Ecological Economics 210 (2023) 107856

Contents lists available at ScienceDirect

Ecological Economics

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

Drivers of PES effectiveness: Some evidence from a quantitative meta-analysis

Legrand D.F. Saint-Cyr [a,*], Lionel Védrine [a,b], Sophie Legras [a], Julie Le Gallo [a], Valentin Bellassen [a]

[a] CESAER UMR1041, INRAE, Institut Agro, Université Bourgogne Franche-Comté, Dijon, France
[b] Université Clermont Auvergne, AgroParis Tech, INRAE, VetAgro Sup, UMR Territoires, Clermont-Ferrand F-63000, France

ARTICLE INFO

Keywords:
Payments for environmental services (PES)
Additionality
Enrolment
Meta-analysis
Effectiveness
Policy design

ABSTRACT

Payments for Environmental or Ecosystem Services (PES) schemes have become a popular tool to address environmental degradation and to promote sustainable management of ecosystem services. We use meta-regression analysis on a sample of 110 individual studies to investigate the determinants of the environmental effectiveness, defined as the probability to increase environmental services (ES) provision, of about 149 PES-schemes implemented worldwide. We find that increased effectiveness of PES schemes is strongly associated with periodical third-party monitoring, generic reference design and to a lesser extent results-based payments. We further study the determinants of PES additionality, defined as direct changes in ES provision induced by the PES scheme, compared to a baseline without PES, on a smaller sample of 41 studies from which we could obtain the necessary data. The results confirm the role of certain design variables, such as monitoring type, and raise a potential trade-off between enrolment and additionality in the assessment of PES effectiveness.

1. Introduction

Payments for Environmental or Ecosystem Services (PES, hereafter) 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 provision of Environmental Services (ES, hereafter) are implemented worldwide, there still exists an academic debate on what PES actually refers to (Derissen and Latacz-Lohmann, 2013). A PES is commonly defined 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 *free market PES*, 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 *regulated demand PES*, 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 environmental effectiveness, which depends on design characteristics as well as on the context of implementation (Wunder et al., 2018).

The effectiveness of PES typically rests on two synergetic pillars: enrolment and additionality. 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, additionality is also paramount: if a faulty design results in payments for what most ES providers would have done anyways, a PES will have little effectiveness even if it manages to enrol a large number of potential ES providers.

Various studies investigate the effectiveness of a number of PES schemes and show mixed evidence (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., Kleijn and Sutherland, 2003; Brouwer et al., 2011; Scheper et al., 2013). Ezzine-de-Blas et al. (2016) is the first to summarize the effect of PES schemes implemented around the world based

* Corresponding author at: CESAER UMR1041, INRAE, Institut Agro, Université Bourgogne Franche-Comté, 26, bd Docteur Petitjean, BP 87999, 21079 Dijon Cedex, France.
E-mail address: legrand.saintcyr@gmail.com (L.D.F. Saint-Cyr).

https://doi.org/10.1016/j.ecolecon.2023.107856
Received 23 July 2021; Received in revised form 29 November 2022; Accepted 17 April 2023

on a broad quantitative meta-analysis.

We go further as the main objective of our paper is to determine, based on a meta-analysis, whether PES schemes achieve significant environmental impacts, measured either as the probability to increase ES provision (effectiveness) or as their additionality. 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 perform the analysis of additionality on a sample of 41 individual studies, less than for effectiveness, due to data availability. We also identify the role of the main PES characteristics on their additionality, and by re-assessing effectiveness on the smaller additionality sample we are able to decompose the drivers of effectiveness between enrolment and additionality. 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.

H1. free market *PES* > regulated demand *PES*.

Free market PES are expected to be more effective than regulated demand PES. This outcome is predicted by contract theory, due to the principal-agent problem (Engel et al., 2008; Ferraro, 2008). In free market PES, the end-consumer of the environmental service is directly involved in the contract and is therefore interested in closely scrutinizing that the expected outcome is achieved. In regulated demand 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 Organization, 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 et al., 2018; Wunder et al., 2008; Engel, 2016). Theoretically, this hypothesis is supported by the classical multi-objectives optimization theory: as soon as additional objectives are added, the solution which maximized the initial ES becomes one point on the Pareto frontier, with other points decreasing the supply of the initial ES.

H3. spatial targeting + .

Spatial targeting is expected to increase the additionality of PES. While the classical solution to capture part of the information rent of ES suppliers is an auction, auctions are not always feasible in practice and spatial targeting can be an alternative. The heterogeneity of PES supply costs often has a spatial structure. Varying payment depending on location is a way for the regulator to pay less than the overall marginal cost where a lower payment is sufficient to trigger action (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 additionality (Wunder et al., 2018; Wünscher et al., 2008; Sims et al., 2014).

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

Result-based payments are expected to be more effective than practice-based payments, because in principle they allow for better targeting of efficient ES providers (Wuepper and Huber, 2022) but also because a majority of farmers prefer the result-based payments conditions over the practice-based payments (Šumrada et al., 2022) This is why the European Commission is currently pushing towards more result-based subsidies in the Common Agricultural Policy (European Commission, 2018). However, the implementation of result-based payments is more difficult and costlier than practice-based payments. Hence, in practice, 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). From the latter point, we expect that the results obtained with result-based payments (included in our meta-analysis) will be equivalent to the results obtained with practice-based payments.

H5. individual reference > generic reference.

Setting counterfactual/reference level of ES provision is another practical design feature addressing the information rent. Implicitly or explicitly, PES value environmental services provided in addition to a baseline level, which may be a generic zero. A generic zero generates windfall profits for ES providers which would have provided some level of ES even without payment. Some PES address this issue by setting a non-zero - and possibly stringent - generic counterfactual level of ES provision, while others set it on a case-by-case basis, based on the individual characteristics of the recipient (Shishlov and Bellassen, 2012). PES are expected to be more additional if the reference level is stringent (non-zero) or not generic (Bento et al., 2015; Cormier and Bellassen, 2013). However, the superiority of individual reference may be balanced out by the possibility for project developers to game regulators where an individual reference has to be estimated (e.g., Dechezleprêtre et al., 2014; Shishlov and Bellassen, 2012).

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 additional than community-level payments hampered by free-riding.

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

PES offering long term contracts to ES providers may be more additional than those with shorter terms, since many ecosystem services may take time to be regenerated (Meyer et al., 2015; Moxey et al., 1999). However, in practice, stated-preference studies have shown that farmers prefer shorter term contracts as a proxy for flexible contractual arrangements (Christensen et al., 2011; Ruto and Garrod, 2009). Hence, we may expect a negative impact of contract length on PES effectiveness if the enrolment effect is stronger than the additional one.

H8a. third-party monitoring > internal monitoring.

H8b. regular monitoring > one-off monitoring.

Monitoring systems that constrain compliance are expected to increase PES additionality (Honey-Roses 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). Third party monitoring and regular monitoring is therefore expected to increase additionality. Regarding effectiveness however, a trade-off between the costs and benefits of increasing monitoring stringency and frequency is expected (Bellassen and Shishlov, 2017), especially as monitoring/verification costs make the bulk of the overall transaction costs of PES (Bellassen et al., 2015) and may deter participation in PES.

The contribution of this article is fourfold. Firstly, we perform a meta-analysis on the impact of PES on the provision of ecosystem services (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 free market and regulated demand PES, expanding from the Ezzine-de-Blas et al. (2016) meta-analysis that focuses on free market 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 additionality.

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 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 additionality. To estimate the impact of PES-schemes characteristics on their effectiveness, we implement a logit specification and 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-schemes characteristics identified in a sample of 110 individual studies that investigate the performance of about 149 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 additionality. 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.

This paper is structured as follows. Section 2 presents the selection process of the documents included in the meta-analysis and the methods used to test the hypotheses. The main results are presented in Section 3. A discussion of the results and some concluding remarks are presented in Section 4.

2. Material and method

2.1. Data collection and design

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.[1] We screened Science Direct (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, 4898 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 synthesis 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, (iii) *ex-ante* analyses and evaluations using simulation methods on fictive PES, and (iv) previous versions of published studies. We analysed the remaining 229 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 110 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 Fig. S1 in the supplementary materials.

The selected studies date from 1983 to 2019 and about 90% are published articles. About 74% of the studies were published from 2010. The studies deal with a wide range of topics and were conducted in different locations (see Fig. 1).[2] A majority of studies worldwide concern PES schemes on biodiversity (27%) and deforestation (26%).[3] The studies conducted in Europe including the United Kingdom (UK) 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 GHG emission (33%) and water quality (21%) 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. The full sample includes 499 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 (see the database in the supplementary materials, and table S7 for some descriptive statistics). Overall, the number of individual observations reporting a positive effect of PES on ES provision is greater than that reporting negative or null effects (Fig. 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 are not significantly different from zero at a 10% level. About 63% of observations reporting a negative effect and 33% of those reporting a positive effect have a p-value larger than 0.1 (as shown in Fig. 2, panel (a), orange bars). Summary statistics of PES outcomes and additionality level are by PES characteristics and the contingency table of PES characteristics as well as results obtained including only papers relying on quasi-experiments are available in the supplementary material.

As stated before, our meta-analysis aims to investigate the impact of PES characteristics on their effectiveness considering both their enrolment and additionality. However, not all the necessary information was

[1] All lists of keywords are presented in the supplementary materials.

[2] It should be noted that a study evaluating the effectiveness of a specific PES may be conducted in more than one country. For example, Daugbjerg, et al., (2011) investigated the impact of subsidies for organic farming and conservation practices implemented in Denmark and the UK.

[3] In our data collection, we differentiated biodiversity study topics from deforestation. As pointed out by a referee, biodiversity and deforestation may be related to each other as the main reasons for wanting to protect forests is to conserve biodiversity. We classified other studies with deforestation as the primary and sole topic, when the study referenced multiple, undiscriminated environmental issues as rationales to preserve the forest cover, and the PES was motivated by preserving the forest cover.

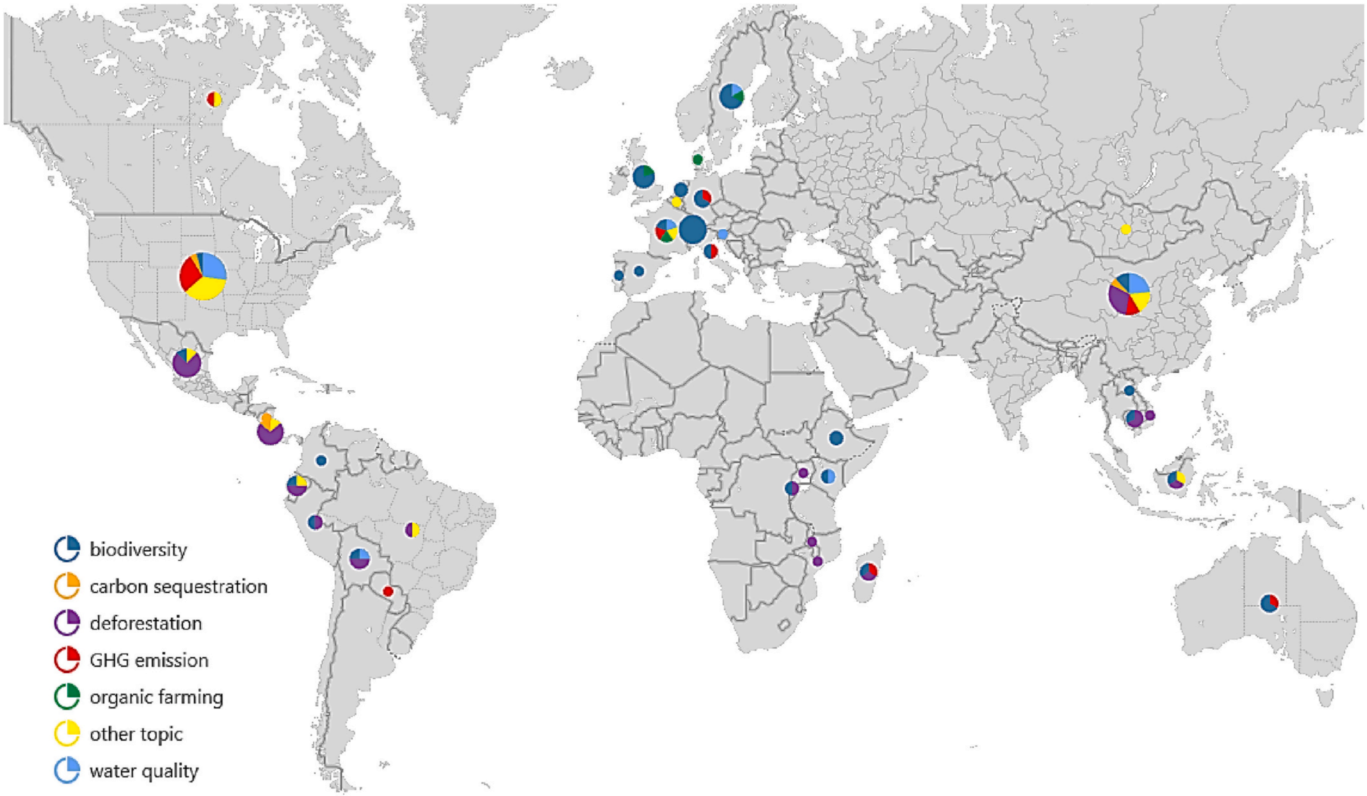
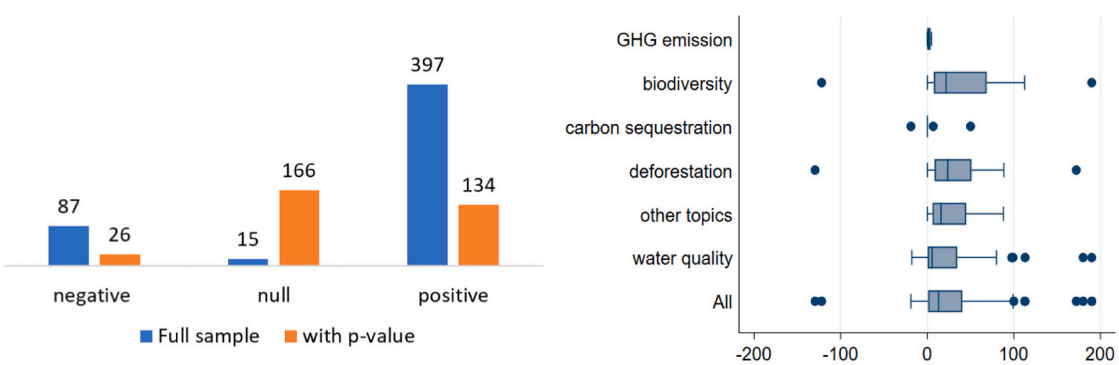


Fig. 1. PES study topics by location.



a. Impact of PES on ES provision b. Additionality level (in %)

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

always explicitly present in the articles selected for this analysis. We thus completed the database in two ways.

Note: Statistics on panel (a) and (b) are for the full sample ($N = 499$) and for the sample with available information on additionality ($N = 174$), 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 with a *p*-value or not in the whole sample (blue) and after recoding status based on reported *p*-value statistics (orange) where the effect is considered to be positive or negative if *p*-value ≤ 0.1 and null otherwise.

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 2. The studies in the sample investigate the effectiveness of 149 PES schemes.[4] 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 3.3.

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

[4] The list of the references used in the meta-analysis and the PES evaluated in the studies are provided in the supplementary materials (see Table S11+ references).

using the delta method (Oehlert, 1992) and the *p*-value statistic considering a normal distribution of the error terms. The level of additionality ranges from −130% to 190% with <8% of negative values (see Fig. 2 panel b). 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. Negative values mean that the PES actually decreased ES provision, an unintended outcome which rarely happens – only 13 observations from 5 studies find negative effects on the ecosystem service studied (particularly on water quality). The PES effect is statistically significantly different for 0 for 8 of them. 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 174 observations for which the additionality level was recorded in the database, we were able to derive standard errors for 90 observations which represent 21 individual studies in the sample.

2.2. 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 2 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.

2.2.1. 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 policymakers, 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 $A = \ exp\ (\sum_1^J \beta_j p_{ij} + \sum_1^K \gamma_k x_{ik})$; 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., 2006).[5] 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.

As mentioned above, a number of studies investigate the effectiveness or the level of additionality of a specific PES considering different dependent variables to measure the variation of the ES level. To avoid double count in our database, we include in the sample used for estimation only results that are different for the PES investigated in the study considering both the direction of the effect and the level of level of significance. The sample used for the logit estimation was thus reduced to 179 observations.

2.2.2. PES additionality

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

$$\widehat{\theta}_i = \theta_0 + \alpha SE_i + \sum\nolimits_1^J \beta_j p_{ij} + \sum\nolimits_1^K \gamma_k x_{ik} + \varepsilon_i \tag{3}$$

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; and θ_0, α, γ_k and β_j are parameters to be estimated; 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 (Fig. 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 Eq. (3). A test for a significant effect of PES schemes is H_0: $\theta_0 = 0$ which is the PET and a valid test for publication bias is H_0: $\alpha = 0$ which is the FAT. We then investigate the effect of PES characteristics on the level of additionality using the full specification of Eq. (3).

The error term in Eq. (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 σ_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 $1/(\sigma_i^2 + \tau^2)$, where σ_i^2 is the variance of the estimates of the estimated effect in the study and τ^2 is the between-study heterogeneity parameter estimated from the sample of studies controlling by the study characteristics. Once again, the standard errors on the estimates from the FAT-PET-MRA are clustered at the study-level as multiple additionality levels may be reported in one study.

For robustness checks, we estimate the parameters from Eq. (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

[5] As pointed out by a referee, some studies remain unpublished because of serious data and/or methodological deficiencies. We hence performed a robustness check including only studies published in peer-reviewed journal.

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., $1/\sigma_i^2$) as weights. To avoid double count, we aggregate the level of additionality from each study by specific PES. The corresponding standard errors are derived using the delta method. The sample used for the FAT-PET-MRA estimation was thus reduced only 41 observations.[6]

2.2.3. Decomposing effectiveness into enrolment and efficiency

In an attempt to distinguish between enrolment drivers and additionality drivers on the same sample, the meta-regression specification used for the assessment of effectiveness on the entire sample (Section 3.2.1) is reiterated on the studies for which an estimate of additionality is available (Section 3.2.2). Indeed, the enrolment effect of a given independent variable can be deduced from the estimates of its effects on effectiveness and additionality. For example, if a variable is associated with higher additionality but not with higher effectiveness, one can deduct that it is likely associated with lower enrolment, explaining why higher additionality does not translate into higher effectiveness.

2.3. Explanatory variables and identification strategy

We use a set of explanatory variables to test the hypotheses stated in Section 2. Following the consideration in Section 2, we test hypotheses H1 to H7 using, respectively, the variables describing PES type (*free market* 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) as well as ES provider (*individual, firm, community*) and contract length (*short-term*, *medium-term* or *long-term*). We use variables that describe the monitoring system (*self*, *third-party* or none) and monitoring frequency (*one-off* or *periodical*) to test hypotheses H8a and H8b. 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).

We group together the study designs and the estimation techniques in one categorical variable that takes three modalities: *quasi-experiments* includes propensity score matching, difference in difference (DID), DID-matching or regression on discontinuity (RD); *other_causal_methods* refers to randomized control trial and other causal methods such as generalized moment method (GMM); and *not_causal* groups all other estimation methods. We also use a dichotomous variable (*aggregate*) for data type that takes the value 1 if the effect-size is estimated using data collected at any spatial level. We group the countries where the PES were implemented by their respective continent except for Australia which we put in North America considering its similarity with the countries of this continent (see Table 1).

To deal with multicollinearity problems, we reduce the number of modalities of the explanatory variables. We divided each PES characteristic in two modalities by recoding the explanatory variables (see Table 1). All the PES characteristics are included in the model specification for the analysis of PES effectiveness while only the variables that allow testing our stated hypotheses are used for the additionality level given the sample size.

[6] Although the number of observations is small relative to the large number of variables (hypotheses) being tested, the FAT-PET-MRA is estimable given the inclusion of the inverse of standard errors as a weighted factor in the model specification, which provides a sufficient degree of freedom for interpretation of the estimated coefficients.

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)	*free_market*	if buyers of ES are free to buy or not (absence of regulatory mandate to purchase any amount of ES)	0.061	0.238
PES objective (H2)	*es_targeted*	if specific ES categories are targeted	0.756	0.430
Eligibility (H3)	*spatial_targeted*	if specific ES providers or specific locations are eligible	0.548	0.500
Payment mode (H4)	*output-based*	if the payment is directly linked to the ES provision	0.399	0.491
Reference design (H5)	*individual*	if the payment is estimated on the basis of individual reference	0.359	0.481
Payment type (H6)	*cash*	if the payment is made in cash	0.853	0.355
Contract length (H7)	*long-term*	if the contract length is >10 years	0.112	0.316
Monitoring system (H8a)	*third-party*	if PES implementation is monitored by public or independent auditor	0.824	0.382
Monitoring frequency (H8b)	*periodical*	if the PES is monitored at different steps of the implementation	0.823	0.382
PES characteristics (controls)				
Payment time	*after-delivery*	if payment is provided after the ES delivery	0.703	0.458
Payment constraint	*none*	if no payment constraint is imposed	0.755	0.431
Payment source	*public*	if the PES is linked to governmental or international programs	0.871	0.336
ES provider	*individual*	if ES are provided by a local community	0.226	0.419
PES characteristics (controls)				
Study country	*Africa*	if the PES is implemented in an African country	0.067	0.251
	Asia	if the PES is implemented in an Asian country	0.141	0.349
	Europe	if the PES is implemented in a European country	0.399	0.491
	Latin_America	if the PES is implemented in a Latin American country	0.198	0.399
Study characteristics				
Study topic	*biodiversity*	if biodiversity is the main topic of the study	0.311	0.462
	carbon_sequestration	if carbon sequestration is the main topic of the study	0.046	0.209
	deforestation	if deforestation is the main topic of the study	0.191	0.394
	GHG_emission	if GHG emission is the main topic of the study	0.127	0.334
	organic_farming	if organic farming is the main topic of the study	0.067	0.251
	water_quality	if water quality is the main topic of the study	0.170	0.375
Publication type	*peer-reviewed*	if peer-reviewed research article	0.905	0.294
Data type	*aggregate*	if the data are collected at any spatial level	0.473	0.500

(*continued on next page*)

Table 1 (*continued*)

Variables	Modalities	Description	Mean	Std. Dev.
Study design	*quasi-experiments*	if propensity score matching, difference in difference (DID) or DID-matching or regression on continuity (RD)	0.127	0.334
Estimation method	*other_causal_methods*	if randomized control trial and other causal method	0.032	0.176

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 Fig. S2 in the supplementary materials.
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 Table S1 in the supplementary materials.

3. Results and discussion

3.1. PES effectiveness

The average predicted probability of a positive effect of PES schemes on the provision of ES varies across study topics (Fig. 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 S2 in the supplementary materials.

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 (Fig. 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.

Considering only the reported effects that are significant at a 5% level, the predicted probabilities of a positive effect are 0.65, 0.48, 0.43 and 0.20 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 are provided in Table 2 for 5% significance level for the reported effects and in Table S3 in the supplementary materials for 1% and 10% levels of significance of the reported effects. The reported directions of the effects of PES schemes on the provision of ES do not differ across publication types. The estimated coefficient of the indicator variable for peer-reviewed studies is not significantly different from zero at a 5% level in all the three settings. 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.[7]

Among the variables retained in the model specification (Table 2), only monitoring (H8a and H8b) conforms to our predictions, whilst payment mode (H4) and reference design (H5) have an opposite impact on PES effectiveness than expected. All other variables used to test the other hypotheses are non significant. Monitoring type and frequency are found to be positively correlated with PES effectiveness (H8a-b). The results show that this probability increases up to 50% when the monitoring of ES provision is undertaken by a third-party (as opposed to the ES provider itself), and by 26% when scheme 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 appears to be the main characteristic that influences the effectiveness 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. This result is also supported by the effect of payment constraint. The PES are found to be less effective when no payment constraint is imposed.

Conversely to our expectation, we find that individual reference is negatively correlated to the probability of PES schemes to increase ES provision (H5). Indeed, the results show that payments based on individual reference may decrease the probability of a positive effect by about 18% 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 (e.g., 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: customized references based on historical provision were grouped with customized 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.

Also, results indicate that output-based payments are more effective than input-based payments, contrary to what we expected (H4). Indeed, result-based PSE increase the probability of achieving a positive outcome by 11%. This result is in line with the current policy trend of fostering result-based schemes, when possible (Wuepper and Huber, 2022), even if the result is imperfectly measured.

However, the PES schemes that target ES providers, either within predefined intervention areas or based on individual characteristics, do not seem to be more likely to increase ES provision in the zones of implementation (H3).

Furthermore, the type (free market vs regulated demand, H1) and specificity of the PES objective (H2) do not seem to play a role on their probability of achieving positive outcomes, neither does the type of payments (cash vs in-kind) (H6) nor 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 levels of correlation – except for PES type and payment source that show high correlation – the chi-square tests are statistically significant (see Table S5 in the supplementary materials).

Contract length (H7) does not impact the probability to increase the effectiveness of the PES schemes. 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 by anticipation that, and multiple renewals of the short- and medium-term contracts may also contribute to increasing the probability of achieving positive outcomes.

[7] The results from the robustness check excluding unpublished articles are reported in Table S4 in the supplementary materials. The results remain similar and confirm the main conclusions.

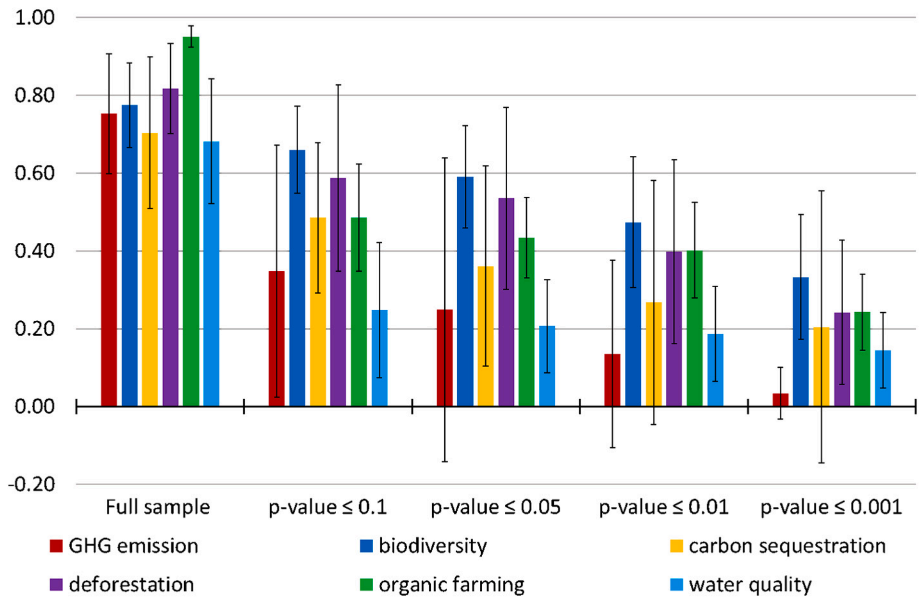


Fig. 3. Average probability of a positive effect of PES schemes on ES provision with a 95% confidence interval by topic from a logistic regression using the full sample and the subsample with a *p*-value statistics, using different *p*-value thresholds for the definition of an effective PES.

Table 2
Marginal effects of PES characteristics on the probability that a PES is effective.

Variables		Marginal effects	Std. Err.
PES characteristics (hypotheses)			
PES type (H1)	*free_market*	0.18	0.24
PES objective (H2)	*es_specific*	0.15	0.12
Eligibility (H3)	*spatial_targeted*	0.07	0.06
Payment mode (H4)	*output_based*	0.11*	0.06
Reference design (H5)	*individual*	−0.18***	0.06
Payment type (H6)	*cash*	0.11	0.13
ES provider	*individual*	0.24*	0.22
Contract length (H7)	*long_term*	−0.16	0.13
Monitoring type (H8a)	*third_party*	0.50***	0.17
Monitoring frequency (H8b)	*periodical*	0.26***	0.16
PES characteristics (controls)			
Payment time	*after_delivery*	−0.04	0.10
Payment source	*public*	0.02	0.10
Payment constraint	*none*	−0.19*	0.27
Study zone	*Africa*	0.19	0.18
	Asia	0.32	0.21
	Europe	0.18	0.18
	Latin_America	0.29	0.11
Study characteristics			
Study topic	*GHG_emission*	0.08	0.22
	biodiversity	0.22***	0.07
	carbon_sequestration	0.02	0.20
	deforestation	−0.16	0.18
	organic_farming	0.33***	0.12
	other_topic	0.32***	0.11
Data type	*aggregate*	−0.24***	0.07
Estimation method	*other_causal_methods*	−0.56***	0.21
	quasi_experiments	0.07	0.12
Publication type	*peer_reviewed*	0.11	0.20
Observations		174	

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

3.2. PES additionality

As presented in Section 3.2 for PES additionality, we estimate the FAT-PET-MRA using a tobit specification to analyse PES additionality in the subsample of studies ($N = 41$) 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 four categories. We focus only on biodiversity, deforestation and water quality, as they account for >90% of the sample, and we group the other four 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

Table 3
Meta-regression of PES-schemes additionality level on PES characteristics.

Variables		Coef.	St. Err.
	intercep	0.171	0.443
	standard_error	0.357	0.370
PES type (H1)	*free market*	0.789***	0.236
PES objective (H2)	*es_specific*	−0.016	0.068
Eligibility (H3)	*spatial_targeted*	0.091*	0.060
Payment mode (H4)	*output-based*	−0.004	0.047
Reference design (H5)	*individual*	−0.104**	0.040
Payment type (H6)	*cash*	0.123	0.105
ES provider (H6)	*individual*	−0.093*	0.052
Contract length (H7)	*long_term*	−0.325***	0.100
Monitoring type (H8a)	*third-party*	0.831***	0.252
Monitoring frequency (H8b)	*periodical*	−0.722***	0.091
Study characteristics			
Study topic	*biodiversity*	−0.119	0.122
	deforestation	0.149	0.088
	water_quality	−0.488***	0.110
Data type	*aggregate*	0.273*	0.129
Estimation methods	*quasi_experiments*	0.055	0.060
	other_causal_methods	0.071	0.078
	var(e.additionality)	0.012***	0.004
Observations		41	
Pseudo/adjusted R-sq		0.795	

Note: var.(e.additionality) is the estimated variance of the regression; standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

model specification is not significant at a 10% level. Both the RE and the WLS specification confirm this result (see Table 3). This suggests that the asymmetry observed in the Funnel plot is due to the high between-study heterogeneity (see Fig. S3 in the supplementary materials). Furthermore, the forest plot (see Fig. S4 in the supplementary materials) shows that the parameter measuring the between-studies heterogeneity (I^2) 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.

Fig. 4 shows that PES schemes achieve different levels of additionality according to the type of ES targeted. 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 at a 10% level.

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

Given our small sample of studies ($N = 41$), 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 results are reported in Table 3. Among the variables retained in the model specification, the type of PES schemes (H1), the reference design (H5), the contract length (H7) and the monitoring type (H8a) and frequency (H8b) play a role on the level of additionality. These results are confirmed by the RE and WLS specifications (see Table S6 in the supplementary materials), except for (H1) with the WLS specification. Furthermore, payment mode does not have an effect on the additionality level, confirming H4, and spatial targeting increases additionality (H3). However, these results are not confirmed by the other specifications (see Table S6 in the supplementary materials), consequently they should be considered with caution. However, it should be noted that the WLS specifications have high levels of VIF compared to the tobit specification and the RE because of the difference in the type of weighted factors used (see Table S7 in the supplementary materials). We therefore have less confidence in the results of these models than in the tobit model. As a robustness check and to deal with the small sample size, we also estimated the tobit model using different PES characteristics as controls. The results tend to confirm our main findings (see Table S9 in the supplementary materials) even if an increase of the degree of multicollinearity (see Table S10 in the supplementary materials).

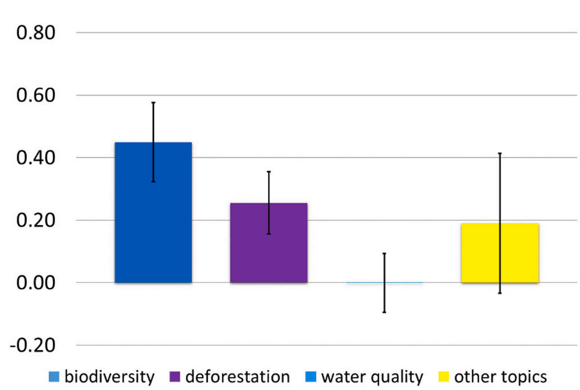


Fig 4. Average additionality level of PES schemes with a 95% confidence interval by topic from a tobit regression.

As expected, PES type is positively correlated to the level of additionality: free market PES are about 79% more additional than regulated demand ones (H1). This points towards the role of voluntariness in achieving additionality. However, the free market vs regulated demand distinction may also capture the effects of other characteristics of the PES schemes, such as funding source and study zone as these variables show a high level of correlation (see Table S8 in the supplementary materials). Indeed, free market PES are usually applied on a smaller scale, targeting less providers, with more local knowledge, which may also explain this higher additionality than regulated demand schemes.

As for effectiveness, individual reference has a negative impact on additionality of PES (H5): the results show that payments based on individual reference may decrease the probability of a positive effect by about 11% compared to PES schemes that use a generic reference. As for effectiveness, strategic behaviour by project developers or our indicator grouping different types of customized references may explain this counter-intuitive result.

Contrary to our expectations (H7), contract length is negatively correlated with the additionality of the PES schemes. Indeed, our results show that long-term contracts reduce the additionality of PES by 33% as compared to shorter-term ones. One possible explanation for such results is that contract length may also capture other characteristics of the PES, since in the restricted sample contract length is highly correlated with PES objective and study area (see Table S8 in the supplementary materials). In this respect, long-term contract PES in our sample are mostly located in developing countries and concern mainly deforestation and water quality, in contexts with poorer farmers that may be more prone to engage in long-term contract to ensure financial stability rather than to provide environmentally friendly practices.

We also find that PES schemes are more additional when monitoring is undertaken by a third-party, by 83%, in line with the results on effectiveness (H8a). However, we find that they are less additional when the implementation is monitored periodically, a reverse result compared to what we find for ES provision (H8b). Let us note that since PES effectiveness depends on both enrolment and additionality level, a different sign on effectiveness and additionality is possible if the sign of their impact is opposite. 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 (Section 4.3). However, there is no sufficient heterogeneity in monitoring frequency in the restricted sample to test for its impact.

Finally, the results seem to confirm our expectations that output-based payments do not increase additionality (H4) and that spatial targeting does (H3). However, those results should be considered with caution since they are not robust to specification (see Table S4 in the supplementary materials). The restricted sample additional effectiveness analysis (Section 4.3) allows us to further discuss the effects of those variables.

3.3. *Enrolment/efficiency trade-offs*

In an attempt to distinguish between enrolment and additionality drivers on the same sample, the meta-regression specification used for the assessment of effectiveness on the entire sample in Section 4.1 is reiterated on the studies used in Section 4.2, that is where an estimate of additionality is available. Effectiveness results obtained for the restricted sample are presented in Table 4.

The only design features that have a clear impact on PES effectiveness in this smaller sample are PES and monitoring types. This confirms the importance of third-party monitoring and free-market type PES to ensure both effectiveness and additionality. PES type is likely also influential in the full sample, but it may be superseded there by funding source, a variable which is highly correlated with PES type (see Table S8 in the supplementary materials). Results tend to confirm H4 in that

Table 4
Effects of PES characteristics on PES effectiveness (restricted sample).

Variables		Marginal effects	Std. Err.
PES characteristics (hypotheses)			
PES type (H1)	*free market*	2.642***	0.441
PES objective (H2)	*es_specific*	0.153	0.217
Eligibility (H3)	*spatial_targeted*	0.165	0.192
Payment mode (H4)	*output-based*	0.069	0.247
Reference design (H5)	*individual*	0.072	0.180
Payment type (H6)	*cash*	–	–
ES provider (H6)	*individual*	–	–
Contract length (H7)	*long_term*	−0.094	0.188
Monitoring type (H8a)	*third-party*	2.651***	0.470
Monitoring frequency (H8b)	*periodical*	–	–
Study characteristics			
Study topic	*biodiversity*	−0.076	0.289
	deforestation	0.492***	0.179
	other_topic	0.275	0.257
	water_quality	–	–
Data type	*aggregate*	−0.547***	0.184
Estimation methods	*quasi_experiments*	−0.424**	0.196
	other_causal_methods	−0.098	0.218
Observations		36	

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

output-based payments do not lead to increased effectiveness compared to input-based ones.

In principle, putting these restricted results on effectiveness in perspective with the results on additionality (Table 3) is a more interesting purpose. 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.

Contract length is clearly associated with a lower additionality but not with effectiveness, indicating a potential trade-off between enrolment and additionality effects. Such results would suggest that a longer-term contract induces less additional practices, a negative effect that could be counteracted by an increased level of enrolment, leading to a non significant aggregate impact on effectiveness. Such results do not support the intuitions developed in H7 that shorter-term contracts are preferred by farmers nor that longer-term contracts are more conducive to additionality. This confirms our intuitions regarding the negative impact of higher contract length on additionality, due to the high correlation between contract length and PES objective and study area.

The lack of impact of payment mode on effectiveness in the restricted sample tends to confirm our intuition stated in H4 about output-based schemes. Neither the additionality level nor enrolment are affected by the possibility to be remunerated based on output produced rather than input used in the restricted sample. This may be due to the few instances of "pure" output-based schemes or the fact that other design features seem to dominate the expected effect of result-based payments. Indeed, payment mode is correlated to other PES features such as PES type, eligibility or reference design (see Table S8 in the supplementary materials).

4. Conclusion

While PES are major instruments available to governmental or private organizations to support the conservation of ecosystems that provide environmental services, their implementation characteristics are 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, this meta-analysis 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 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 interpreted as more robust than those on additionality, given the smaller available sample with additionality measurements and the associated econometric limitations.

Type of ES providers, payment constraint, reference design, monitoring system and the environmental issue at stake 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 contract length seem to be the main factors that influence their additionality, along with reference design and monitoring. 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 do not imply a stronger additionality than practice-based payments (although they influence the probability of a PES to have a significant effect). 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 hydrid schemes, with baseline management requirements, or in which the result-based payment is optional above a baseline practice-based payment. Then, other PES 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 Common Agricultural Policy (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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is part of the LIFT ('Low-Input Farming and Territories – Integrating knowledge for improving ecosystem-based farming') project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 770747.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolecon.2023.107856.

References

Alston, L.J., Andersson, K., Smith, S.M., 2013. Payment for environmental services: hypotheses and evidence. Ann. Rev. Resour. Econ. 5, 139–159.

Antle, J., Capalbo, S., Mooney, S., Elliott, E., Paustian, K., 2003. Spatial heterogeneity, contract design, and the efficiency of carbon sequestration policies for agriculture. J. Environ. Econ. Manag. 46, 231–250.

Behaghel, L., Macours, K., Subervie, J., 2019. How can randomised controlled trials help improve the design of the common agricultural policy? Eur. Rev. Agric. Econ. 46, 473–493.

Bellassen, V., Shishlov, I., 2017. Pricing monitoring uncertainty in climate policy. Environ. Resour. Econ. 68, 949–974.

Bellassen, V., Stephan, N., Afriat, M., Alberola, E., Barker, A., Chang, J.-P., others, 2015. Monitoring, reporting and verifying emissions in the climate economy. Nat. Clim. Chang. 5, 319–328.

Bento, A.M., Kanbur, R., Leard, B., 2015. Designing efficient markets for carbon offsets with distributional constraints. J. Environ. Econ. Manag. 70, 51–71.

Berkey, C.S., Hoaglin, D.C., Mosteller, F., Colditz, G.A., 1995. A random-effects regression model for meta-analysis. Stat. Med. 14, 395–411.

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 Dev. 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. Environ. Conserv. 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.

Canton, J., De Cara, S., Jayet, P.-A., 2009. Agri-environmental schemes: adverse selection, information structure and delegation. Ecol. Econ. 68, 2114–2121.

Christensen, T., Pedersen, A.B., Nielsen, H.O., Mørkbak, M.R., Hasler, B., Denver, S., 2011. Determinants of farmers' willingness to participate in subsidy schemes for pesticide-free buffer zones—a choice experiment study. Ecol. Econ. 70 (8), 1558–1564.

Commission, E.U., 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.

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.

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.

Derissen, S., Latacz-Lohmann, U., 2013. What are PES? A review of definitions and an extension. Ecosyst. Serv. 6, 12–15.

Doucouliagos, H., Stanley, T.D., 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. Br. J. Ind. Relat. 47, 406–428.

Engel, S., 2016. The devil in the detail: a practical guide on designing payments for environmental services. Int. Rev. Environ. Resour. Econ. 9, 131–177.

Engel, S., Pagiola, S., Wunder, S., 2008. Designing payments for environmental services in theory and practice: an overview of the issues. Ecol. Econ. 65, 663–674.

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, e0149847.

Ferraro, P.J., 2008. Asymmetric information and contract design for payments for environmental services. Ecol. Econ. 65, 810–821.

Hardy, R.J., Thompson, S.G., 1998. Detecting and describing heterogeneity in meta-analysis. Stat. Med. 17, 841–856.

Herzon, I., Birge, T., Allen, B., Povellato, A., Vanni, F., Hart, K., others, 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. Environ. Conserv. 36, 120–128.

Kaiser, J., Haase, D., Krueger, T., 2021. Payments for ecosystem services: a review of definitions, the role of spatial scales, and critique. Ecol. Soc. 26.

Kleijn, D., Sutherland, W.J., 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? J. Appl. Ecol. 40, 947–969.

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? J. Clin. Epidemiol. 56, 44–51.

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. J. Environ. Manag. 157, 146–159.

Moxey, A., White, B., Ozanne, A., 1999. Efficient contract design for agri-environment policy. J. Agric. Econ. 50, 187–202.

Muñoz-Piña, C., Guevara, A., Torres, J.M., Braña, J., 2008. Paying for the hydrological services of Mexico's forests: analysis, negotiations and results. Ecol. Econ. 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. Ecol. Econ. 69, 1202–1208.

Oehlert, G.W., 1992. A note on the delta method. Am. Stat. 46, 27–29.

Pattanayak, S.K., Wunder, S., Ferraro, P.J., 2010. Show me the money: do payments supply environmental services in developing countries? Rev. Environ. Econ. Policy 4, 254–274.

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

Ruto, E., Garrod, G., 2009. Investigating farmers' preferences for the design of agri-environment schemes: a choice experiment approach. J. Environ. Plan. Manag. 52 (5), 631–647.

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

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. Conserv. Biol. 28, 1151–1159.

Stanley, T.D., Doucouliagos, H., 2017. Neither fixed nor random: weighted least squares meta-regression. Res. Synth. Methods 8, 19–42.

Stanley, T.D., Doucouliagos, H., Ioannidis, J.P., 2017. Finding the power to reduce publication bias. Stat. Med. 36, 1580–1598.

Šumrada, T., Japelj, A., Verbič, M., Erjavec, E., 2022. Farmers' preferences for result-based schemes for grassland conservation in Slovenia. J. Nat. Conserv. 66, 126143.

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. Agron. J. 98, 603–615.

Wuepper, D., Huber, R., 2022. Comparing effectiveness and return on investment of action-and results-based agri-environmental payments in S witzerland. Am. J. Agric. Econ. 104 (5), 1585–1604.

Wunder, S., 2015. Revisiting the concept of payments for environmental services. Ecol. Econ. 117, 234–243.

Wunder, S., Engel, S., Pagiola, S., 2008. Taking stock: a comparative analysis of payments for environmental services programs in developed and developing countries. Ecol. Econ. 65 (4), 834–852.

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. Nat. Sustain. 1, 145–150.

Wünscher, T., Engel, S., Wunder, S., 2008. Spatial targeting of payments for environmental services: a tool for boosting conservation benefits. Ecol. Econ. 65, 822–833.