the dependent variable set to 1 if the p-value statistic is ≤ 0.1, ≤ 0.05, ≤ 0.01 and ≤ 0.001, respectively. *** p<0.01, ** p<0.05, * p<0.1.

Table S2: Marginal effects of PES characteristics effect-size direction from logistic regression by level of significance for the reported effect

Variables		Full sample	p-value ≤ 0.1	p-value ≤ 0.05	p-value ≤ 0.01	p-value ≤ 0.001
PES characteristics (hypotheses)						
PES type (H1)	*Coasean*	0.033	0.100	0.068	0.047	0.102
		(0.207)	(0.247)	(0.235)	(0.219)	(0.231)
PES objective (H2)	*es_specific*	0.073	-0.022	0.024	0.017	0.102
		(0.079)	(0.134)	(0.107)	(0.117)	(0.236)
Eligibility (H3)	*spatial_targeted*	0.086	0.188**	0.143*	0.144**	0.058
		(0.059)	(0.074)	(0.076)	(0.062)	(0.063)
Payment mode (H4)	*output-based*	-0.087	0.061	0.018	-0.030	-0.046
		(0.054)	(0.070)	(0.074)	(0.072)	(0.083)
Reference design (H5)	*individual*	-0.040	-0.159**	-0.172**	-0.124	-0.153**
		(0.052)	(0.072)	(0.069)	(0.081)	(0.072)
Payment type (H6)	*cash*	0.052	0.097	0.125	0.073	0.317
		(-0.208)	(-0.367)	(-0.285)	(-0.262)	(-0.291)
	cash&in-kind	0.030	0.253	0.191	0.111	0.195
		(0.222)	(0.311)	(0.230)	(0.179)	(0.175)
Contract length (H7)	*long-term*	0.032	-0.096	0.049	0.202	0.311**
		(-0.123)	(-0.174)	(-0.158)	(-0.164)	(-0.158)
	medium-term	0.046	-0.064	0.066	0.164	0.064
		(0.083)	(0.132)	(0.107)	(0.116)	(0.113)
Monitor (H8)	*third-party*	-0.150*	0.281*	0.594***	0.622***	0.491**
		(0.080)	(0.166)	(0.193)	(0.207)	(0.226)
Monitoring frequency (H8)	*periodical*	0.234**	0.394***	0.481***	0.645***	0.538
		(0.098)	(0.127)	(0.116)	(0.209)	(0.467)
PES characteristics (controls)						
Payment source	*public*	-0.048	0.224	-0.018	-0.233	-0.211
		(0.144)	(0.198)	(0.182)	(0.177)	(0.135)
Payment constraint	*none*	-0.197**	-0.024	-0.173*	-0.390***	-0.356***
		(0.087)	(0.105)	(0.105)	(0.097)	(0.082)
ES provider	*community*	-0.048	-0.144	-0.154	-0.043	0.127
		(-0.125	(-0.174)	(-0.140)	(-0.122)	(-0.095)

Note: The dependent variable (ES increase) takes the value 1 if the PES effect is positive and zero otherwise; estimation in columns use the full sample and subsamples with the dependent variable set to 1 if the p-value statistic is ≤ 0.1, ≤ 0.05, ≤ 0.01 and ≤ 0.001, respectively. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S2 (Continued)

Variables		Full sample	p-value ≤ 0.1	p-value ≤ 0.05	p-value ≤ 0.01	p-value ≤ 0.001
PES characteristics (controls)						
ES provider	*firm*	-0.164	-0.164	-0.257	-0.021	0.239*
		(0.113)	(0.159)	(0.159)	(0.169)	(0.135)
Study country	*Africa*	0.044	0.051	0.016	-0.104	0.041
		(-0.140)	(-0.213)	(-0.169)	(-0.182)	(-0.193)
	Asia	0.108	0.042	0.182	0.306*	0.133
		(0.098)	(0.166)	(0.164)	(0.158)	(0.188)
Study characteristics						
Study topic	*biodiversity*	-0.067	-0.029	0.102	0.432***	0.281**
		(-0.082)	(-0.112)	(-0.121)	(-0.135)	(-0.116)
	carbon_sequestration	-0.153	-0.196	-0.219	0.100	0.070
		(0.122)	(0.255)	(0.221)	(0.219)	(0.275)
	deforestation	-0.132	-0.101	-0.150	0.110	0.205
		(0.119)	(0.165)	(0.178)	(0.180)	(0.237)
	organic_farming	0.271***	0.066	0.200	0.534***	0.357**
		(0.103)	(0.119)	(0.137)	(0.162)	(0.142)
	Water quality	-0.132	-0.278**	-0.184	0.163	0.118
		(0.094)	(0.127)	(0.142)	(0.165)	(0.130)
Publication type	*peer_reviewed*	0.072	-0.140	-0.070	-0.170	-0.098
		(0.111)	(0.185)	(0.198)	(0.157)	(0.166)
Data type	*aggregate*	0.005	-0.113	-0.190**	-0.141*	0.063
		(0.060)	(0.089)	(0.086)	(0.085)	(0.069)
Study design	*best_design*	0.352***	0.272**	0.353***	0.314**	0.222***
		(0.125)	(0.116)	(0.124)	(0.125)	(0.076)
Estimation method	*robust_method*	-0.110*	0.004	-0.145	0.027	0.047
		(0.063)	(0.094)	(0.101)	(0.093)	(0.100)
	Observations	490	285	285	285	285
	Pseudo R-sq	0.149	0.186	0.208	0.232	0.243

Note: The dependent variable (ES increase) takes the value 1 if the PES effect is positive and zero otherwise; estimation in columns use the full sample and subsamples with the dependent variable set to 1 if the p-value statistic is ≤ 0.1, ≤ 0.05, ≤ 0.01 and ≤ 0.001, respectively. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S3: Cramer's V coefficient (V) measuring the relative strength of association between the PES characteristics included in the model of effect direction

	PES type	PES objective	Eligibility	Contract length	Payment source	Payment mode	Reference design	Payment constraint	Payment type	Payment time	ES provider	Monitor	Monitoring frequency
PES objective	0.08*												
Eligibility	0.04	0.03											
Contract length	0.04	0.28***	0.18***										
Payment source	0.63***	0.02	0.04	0.04									
Payment mode	0.13***	0.05	0.17***	0.04	0.20***								
Reference design (H)	0.03	0.20***	0.02	0.07	0.10**	0.08*							
Payment constraint	0.01	0.03	0.11**	0.01	0.05	0.01	0.02						
Payment type	0.12***	0.21***	0.2***	0.13***	0.08*	0.14***	0.06	0.11					
Payment time	0.06	0.26***	0.14***	0.11*	0.07	0.20***	0.49***	0.24***	0.31				
ES provider	0.14***	0.12***	0.13***	0.01	0.33***	0.36***	0.02	0.12***	0.00	0.07			
Monitor	0.17***	0.10**	0.11**	0.06	0.17***	0.09**	0.05	0.13***	0.06	0.35***	0.08*		
Monitoring frequency	0.05	0.20***	0.07	0.10*	0.11*	0.11**	0.12**	0.20***	0.06	0.28***	0.10*	0.59***	
Study zone	0.28***	0.22***	0.24***	0.49***	0.40***	0.20***	0.13**	0.20***	0.28***	0.23***	0.24***	0.07	0.12***

Note: The association between two PES characteristics is considered as little, low, moderate, and high if Cramer's V coefficient ranges from 0 to 0.1, 0.1 to 0.3, 0.3 to 0.5 and 0.5 to 1, respectively. *** p<0.01, ** p<0.05, * p<0.1.

Table S4: Meta-regression of PES-schemes additionality level on study characteristics

Variables	Tobit	RE	WLS
Intercept	-0.18***	-0.13**	-0.10
	(0.05)	(0.06)	(0.12)
Standard Error	0.22	0.40	1.24
	(0.35)	(0.27)	(1.22)
Study topics			
biodiversity	0.45***	0.38***	0.25
	(0.06)	(0.05)	(0.04)
deforestation	0.26***	0.19**	0.15*
	(0.07)	(0.07)	(0.07)
other_topic	0.19*	0.16	0.38
	(0.00)	(0.06)	(0.00)
Study characteristics			
aggregate_data	0.13**	0.14***	0.10
	(0.05)	(0.05)	(0.12)
best_design	0.02	0.02	-0.24
	(0.08)	(0.8)	(0.14)
robust_method	0.25**	0.23***	0.30
	(0.10)	(0.09)	(0.22)
var(e.additionality)	0.05***		
	(0.02)		
N	90	90	90
R2/pseudo R2	0.39	0.38	0.31

Note: all the dependent variables were selected based on the VIF using the tobit model. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S5: Cramer's V coefficient (V) measuring the relative strength of association between the PES characteristics included in the model of additionality level

	PES type	PES objective	Eligibility	Contract length	Payment source	Payment mode	Reference design	Payment constraint	Payment type	Payment time	ES provider	Monitor	Monitoring frequency
PES objective	0.08												
Eligibility	0.02	0.17**											
Contract length	0.03	0.48***	0.23***										
Payment source	0.64***	0.07	0.08	0.01									
Payment mode	0.21***	0.02	0.23***	0.02	0.28***								
Reference design	0.16**	0.26***	0.09	0.2***	0.32***	0.16**							
Payment constraint	0.01	0.04	0.14**	0.02	0.11	0.43***	0.19***						
Payment type	0.13*	0.16**	0.27***	0.01	0.11	0.17**	0.25***	0.01					
Payment time	0.12	0.11	0.22***	0.16*	0.02	0.18**	0.35***	0.09	0.71				
ES provider	0.11	0.26***	0.16**	0.17**	0.48***	0.17**	0.2***	0.12*	0.2***	0.24***			
Monitor	0.05	0.14**	0.23***	0.04	0.14**	0.13*	0.15**	0.1	0.18**	0.34***	0.03		
Monitoring frequency	0.04	0.14	0.34***	0.17*	0.08	0.31***	0.22***	0.51***	0.14	0.25***	0.11	0.58***	
Study zone	0.34***	0.34***	0.24***	0.54***	0.44***	0.14	0.33***	0.23***	0.3***	0.26***	0.36***	0.05	0.14

Note: The association between two PES characteristics is considered as little, low, moderate, and high if Cramer's V coefficient ranges from 0 to 0.1, 0.1 to 0.3, 0.3 to 0.5 and 0.5 to 1, respectively. *** p<0.01, ** p<0.05, * p<0.1.