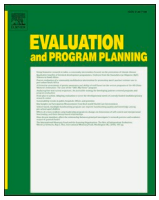
Evaluation and Program Planning 116 (2026) 102751

Contents lists available at ScienceDirect

Evaluation and Program Planning

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



Evaluating policy impact under sparse and staggered adoption. A synthetic difference-in-differences application to EU rural development measures

Roberto Esposti

Department of Economics and Social Sciences, Università Politecnica delle Marche, Ancona, Italy

ARTICLE INFO

JEL Classifications:
C21
C23
O18
Q12
Q18

Keywords:
Causal Inference in Policy Evaluation
Synthetic Difference-in-Differences
Staggered Treatments
Rural Development Policy
Farmers' Decision-Making

ABSTRACT

This paper investigates how farmers respond to targeted policy measures under the EU Common Agricultural Policy (CAP), focusing on whether voluntary adoption, often driven by private motivations, also leads to outcomes of societal interest. To address the methodological challenges posed by staggered and sparse treatment adoption, the study employs a Synthetic Difference-in-Differences (SDID) approach. A theoretical framework is developed to distinguish between private and societal outcomes of policy adoption. The empirical analysis uses a balanced panel of Italian farms from 2014 to 2022 and focuses on selected second-pillar CAP measures. Results reveal that while some measures significantly affect private outcomes (e.g., farm income or productivity), their impact on societal outcomes (e.g., environmental indicators) is weaker and more volatile. The paper discusses key challenges in identifying and estimating effects with few, heterogeneous treated units and staggered policy uptake. While SDID is well-suited to such contexts, its real-world application may face practical limitations.

1. Introduction

This paper explores the application of Causal Inference (CI) logic to policy assessment. Conceptually, CI aligns well with the core questions of policy evaluation, interpreting the impact of a policy as a treatment effect (TE) and framing policy implementation as a natural or quasi-experiment (Cerulli, 2015). Several recent contributions, such as Athey and Imbens (2017), Fougère and Jacquemet (2021), and Schnake-Mahl et al. (2025), among others, have emphasized that observational data and complex policy environments constrain the applicability of standard CI methods. These constraints make such methods, in practice, highly demanding and often unfeasible. Specifically, regarding the policy measures examined in this study, Esposti (2022) highlights significant methodological challenges in assessing the impact of EU Common Agricultural Policy (CAP) support on the farming sector, particularly in identifying credible counterfactuals and isolating treatment effects.

The challenge becomes even more severe for policies where implementation limits any control over treated units (i.e., those subject to the policy) and, consequently, over potential controls (i.e., those not subject to it). This situation often arises when adoption is voluntary and involves a varying number of heterogeneous agents. In such cases, conducting an evaluation within a TE framework becomes empirically complex. These difficulties often manifest as violations of the identifying assumptions required by CI methods. Beneath these identification issues lie deeper theoretical and practical challenges, namely, how actual policies can be meaningfully interpreted as quasi-experimental designs. Properly aligning a policy with such a framework requires a conceptual model that traces the causal pathway from the policy intervention to the outcome.

This conceptual framework entails significant methodological implications. First, the decision-making process of potential adopters is typically driven by private benefits rather than the societal goals for which the policy was designed. Consequently, policy evaluations should consider two distinct outcomes, and corresponding treatment effect (TE) estimates, to capture both individual and collective dimensions of impact. Second, since adoption under these policies is voluntary, entry into treatment is staggered over time, often beginning with a small number of early adopters. Third, adopters, particularly early ones, tend to exhibit distinctive characteristics, especially in the farming sector, which is marked by substantial heterogeneity across units.

This paper focuses on the methodological challenges posed by such policy designs. Many of these challenges are well recognized in the CI literature and have been addressed in various empirical studies. In particular, a recently proposed method, Synthetic Difference-in-Differences (SDID), has been developed to tackle these issues within a

E-mail address: r.esposti@staff.univpm.it.

https://doi.org/10.1016/j.evalprogplan.2026.102751
Received 30 January 2025; Received in revised form 28 November 2025; Accepted 4 January 2026
Available online 6 January 2026

unified framework (Arkhangelsky et al., 2021). The present study adopts the SDID approach to demonstrate both its strengths and its potential limitations when applied to real-world policy contexts characterized by sparse and staggered adoption.

As an illustrative case, this study focuses on selected measures from the second pillar of the EU CAP, known as the Rural Development Policy (RDP), implemented during the 2014–2020 programming period and extended to 2022 (Esposti, 2022). The aim is to evaluate their impact on both farm-level performance and broader societal outcomes. To this end, policy measures are systematically reclassified according to the societal objectives they pursue. The empirical analysis draws on a balanced panel of Italian farms from the Farm Accountancy Data Network (FADN), covering 2014–2022. Italy offers a compelling case due to its pronounced heterogeneity in farming conditions, which shapes both policy uptake and farmers' responses (Coderoni & Esposti, 2018; Baldoni et al., 2021).

Building on this dataset and adopting SDID estimation, the paper contributes to the literature on the evaluation of regional development and environmental policies, with a focus on the EU RDP. These policies address "grand societal challenges" such as sustainability, climate change mitigation, and social inclusion, placing them at the intersection of ambitious goals and complex implementation dynamics (OECD, 2019; European Commission, 2021). Yet, evaluation studies often overlook how heterogeneous behavioral responses can undermine policy effectiveness and must be explicitly accounted for (UCL Institute for Innovation and Public Purpose, 2020). This study seeks to fill that gap.

The remainder of the paper is structured as follows. Section 2 reviews the main challenges and relevant literature concerning the evaluation of policies characterized by granular implementation and sparse adoption, such as the EU RDP measures examined in this study. Section 3 outlines the theoretical framework underlying farmers' behaviour in response to these policy measures and discusses the resulting societal value of the policy within a treatment effect logic. Section 4 presents the sample and dataset used, and describes the research design, with particular attention to the reclassification of EU RDP measures and the resulting definition of treatment and outcome variables. Section 5 details the SDID estimation approach, highlighting both its potential, reflected in the range of results it can produce, and its limitations, which stem from the complex set of identifying assumptions it requires. Section 6 reports and discusses the results, while Section 7 highlights key lessons, particularly regarding policy design, and concludes by suggesting potential directions for future research in the field.

2. Challenges in program evaluation

2.1. Framing within the policy evaluation literature

A key characteristic of the regional development and environmental policies (and of RDP measures in particular) is the tension between their high-level objectives and the decentralized, voluntary nature of their implementation. In such case, policy uptake is typically driven by a multitude of small, heterogeneous agents whose decisions are shaped by diverse private incentives and constraints. This creates a fundamental disconnection between the societal goals of the policy and the individual motivations of adopters, which can lead to misaligned expectations in both ex ante and ex post evaluations (Falconer et al., 2001; Espinosa-Goded et al., 2010).

Agents respond to policy incentives based on their individual interests which are shaped by idiosyncratic factors, including timing. The resulting private outcome often diverge from the intended societal outcomes emerging from the aggregate behaviour of these agents and that have to be evaluated in terms of collective welfare in response (Coderoni et al., 2025). The distinction between the private and societal value of a policy is often presented as an alternative perspective in evaluation theory and practice (UCL Institute for Innovation and Public Purpose, 2020). In reality, the two dimensions coexist and must be considered jointly.

This paper contributes to the evaluation literature in two key directions. First, it proposes a conceptual framework that explicitly links the societal objectives of policies like the EU RDP with the voluntary adoption behaviour of individual agents. This framework highlights the importance of accounting for unit-level and temporal heterogeneity, as well as the private rationales behind participation decisions. Second, building on this foundation, the paper introduces a methodological approach that distinguishes between private effects (on adopters) and societal effects (on policy goals). To estimate these, it employs a design-based strategy tailored to the specific features of such policies: numerous small units, staggered and sparse adoption, and significant heterogeneity across time and space.

This study can be positioned within the conceptual framework of the *evaluation theory tree* proposed by Lemire et al. (2020), which builds on earlier work by Alkin (2004) and Christie and Alkin (2008). This framework organizes the diverse traditions of evaluation theory into three main branches: (i) the *methods* branch, which focuses on the methodological tools for causal inference and impact assessment; (ii) the *use* branch, which emphasizes the utilization of evaluation findings by decision-makers; and (iii) the *valuing* branch, which centers on the normative and ethical dimensions of evaluation.

Within this structure, the present contribution primarily contributes to the methods branch by advancing a theoretically grounded quasi-experimental design tailored to observational data in complex policy settings. However, it also engages with the valuing branch by examining how the divergence between private motivations and public goals influences the interpretation of outcomes. Finally, it contributes to the use branch by offering practical insights for policymakers aiming to improve the design of such interventions.

The EU RDP serves as an ideal empirical setting. It exemplifies the challenges discussed while offering a diverse set of measures under a unified framework. In Lemire et al.'s terms, it is "complicated" but not "complex": stable and well-defined, yet implemented through varied instruments targeting heterogeneous agents. This makes it a fertile ground for testing the proposed framework and methodology.

2.2. The main challenges from a Causal Inference perspective

As Lemire et al. (2020) highlight, within the "evaluation theory tree," the "methods" branch has seen the most significant expansion in recent decades. In particular, CI approaches have gained traction in policy evaluation. However, this rapid methodological development, often originating from disciplines distant from applied policy analysis, has sometimes led to the uncritical application of CI tools in policy contexts, obscuring key conceptual and practical implications of conceptualizing policies as "treatments".

A central concern lies in the potential mismatch between the assumptions underpinning CI methods and the complex realities of policy implementation. Robust quasi-experimental designs typically require clearly defined treatments, well-measured outcomes, and a clear mapping between them, including distinct treated and control units. These conditions are more easily met when policies are highly granular, that is, tailored to specific beneficiary characteristics. Conversely, CI methods rely on identification assumptions (see also Section 5.2) such as the overlap condition, and this latter is more easily satisfied when the pool of potential adopters is broadly defined.

This creates a fundamental tension. One the one hand, granular policies enhance construct validity and treatment relevance, but may undermine identification due to limited overlap. On the other hand, less granular policies may better satisfy the overlap condition, but complicate the definition of expected outcomes and the treatment-effect linkage due to the potential ambiguity about the outcome variables on which policy evaluation has to be grounded. This trade-off poses a major challenge for applying CI tools to policy evaluation and demands careful attention in both research design and interpretation. The present paper

focuses on those policies whose degree of granularity may interfere with the actual applicability of the CI logic to their evaluation.

From a methodological standpoint, the compatibility between granular policy interventions and CI logic depends on the extent to which policymakers and analysts can control or observe the assignment of treated and control units. This raises two interrelated challenges: the sparse and heterogeneous nature of adoption, and the voluntary, staggered timing of treatment entry.

First, adopters may be few and highly diverse, a condition we refer to as sparse adoption, driven by latent characteristics such as capacity, motivation, and beliefs (Esposti, 2024), which are difficult to observe and control for within a CI framework. Second, voluntary participation complicates identification by blurring the line between factors influencing adoption and those driving impact. Moreover, voluntariness often entails variability in the timing of adoption, leading to staggered treatment (Sun & Abraham, 2021). This staggered entry may be further affected by administrative or budgetary constraints that delay or prevent adoption among willing units (Baldoni & Esposti, 2023). As a result, the number of treated units increases progressively over time, and the pool of untreated units becomes time-varying, posing additional challenges for constructing valid control groups.

The standard identification strategy in panel data, Difference-in-Differences (DID), may fail In settings with sparse adoption due to violations of the Parallel Trends Assumption (PTA), particularly when cohort-specific trends are not adequately addressed (see also Section 5.2). Recent methodological advances offer two main solutions. One involves using not-yet-treated units as controls (e.g., Goodman-Bacon, 2021; Sun & Abraham, 2021; Baldoni & Esposti, 2025), which can mitigate issues related to staggered entry, though its effectiveness diminishes when the number of adopters is limited. Alternatively, the Synthetic Control Method (SCM) (Abadie et al., 2010; 2015; Abadie, 2021) has been proposed for settings with few treated units, but it struggles with staggered treatment and cohort-specific effects. To overcome these limitations, Arkhangelsky et al. (2021) introduced a hybrid approach, Synthetic Difference-in-Differences (SDID) that can be extended also to cases of staggered adoption. As discussed by Clarke et al. (2024), SDID combines the strengths of SCM and DID by tailoring synthetic controls to each entry cohort, enabling more flexible counterfactual construction.

SDID estimation is particularly effective in addressing non-parallel pre-treatment trends, which frequently arise in micro-level policy evaluations due to heterogeneity among individual agents. By constructing a weighted combination of untreated units that closely matches the treated unit's pre-intervention trajectory, SDID yields more credible counterfactuals. It is especially well-suited to contexts like farming, where time-varying, unobserved factors, such as capacity, motivation, and external constraints, are prevalent. One of SDID's main advantages is its relaxation of the Parallel Trends Assumption: unlike traditional DID, it does not require treated and control units to follow parallel trends, making it particularly appropriate for farm-level data with diverse and non-parallel trajectories.

The theoretical advantages offered by the SDID approach under sparse adoption, however, must contend with its application in non-experimental settings, particularly in rural development policies, where significant practical challenges may arise. As detailed in Section 5.2, these challenges stem from the strong identification assumptions that still underpin SDID estimation. Their validity and reliability must be critically assessed to ensure credible inference.

2.3. The policy under analysis: EU RDP measures

The evaluation of the EU Rural Development Policy (RDP), and its specific measures, exemplifies many of the challenges discussed above. For this reason, the RDP is the focus of the present analysis.

Since its inception in 1962, the Common Agricultural Policy (CAP) has supported the agricultural sector and farmers' income. Over time, however, the CAP has undergone major reforms: market support has gradually been replaced by direct income support, while structural measures aimed at modernising agriculture have been progressively introduced and strengthened. With the Agenda 2000 reform (Sotte & Brunori, 2025), these structural measures were grouped under the second pillar of the CAP, known as the EU Rural Development Policy (RDP). The RDP aims to help farmers adapt to evolving societal needs and policy objectives, typically summarised as competitiveness, diversification, and the provision of public goods, mostly environmental ones (Swales, 2007; Camaioni et al., 2016; Esposti, 2022).

Despite successive reforms, the RDP continues to face criticism, also motivated by the lack of robust and credible evidence on its actual impacts (Esposti & Sotte, 2013; Mennig, 2024). A key source of criticism is that many RDP measures still primarily function as income support, with little or no behavioural implications for beneficiaries (Esposti, 2022). These can be referred to as *non-behavioural* or *conservative measures*, analogous to *passive measures* in Esposti (2022). Their evaluation is inherently problematic: since no behavioural change is expected upon adoption, no observable outcome can be directly attributed to the policy. In such cases, the CI logic becomes inapplicable.

Conversely, certain RDP measures are *behavioural* or *non-conservative* (*active*) in nature (Esposti, 2022). While diverse in form, they share a common logic: farmers voluntarily participate in measures that resemble a contract between individual producers and society, receiving financial support conditional on producing measurable outcomes of societal interest. These behavioural changes, such as shifts in production practices, can lead to observable variations in farm performance, which may be interpreted as the treatment effect (TE) of the policy intervention. In principle, this active nature makes such measures more amenable to evaluation within a CI framework. In practice, however, these evaluations remain challenging, and empirical assessments are still relatively rare and often limited in scope (European Court of Auditors, 2013; Andersson et al., 2017; Castaño et al., 2019).

The underlying reason for these challenges lies in issues discussed earlier. A major problem with non-conservative RDP measures is their lack of granularity. While they aim to support specific behavioural responses often linked to farm characteristics, they are implemented as general policies with minimal entry restrictions. As a result, most farmers are potential beneficiaries and can opt into the policy at any time, making adoption highly variable and complicating evaluation. The apparently contradictory fact that some RDP measures had very few adopters, especially in the early years (see Section 4), is not due to strong policy selectivity. Rather, it reflects a lack of convenience for farmers or the technical and bureaucratic complexity of adoption, often a combination of both. In any case, sparse adoption poses challenges for both identification and estimation, due to limited statistical power and reduced ability to control for confounding. From a policy interpretation perspective, it also raises concerns about the generalizability of results.

Ultimately, in the case of granular RDP measures, strict eligibility criteria may hinder the identification of suitable counterfactuals. At the same time, broader or less restrictive access conditions may facilitate overlap and thus support identification, by expanding the pool of potential adopters and increasing the likelihood of finding comparable control units. This would, in principle, make the application of CI logic more feasible. However, in practice, heterogeneity among adopters complicates the definition of both treatment conditions and, more importantly, outcome variables. Unit- and time-specific heterogeneity in farmers' responses poses a challenge regardless of the measure considered.

Given these characteristics, the SDID approach appears particularly well-suited for evaluating non-conservative (behavioural) RDP measures. Accordingly, this paper applies SDID to a selection of such measures over the 2014–2022 period, aiming to assess whether, and to what extent, this method can yield informative insights into policy effects. The group of RDP measures considered in this study, along with the definition of their respective treatment sets and outcome variables, is

detailed in Section 4.1. Differences across measures may also offer comparative perspectives on the strengths and limitations of SDID in non-experimental settings.

3. Theoretical framework: farmers' decision-making and policy indicators

Following the preceding discussion, any evaluation exercise grounded in CI must rely on a quasi-experimental design that is underpinned by a coherent theoretical representation of farmers' behaviour and its link to societal outcomes. In line with this, we develop a theoretical framework centred on farmers' decision-making processes and their resulting performance. This framework distinguishes between two types of outcomes: the private outcome, which directly influences the farmer's choices, and the societal outcome, which may be perceived as incidental by the farmer but is crucial for evaluating the policy's effectiveness and efficiency.

3.1. A theory of farms response

Consider a panel of N agents or, considering the present application, production units (farms) observed over Z time periods. For any i-th farm ($i=1,\ldots,N$) and t-th time period ($t=1,\ldots,Z$), an aggregated multi-input multi-output production technology can be represented by the feasible production set $\mathscr{T}_{it}\subset\mathbb{R}^M$. $\mathscr{T}_{it}$ is farm specific and contains all sources of heterogeneity in farmers' decision-making in terms of both treatment participation and production choices (Esposti, 2024). $\mathscr{T}_{it}$ is shaped by all the specific features of the i-th farms, depending on both external and internal factors, that we generally indicate with the $(Q\times 1)$ vector $\boldsymbol{X}_{it}$. Given $\mathscr{T}_{it}$, a $(M\times 1)$ vector of netputs $\boldsymbol{Y}_{it}=(Y_{it1},\ldots,Y_{itM})'$ is feasible if $\boldsymbol{Y}_{it}\in\mathscr{T}_{it}$.[1] This netput vector, $\boldsymbol{Y}_{it}$, contains both farm outputs (with a positive sign) and inputs used (negative).

Each farm is offered a binary policy $T=0,1$. Let T_{it} be the indicator of adoption of the policy measure: if the i-th farm adopts the measure at time t, then $T_{it}=1$; otherwise, $T_{it}=0$. Moreover, denote with S_{it} the amount of financial support granted under $T_{it}=1$. Therefore, the treatment is expected to induce specific production choices ($\boldsymbol{Y}_{it}$). Non-conservative policy measures can be defined as those targeting specific farming activities that affect farms' performance and, consequently, some societal objective. Consequently, for a given farm, a treatment can be univocally mapped into production choices ($T_{it}\leftrightarrow\boldsymbol{Y}_{it,k}$). It is then possible to express these choices as a function of the treatment itself, plus those abovementioned exogenous farm-specific characteristics $\boldsymbol{X}_{it}$ (or confounding variables) influencing farmers' behaviour: $\boldsymbol{Y}_{it}=f(T_{it},\boldsymbol{X}_{it})$ where $f(\cdot)$ is a vector-valued function.

Within this theoretical framework and a binary TE logic, if both $\boldsymbol{Y}_{it}$ and T_{it} are observed $\forall i\in N$ and $\forall t\in T$, two different TEs can be identified. The first TE concerns the outcome of interest for the farmers, that is, their farm profit (or net income; see Section 4), π_{it}. Ultimately, what matters from the farmers' perspective is whether policy measure adoption, with the associated policy support S_{it}, improves their profit. Let us generically express the farm-specific profit function as:

$$\pi_{it}=p[f(T_{it},\boldsymbol{X}_{it})] \tag{1}$$

where $p(\cdot)$ is a single-valued function. Consequently, the first TE associated with the generic policy will be π_{it}, that is:

$$\pi_{it}=[(\pi_{it}|T_{it}=1,\boldsymbol{X}_{it})-(\pi_{it}|T_{it}=0,\boldsymbol{X}_{it})]=p[f(T_{it},\boldsymbol{X}_{it})] \tag{2}$$

where T_0represents the baseline case, that is, the non-treatment condition and $f(\cdot)=\boldsymbol{Y}_{it}=[(\boldsymbol{Y}_{it}|T_{it}=1,\boldsymbol{X}_{it})-(\boldsymbol{Y}_{it}|T_{it}=0,\boldsymbol{X}_{it})]$. It is worth noticing that $\Delta\pi_{it}$ here does not include the policy support, S_{it}, thus it is also referred to as before-support farm net income.[2]

3.2. The definition of the societal value of the response

The impact of the policy measure from the farmers' perspective, referred to as the private TE, often diverges from the actual objective of the policy. This necessitates the consideration of a second TE, one that reflects the societal outcome of interest to policymakers consisting is some improved performance of the treated (i.e., supported) farms. While the private TE captures the individual benefit that motivates adoption, the societal TE represents the aggregate effect that determines the policy's effectiveness and efficiency from a collective welfare standpoint.

This improved performance from a societal perspective can be either a negative (for instance, lower use of polluting inputs)[3] or a positive (for instance, some biodiversity indicator)[4] variation. Assume that for any i-th farm the production choice expressed by $\boldsymbol{Y}_{it}$ univocally determines its societal performance $\boldsymbol{G}_{it}$, where $\boldsymbol{G}_{it}$ is a $(P\times 1)$ vector of P indicators of societal interest. Therefore, production choices can be univocally mapped into these indicators ($\boldsymbol{Y}_{it}\leftrightarrow\boldsymbol{G}_{it}$). If the interest is in a single performance, G_{it} will be a scalar term. Therefore, it will be possible to express the farm-level societal performance as a function of what determines the production choice $\boldsymbol{Y}_{it}$. It will be $G_{it}=g(T_{it},\boldsymbol{X}_{it})$ where $g(\cdot)$ is a single-valued function. The second TE will thus be G_{it}:

$$G_{it}=[(G_{it}|T_{it}=1,\boldsymbol{X}_{it})-(G_{it}|T_{it}=0,\boldsymbol{X}_{it})]=g(T_{it},\boldsymbol{X}_{it}) \tag{3}$$

Given the definition of π_{itit} as a before-support net income, we can argue that, in general terms, the farmers' voluntary choice of the treatment ($T_{it}=1$) implies $(\pi_{it}+S_{it})\geq 0$. However, since we admit here that both π_{it} and G_{it} are heterogenous, namely farm-specific, we cannot exclude that non-monetary motivations of farmers' policy adoption might also justify a non-positive profit outcome of the treatment ($\pi_{it}+S_{it})\leq 0$ (thus, $\pi_{it}<0$) (Esposti, 2024). Consequently, the actual social cost of the policy measure is $(S_{it}-\Delta\pi_{it})$.

On the basis of the two TEs identified above, it is thus possible to express the efficiency of the policy under investigation. This consists in computing an indicator expressing how much it costs to the society as a whole to obtain a unit-variation of the farm-level societal performance G_{it}. Here, we generally call this indicator the *Social Cost of Farm Performance* (SCFP) and it is calculated as:

$$SCFP_{it}=\frac{(S_{it}-\Delta\pi_{it})}{\Delta G_{it}} \tag{4}$$

The inverse of this *SCFP* can be thus interpreted as an indicator of policy efficiency, that is, how much societal outcome is obtained from a unit policy expenditure.

As the outcome variables in Eq. (4), $\Delta\pi_{it}$ and ΔG_{it}, could also be expressed in relative (i.e., size-independent) terms (see Section 4 for details), (4) becomes:

$$SCFP_{it}=\frac{\left(\frac{S_{it}}{R_{it}}-\frac{\Delta\pi_{it}}{R_{it}}\right)}{\frac{\Delta G_{it}}{R_{it}}} \tag{5}$$

where R_{it} indicates the farm size possibly expressed either as physical size (e.g., Utilised Agricultural Area or unit of family labour) or economic size (e.g., Gross Production Value or Value Added). If different indicators of the farm size are used to relativize the two outcome variables, say R_{it} and V_{it} respectively, (5) becomes:

[1] Therefore, $\mathscr{T}_{it}$ is not a single M-dimensional vector but a subset of $\mathbb{R}^M$ that represents the collection of all possible netput vectors for farm i at time t.

[2] As clarified in Section 4, this is the definition of the outcome variable adopted in the present empirical analysis and it is also the variable provided within the FADN dataset.

[3] This is the case of LSU/UAA for measures P5 (see Table 1).

[4] This is the case of FA/UAA for measures P6 (see Table 1).

$$SCFP_{it} = \frac{\left(\frac{S_{it}}{R_{it}} - \frac{\Delta\pi_{it}}{Z_{it}}\right)}{\frac{\Delta G_{it}}{R_{it}}\frac{V_{it}}{Z_{it}}} \quad (6)$$

As detailed in Sections 4 and 5, (5) and (6) also offer the additional benefit of mitigating the dispersion of outcome variables resulting from substantial farm size heterogeneity. This may significantly improve the statistical robustness of the estimation of the TE. Expressing outcome variables in relative terms, however, does not exclude the possible impact of the farm size on the TE calculation due to scale economies. Consequently, variables expressing the size can still be part of the $\boldsymbol{X}_{it}$ set.

There is a further motivation to express the variables under consideration in relative terms, thus making them relatively size-independent. Eq. (4) suggests that treatment intensity S_{it} may differ across farms implying a multi-valued treatment. However, for all policies under consideration, the monetary support S_{it}, though farm-specific, actually depends on farm size and is largely proportional to it. Therefore, as noted in previous studies (Coderoni et al., 2024), when variables are expressed in relative terms, this proportionality reduces the variability of treatment intensity across farms, making the binary specification a reasonable approximation.

Nevertheless, this dichotomization is not neutral: it may obscure intensity-related effects and introduce interpretability challenges (Callaway et al., 2024). Within the present conceptual framework, subsidy intensity could be modeled as a continuous treatment. However, doing so would complicate the design of an appropriate quasi-experimental approach and undermine identification strategies. While continuous treatment frameworks exist (Wooldridge, 2023; Chernozhukov et al., 2024), their implementation under sparse and staggered adoption remains highly demanding. This limitation calls for caution in interpreting results and suggests that future research should explore continuous or multi-valued treatment frameworks for the kind of policy measures considered here.

4. Data and research design

4.1. RDP measures under study: reclassification and aggregation

The 2014–2022 EU RDP, which constitutes the second pillar of the CAP, is organized around six priorities and eighteen Focus Areas. These are implemented through twenty measures and sixty-six sub-measures. One of these (Measure 18), specifically supported Croatia's accession in 2014,therefore it is excluded from the present analysis. Consequently, the measures relevant to this study amount to nineteen, comprising sixty-five sub-measures. A detailed overview of this policy architecture is provided in Annex 1 (Table A1).

Treating each measure and sub-measure as a distinct "treatment" results in a very limited number of treated units, with implications for identification and inference. Furthermore, from a policy perspective, results tied to narrowly defined measures may lack interpretability and broader relevance. In practice, some measures differ primarily in their target populations, often heterogeneous farm types, while sharing similar mechanisms of action. Evaluating them jointly allows for more generalizable insights into policy effectiveness.

For these reasons, a reorganization of the policy instruments is warranted. As outlined in Section 2, the reclassification and aggregation process is driven by a focus on behavioural (i.e., non-conservative or active) measures. Therefore, of the sixty-five sub-measures, non-behavioural ones are first excluded, and the remaining ones are then reclassified and grouped. Previous studies have proposed similar aggregations, often based on shared objectives reflected in the overarching priorities (Sotte, 2009; Camaioni & Sotte, 2010; Sotte, 2012). In the present study, the reclassification is grounded in the theoretical framework discussed in Section 3, and is based on the expected behavioural responses of farmers and the resulting societal outcomes, i.e., the performance indicators that justify public support.

Ultimately, the behavioural measures are grouped into six broader policy categories or "treatments" (P1–P6), each corresponding to a distinct expected societal impact: P1: Information, Education, Training; P2: Food Chain and Product Quality; P3: Structural Agricultural Investments; P4: Structural Non-Agricultural Investments; P5: Animal Production (or Livestock) Management; P6: Forest Management. Each treatment is intended as binary, that is, the treatment status assumes value 1 or 0 ($T_{it} = 1;0$),[5] depending on whether, at time t, the farm has adopted or not one of the measures included in respective policy aggregate (P1-P6). Table 1 details the allocation of the RDP non-conservative sub-measures across these six general policy categories (see Annex 1 for codification and detailed definitions of the sub-measures).

It is important to anticipate a potential drawback of this reclassification and aggregation. With the exception of P5, most policy categories implicitly encompass multiple measures which, in CI terminology, represent distinct variants of the treatment. On the one hand, grouping RDP measures based on their shared societal objectives, i.e., the actual policy intent, is consistent with the theoretical framework and addresses practical concerns, particularly the sparse adoption of individual measures. On the other hand, the SDID approach relies on the validity of the Stable Unit Treatment Value Assumption (SUTVA), which may be violated when multiple versions of a treatment are present.

Sections 5.2 and 6.4 further elaborate on this potential tension between the methodological requirements of SDID and the practical rationale for aggregating policy measures.

4.2. The observational dataset and the treatment sets

In this study we use information from the Italian FADN dataset. Italian agriculture is often considered an interesting case study for the wide heterogeneity of its farming conditions and traditions which inevitably affect farmers' decision-making and farm performance ((Esposti, 2017a,b; Coderoni & Esposti, 2018; Coderoni et al., 2024). The Italian FADN sample consists of a representative collection of commercial farms observed over the CAP regime under consideration (2014–2020 then extended to 2022). It comes as an unbalanced panel consisting of 10573 observations in 2014, 9569 observations in 2015, 10153 in 2016, 10792 in 2017, 10769 farms in 2018, 10805 observations in 2019, 10764 observations in 2020, 11040 observations in 2021 and 11084 observations in 2022. Though the programming period is 2014–2022, in the year 2014 the only treated units concern payments from the previous programming period. Therefore, the actual period of policy implementation of interest (and the consequent sample) is 2015–2022.[6]

Two additional processing steps are required on the observational dataset, both dictated by the SDID estimation approach (see Section 5). First, the method requires a balanced panel sample. Second, each unit must have at least two years of pre-treatment observations. Accordingly, two balanced panels are constructed for the remainder of the analysis. The first covers the period 2013–2022 and includes policy measures for which some units entered treatment as early as 2015 (i.e., policies P1, P3, P5, and P6 in Table 1). The second panel spans 2015–2022 and includes measures for which early adopters entered treatment only in 2017 (i.e., policies P2 and P4).

Panel 2013–2022 contains 2605 units observed over 10 years, thus 26050 observations overall; for panel 2015–2022 the units are 3899 corresponding to 31192 observations overall. It could obviously be possible to align the evaluation of all policy measures on period 2015–2022 with the respective balanced sample. At the cost of losing

[5] For a justification of this dichotomization, see Section 3.2.

[6] At the time of writing, validated FADN data for 2023 had yet to be released and, in any case, this year would refer, at least partially, to the new CAP regime (2023–2027) that started on 1st January 2023.

Table 1
List of EU RDP 2014-2022 measures included and excluded in the study with association with respective outcome variables.

Priority & Focus Area	Treatment	RDP Measures (active behavioural response) (Annex 1)	Theme/Reclassification	Period	Outcome variable		Reference indicators (Annex 2)
					Private	Societal	
Pr1 & 1A-1C	P1	1.1, 2.1, 2.2, 16.1	Information, Education, Training	2013–2022	NI/ FAWU	GVA/AWU, GVA/ UAA	I2, I3
Pr3 & 3 A	P2	3.1, 3.2, 16.2, 16.4	Food chain, Quality	2015–2022	NI/ FAWU	GVA/SO	R2, R4, T6, I1, I2
Pr2 & 2 A	P3	4.1–4.4, 5.1, 5.2	Structural agricultural investments	2013–2022	NI/ FAWU	Inv/AWU Inv/SO	C28, R1, T4
Pr2 & 5A-5C, 6 A	P4	6.4, 16.9	Structural non-agricultural investments	2015–2022	NI/ FAWU	OGA/SO, OGA/AWU, Other Inv/SO	C28, R1, R14, R15, T4
Pr3 & 5E	P5	14	Animal/livestock production management	2013–2022	NI/ FAWU	LSU/AWU LSU/UAA VE/SO	C21, C45, O8, R16, T17, I7
Pr4–5 & 4 A, 5E	P6	15.1–15.2, 8.1–8.6	Forests	2013–2022	NI/ FAWU	FA/UAA	C38, R6, R18, R9, R11, R20, T8, T11, T13, T19, I9

LEGEND: AWU = Annual Working Units; FA = Forest Area; FAWU= Family Annual Working Units. GVA = Gross Value Added; Inv = Investments; LSU = Livestock Units; NI=before-support Net farm Income; OGA = Other Gainful Activities; SO = Gross Production Value; UAA = Utilised Agricultural Area; VE = Veterinary and sanitary Expenditure.

more than one thousand observations,[7] going back to 2013 has the advantage to include a longer period under the treatment. As will be clarified in Sections 5 and 6, this brings about more robustness in the identification of the TE especially when TE itself shows wide variability across cohorts.

Details on the numerosity of the two balanced panels across the policy measures and the entry years are provided in Table 2. Fig. 1 displays the progressive entries into the treatment during the period of observation.[8] It can be argued that three types of policies emerge in terms of numerosity of treatment adoption and entry timing: very few units and a little staggered entry in P1 and P4 (measures 1.1, 2.1, 2.2; 6.4; 16.1, 16.9); few units but largely staggered entry mostly in P2 and P6 (measures 3.1–3.2; 8.1–6); many units and largely staggered treatment in P3 and P5 (measures 4.1–5.2; 14).

4.3. Definition of the outcome variables

The proper definition of the outcome and confounding (**X**) variables is driven by the theoretical framework illustrated in Section 3. It thus follows that two distinct outcome variables have to be simultaneously considered in any policy measure assessment: the private and the societal outcome. Table 1 also associates the private and societal outcome variables with the six treatment groups. The private outcome is always the Net Farm Income (NI), that is, variable π of Section 3, to be intended as net of RDP measure support. One practical problem in using net income as an outcome variable is that it is highly size-dependent. This dependency can hardly be controlled by including size variables in **X**, as it cannot be excluded that a size bias remains and, in any case, this would not eliminate the very large variability of this private outcome. Therefore, to make π size-independent, NI is here divided by the annual units of farm family work (FAWU). After all, per capita farm income is what the farmers really care about, and what drives their choices. Therefore, it seems the most appropriate definition of the private outcome variable within the present study.

The societal outcome variable (variable G of Section 3) must be selected according to what is explicitly established by the policy itself. On this aspect an imponent effort was continuatively and incrementally made by EU institutions (European Commission, 2012; European Commission, 2015a; European Commission, 2015b; European Commission, 2016; European Commission, 2018). These documents clarify that RDP measures are expected to chase very general purposes defined by the three fundamental objectives of the "Europe 2020" strategy (smart, sustainable, inclusive growth) and the six key priorities of the 2014–2022 RDP. Some measures can be univocally associated with a single priority while others refer to several of them, as in the case of the so-called horizontal measures (Measure 20, for instance). In turn, any priority entails several (from two to five) focus areas for a total number of 19 focus areas.

The evaluation approach proposed by the European Commission for the 2014–2020 Rural Development Programs includes a comprehensive set of Common Monitoring and Evaluation Framework (CMEF) indicators. The CMEF outlines the comprehensive approach for monitoring and evaluating the performance of the RDP. The CMEF provides key information on the implementation of the RDP through the use of indicators and sub-indicators (European Commission, 2014; 2015b; 2019). These indicators are categorized into three main types. Output Indicators (OI) measure the direct products of the program's activities.[9] Result Indicators (RI) assess the immediate effects of the program's activities on the target groups.[10] Impact Indicators (II) evaluate the broader, longer-term impacts of the RDP on the economic, social, and environmental dimensions of rural areas.[11] The European Commission also included a further set, that of Target Indicators (TI), for the evaluation of the 2014–2020 Rural Development Program. Target indicators are specific, measurable goals set at the beginning of the program to define expected achievements by the end of the programming period.[12]

[7] Part of the loss of observations in the balanced panel going back from 2015 to 2013 depends on the fact that several farms, which would remain in the balanced sample, still report the adoption of analogous or similar measures for years 2014 and 2013, though actually referring to the previous CAP programming period (2007–2013). Since at least two before-treatment observations are needed, and in order to avoid overlap between two different CAP regimes, these units have been removed from the balanced panel.

[8] Cases of multiple treatments (i.e., units receiving simultaneously two or more of the six treatments here considered) cannot be excluded, in principle. However, as clearly emerging from Table 2, the numerosity of adopters is usually quite low and this possibility of simultaneous adoption is ruled out in the two balanced panels under study. In any case, the presence of multiple treatments would not necessarily undermine the analysis if the assumption of independence across treatments is maintained (see Section 5.2).

[9] Examples include: number of operations supported; total public expenditure; number of beneficiaries.

[10] Examples include: jobs created in supported projects; percentage of agricultural land under management contracts to improve biodiversity; percentage of the rural population covered by local development strategies.

[11] Examples include: increase in agricultural productivity; reduction in greenhouse gas emissions; Improvement in water quality; reduction in the rate of population decline in rural areas.

[12] Additionally, the CMEF framework also includes Context Indicators (CI): These provide information about the general conditions and trends in the rural areas and are used for contextual analysis.

Table 2
Number of treated farms by policy in the two alternative panels during the period (2015–2022).

Treatment	2015	2016	2017	2018	2019	2020	2021	2022	Never treated units[a]	Total panel
Unbalanced Panel										
P1	3	3	3	4	6	7	8	9	9566–11075	
P2	0	2	10	32	52	63	69	74	9569–11010	
P3	6	38	66	134	218	307	366	402	9563–10726	
P4	0	0	2	4	7	14	17	19	9569–11065	
P5	28	268	382	420	482	455	494	511	9541–10573	
P6	1	33	34	49	70	74	77	81	9568–11003	
Balanced Panel										
P2	0	0	3	10	17	32	38	43	3856	3899
P4	0	0	1	1	2	4	4	4	3895	
Balanced Panel (2013–2022)										
P1	2	2	2	2	2	3	3	3	2602	2605
P3	2	13	21	41	55	65	80	98	2507	
P5	3	79	113	130	142	148	153	160	2445	
P6	1	14	17	25	32	31	36	40	2565	

[a] In the unbalanced panel the yearly minimum and maximum number of untreated units during the period are reported

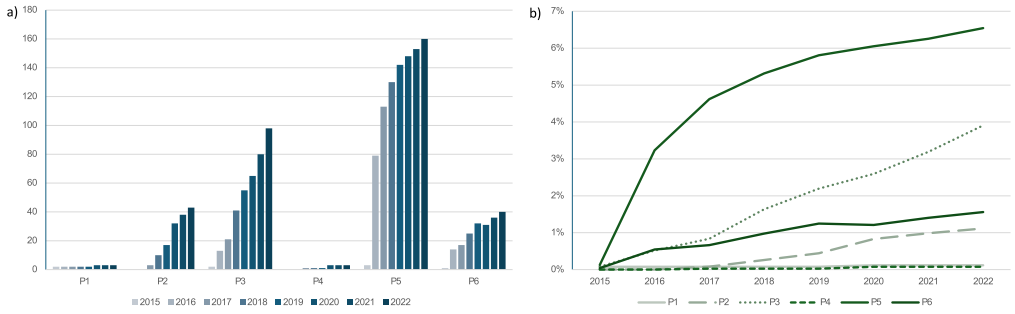


Fig. 1. Evolution of the treated farms across the period under consideration in: a) units; b) as percentage on the balanced panel.

For more details, see also Table A2 and Fig. A1 in Annex 2.

These sets of indicators are obviously informative for the present investigation in order to properly identify the expected societal outcome of the different RDP measures. Nonetheless, this plethora of indicators put forward by the EU Commission for the RDP evaluation does not allow an univocal identification of the societal outcome variable to be considered in investigating the farm-level response to policy adoption. Not only this outcome necessarily differs across measures. More importantly, for several policies, more than one single societal outcome can be legitimately associated with the measure under consideration. In principle, RI and TI are closer to this idea as both measure direct, short-term outcomes and concentrate on the involvement in terms of land and farms. However, most RI and TI are not expressed at the farm level, thus they do not capture farmers' choices and behaviour that eventually link to the ultimate societal objective.

What we need here is some mixture between all the abovementioned sets of indicators, though at the different levels. The set of societal outcome variables here considered is detailed in Table 1 that also reports, in the last column, the closer reference indicators of the CMEF terminology associated to each of the selected societal outcome variables. Some of these latter are, in fact, mid-way between impact and result indicators. They are measure-specific and may be multiple leading to respective alternative estimates of the societal TE. In order to facilitate the comparison across largely heterogenous farm structure, this impact or result indicator is preferably referred to a variable expressing the farm size like labour units (expressed by Annual Working Units, AWU), physical size (Utilised Agricultural Area, UAA) or economic size (Gross Production Value, GPV; Gross Value Added, GVA; Standard Output, SO).

Although all adopted outcome indicators are drawn from the abovementioned RDP Indicator Fiches, our selection also reflects the logic recommended by international guidelines, outlooks, and surveys. In line with established frameworks for agricultural and rural development policy evaluation, such as those issued by FAO and OECD (DeBoe, 2020; FAO, 2024; OECD, 2024), the chosen variables integrate measures of farm economic performance (e.g., GVA ratios) and income (e.g., NI ratios), resource efficiency (e.g., AWU, UAA), and diversification (e.g., OGA indicators). With specific reference to the outcome variables linked to the individual policy measures, we also draw on recent assessment exercises in the literature.

Farm income per labour unit (NI/AWU or NI/FAWU) is widely used in Eurostat accounts (Eurostat, 2024), in farm income surveys (Wageningen Economic Research, 2024), as well as in Figurek et al. (2024) and Prigoreanu et al. (2025). GVA/AWU is a standard indicator of labour productivity discussed and used by Coca et al. (2023), Figurek et al. (2024), and Popescu et al. (2021), while GVA/UAA is recognized as a measure of production intensity (Prigoreanu et al., 2025). The latter study also employs indicators such as GVA/SO to reflect economic return relative to production scale, and investment ratios like Inv/AWU and Inv/SO. Diversification measures (OGA/AWU, OGA/SO) appear in Garcia-Cornejo et al. (2020) and Prigoreanu et al. (2025). Livestock indicators (LSU/AWU, LSU/UAA) feature in FAO analyses and OECD Agri-Environmental Indicators (OECD, 2025). VE/SO is used by FAO to monitor public spending on veterinary services and assess their effectiveness and sustainability. Finally, FA/UAA is relevant for agroforestry and land-use analysis in OECD Agri-Environmental Indicators (OECD, 2025).

4.4. Covariate set

The confounding variables are selected to capture the key features of the unknown farm-specific technology (Coderoni et al., 2024; Esposti, 2024). Following Brown et al. (2021) and Stetter et al. (2022), these farm attributes can be assembled into four groups: economic factors (i.

e., factor endowment); socio-demographic characteristics (of the farm's holder and workforce); environmental factors (mostly geographical but also including other forms of policy support); and idiosyncratic characteristics (such as ability, knowledge, motivations, beliefs, and values of the farm's holder and workforce, as well as unobserved environmental features such as agronomic characteristics and fertility).[13]

Among economic factors, in order to express the possible presence of scale economies, a size variable is included (SIZE). Size here enters as a dimensional class, that is, an ordered categorical variable[14] and it is preferred to a continuous variable expressing the economic size (for instance, farm's GPV or SO) as it is much more stable over time. To take the multioutput nature of technology into account, the farm's production specialization (PRO) is included. Also in this case a (unordered) categorical variable is preferred as the share of specific products on the farm's GPV or SO would be highly unstable.[15] Moreover, as this PRO variable does not consider forestry production, an additional variable (FOR) expressed as the ratio of the farm's forest area on the utilised agricultural area (UAA) is included.[16] To represent the multiple production factors, a four-input technology (i.e., capital (K), labour (L), energy (E), materials (M)) is considered. Capital is expressed by the total farm's machinery power (KW), by the UAA and by the livestock units (LSU). Labour is expressed by annual working units (AWU). Energy is expressed by the total expenditure for energy (fuels included) (ENE), while materials are expressed by the total expenditure for fertilizers, pesticides and animal feed (MAT).

To complete the definition of the farm's factor endowment, two further aspects are considered. First of all, input quality matters and this is particularly true for labour. To take this aspect into account, the farm holder's age is included (AGE) together with the dummy expressing the farmer's gender (GEN, where female=1). In addition, to be consistent with the size-independent outcome variable, all the non-categorical input variables are expressed as intensities, that is, divided by the farm's SO. For instance, AWU is computed as the annual working units per unit of SO.

A further set of confounding variables is intended to take the farm's geographical context into account. In particular, the farm's average altitude (ALT), latitude (LAT) and longitude (LON) are included among covariates. We include among confounders also the overall CAP first pillar payments received by the farms, also divided by the farm's SO (PIL1). This variable reflects a combination of several farm-specific factors. In the Italian case, the 2014–2022 CAP regime implies that PIL1 partially depends on the farm's specialization (due to the historical base of the first pillar basic payment), on geographical location (due to internal convergence) and on the farmer's characteristics (for instance, due to the additional payment for young farmers). But PIL1 also affects farmers' decision-making due to the associated cross-compliance requirements (Sotte & Brunori, 2025).

The $(P \times 1)$ vector $\mathbf{X}$ eventually includes the following 11 variables $(Q = 11)$: SIZE, PRO, FOR, KW, UAA, LSU, AWU, ENE, MAT, AGE, GEN, ALT, LAT, LON, PIL1. As shown in Table A4 (Annex 4), most of these variables exhibit limited variation over time. In fact, several are clearly time-invariant (SIZE, PRO, FOR, AGE, GEN, ALT, LAT, LON, PIL1),[17] while others are time-variant (KW, UAA, LSU, AWU, ENE, MAT). The time-invariant variables reflect structural characteristics of the farm and remain stable over time, making the specific year used to represent them irrelevant. Conversely, the time-variant variables are at least partially influenced by the farmer's annual production decisions. Since these are behavioural in nature, it cannot be excluded that they are affected by the decision to adopt a treatment.

To mitigate this potential endogeneity, we adopt a common strategy: all covariates are included in the estimation as biennial averages, calculated over the two pre-treatment periods, 2013–2014 and 2015–2016, for the two balanced panels, respectively. Section 6.4 further discusses this strategy, also referred to as *lag identification* (Bellemare et al., 2017), and its implications for the validity of the Conditional Independence Assumption.

5. Identification and estimation approach

As discussed in Section 2, the main challenge from an identification and estimation standpoint posed by policies such as the RDP measures under investigation lies in both unit and time heterogeneity. Specifically, early adopters are few and often exhibit atypical characteristics, while late adopters enter the treatment at different points in time. Beyond the intrinsic heterogeneity of farms, variation in treatment response may also depend on the external conditions prevailing at the time of policy adoption. As a result of this staggered treatment structure, the response to the intervention is cohort-specific, meaning that it varies depending on the timing of entry. Consequently, the estimated average treatment effect (TE) reflects not only variation across units, but also across cohorts.

In this analysis, we assume that once a unit enters treatment, it remains exposed thereafter, an assumption commonly referred to as *absorbing treatment*. It is important to note that while staggered adoption allows for cohort-specific treatment effects (TE), it does not necessarily imply time-varying effects within each cohort (Lechner & Miquel, 2010; Sun & Abraham, 2021; Baldoni & Esposti, 2025). In this study, we explicitly exclude the possibility of within-cohort temporal variation in TE and focus solely on differences across cohorts, which arise from the timing of treatment adoption.

Estimation approaches such as DID and SCM may face limitations under these conditions. In contrast, Synthetic Difference-in-Differences (SDID) offers a flexible framework that combines key features of both DID and SCM, potentially addressing the challenges posed by staggered adoption and cohort heterogeneity. One of the main intended contributions of this study is to assess the performance of SDID estimation in the context of RDP measures. More specifically, we aim to examine how the results obtained via SDID compare to those produced by DID and SCM, not only highlighting any substantial differences, but also exploring potential reasons behind these discrepancies across policy measures.[18]

5.1. The Synthetic DID estimation approach

The SDID procedure requires, as input, a balanced panel, a binary treatment, well-defined outcome variables and a set of confounders. The information provided by the dataset illustrated in Section 4 is thus fully suitable for this estimation approach. Though binary, the treatment variable can be time-variant (i.e., $T_{it} = 0; 1$) since staggered adoption is admitted (Athey & Imbens, 2022). In the particular setting of the SDID estimation, a treated unit can be included in the estimation only if it has at least two pre-treatment periods. These are necessary to allow the unit to serve as a control. For this reason, as anticipated in Section 4, two different time periods are considered (either 2013–2022 or 2015–2022)

[13] See also Eichorn et al. (2024), Fig. 1 for a slightly different articulation of the covariates affecting farmers' policy adoption and outcome.

[14] Farms are distributed across classes from 1 to 7 (variables from SIZE1 to SIZE7) where class 1 represents the smallest size and class 7 the largest.

[15] Five farm types are considered: Livestock farms (PRO1), Livestock&crops farms (PRO2), Annual-crops farms (PRO3), Perennial-crops farms (PRO4), Other farms (PRO5).

[16] Variables FOR and LSU are excluded from the covariate set whenever they enter in one of the outcome variables (see Table 1).

[17] Although variable AGE is not strictly time-invariant, it behaves as a deterministic trend and can therefore be considered exogenous for the purposes of the analysis.

[18] There is some similarity between the SDID approach and the Causal Machine Learning (CML) estimation that has recently gained popularity (Coderoni et al., 2024; Esposti, 2024) in the sense that, though differently, both use data to train systems (the synthetic pre-treatment counterfactuals in the SDID approach) and use information very flexibly (via weights in SDID estimation).

depending on the policy measure under consideration (see Table 1).

Estimation proceeds as described in Arkhangelsky et al. (2021) and Clarke et al. (2024). Basically, the average TE on the treated (ATT) estimates are generated from a Two-Way Fixed Effect (TWFE) regression specified as follows:

$$Y_{it} = a + \alpha_i + \lambda_t + \tau_{p(s)} T_{it} + \mathbf{X}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{7}$$

Where: Y_{it} indicates the generic outcome variable that here alternatively takes the form of the private ($\Delta\pi_{it}$) or societal (ΔG_{it}) performance discussed in Section 3; a expresses a time-invariant and farm-invariant constant term; α_i is the farm-specific time-invariant term; λ_t is the shared year-specific term; $\tau_{p(s)}$ is the ATT, the estimand, for the private (p) and societal (s) outcome; $\boldsymbol{\beta}$ is the (Px1) vector of the time-invariant and farm-invariant coefficients associated to the exogenous covariates, $\mathbf{X}_{it}$; ε_{it} is the conventional spherical disturbance assumed i.i.d. $\sim N(0,\sigma^2)$.[19]

The SDID estimation looks for the parameter values that minimize the sum over the whole panel of squared residuals of (7) with N unit-specific (ω_i) and Z time-specific (γ_t) weights:

$$\left(\widehat{\tau}_{p(s)}, \widehat{a}, \widehat{\alpha}_i, \widehat{\lambda}_t, \widehat{\boldsymbol{\beta}}\right) = \underset{\tau_{p(s)},a,\alpha_i,\lambda_t,\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \times \sum_{t=1}^{Z} [Y_{it} - \left(a + \alpha_i + \lambda_t + \tau_{p(s)} T_{it} + \mathbf{X}'_{it}\boldsymbol{\beta}\right)^2] \omega_i \gamma_t \right\} \tag{8}$$

The presence of farm-specific FE implies that SDID seeks to match treated and control units on pre-treatment trends, and not necessarily on both pre-treatment trends and levels, allowing for a constant difference between treatment and control units. Consistently with the SCM logic, the selection of farm weights, ω_i, aims to ensure that comparison is made between treated units and controls approximately following parallel trends prior to the policy adoptions. The selection of time weights, γ_t, aims to give more relevance to pre-treatment periods which are more similar to post-treatment periods, in the sense of finding a constant difference between each control unit's post treatment average, and pre-treatment weighted averages across all selected controls. Here, these unit-specific and time-specific weights, ω_i and γ_t, are found following the same procedure proposed by Arkhangelsky et al. (2021, pp. 4091-4092) consisting of an iterative constrained optimization process that alternates between finding the unit-specific weights and time-specific weights minimizing the pre-treatment discrepancy between the treated unit and the synthetic control.

The minimization problem described in Eq. (8), which underlies the SDID estimation procedure, may entail a substantial computational burden, particularly when the number of covariates is large, the sample size is considerable, and the number of treated units is high. To address this, Arkhangelsky et al. (2021) propose a two-stage estimation strategy. In the first stage, the outcome variable Y_{it} is regressed on the covariates $\mathbf{X}_{it}$, yielding consistent estimates of the covariate coefficients $\widehat{\boldsymbol{\beta}}$ and corresponding residuals $\left(Y_{it} - \mathbf{X}'_{it}\widehat{\boldsymbol{\beta}}\right)$.[20] In the second stage, these residuals replace the original outcome variable Y_{it} in the SDID minimization procedure (Eq. 8), effectively removing the term $\mathbf{X}'_{it}\boldsymbol{\beta}$ from the objective function. This approach simplifies computation while preserving the consistency of the estimator.

Though originally designed for the block treatment, where all treated units enter the treatment in the same year, this SDID estimation approach can be extended to a staggered adoption case, as in the present study (Arkhangelsky et al., 2021; Clarke et al., 2024). This extension is achieved by specifying a (NxZ) adoption matrix ($\mathbf{T}$) whose element T_{it} indicates whether at time t the i-th unit is treated (T_{it}=1) or not (T_{it}=0). In turn, $\mathbf{T}$ can be broken down into (Z-2) adoption-date-(or cohort-) specific matrices, with Z columns and a number of rows equal to the number of units entering the treatment in that specific year (thus, units belonging to the same cohort). With this adjustment, the SDID estimation illustrated above provides cohort-specific ATT estimates, as well as cohort-specific unit-specific and time-specific weights (ω_i and γ_t). It follows that also the ATT over the whole sample can be computed by calculating a weighted average of the cohort-specific ATT, where weights are assigned on the basis of the relative number of treated units and time periods in each cohort in order to give more relevance to larger and early-entry cohorts.

Arkhangelsky et al. (2021) demonstrate that the abovementioned ATT estimators are asymptotically normal. This allows constructing confidence intervals provided that a consistent estimation of the respective variance is available. Arkhangelsky et al. (2021) propose a block (or clustered) bootstrap procedure for inference in SDID estimation. While this method exhibits strong statistical properties, it can be computationally intensive and may yield less reliable variance estimates when the number of treated units is small. In such cases, particularly when only a single treated unit is available, an alternative placebo-based (or permutation) inference procedure is recommended. This latter approach, detailed in Clarke et al. (2024), constructs a distribution of placebo treatment effects by iteratively assigning the treatment to control units and re-estimating the SDID model. The observed treatment effect is then compared to this empirical distribution to assess its significance. This method does not rely on asymptotic approximations and provides a more robust inference framework under small-sample conditions.

It seems interesting to stress how and to what extent this SDID estimation differs compared to the standard DID and SCM procedures. DID estimates can be obtained from (8) by simply assigning equal weights to all units and time periods, whereas the SCM estimation can be obtained from (8) by maintaining optimally chosen farm-specific weights (ω_i) and assigning equal weights to all years. Moreover, in (8) the SCM omits the farm-specific fixed effects α_i, thus implying that the synthetic control and treated units should maintain approximately an equivalent pre-treatment level, as well as trend. Therefore, the SDID estimation offers greater flexibility than the DID and SCM procedures.

Ultimately, DID can be seen as a special case of the SCM, where all unit-specific weights are equal. Conversely, DID accounts for unit and time fixed effects, whereas SCM does not. This distinction may help explain why SCM often outperforms traditional DID when the number of treated units is small. The SDID approach seeks to combine the strengths of both methods while mitigating their respective limitations. By flexibly balancing unit weighting and fixed effects adjustment, SDID offers a pragmatic compromise tailored to the structure of the data and the specific policy context (Lu, 2021). Applying SDID to multiple policy measures over the same time period, as done in the present study, may thus yield valuable insights into the conditions under which the advantages of SDID become particularly evident.

5.2. *Identification issues and assumptions: pros and cons of the SDID method*

The SDID method integrates elements of the traditional DID

[19] To avoid perfect collinearity, as standard in fully saturated FE models, one α_i and one λ_t are set at zero.

[20] Following Clarke et al. (2024), in this first stage all covariates are standardized in order to resize the variability of very-high-variance confounders while capturing the same underlying variation in covariates. In addition, two alternative approaches can be followed to estimate β parameters. One adopts an interactive optimization procedure which additionally allows for the efficient calculation of optimal weights ω_i and γ_t. This approach, however, can be highly computationally demanding and time consuming and may incur the risk of converging towards local minima. Alternatively, these parameters can be estimated much faster in a single step via OLS. This second procedure is adopted here.

approach with the SCM, aiming to improve causal inference in contexts characterized by sparse treatment adoption, conditions frequently encountered in real-world policy evaluations. This hybrid structure represents the core strength of the method. As with any causal inference strategy, however, the validity of SDID relies on a set of identification assumptions (Imbens & Rubin, 2015; Arkhangelsky et al., 2021; Clarke et al., 2024): the Conditional Independence Assumption (CIA), the Overlap Condition, and the Stable Unit Treatment Value Assumption (SUTVA). The validity of these assumptions depends less on the estimation method itself and more on the specific characteristics of the policy under study and the structure of the available data. Therefore, in applied settings their plausibility must be carefully assessed and, where possible, empirically verified.

The CIA, also referred to as *unconfoundedness* or *no selection on unobservables*, states that, conditional on observed covariates, the potential outcomes are independent of treatment assignment. In panel data settings, this can be formally expressed as:

$$Y_{it}(0), \quad Y_{it}(1) \perp T_{it} | \mathbf{X}_{it} \tag{9}$$

where $Y_{it}(1)$ and $Y_{it}(0)$ indicate the potential outcome of the i-th farm at time t with and without the treatment, respectively (Imbens & Rubin, 2015).

In this context, the CIA implies that, given covariates $\mathbf{X}_{it}$ (and possibly unit and time fixed effects), treatment assignment is not driven by unobserved factors that also affect the outcome. A testable implication of CIA in panel data settings is the conditional PTA, which can be expressed as:

$$E[Y_{it}(0) - Y_{it-1}(0) | T_{it} = 1, \mathbf{X}_{it}] = E[Y_{it}(0) - Y_{it-1}(0) | T_{it} = 0, \mathbf{X}_{it}] \tag{10}$$

SDID relaxes this condition by requiring parallel trends conditional on synthetic control weights. That is, in the absence of treatment, the treated unit and its synthetic counterpart are assumed to follow similar outcome dynamics. This more flexible matching of pre-treatment trajectories allows SDID to better approximate the counterfactual for each treated unit, even when treatment effects are heterogeneous. In other words, SDID facilitates the satisfaction of the conditional PTA by construction. Compared to DID estimations, it permits a violation of the PTA in aggregate data. Compared to the SCM, it aims to optimally weigh time periods when considering counterfactual outcomes and allows for level differences between treatment and control groups.

Specifically, in the absence of treatment, SDID assumes that the treated unit and its synthetic counterpart, constructed from a weighted combination of control units, would have followed similar outcome trajectories. This conditional formulation allows SDID to flexibly match pre-treatment dynamics, thereby improving the approximation of counterfactual outcomes even in the presence of heterogeneous treatment effects. Therefore, by construction, SDID facilitates the satisfaction of a conditional PTA, which is more robust than the aggregate-level PTA required by conventional DID estimators. Unlike DID, SDID tolerates violations of parallel trends in the raw data, provided that the synthetic control achieves balance. Compared to the SCM, SDID introduces optimal weighting over time periods, enhancing the estimation of counterfactuals and allowing for level differences between treated and control units. This temporal weighting makes SDID particularly effective in settings where treatment effects vary over time or across units.

However, two important caveats should be emphasized. First, in SDID estimation, the conditional Parallel Trends Assumption (PTA) requires that the synthetic control, constructed via unit and time weights, accurately reproduces the counterfactual pre-treatment trajectory of the treated units. This condition may be compromised if pre-treatment dynamics are poorly matched, if the pre-treatment period is short or noisy, or if the synthetic weights are unstable. These issues are particularly pronounced when the number of treated units is very small. In the extreme case of a single treated unit, the entire inference and identification strategy hinges on the quality of the synthetic match. Any mismatch or instability in weights can significantly undermine the credibility of the estimated counterfactual. In such settings, conducting careful robustness checks is advisable to ensure the reliability of the SDID estimates (Arkhangelsky et al., 2021).

Second, and more critically, even under SDID estimation, assumptions (9) and (10) still require that, conditional on observed covariates and synthetic weights, potential outcomes are independent of treatment assignment. This implies that no unobserved variable simultaneously influences both the outcome and the treatment, and that the covariates $\mathbf{X}_{it}$ are strictly exogenous, i.e., not influenced by either the treatment or the outcome, directly or indirectly. In panel data settings, this exogeneity condition may be problematic for time-varying covariates, as they may reflect farmers' production choices that are themselves influenced by treatment status and outcomes. A commonly adopted solution, also applied in this study (see Section 4), is *lag identification*, which consists in conditioning on lagged values of the covariates, observed sufficiently before treatment to ensure their exogeneity, in practice as they were time-independent. Ideally, under this approach, the CIA could be reformulated as $Y_{it}(0),\ Y_{it}(1) \perp T_{it} | \mathbf{X}_{i}$.

Bellemare et al. (2017) caution against the uncritical use of lag identification, noting that it can purge endogeneity only under *sequential exogeneity*, i.e., the error term must be uncorrelated with all past values of the covariates. In practice, this condition requires: (i) serial correlation in the explanatory variables, and (ii) absence of serial correlation in the unobserved sources of endogeneity. These conditions are difficult to verify and occasionally satisfied in real-world policy evaluations. Nevertheless, it is important to note that Bellemare et al.'s concerns apply primarily to time-varying, potentially endogenous covariates. In contrast, time-invariant covariates, such as structural or demographic characteristics, are generally considered exogenous, as they are not subject to temporal shocks or reverse causality.

In this study, lag identification is applied with caution. As shown in Table A4 (Annex 4), most covariates in $\mathbf{X}_{it}$ are structural and thus time-invariant or nearly constant. Therefore, for these variables, the risk of endogeneity is minimal. However, this risk cannot be excluded for time-varying covariates, and further investigation is warranted (see Section 6.4).

The Overlap Condition, also commonly known as *common support* (or *positivity*), implies that, for most observational units (in this case, farms), both treatment and non-treatment are feasible alternatives. More formally, it requires that for every combination of covariates, the probability of receiving either treatment level is strictly positive: $0 < \Pr(T_{it} = 1 | \mathbf{X}_{it}) < 1$. This assumption ensures that treated and untreated units are comparable within strata of the covariates $\mathbf{X}_{it}$, thereby limiting extrapolation and reducing the risk of biased estimates of the ATT. In the context of SDID, however, overlap takes a slightly different form. Rather than relying on probabilistic treatment assignment, SDID requires that a weighted combination of control units can credibly reproduce the pre-treatment trajectory of treated units. This operational condition, which could be designated, for simplicity, as *synthetic overlap*, is central to the credibility of the synthetic counterfactual (Arkhangelsky et al., 2021; Clarke et al., 2024).

However, when evaluating granular policy interventions with sparse and staggered adoption, satisfying this synthetic overlap becomes more challenging. As the feasibility of constructing a reliable synthetic control depends on the richness of the donor pool and the stability of pre-treatment dynamics, if treated units are few and exhibit unique characteristics, it becomes difficult to find suitable counterparts among the untreated group. One of the key strengths of the SDID estimator lies in its ability to address this issue. By constructing synthetic controls through reweighting untreated units to closely match the pre-treatment outcomes of treated units, SDID effectively restricts comparisons to regions of common support. This synthetic matching enhances the credibility of the estimated effects, provided that sufficient similarity exists in the covariate distributions between treated and control units.

To empirically assess the validity of the overlap assumption,

Table A4 (Annex 4) presents the results of Welch's *t*-tests comparing the means of covariates between treated and untreated farms.[21] While the test can be conducted for any year with at least two treated units, we compute the sample averages and variances over the whole periods as it includes a large enough number of treated units, thereby enhancing the robustness of the statistical comparison. The results indicate that, for most policy interventions (P1–P6), the distributions of covariates are sufficiently similar, with no statistically significant differences in the majority of cases. Very few exceptions are observed for policies with very few adopters and for covariates that exhibit substantial temporal variation. Overall, these findings support the validity of the overlap condition in the context of the present analysis.

A more challenging identification condition in this application is the SUTVA. Broadly, SUTVA requires that the treatment effect on a given unit is independent of the external context, except for variables explicitly controlled for in $\mathbf{X}_{it}$. It is commonly interpreted as the absence of spillover effects or interference, meaning that the treatment assigned to one unit does not influence the outcomes of other, potentially neighbouring, units. This aspect is particularly relevant in agricultural settings (Baldoni and Esposti, 2023; Esposti, 2024), but is not explicitly addressed here, as several spatial features with potential influence are already controlled for in $\mathbf{X}_{it}$ (see Section 4).

SUTVA also entails a second, often overlooked, condition: the absence of multiple, possibly hidden, versions of the treatment (VanderWeele & Hernán, 2013). This implies that the treatment must be consistently defined across units to ensure that the estimated TE corresponds to a well-specified policy intervention. In our case, this condition is challenged by the fact that most policy aggregates (P1–P6), except for P4, consist of multiple underlying measures. Although the aggregation is theoretically and practically motivated (see Section 4), it may conflict with SUTVA. Nevertheless, SUTVA may still hold if the different versions are functionally equivalent, i.e., they produce similar effects across units. This assumption underpins our aggregation strategy, based on the expectation that the measures elicit comparable responses from farmers.

Even if full equivalence is uncertain, SUTVA remains valid provided that the version assigned to one unit does not affect the outcome of another. In such cases, the estimated TE can be interpreted as a consistent estimate of the weighted average of the effects of the component versions. This condition appears to be satisfied in our data. Although multiple treatments cannot be excluded in principle, the low number of adopters in the two balanced panels effectively rules out this possibility. For none of the six policy aggregates do we observe farms receiving support for more than one measure within the same aggregate. This suggests that the measures are mutually exclusive in practice, and their adoption can be considered independent, with no interference across versions. Still, the validity of SUTVA remains a critical and debatable issue in this context. For this reason, it is further explored in Section 6.4.

A final identification condition worth highlighting, though often overlooked, is the *no-anticipation assumption*. This assumption requires that units do not alter their behaviour in response to the expectation of future treatment. If units adjust their actions in anticipation of receiving support, this may introduce bias into the estimated treatment effects. More formally, it implies that the potential outcome in the absence of treatment at time t, conditional on covariates $\mathbf{X}_{it}$, is equal to the potential outcome at a future time $t+s$, even for units that eventually receive treatment: $Y_{it}(0)|\mathbf{X}_{it} = Y_{it+s}(0)\,|T_{it} = 1, \mathbf{X}_{it}$.

This issue is particularly relevant in panel data settings with staggered adoption, where knowledge of future treatment assignments may influence current decisions. In principle, such anticipatory behaviour cannot be excluded in the case of the EU Rural Development Programme (RDP) measures under investigation. As discussed in Esposti (2022), the sequence of programming periods (e.g., 2007–2013 preceding the one analysed here) may provide potential beneficiaries with information about forthcoming measures, prompting them to adjust their behaviour to increase the likelihood of receiving support. Even within a single programming period, staggered adoption may allow late adopters to anticipate and modify production decisions to maximize expected benefits. At the same time, however, the adoption of RDP measures involves formal application and approval procedures, which introduce uncertainty and delays. These factors may discourage anticipatory behaviour, thereby mitigating the risk of violating the assumption.

Although the violation of the no-anticipation assumption cannot be excluded in this empirical setting, its investigation lies beyond the scope of the present analysis. A rigorous assessment would require modelling farmers' behaviour as a dynamic adaptation process, and interpreting the treatment effect as time-variant, thus incorporating a dynamic component. This approach introduces significant methodological complexity. Recent studies have begun to explore this direction (Baldoni & Esposti, 2023; 2025), and it certainly represents a promising avenue for future research in this field.

6. Results

6.1. SDID estimates

Table 3 reports ATT estimates for both the private and societal outcome variables over the whole period (namely, over all cohorts) for the six policy treatments.[22] The standard error of the ATT estimate is computed using the bootstrap procedure described in Arkhangelsky et al. (2021) and using 50 replications. Following Clarke et al. (2024), for treatments in which there are cohorts with only one treated unit (P4 and P6) the bootstrap procedure is replaced by the placebo inference (see Section 5.1).

It clearly emerges that for some of them (P1 and P2, in particular) no

Table 3
Private and societal ATT by treatment (estimated standard errors in parenthesis).

	Private ATT ($\Delta\pi$) (.000€)	Societal ATT (ΔG)[a] (1)	Societal ATT (ΔG)[a] (2)	Societal ATT (ΔG)[a] (3)
P1	32.80 (31.1)	192.2 (658.3) [GVA/AWU]	519.6 (521.9) [GVA/UAA]	-
P2	-0.244 (5.27)	0.152 (0.859) [GVA/SO]	-	-
P3	11.06 (5.19)**	812.5 (304.6)** [INV/AWU]	0.005 (0.002)** [INV/SO]	-
P4	37.63 (12.3)**	2.757 (1.359)** [OGA/SO]	5581 (4314) [OGA/AWU]	-
P5	8.443 (3.861)**	-0.084 (0.146) [LSU/UAA]	-1.468 (0.763)* [LSU/AWU]	-0.006 (0.003)** [VE/SO]
P6	38.30 (12.93)**	0.045 (0.024)* [FA/UAA]	-	-

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a] Societal outcome variable in square brackets.

[21] Welch's *t*-test is preferred over the standard *t*-test when sample sizes or variances differ, as it provides more reliable results without assuming equal variances.

[22] All results have been obtained with the SDID command ran in STATA 18.5 software. Annex 3 reports the estimates of the covariates' coefficients (β in Eqs. (7) and (8)) for a selected treatment, P4. Since these estimates are cohort-specific, Annex 3 only reports coefficient estimates for P4 as it only presents two cohorts (2017 and 2020) but still shows a statistically significant overall ATT for both the private and societal outcome variables. Therefore, Table A3 reports four coefficient vectors, two for each cohort and two for each outcome variable. Estimates for all other treatments are available upon request.

estimated ATT is statistically different from zero, meaning that there is no clear evidence of an impact of these RDP measures on both the private and societal performance. This result could have two different interpretations. The first reading is that, in the case of the private outcome, the adoption of these policies does not bring about any significant before-support income gain for the farm. In other words, these measures would not be adopted if an economic incentive (the monetary support associated to the measure) were not granted. In the case of the societal outcome, according to this interpretation, the lack of a significant impact would indicate that, beside the alleged and declared policy measures, these measures simply turn out to be an income support to farmers', so they consist in a purely income redistributive transfer from the rest of the society to the farmers without any appreciable counterpart or justification for the society itself.

Though this interpretation is not new within the literature on the real underlying motivations of the RDP measures (Esposti, 2022), it should be acknowledged that these results for P1 and P2 can be more prudently interpreted in terms of an inconclusive evidence on the actual impact of these measures on farmers' behaviour and performance. Inconclusiveness that can be arguably attributable to the lack of more robust information due to the limited number of post-treatment years,[23] the limited number of treated units, especially in the early years, and, as will be discussed below, the highly year-by-year volatility of the outcome variables in use. In fact, while this latter interpretation seems to be reasonable for P2 where both the private and societal ATT show small magnitude, it seems more questionable for P1. In this case, both outcome variables show a response of remarkable magnitude (more than 35 thousand € per unit of FAWU in the case of the private outcome), but the high variance associated to these estimates makes this evidence inconclusive.

For policies P3, P4, P5 and P6, all estimated private ATT are significantly different from zero and positive. They are all quite consistent in magnitude as estimates range between about 8.5 thousand € per unit of FAWU (for P5) to about 38 thousand € per unit of FAWU (for P6). On the one hand, this seems consistent with the rationale of farmers' decision-making as illustrated in Section 3: a negative impact on the private outcome, if larger than the underlying policy support, would imply a negative net income as a consequence of the policy adoption. On the other hand, however, the evident policy implication of these results is that for all these measures farmers obtain a monetary advantage that goes beyond the policy support itself. In other words, and in principle, the societal outcome could have been obtained even without the policy support as farmers would still have found an advantage to produce that response.

In the case of these four policies, also results (ATT) for the societal outcome seem quite robust though slightly less statistically significant. In all cases, the societal outcome moves in the expected direction (positive for P3, P4 and P6 and negative for P5). For P4 and P5, however, one of the candidate societal outcomes does not show a statistically significant difference from 0. In any case, it can be concluded that these measures demonstrate to be effective in inducing the response of societal interest, but their efficiency is questionable since this response would have been obtained, in principle, even with a lower (possibly zero) public expenditure.

As discussed, besides these ATT averaged over the whole period and cohorts, SDID estimation also returns cohort-specific ATT providing interesting evidence on the volatility of farmers' response to these measures across the different adoption times. Table 4 provides evidence about this time heterogeneity by reporting the cohort-specific ATT estimates. Table 4 confirms the cohort-dependency of the ATT estimates, both private and societal. These latter, in particular, are not only volatile, thus instable, in magnitude but also in statistical significance and, sometimes, even in sign. This volatility may eventually explain the weak (or lack of) statistical significance of the ATT computed over the whole period at least for some policy measures.

More in detail, in the case of P1 only the first cohort (2015) shows a statistical significance for private and societal ATT, while significance is found only in the last two cohorts for policy P2. P4 reports statistical significance for both available cohorts and for both the private and societal outcome variables. The other policies (P3, P5, P6), for both outcomes, alternate years showing significant ATT with others in which significance is lost though the sign remains the same (with the only exception of P5).

While Table 4 demonstrates the heterogeneity of the cohort-specific ATT estimates, Figs. 2–7 display one of the major causes of this heterogeneity, namely, the large volatility of outcome variables over time.[24] Heterogeneity across cohorts, therefore, lies not only in differences in ATT, but also in the volatility of outcome variables, which inevitably affects the cohort-specific ATT estimation. This evidently holds true for the private outcome which is always the same across all measures. But it also seems true for the different societal outcomes across measures.

These graphs show the trends of the outcome variables for both the treated units and the synthetic controls before and after the entry. This makes it possible to visualize the impact of the treatment but also the possible violation of the PTA. The weights used to average pretreatment time periods are also reported at the bottom of the graphs (shaded areas) to emphasize how the SDID estimation flexibly makes use of the pre-treatment observations. The estimated effect can be detected by the deviation from a parallel trend between treated and control unis after the treatment. It is worth noticing that, as stressed in Arkhangelsky et al. (2021, Figure 1), the ATT applies to the year of the treatment but since then it deviates from the trend for any year after the treatment, so the deviation from parallelism applies over the whole post-treatment period.

Therefore, for instance, a positive ATT implies that the treated units either get closer or even overcome the synthetic control if it was underperforming in the pre-treatment period (see Fig. 2b) or divaricates from the synthetic control if it was overperforming in the pre-treatment period (see Fig. 4b). In fact, all possible cases are found. There are cases for which the treatment makes the treated units closer to the controls (Fig. 2b), and other cases in which they divaricate (Fig. 6a). They also signals both positive (all figures referring to the private outcome variable) and negative (Fig. 6b) effects consistently with what reported in Table 4. Another interesting consideration about the volatility of outcome variables before and after the treatment is that one often disregarded (and often unintended) effect of a policy may be to stabilize (or destabilize) the outcome variable beside affecting its level.

The key message here is that, unlike Arkhangelsky et al. (2021) and Clarke et al. (2024), the time dimension does not bring about a clear and regular dynamics but an highly volatile and unstable movement of outcome variables. This might not surprise when farm-level data are in use. However, it implies that the DID logic, being it based on the PTA, is by itself less helpful. Moreover, this may explain why in recent studies concerning farm-level data the time dimension has been collapsed into a cross-section dataset in order to get rid of this volatility (Baldoni & Esposti, 2023; Coderoni et al., 2024). Alternatively, as done in the present study, the SDID estimation can more flexibility take care of the time dimension without assuming unlikely parallel trends.

[23] It is worth noticing, however, that his lack of statistical significance is found for both period 2013–2022 (P1) and period 2015–2022 (P2).

[24] Only exemplary cohorts are reported. They are selected by looking at the cohort-specific ATT (significance and magnitude) but, above all, concentrating on cohorts in the middle of the period to better visualize some years before and some years after, to appreciate volatility pre and post-treatment but also allowing for an appreciable number of treated units. All other years are available upon request.

Table 4
Cohort-specific (i.e., by year of entry in the treatment) private and societal ATT (standard errors in parenthesis).

	P1[a]	P2	P3[b]	P4[c]	P5[d]	P6
Private ATT ($\Delta\pi$) (.000€)						
2015	35.3 (17.3)**	-	36.4 (33.5)	-	7.625 (8.601)	25.130 (12.060)**
2016	-	-	34.08 (16.8)**	-	22.03 (11.65)**	15.39 (9.024)*
2017	-	18.1 (25.7)	8.06 (15.3)	2.606 (1.293)**	13.69 (13.81)	13.16 (13.61)
2018	-	-5.81 (12.9)	3.53 (18.0)	-	14.19 (13.187)	36.11 (14.68)**
2019	-	-1.18 (6.76)	2.07 (17.1)	-	7.261 (7.283)	32.39 (6.940)**
2020	19.5 (30.3)	3.21 (5.54)	11.5 (5.51)**	74.91 (34.42)**	8.057 (7.629)	35.536 (19.43)*
2021	-	7.59 (3.96)*	17.4 (9.25)*	-	14.09 (5.581)**	32.95 (26.97)
2022	-	4.82 (2.33)**	7.18 (15.8)	-	13.07 (6.535)**	18.42 (7.144)**
Total ATT	32.80 (31.1)	-0.244 (5.27)	11.06 (5.19)**	37.63 (12.3)**	8.443 (3.861)**	38.30 (12.93)**
Societal ATT (ΔG)						
2015	7110 (1973)**	-	182.4 (294.8)	-	-0.007 (0.004)	0.014 (0.012)
2016	-	-	1871 (1131)	-	0.000 (0.001)	0.071 (0.029)**
2017	-	-0.182 (0.396)	2378 (1069)**	5.598 (0.416)**	0.001 (0.003)	-0.014 ((0.021)
2018	-	-0.276 (3.41)	104.3 (131.2)	-	-0.002 (0.001)**	0.021 (0.026)
2019	-	0.048 (0.055)	198.1 (259.0)	-	-0.003 (0.001)**	0.078 (0.039)**
2020	-6707 (7969)	0.019 (0.088)	525.9 (286.3)*	0.823 (0.443)*	-0.002 (0.001)**	0.013 (0.018)
2021	-	0.225 (0.045)**	1526 (755.3)**	-	-0.003 (0.001)**	0.002 (0.015)
2022	-	0.141 (0.040)**	149.2 (328.8)	-	-0.003 (0.002)*	0.035 (0.019)*
Total ATT	519.6 (521.9)	0.152 (0.859)	812.5 (304.6)**	2.757 (1.359)**	-0.006 (0.003)**	0.045 (0.024)*

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a] Societal outcome: GVA/UAA;
[b] Societal outcome: INV/AWU;
[c] Societal outcome: OGA/SO;
[d] Societal outcome: VE/SO

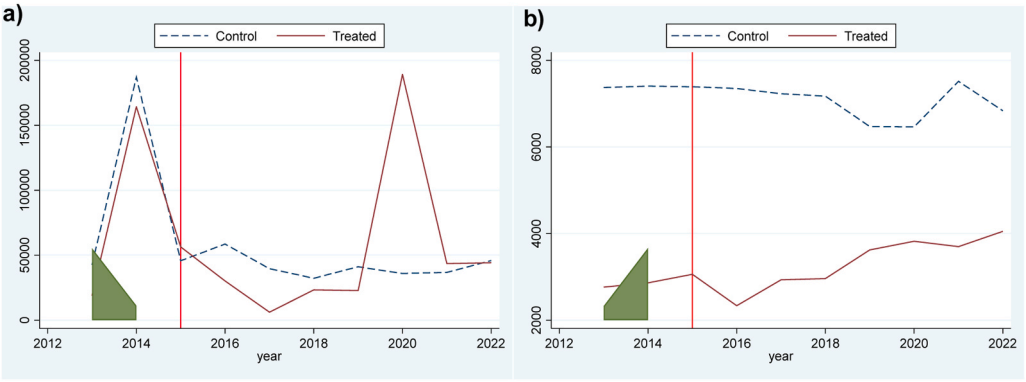


Fig. 2. Outcome trends for units entering treatment P1 in year (cohort) 2015: a) private outcome (NFI/FAWU); b) societal outcome (GVA/UAA). The shaded area indicates the time-specific weights.

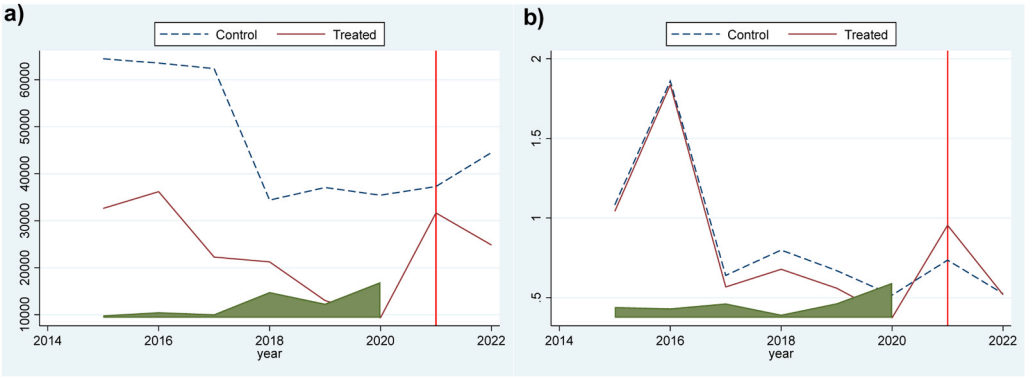


Fig. 3. Outcome trends for units entering treatment P2 in year (cohort) 2021: a) private outcome (NFI/FAWU); b) societal outcome (GVA/SO). The shaded area indicates the time-specific weights.

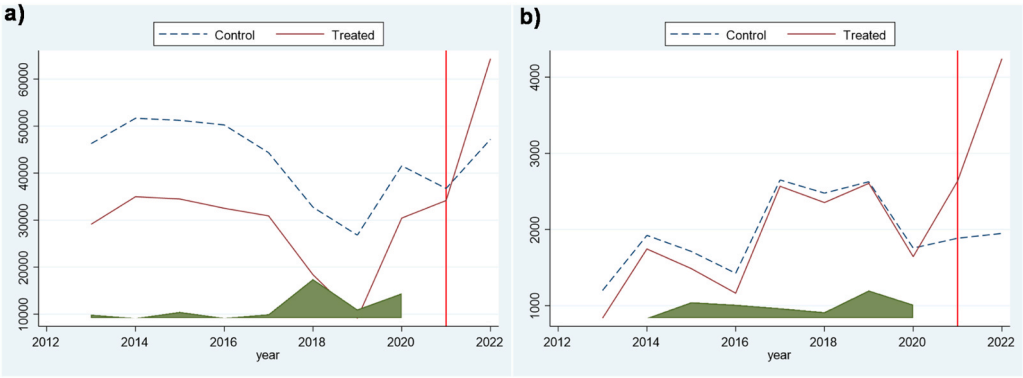


Fig. 4. Outcome trends for units entering treatment P3 in year (cohort) 2021: a) private outcome (NFI/FAWU); b) societal outcome (INV/UAA). The shaded area indicates the time-specific weights.

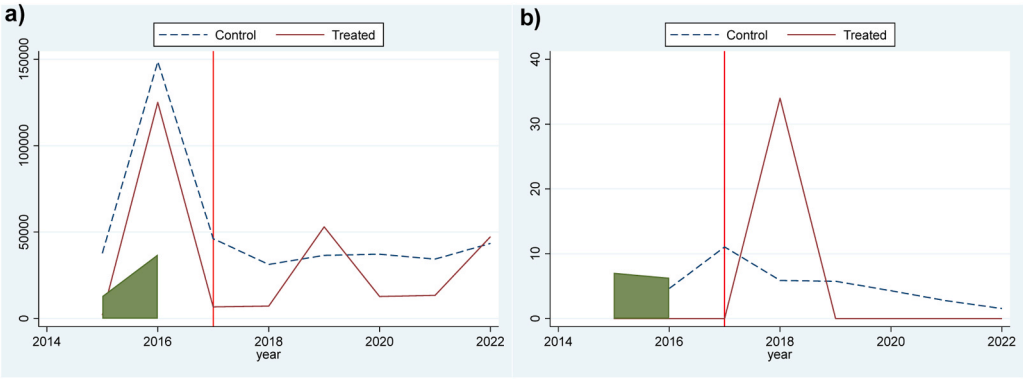


Fig. 5. Outcome trends for units entering treatment P4 in year (cohort) 2017: a) private outcome (NFI/FAWU); b) societal outcome (OGA/SO). The shaded area indicates the time-specific weights.

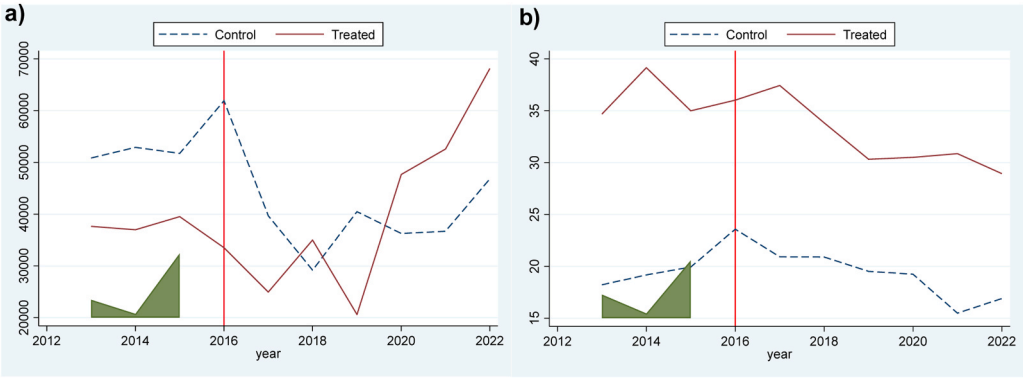


Fig. 6. Outcome trends for units entering treatment P5 in year (cohort) 2016: a) private outcome (NFI/FAWU); b) societal outcome (LSU/AWU). The shaded area indicates the time-specific weights.

6.2. *Comparison with DID and SCM results*

Table 5 and Figs. 8–11 compare SDID ATT estimates with the respective DID and SCM counterparts for a selected group of measures. Table 5 reports both private and societal ATT over the whole period while, for the sake of space limitations, figures concern only the private outcome variable for the policy measures showing a statistically significant SDID ATT (P3, P4 P5 and P6; see Table 4). We select only one cohort per policy treatment with a statistically significant ATT but different across measures in order to represent all the periods considered.[25] The objective here is only to show how the SDID looks for a compromise between DID and SCM estimations and this compromise depends on the specific characteristics of the treatment under consideration, in particular the number of treated units.

The volatility of the outcome variables discussed above implies that the validity of DID estimates may be questionable since PTA barely holds true. In principle, the advantage of the SCM on DID estimation concerns cases with very few (early) adopters (such as P1 and P4). Table 5

[25] All other figures are available upon request.

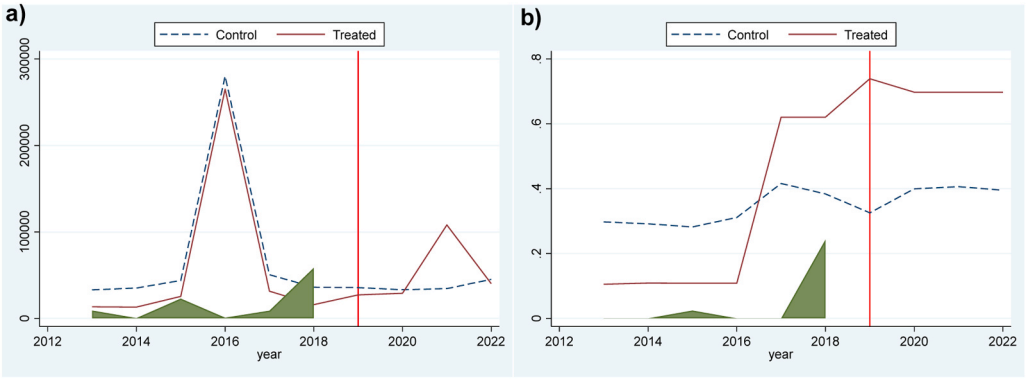


Fig. 7. Outcome trends for units entering treatment P6 in year (cohort) 2019: a) private outcome (NFI/FAWU); b) societal outcome (FA/UAA). The shaded area indicates the time-specific weights.

Table 5
Comparison of SDID, DID and SCM ATT estimates by policy (standard errors in parenthesis).

	P1[a]	P2	P3[b]	P4[c]	P5[d]	P6
Private ATT ($\Delta\pi$) (.000€)						
SDID	32.80 (31.1)	-0.244 (5.27)	11.06 (5.19)**	37.63 (12.3)**	8.443 (3.861)**	38.30 (12.93)**
DID	19.91 (29.5)	-0.679 (5.22)	8.49 (9.36)	15.22 (19.6)	10.23 (2.35)**	38.47 (20.9)*
SCM	27.70 (64.1)	-1.052 (3.11)	5.07 (8.04)	16.48 (13.6)	5.30 (5.21)	46.04 (37.13)
Societal ATT (ΔG)						
SDID	519.6 (521.9)	0.152 (0.859)	812.5 (304.6)**	2.757 (1.359)**	-0.006 (0.003)**	0.045 (0.024)*
DID	381.4 (463.0)	-1.44 (2.55))	849.8 (936.0)	2.710 (2.113)	-0.005 (0.002)**	0.161 (0.138)
SCM	390.3 (808.6)	0.960 (2.86)	307 1(504.5)	1.789 (1.889)	-0.0065 (0.077)	0.053 (0.057)

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a] Societal outcome: GVA/UAA;
[b] Societal outcome: INV/AWU;
[c] Societal outcome: OGA/SO;
[d] Societal outcome: LSU/AWU

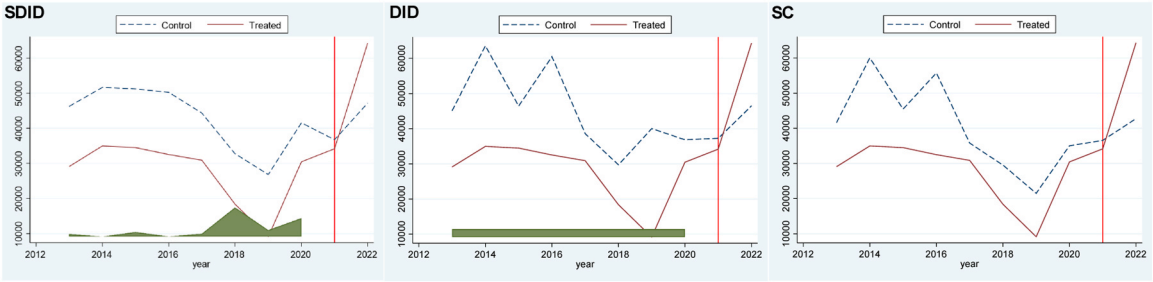


Fig. 8. Comparison of SDID, DID and SCM estimates for P3, entry year (cohort) 2021.

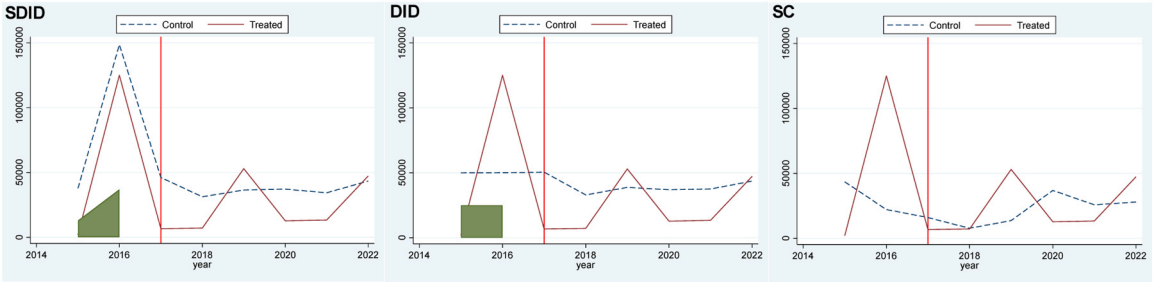


Fig. 9. Comparison of SDID, DID and SCM estimates for P4, entry year (cohort) 2017.

confirms the SDID advantage over both SCM and DID estimations but it seems more evident compared to SCM. For both cases, SDID shows higher efficiency: no SCM ATT estimate is statistically different from 0, while few (P5 and P6, in particular) are significant under the DID estimation. The flexible use of the panel informative set made by SDID estimation evidently brings about this gain in efficiency particularly in admitting different trends between treated units and controls (like SCM and unlike DID). To stress this, Figs. 8–11 visually compare SDID, DID and SC as in Arkhangelsky et al. (2021). DID and SCM alone might not work properly, while flexibly combining them in the more data-driven

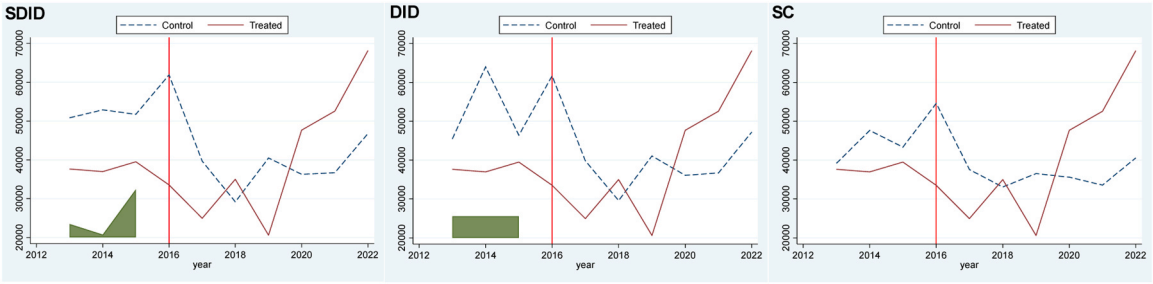


Fig. 10. Comparison of SDID, DID and SCM estimates for P5, entry year (cohort) 2016.

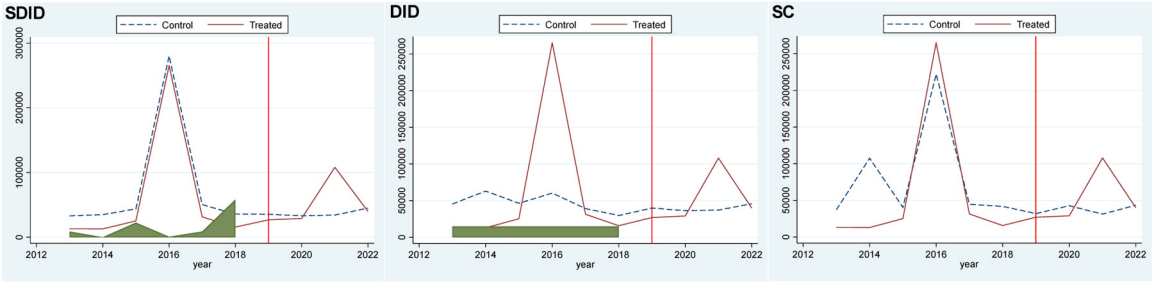


Fig. 11. Comparison of SDID, DID and SCM estimates for P6, entry year (cohort) 2019.

SDID estimation allows for a more efficient use of the available information within the panel.

6.3. Policy efficiency

Table 6 reports evidence on policy efficiency using the SCFP indicator illustrated in Section 3 (Eq. (6)). It is a sort of back-of-the-envelope calculation based on the average ATT estimates in Table 3 and on the average expenditure per unit of FAWI, SO or UAA as reported in the first three columns of Table 6. Calculation is performed on selected policies for which the private and societal ATT are statistically different from 0 over the whole period.

Direct comparisons of SCFP values across different policy aggregates are not always feasible, as they may reflect distinct societal objectives. Nevertheless, the indicator offers valuable insights. As discussed in Section 3.2, the inverse of SCFP can be interpreted as a measure of policy efficiency. In this context, we may draw an analogy with the classical macroeconomic output multiplier, which measures the additional Gross Domestic Product (GDP) generated by one unit of public expenditure.

Following this analogy, in the case of P3 it can be concluded that about 50€ (or more) of social cost per AWU, or unit of SO, is needed to induce an additional investment of 1€ per AWU, or unit of SO. This would indicate rather low efficiency that appears significantly improved in the case of P4 for which, using the same societal outcomes, we observe a much lower (less than 10€) social cost necessary to induce an additional € of OGA. For P5 and P6 this efficiency indicator actually refers to physical outcomes but also in these cases a low efficiency seems to emerge. More than 5 thousand € of social cost per AWU, or unit of UAA, is required to have a one-unit reduction of animal units, i.e. LSU (P5), or to have a one-unit increase of forest area, i.e. FA (P6). The low efficiency of P5 is confirmed also in the case of VE for which about 17.5€ per AWU, or unit of SO, is needed to induce a one-unit reduction of veterinary expenditure (VE).

The generalised low efficiency emerging from these results would seem to suggest substantial room for improving policy effectiveness. Nevertheless, these findings clearly warrant further confirmation in future studies.

Table 6
Social Cost of Farm Performance (SCFP) for selected societal ATT by policy.

	Avg. S_i on FAWU[a]	Avg. S_i on SO[a]	Avg. S_i on UAA[a]	Societal ATT (ΔG)[b] (1)	Societal ATT (ΔG)[b] (2)	Societal ATT (ΔG)[b] (3)
P3	38668.8	0.384	2135.8	47.58 [INV/AWU]	76.77 [INV/SO]	-
P4	46945.8	0.466	2593.0	1.697 [OGA/SO]	8.407[c] [OGA/AWU]	-
P5	10596.2	0.105	585.3	-6963[c] [LSU/UAA]	-7214 [LSU/AWU]	-17.52 [VE/SO]
P6	4628.6	0.046	255.7	5649 [FA/UAA]	-	-

[a] The average refers only to treated units.
[b] Societal outcome variable in square brackets.
[c] Not statistically significant ATT

6.4. Some validation exercises

As detailed in Section 5.2, the reliability of the SDID estimates presented above depends on several critical identification assumptions. Strictly speaking, violations of these assumptions cannot be directly tested. In some cases, as the overlap condition in the present study (see Table A4), some distributional evidence on the available dataset can provide support in this direction. Other violations, as the presence of spatial spillovers and of anticipation effects, require more sophisticated, and often unviable, approaches (Baldoni & Esposti, 2023; 2025). For other key assumptions, however, it is still possible to conduct validation exercises aimed at assessing the sensitivity of the results to potential violations. This section focuses on such exercises, particularly in relation to two assumptions that may be especially vulnerable in the present context: the Conditional Independence Assumption (CIA) and the Stable Unit Treatment Value Assumption (SUTVA).

Specifically, we explore the robustness of the estimated effects with respect to three potential sources of bias: Violation of the CIA due to the use of potentially endogenous lagged covariates; Violation of the CIA

Table 7
SDID estimates of the ATT for P5 with a sub-set of covariates (only time-invariant) covariates and with placebo outcome variables. All cohorts (standard errors in parenthesis).

Only time-invariant covariates:	
Private ATT ($\Delta\pi$) (.000€)	7.838 (3.250)**
Societal ATT (ΔG) (1)[a]	-0.039 (0.033)
Societal ATT (ΔG) (2)[b]	-1.573 (0.838)*
Societal ATT (ΔG) (3)[c]	-0.008 (0.003)**
SCFP (1)[a]	-7497[b]
SCFP (2)[b]	-6250
SCFP (3)[c]	-14.23
Placebo outcome variables:	
ULT (€/Working Hour)	-0.042 (0.056)
ACR (mm per year)	-23.6 (127.0)

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a] Societal outcome: LSU/UAA; [b]Societal outcome: LSU/AWU; [c]Societal outcome: VE/SO
[b] Not statistically significant ATT

due to unobserved farm or farmer characteristics that influence both treatment assignment and outcomes; Violation of the SUTVA due to the presence of multiple, possibly heterogeneous, versions of the treatment. These exercises aim to provide additional support for the credibility of the identification strategy adopted in this study.

As a first validation exercise, we re-estimate the SDID model by excluding lagged covariates, namely, using a subset of covariates that includes only time-invariant variables, as defined in Section 4.4. These variables can be considered deterministic and therefore exogenous by definition. The results of this estimation are reported in the upper part of Table 7. Due to space limitations, we focus on policy P5, which does not involve the aggregation of different measures and is therefore not subject to potential SUTVA violations.[26] It emerges that the estimated ATT over the whole period,[27] both private and societal, differ from those obtained using the full set of covariates, indicating that time-variant variables do play a role. However, this difference does not substantially alter the overall policy evaluation: the sign and statistical significance of the ATT are preserved, and the magnitude of the difference remains limited. Consequently, also the SCFP shows only minor deviations from the results obtained with the full covariate set, with differences consistently below 20 %. We can therefore conclude that the results are relatively robust to potential violations of the lag identification strategy involving time-variant covariates.

The second validation exercise addresses a potential violation of the Conditional Independence Assumption (CIA) that may occur independently of the presence of endogenous covariates, specifically, when selection on unobservables is at play. A commonly adopted strategy to test for the influence of unobserved confounders is the placebo outcome test (Eggers et al., 2024; Baldoni & Esposti, 2025). A placebo outcome is a variable that, while possibly correlated with observed covariates, should not be causally affected by the treatment under study. If treatment assignment is truly independent of unobserved factors, then it should have no effect on an outcome that is not causally linked to the treatment.[28] Therefore, a valid placebo outcome must be demonstrably independent of both the treatment and any observed or unobserved variables that influence the actual outcome of interest, whether private or societal. Suitable candidates for placebo outcomes are variables that depend primarily on the external environment, such as geographical, social, or political conditions, that may vary across farms and over time, but are not influenced by farmers' behavioural choices.

Two placebo outcome variables are used. Following Liu et al. (2024), we consider taxes, specifically, payroll taxes, defined as taxes and social security contributions paid on labour. To ensure that this variable is independent of farm size and labour intensity, we compute the Unit Labour Tax (ULT) by dividing the total amount of labour-related taxes by the total number of hours worked. As an additional robustness check, and following Baldoni & Esposti, (2023; 2025), we also consider a second placebo outcome variable: the Annual Cumulated Rainfall (ACR), which captures contextual climate and weather conditions. These environmental factors are unquestionably independent of policy adoption, as well as of farm characteristics and production decisions, although they are related to geographical location. The SDID estimation is replicated using these placebo outcomes, ULT and ACR, in place of the private and societal outcome variables, across all six policy measures under consideration, while maintaining the same dataset and identification strategy.

The results obtained using these placebo outcomes are reported in the lower part of Table 7. For the same reasons mentioned above, we focus here on policy P5.[29] The ATT estimates derived from both placebo variables are not statistically significant and are substantially lower in magnitude compared to those obtained using the actual private and societal outcome variables (see Table 4). These findings can be interpreted as indirect evidence supporting the validity of the CIA, suggesting that selection on unobservables is unlikely to be a major source of bias in this context.

The third validation exercise addresses a potential violation of the Stable Unit Treatment Value Assumption (SUTVA), arising from the presence of multiple versions of the treatment. When detailed information is available on specific treatment variants (e.g., individual RDP measures), the SDID approach can be stratified by measure, allowing for the estimation of separate ATT values. Significant differences across these estimates would indicate a violation of the assumption of functional equivalence among variants. Due to space limitations, this sensitivity analysis is conducted only for policy P4, which represents the simplest case of multiple treatments, comprising just two variants (Measures 6.4 and 16.9; see Table 1). Moreover, P4 is characterized by very sparse adoption, with few total adopters and even fewer early adopters (see Table 2).[30] As previously discussed, aggregation is often justified precisely to mitigate issues of sparsity. P4 therefore provides a meaningful case to examine the trade-off between reducing sparsity through aggregation and the risk of violating SUTVA.

The upper part of Table 8 shows the uneven distribution of adopters between the two P4 measures. The lower part of the table compares the ATT estimates for both private and societal outcomes under the two individual variants and for the aggregated policy. The point estimates appear relatively consistent across the two variants and the aggregate, but their statistical significance generally decreases when disaggregated. This suggests that, at least in this case, aggregating the two variants into a single policy does not raise major concerns regarding SUTVA violations, while it allows for improved statistical precision.

Moreover, as discussed in Section 5.2, although this evidence cannot entirely rule out violations of SUTVA, concerns are substantially mitigated when treatment variants are mutually exclusive and adopted independently. This supports the interpretation of the ATT for the aggregated policy as a weighted average of the ATTs for its components. Aggregation thus appears to preserve internal validity while enhancing statistical efficiency. Since the variants share the same private and societal objectives, the ATT estimated for the aggregate policy remains informative regarding the overall effectiveness and efficiency of the intervention.

[26] Results for all other policies (P1, P2, P3, P5 P6) are available upon request. They are substantially concordant with what obtained for P5.
[27] Results by cohorts are available upon request.
[28] Placebo tests could also be performed via a placebo treatment, namely a placebo treatment test assigns a fake or irrelevant treatment to assess whether the model incorrectly detects an effect.
[29] Results obtained for all other cases are available upon request.
[30] Results obtained for all other multiple-variant policy aggregates (P1, P2, P3, P6) are available upon request. Overall, they confirm what obtained for P4.

7. Lessons learnt and concluding remarks

This study offers several lessons, both methodological and policy-related, stemming from the application of the SDID method to the evaluation of actual policies characterized by limited granularity and sparse adoption.

The first key insight concerns the theoretical flexibility of SDID compared to traditional DID and SCM. While SDID is often praised for its ability to accommodate unit and time heterogeneity in policy adoption, our results suggest that this advantage may be overstated. Its effectiveness critically depends on the quantity and quality of pre-treatment information available.

Related to this, a second insight concerns to the potential of SDID to uncover TE heterogeneity across cohorts. In practice, cohort-specific estimates may lack robustness and offer limited policy relevance, especially in contexts, such as the farming sector, where outcome variability is naturally high. In such cases, these estimates may reflect noise rather than meaningful heterogeneity, making the average ATT computed across all cohorts more reliable indicator.

Third, despite its methodological innovations, SDID still relies on a set of identification assumptions. Apart from relaxing the PTA required by DID, SDID does not inherently resolve the challenges of causal inference. In non-experimental applications, constructing a valid quasi-experimental design remains difficult due to data limitations and the complexity of policy implementation. Therefore, it is essential to complement SDID estimates with a comprehensive set of validation exercises to assess robustness and sensitivity to potential violations of these assumptions.

Beyond methodological considerations, a central lesson of this study lies at the intersection of evaluation technique and policy analysis. Policy evaluation should not be confined to measuring effectiveness, i.e., the estimated ATT, but must also consider efficiency, which reflects the societal cost of achieving policy objectives. Assessing efficiency requires evaluating both private and public costs and benefits, thereby enabling a more comprehensive understanding of policy performance.

From this broader perspective, several policy lessons emerge. The results indicate that private benefits from policy adoption are evident across most Rural Development Programme (RDP) measures analysed. While variability across cohorts is observed, it appears to stem more from the inherent instability of outcome variables than from the differential impact of the policies themselves. This suggests that adoption is primarily driven by monetary incentives, and that the financial support provided is sufficient to motivate farmers. It can also be concluded that the observed sparsity in adoption likely reflects practical and administrative barriers rather than insufficient incentives.

At the same time, result indicate that making adoption easier and more attractive does not automatically translate into societal benefit. From a societal perspective, for several RDP measures the quality and quantity of farmers' responses upon adoption fall short of expectations. This gap between intended and actual outcomes, therefore low effectiveness, has been widely discussed in the literature (Esposti & Sotte, 2013; Sotte & Brunori, 2025).

To realign the private (i.e., farmers') and societal objectives of EU Rural Development Programme (RDP) measures, recent literature has emphasized the need to shift from action-based to result-based (or performance-based) policy design (Eichorn et al., 2024; Targetti et al., 2024). The rationale behind this shift is that action-based approaches often fail to account for the heterogeneity of farming conditions, leading to inconsistent outcomes despite similar actions.

One of the key findings of this study is that, although result-based approaches promise greater effectiveness and efficiency, they also introduce significant challenges. Even when policies are designed around outcomes, they still face the same conceptual and empirical challenges in evaluation, especially in constructing appropriate quasi-experimental strategies to identify and estimate treatment effects. The lesson here is that improving the design of RDP measures requires a stronger linkage between policy support and the societal outcomes of interest. Achieving this linkage demands a deeper understanding of farmers' decision-making processes and, consequently, better tailoring and targeting of measures to specific farmer characteristics and needs.

Among the RDP measures considered, those showing higher effectiveness tend to be associated with either very specific production orientations, such as animal production and related management practices (P5: Animal Production (or Livestock Management), or with clearly defined and verifiable structural investments at the farm level, regardless of production orientation (P3: Structural Agricultural Investments). However, the efficiency of policy expenditure appears to be more problematic. In particular, efficiency is notably low for the

Table 8
Different variants of P4: numerosity and ATT estimates. All cohorts (standard errors in parenthesis).

	Measure 6.4	Measure 16.9	P4 (total)
Number of adopters (unbalanced panel):			
2015	0	0	0
2016	0	0	0
2017	1	1	2
2018	3	1	4
2019	5	2	7
2020	11	3	14
2021	14	3	17
2022	16	3	19
Number of adopters (balanced panel):			
2015	0	0	0
2016	0	0	0
2017	1	0	1
2018	1	0	1
2019	1	1	2
2020	2	2	4
2021	2	2	4
2022	2	2	4
Private ATT ($\Delta\pi$) (.000€)	30.22 (17.9)	27.64 (20.3)	37.63 (12.3)**
Societal ATT (ΔG) (1)[a]	2.065 (1.98)	3.184 (1.696)*	2.757 (1.359)**
Societal ATT (ΔG) (2)[b]	9934 (7320)	4853 (6465)	5581 (4314)
SCFP (1)[a]	1.819	1.087	1.697
SCFP (2)[b]	3.774	7.152	8.407

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a] Societal outcome: OGA/SO;
[b] Societal outcome: OGA/AWU

aforementioned measures, as well as for those aimed at influencing the physical endowment of farms, such as reducing the number of animals (LSU) or increasing forest area (FA). Appreciable efficiency in terms of social cost emerges only for measures designed to encourage a shift in farm business towards other gainful activities (P4: Structural Non-Agricultural Investments).

The extent to which these results and the consequent lessons can go beyond the specific case of EU RDP and be generalized to all policies with similar problems in terms of granularity and adoption sparsity evidently deserves further empirical investigation. More generally, confirming these conclusions and, more importantly, refining them in order to inform policy decision making in a more concrete way, points to further research in the field. This may concentrate in repeating the analysis with better data and more sophisticated methodological approaches. Larger panel dataset, particularly along the time dimension allowing for longer pre and post-treatment observation periods, may reduce that apparent heterogeneity actually depending on the "natural" instability of outcomes in farming while making the actual heterogeneity in response to policy more clearly emerge. This may also be the direction to follow in terms of better methodologies. Flexibility brough about by the SDID approach could be combined by an explicit consideration of the progressive adaptation of farmers to policy adoption thus implying time-variant treatment effect (Goodman-Bacon, 2021; Sun & Abraham, 2021). Also accounting for strong specificity of few adopters can be improved, compared to SDID estimation, by more refined and flexible approaches on which current research is becoming a very active area (Chen et al., 2025).

CRediT authorship contribution statement

ESPOSTI Roberto: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

ANNEX 1

Table A1
List of measures of the EU RDP 2014–2020/2022 and correspondence with priorities and focus areas

Code	Title
Priority	
Pr1	Fostering knowledge transfer and innovation in agriculture, forestry and rural areas
Pr2	Enhancing farm viability and competitiveness of all types of agriculture in all regions and promoting innovative farm technologies and the sustainable management of forests
Pr3	Promoting food chain organisation, including processing and marketing of agricultural products, animal welfare and risk management in agriculture
Pr4	Restoring, preserving and enhancing ecosystems related to agriculture and forestry
Pr5	Promoting resource efficiency and supporting the shift towards a low carbon and climate resilient economy in agriculture, food and forestry sectors
Pr6	Promoting social inclusion, poverty reduction and economic development in rural areas
Focus area	
1 A	Fostering innovation, cooperation, and the development of the knowledge base in rural areas
1B	Strengthening the links between agriculture, food production and forestry and research and innovation, including for the purpose of improved environmental management and performance
1 C	Fostering lifelong learning and vocational training in the agricultural and forestry sectors
2 A	Improving the economic performance of all farms and facilitating farm restructuring and modernisation, notably with a view to increasing market participation and orientation as well as agricultural diversification
2B	Facilitating the entry of adequately skilled farmers into the agricultural sector and, in particular, generational renewal
3 A	Improving competitiveness of primary producers by better integrating them into the agri-food chain through quality schemes, adding value to agricultural products, promotion in local markets and short supply circuits […] inter-branch organisations
3B	Supporting farm risk prevention and management
4 A	Restoring, preserving and enhancing biodiversity, including in Natura 2000 areas, and in areas facing natural or other specific constraints and high nature value farming, as well as the state of European landscapes
4B	Improving water management, including fertiliser and pesticide management
4 C	Preventing soil erosion and improving soil management
5 A	Increasing efficiency in water use by agriculture
5B	Increasing efficiency in energy use in agriculture and food processing
5 C	Facilitating the supply and use of renewable sources of energy, of by products, wastes, residues and other non food raw material for the purposes of the bio-economy
5D	Reducing green house gas and ammonia emissions from agriculture
5E	Fostering carbon conservation and sequestration in agriculture and forestry
6 A	Facilitating diversification, creation and development of small enterprises, as well as job creation
6B	Fostering local development in rural areas
6 C	Enhancing the accessibility, use and quality of information and communication technologies (ICT) in rural areas
Measures	
M01	M01 - Knowledge transfer and information actions (art 14)
M02	M02 - Advisory services, farm management and farm relief services (art 15)
M03	M03 - Quality schemes for agricultural products and foodstuffs (art 16)
M04	M04 - Investments in physical assets (art 17)
M05	M05 - Restoring agricultural production potential damaged by natural disasters and catastrophic events and introduction of appropriate prevention actions (art 18)
M06	M06 - Farm and business development (art 19)
M07	M07 - Basic services and village renewal in rural areas (art 20)
M08	M08 - Investments in forest area development and improvement of the viability of forests (art 21–26)
M09	M09 - Setting-up of producer groups and organisations (art 27)
M10	M10 - Agri-environment-climate (art 28)
M11	M11 - Organic farming (art 29)
M12	M12 - Natura 2000 and Water Framework Directive payments (art 30)
M13	M13 - Payments to areas facing natural or other specific constraints (art 31)
M14	M14 - Animal welfare (art 33)

(*continued on next page*)

Table A1 (*continued*)

Code	Title
M15	M15 - Forest environmental and climate services and forest conservation (art 34)
M16	M16 - Co-operation (art 35)
M17	M17 - Risk management (art 36)
M18	M18 - Financing of complementary national direct payments for Croatia (art 40)
M19	M19 - Support for LEADER local development (CLLD – community-led local development) (art 35 Regulation (EU) No 1303/2013)
M20	M20 - Technical assistance Member States (art 51–54)
M113	M113 - Early retirement
M131	M131 - Meeting standards based on Community legislation
M341	M341 - Skills acquisition, animation and implementation
Sub-measures	
M01.1	1.1 - support for vocational training and skills acquisition actions
M01.2	1.2 - support for demonstration activities and information actions
M01.3	1.3 - support for short-term farm and forest management exchange as well as farm and forest visits
M02.1	2.1 - support to help benefiting from the use of advisory services
M02.2	2.2 - support for the setting up of farm management, farm relief and farm advisory services as well as forestry advisory services
M02.3	2.3 - support for training of advisors
M03.1	3.1 - support for new participation in quality schemes
M03.2	3.2 - Support for information and promotion activities implemented by groups of producers in the internal market
M04.1	4.1 - support for investments in agricultural holdings
M04.2	4.2 - support for investments in processing/marketing and/or development of agricultural products
M04.3	4.3 - support for investments in infrastructure related to development, modernisation or adaptation of agriculture and forestry
M04.4	4.4 - support for non-productive investments linked to the achievement of agri-environment-climate objectives
M05.1	5.1 - support for investments in preventive actions aimed at reducing the consequences of probable natural disasters, adverse climatic events and catastrophic events
M05.2	5.2 - support for investments for the restoration of agricultural land and production potential damaged by natural disasters, adverse climatic events and catastrophic events
M06.1	6.1 - business start up aid for young farmers
M06.2	6.2 - business start up aid for non-agricultural activities in rural areas
M06.3	6.3 - business start up aid for the development of small farms
M06.4	6.4 - support for investments in creation and development of non-agricultural activities
M06.5	6.5 - payments for farmers eligible for the small farmers scheme who permanently transfer their holding to another farmer
M07.1	7.1 - support for drawing up and updating of plans for the development of municipalities and villages and of protection and management plans relating to N2000/HNV areas
M07.2	7.2 - support for investments in the creation/improvement of all types of small scale infrastructure, including investments in renewable energy and energy saving
M07.3	7.3 - support for broadband infrastructure and provision of access to broadband and public e-government
M07.4	7.4 - support for investments in the setting-up, improvement or expansion of local basic services for the rural population
M07.5	7.5 - support for investments for public use in recreational infrastructure, tourist information and small scale tourism infrastructure
M07.6	7.6 - support for studies/investments associated with the maintenance, restoration and upgrading of the cultural and natural heritage of villages, landscapes and HNV sites
M07.7	7.7 - support for investments targeting the relocation of activities and conversion of buildings, to improve the quality of life or increasing their environmental performance
M07.8	7.8 – others
M08.1	8.1 - support for afforestation/creation of woodland
M08.2	8.2 - support for establishment and maintenance of agro-forestry systems
M08.3	8.3 - support for prevention of damage to forests from forest fires and natural disasters and catastrophic events
M08.4	8.4 - support for restoration of damage to forests from forest fires and natural disasters and catastrophic events
M08.5	8.5 - support for investments improving the resilience and environmental value of forest ecosystems
M08.6	8.6 - support for investments in forestry technologies and in processing, mobilising and marketing of forest products
M09.1	9.1 - setting up of producer groups and organisations in the agriculture and forestry sectors
M10.1	10.1 - payment for agri-environment-climate commitments
M10.2	10.2 - support for conservation and sustainable use and development of genetic resources in agriculture
M11.1	11.1 - payment to convert to organic farming practices and methods
M11.2	11.2 - payment to maintain organic farming practices and methods
M12.1	12.1 - compensation payment for Natura 2000 agricultural areas
M12.2	12.2 - compensation payment for Natura 2000 forest areas
M12.3	12.3 - compensation payment for agricultural areas included in river basin management plans
M13.1	13.1 - compensation payment in mountain areas
M13.2	13.2 - compensation payment for other areas facing significant natural constraints
M13.3	13.3 - compensation payment to other areas affected by specific constraints
M14.1	14.1 - payment for animal welfare
M15.1	15.1 - payment for forest -environmental and climate commitments
M15.2	15.2 - support for the conservation and promotion of forest genetic resources
M16.0	16.0 – others
M16.1	16.1 - support for the establishment and operation of operational groups of the EIP for agricultural productivity and sustainability
M16.2	16.2 - support for pilot projects, and for the development of new products, practices, processes and technologies
M16.3	16.3 - co-operation among smalls operators in organising joint work processes and sharing facilities and resources, and for developing/marketing tourism
M16.4	16.4 - support for co-operation among supply chain actors for the establishment and development of short supply chains and local markets, and for promotion activities
M16.5	16.5 - support for joint action undertaken to mitigate or adapt to climate change, and for joint approaches to environmental projects and practices
M16.6	16.6 - support for cooperation among supply chain actors for sustainable provision of biomass for use in food and energy production and industrial processes
M16.7	16.7 - support for non-CLLD local development strategies
M16.8	16.8 - support for drawing up of forest management plans or equivalent instruments
M16.9	16.9 - support for diversification of farming activities into activities concerning health care, social integration, community-supported agriculture and education about environment/food
M17.1	17.1 - Crop, animal and plant insurance premium
M17.2	17.2 - Mutual funds for adverse climatic events, animal and plant diseases, pest infestations and environmental incidents

(*continued on next page*)

Table A1 (*continued*)

Code	Title
M17.3	17.3 - Income stabilisation tool
M18.1	18 - Financing of complementary national direct payments for Croatia
M19.1	19.1 - Preparatory support
M19.2	19.2 - Support for implementation of operations under the community-led local development strategy
M19.3	19.3 - Preparation and implementation of cooperation activities of the local action
M19.4	19.4 - Support for running costs and animation
M20.1	20.1 - support for technical assistance (other than NRN)
M20.2	20.2 - support for establishing and operating the NRN

Source: European Commission

ANNEX 2

Table A2
Different sets of CMEF indicators

Code	Description
Context Indicators (CI)	
C1	Population
C2	Age Structure
C3	Territory
C4	Population Density
C5	Employment Rate
C6	Self-employment rate
C7	Unemployment rate
C8	GDP per capita
C9	Poverty rate
C10	Structure of the economy (GVA)
C11	Structure of Employment
C12	Labour productivity by economic sector
C13	Employment by economic activity
C14	Labour productivity in agriculture
C15	Labour productivity in forestry
C16	Labour productivity in the food industry
C17	Agricultural holdings (farms)
C18	Agricultural Area
C19	Agricultural area under organic Farming
C20	Irrigated Land
C21	Livestock units
C22	Farm labour force
C23	Age structure of farm managers
C24	Agricultural training of farm managers
C25	Agricultural factor income
C26	Agricultural Entrepreneurial Income
C27	Total factor productivity in agriculture
C28	Gross fixed capital formation in agriculture
C29	Forest and other wooded land (FOWL) (000)
C30	Tourism infrastructure
C31	Land Cover
C32	Areas with Natural Constraints
C33	Farming intensity
C34	Natura 2000 areas
C35	Farmland Birds index (FBI)
C36	Conservation status of agricultural habitats (grassland)
C37	HNV Farming
C38	Protected Forest
C39	Water Abstraction in Agriculture
C40	Water Quality
C41	Soil organic matter in arable land
C42	Soil Erosion by water
C43	Production of renewable Energy from agriculture and forestry
C44	Energy use in agriculture, forestry and food industry
C45	GHG emissions from agriculture
Output Indicators (TI)	
O1	Total public expenditure
O2	Total investment
O3	Number of actions/operations supported
O4	Number of holdings/beneficiaries supported
O5	Total area (ha)
O6	Physical area supported (ha)
O7	Number of contracts supported
O8	Number of Livestock Units supported (LU)
O9	Number of holdings participating in supported schemes

(*continued on next page*)

Table A2 (*continued*)

Code	Description
O10	Number of farmer benefiting from pay-outs
O11	Number of training days given
O12	Number of participants in trainings
O13	Number of beneficiaries advised
O14	Number of advisor trained
O15	Population benefiting of improved services/infrastructures
O16	Number of EIP groups and operations supported and number/type of partners in EIP groups
O17	Number of cooperation operations supported (other than EIP)
O18	Population covered by LAG
O19	Number of LAGs selected
O20	Number of LEADER projects supported
O21	Number of cooperation project supported
O22	Number and type of project promoters
O23	Unique number of LAG involved in cooperation project
O24	Number of thematic and analytical exchanges set up with the support of NRN
O25	Number of NRN communication tools
O26	Number of ENRD activities in which the NRN has participated
Result Indicators (TI)	
R1	Percentage of agricultural holdings with RDP support for investments in restructuring or modernisation (focus area 2 A)
R2	Change in Agricultural output on supported farms/AWU (Annual Work Unit) (focus area 2 A)*
R3	Percentage of agricultural holdings with RDP supported business development plan/investments for young farmers (focus area 2B)
R4	Percentage of agricultural holdings receiving support for participating in quality schemes, local markets and short supply circuits, and producer groups/organisations (focus area 3 A)
R5	Percentage of farms participating in risk management schemes (focus area 3B)
R6	Percentage of forest or other wooded areas under management contracts supporting biodiversity (focus area 4 A)
R7	Percentage of agricultural land under management contracts supporting biodiversity and/or landscapes (focus area 4 A)
R8	Percentage of agricultural land under management contracts to improve water management (focus area 4B)
R9	Percentage of forestry land under management contracts to improve water management (focus area 4B)
R10	Percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion (focus area 4 C)
R11	Percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion (focus area 4 C)
R12	Percentage of irrigated land switching to more efficient irrigation systems (focus area 5 A)
R13	Increase in efficiency of water use in agriculture in RDP supported projects (focus area 5 A)*
R14	Increase in efficiency of energy use in agriculture and food-processing in RDP supported projects (focus area 5B)*
R15	Renewable energy produced from supported projects (focus area 5 C)*
R16	Percentage of LU (Live-stock Unit) concerned by investments in live-stock management in view of reducing GHG (Green House Gas) and/or ammonia emissions (focus area 5D)
R17	Percentage of agricultural land under management contracts targeting reduction of GHG and/or ammonia emissions (focus area 5D)
R18	Reduced emissions of methane and nitrous oxide (focus area 5D)*
R19	Reduced ammonia emissions (focus area 5D)*
R20	Percentage of agricultural and forest land under management contracts contributing to carbon sequestration or conservation (focus area 5E)
R21	Jobs created in supported projects (focus area 6 A)
R22	Percentage of rural population covered by local development strategies (focus area 6B)
R23	Percentage of rural population benefiting from improved services / infrastructures (focus area 6B)
R24	Jobs created in supported projects (Leader) (focus area 6B)
R25	Percentage of rural population benefiting from new or improved services / infrastructures (Information and Communication Technology - ICT) (focus area 6 C)
Target Indicators (TI)	
T1	Percentage of expenditure under Articles 14, 15 and 35 of Regulation (EU) No 1305/2013 in relation to the total expenditure for the RDP (focus area 1 A)
T2	Total number of cooperation operations supported under the cooperation measure (Article 35 of Regulation (EU) No 1305/2013) (groups, networks/clusters, pilot projects…) (focus area 1B)
T3	Total number of participants trained under Article 14 of Regulation (EU) No 1305/2013 (focus area 1 C)
T4	Percentage of agricultural holdings with RDP support for investments in restructuring or modernisation (focus area 2 A)
T5	Percentage of agricultural holdings with RDP supported business development plan/investments for young farmers (focus area 2B)
T6	Percentage of agricultural holdings receiving support for participating in quality schemes, local markets and short supply circuits, and producer groups/organisations (focus area 3 A)
T7	Percentage of farms participating in risk management schemes (focus area 3B)
T8	Percentage of forest/other wooded area under management contracts supporting biodiversity (focus area 4 A)
T9	Percentage of agricultural land under management contracts supporting biodiversity and/or landscapes (focus area 4 A)
T10	Percentage of agricultural land under management contracts to improve water management (focus area 4B)
T11	Percentage of forestry land under management contracts to improve water management (focus area 4B)
T12	Percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion (focus area 4 C)
T16	Total investment in renewable energy production (focus area 5 C)
T13	Percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion (focus area 4 C)
T14	Percentage of irrigated land switching to more efficient irrigation system (focus area 5 A)
T15	Total investment for energy efficiency (focus area 5B)
T17	Percentage of LU concerned by investments in live-stock management in view of reducing GHG and/or ammonia emissions (focus area 5D)
T18	Percentage of agricultural land under management contracts targeting reduction of GHG and/or ammonia emissions (focus area 5D)
T19	Percentage of agricultural and forest land under management contracts contributing to carbon sequestration and conservation (focus area 5E)
T20	Jobs created in supported projects (focus area 6 A)
T21	Percentage of rural population covered by local development strategies (focus area 6B)
T22	Percentage of rural population benefiting from improved services/infrastructures (focus area 6B)
T23	Jobs created in supported projects (Leader) (focus area 6B)
T24	Percentage of rural population benefiting from new or improved services/infrastructures (ICT) (focus area 6 C)
Impact Indicators (II)	
I01	Agricultural entrepreneurial income
I02	Agricultural factor income
I03	Total factor productivity in agriculture

(*continued on next page*)

Table A2 (*continued*)

Code	Description
I04	EU commodity price variability
I05	Consumer price evolution of food products
I06	Agricultural trade balance
I07	Emissions from agriculture
I08	Farmland bird index
I09	High nature value (HNV) farming
I10	Water abstraction in agriculture
I11	Water quality
I12	Soil organic carbon in arable land
I13	Soil erosion by water
I14	Rural employment rate
I15	Degree of rural poverty
I16	Rural GDP per capita

Source: European Commission

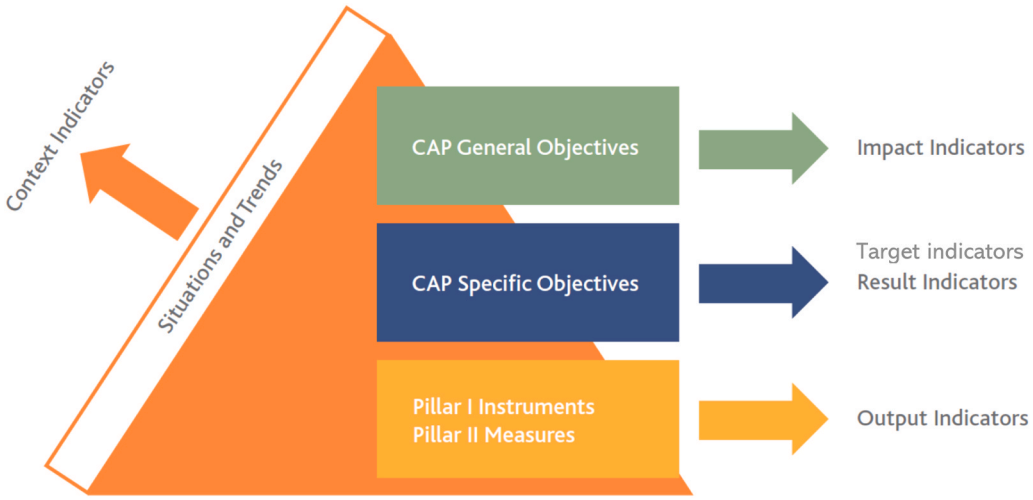


Figure A1. Articulation of the different sets of CMEF indicators with reference to the CAP objectives
Source: Adaptation from European Commission (2019).

ANNEX 3

Table A3
SDID Estimates of confounders' coefficients (**β** in Eqs. (7) and (8)) for both outcome variables under policy P4 (and the two respective cohorts)

Covariates (**X**)	Cohort 2017		Cohort 2020	
	Private ATT ($\Delta\pi$) (.000€)	Societal ATT (ΔG)[a]	Private ATT ($\Delta\pi$) (.000€)	Societal ATT (ΔG)[a]
SIZE1	-3733.4	1299.7	-4650.9	-7755.2
SIZE2	-3962.5	-2313.8	-7455.2	-333.4
SIZE3	2024.7	695.8	-14734.5	-41211.4
SIZE4	4846.3	1607.6	-12533.3	-5332.8
SIZE5	1829.1	-1792.6	-38987.1	-406.1
SIZE6	24587.8	36834.1	8718.8	5058.4
PRO1	743.1	581.2	-14008.7	5422.1
PRO2	-787.7	-229.3	5281.4	4785.1
PRO3	-1086.8	-211.1	2544.7	7447.6
PRO4	1431.2	61.2	6582.5	5959.5
FOR/SO	286.2	-414.7	11574.6	-787.1
KW/SO	556.4	10219.1	-72794.9	-4703.5
UAA/SO	2699.7	1132.6	-1783.0	-1027.7
LSU/SO	7776.3	17650.5	584991.6	13746.8
AWU/SO	12845.0	14644.3	21047.0	4261.0
ENE/SO	1200.3	1911.3	276.4	131.8
MAT/SO	1364.9	1569.4	-3323.8	18535.6
AGE	-351.3	-660.1	-1479.2	6466.9
GEN	-4558.8	109.3	-9744.7	-4369.7
ALT	-715.8	23.69	2104.5	-58.48
LAT	-3.645	4.355	4.120	3.058
LON	-0.186	-0.174	0.487	-0.289
PIL1/SO	4762.41	1904.3	1402.2	8849.1

[a]Societal outcome: OGA/AWU

ANNEX 4

Table A4
Average yearly variation and Welch's *t*-test p-values comparing treated and control units across all model covariates. Whole period

	Avg. yearly variation rate (%)[a]	Welch's *t*-test p-values[b]					
		P1	P2	P3	P4	P5[d]	P6
SIZE1	0.00	0.910	0.113	0.937	0.117	0.323	0.488
SIZE2	0.00	0.686	0.383	0.419	0.439	0.327	0.973
SIZE3	-1.50	0.857	0.162	0.350	0.360	0.648	0.511
SIZE4	1.70	0.379	0.641	0.829	0.724	0.671	0.695
SIZE5	0.00	0.613	0.729	0.913	0.939	0.420	0.609
SIZE6	0.00	0.305	0.294	0.140	0.955	0.830	0.347
PRO1	0.00	0.343	0.266	0.491	0.448	0.570	0.256
PRO2	0.00	0.334	0.219	0.580	0.233	0.430	0.534
PRO3	0.00	0.222	0.321	0.174	0.492	0.300	0.733
PRO4	0.00	0.448	0.135	0.468	0.609	0.850	0.860
FOR/SO	0.08	0.446	0.365	0.518	0.426	0.836	0.615
KW/SO	-2.30	0.662	0.792	0.978	0.717	0.326	0.703
UAA/SO	-1.26	0.135	0.051*	0.156	0.098*	0.164	0.172
LSU/SO	0.95	0.148	0.071*	0.168	0.220	0.113	0.063*
AWU/SO	-1.62	0.053*	0.114	0.322	0.576	0.693	0.112
ENE/SO	3.07	0.719	0.528	0.260	0.538	0.613	0.365
MAT/SO	-2.21	0.098*	0.183	0.113	0.448	0.504	0.563
AGE	1.79	0.589	0.086*	0.248	0.299	0.366	0.166
GEN	0.00	0.490	0.516	0.332	0.574	0.739	0.488
ALT	0.00	0.357	0.657	0.728	0.834	0.522	0.511
LAT	0.00	0.517	0.195	0.124	0.187	0.261	0.288
LON	0.00	0.795	0.156	0.148	0.184	0.232	0.493
PIL1/SO	1.21	0.696	0.220	0.489	0.472	0.464	0.753

*,**: Statistically significant at 10 % and 5 % level, respectively.
[a]Average variation rates refer to period 2013–2022. For dummy variables it expresses the ratio between the changes in the status (from 0 to 1 and viceversa) on the total number of observations.
[b]In each policy (P), the test is performed on the whole period reported in Table 1.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature, 59*(2), 391–425. ⟨https://www.aeaweb.org/articles?id=10.1257/jel.20191450⟩.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science, 59*(2), 495–510.

Alkin, M. C. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage Publication.

Andersson, A., Höjgård, S., & Rabinowicz, E. (2017). Evaluation of results and adaptation of EU rural development programmes. *Land Use Policy, 67*, 298–314.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review, 111*(12), 4088–4118.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3–32.

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics, 226*(1), 62–79.

Baldoni, E., Coderoni, S., & Esposti, R. (2021). Immigrant workforce and agriculture productivity: Evidence from Italian farm-level data. *European Review of Agricultural Economics, 48*(4), 805–834.

Baldoni, E., & Esposti, R. (2023). Estimating the impact of policies under spatial interference. *The case of CAP support to organic farming. Quaderno di Ricerca n. 475, Dipartimento di Scienze Economiche e Sociali*. Università Politecnica delle Marche.

Baldoni, E., & Esposti, R. (2025). The impact of environmental policies on adopters under general interference. *The case of EU support to organic farming. Quaderno di Ricerca n. 497, Dipartimento di Scienze Economiche e Sociali*. Università Politecnica delle Marche.

Bellemare, M. F., Masaki, T., & Pepinsky, T. B. (2017). Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics, 79*(3), 949–963.

Brown, C., Kovács, E., Herzon, I., Villamayor-Tomas, S., Albizua, A., Galanaki, A., Grammatikopoulou, I., McCracken, D., Alkan Olsson, J., & Zinngrebe, Y. (2021). Simplistic understandings of farmer motivations could undermine the environmental potential of the common agricultural policy. *Land Use Policy, 101*, Article 105136.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2024). *Difference-in-differences with a Continuous Treatment*. Working Paper No. 32117. Cambridge (MA): NBER.

Camaioni, B., Esposti, R., Pagliacci, F., & Sotte, F. (2016). How does space affect the allocation of the EU rural development policy expenditure? A spatial econometric assessment. *European Review of Agricultural Economics, 43*(3), 433–473.

Camaioni, B., & Sotte, F. (2010). Un primo bilancio della politica di sviluppo rurale in Europa. *Agriregionieuropa, 6*(20), 41–45.

Castaño, J., Blanco, M., & Martinez, P. (2019). Reviewing counterfactual analyses to assess impacts of EU rural development programmes: What lessons can be learned from the 2007–2013 ex-post evaluations. *Sustainability, 11*(4), 1105. https://doi.org/10.3390/su11041105

Cerulli, G. (2015). *Econometric Evaluation of Socio-Economic Programs: Theory and Applications*. Berlin: Springer.

Chen, Q., Xiao, Z., & Yao, Q. (2025). Quantile control via random forest. *Journal of Econometrics, 249*(Part A), Article 105789.

Chernozhukov, V., Fernández-Val, I., Han, S., & Wüthrich, K. (2024). Estimating causal effects of discrete and continuous treatments with binary instruments. *arXiv Preprint arXiv, 2403*, 05850. ⟨https://arxiv.org/abs/2403.05850⟩.

Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation, 34*(3), 131–135.

Clarke, D., Pailañir, D., Athey, S., & Imbens, G. (2024). On synthetic difference-in-differences and related estimation methods in stata. *The Stata Journal, 2*(4), 557–598.

Coca, O., Creangă, D., Viziteu, S., Brumă, I. S., & Ştefan, G. (2023). Analysis of the determinants of agriculture performance at the European Union Level. *Agriculture, 13*(3), 616.

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

Coderoni, S., Esposti, R., & Varacca, A. (2024). How differently do farms respond to agri-environmental policies? A probabilistic machine-learning approach. *Land Economics, 100*(2), 370–397. https://doi.org/10.3368/le.100.2.060622-0043R1

Coderoni, S., Esposti, R., Varacca, A. (2025). Evaluating the Dual Impact of Organic Farming Support: Private Response and Societal Outcome for GHG Emission Abatement. A Causal Machine Learning assessment. Unpublished manuscript, Department of Economics and Social Sciences, Università Politecnica delle Marche.

DeBoe, G. (2020). Impacts of agricultural policies on productivity and sustainability performance in agriculture: A literature review. OECD Food. *Agriculture and Fisheries Papers, No. 141*. Paris: OECD Publishing.

Eggers, A. C., Tuñón, G., & Dafoe, A. (2024). Placebo Tests for Causal Inference. *American Journal of Political Science, 68*(3), 1106–1121.

Eichorn, T., Schaller, L., Hamunen, K., & Runge, T. (2024). Exploring acro-environmental factors influencing adoption of result-based and collective agri-environmental measures: a PESTLE approach based on stakeholder statements. *Bio-Based and Applied Economics, 13*(1), 47–91.

Espinosa-Goded, M., Barreiro-Hurlé, J., & Ruto, E. (2010). What do farmers want from agri-environmental scheme design? A choice experiment approach. *Journal of Agricultural Economics, 61*(2), 259–273.

Esposti, R. (2017a). The Empirics of Decoupling: Alternative Estimation Approaches of the Farm-level Production Response. *European Review of Agricultural Economics, 44*(3), 499–537.

Esposti, R. (2017b). The heterogeneous farm-level impact of the 2005 CAP-first pillar reform: a multivalued treatment effect estimation. *Agricultural Economics, 48*(3), 373–386.
Esposti, R. (2022). The Coevolution of Policy Support and Farmers Behaviour and Performance. An investigation on Italian agriculture over the 2008-2019 period. *Bio-Based and Applied Economics, 11*(3), 231–264.
Esposti, R. (2024). Non-Monetary Motivations of Agri-Environmental Policy Adoption. A Causal Forest Approach. *Journal of Environmental Management, 352*, Article 119992. https://doi.org/10.1016/j.jenvman.2023.119992
Esposti, R., & Sotte, F. (2013). Evaluating the Effectiveness of Agricultural and Rural Policies: An Introduction. *European Review of Agricultural Economics, 40*(4), 535–539.
European Commission (2012). Strategic Programming and Monitoring and Evaluation for RDPs 2014-2020. Joint ENRD Coordination Committee and Evaluation Expert Committee workshops, Newsletter 14th and 15th March, Brussels.
European Commission. (2014). *Getting the most from your RDP: Guidelines for the ex-ante Evaluation of 2014-2020 RDPs*. Brussels: European Evaluation Network for Rural Development.
European Commission. (2015a). *Establishing and implementing the Evaluation Plan of 2014-2020 RDPs*. Brussels: Guidelines. Directorate-General for Agriculture and Rural Development – Unit E.4.
European Commission. (2015b). *Common Evaluation Questions for Rural Development Programmes 2014-2020* (Working Paper). Brussels: Directorate-General for Agriculture and Rural Development – Unit E.4.
European Commission. (2016). *Guidelines: Assessment of RDP results: how to prepare for reporting on evaluation in 2017*. Brussels: Directorate-General for Agriculture and Rural Development – Unit E.4.
European Commission. (2018). Guidelines. Assessing RDP achievements and impacts in 2019. Directorate-General for Agriculture and Rural. *Development – Unit C*. Brussels: 4 (2018).
European Commission (2019). Interactive CAP Indicator Dashboards: transparency and growth for a more sustainable tomorrow's. Directorate-General for Agriculture and Rural Development, European Evaluation Helpdesk for Rural Development, Rural Evaluation NEWS, Number 12, Brussels.
European Commission (2021). *Evaluation of the CAP measures applied to the environment*. Final Report.
European Court of Auditors (2013). Can the Commission and the Member States Show that EU Budget Allocated to the Rural Development Policy Is Well Spent? Publications Office of the European Union: Luxembourg, Available online: ⟨https://publications.europa.eu/en/publication-detail/-/publication/5f439781-0432-4473-8e91-c1697bd475f7/language-en⟩ (accessed on 1 August 2025).
Eurostat. (2024). *Agricultural real factor income per AWU*. Luxembourg: Eurostat. ⟨https://ec.europa.eu/eurostat⟩.
Falconer, K., Dupraz, P., & Whitby, M. (2001). An investigation of policy administrative costs using panel data for the EU. *Journal of Agricultural Economics, 52*(1), 83–103.
FAO. (2024). *Impact evaluation of agricultural interventions – A technical guide*. Rome: FAO.
Figurek, A., Semenova, E. I., Rocha, J. M. F., Thrassou, A., & Uzunboylu, N. (2024). Economic indicators in cereal production in the EU. In A. Galati, D. Vrontis, A. Thrassou, & M. Fiore (Eds.), *Agribusiness innovation and contextual evolution: Strategic, managerial and marketing advancements, II*. London: Palgrave Macmillan.
Fougère, D., & Jacquemet, N. (2021). Policy Evaluation Using Causal Inference Methods. In N. Hashimzade, & M. Thornton (Eds.), *Handbook of Research Methods and Applications in Empirical Microeconomics* (pp. 294–324). Cheltenham (UK): Edward Elgar Publishing.
Garcia-Cornejo, B., Perez-Mendez, J. A., Roibas, D., & Wall, A. (2020). Efficiency and Sustainability in Farm Diversification Initiatives in Northern Spain. *Sustainability, 12*, 3983.
Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277.
Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge, UK: Cambridge University Press.
[11.1] Lechner, M., & Miquel, R. (2010). Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions. *Empirical Economics, 39*(1), 111–137.
Lemire, S., Peck, L. R., & Porowski, A. (2020). The Growth of the Evaluation Tree in the Policy Analysis Forest: Recent Developments in Evaluation. *Policy Studies Journal, 48*(S1), S47–S70.
Liu, J.-Y., Lee, B., & Chang, H.-H. (2024). Place-based rural development programs and the labour allocation of farm households. *China Agricultural Economic Review, 16*(4), 683–711.
Lu, J. (2021). Synthetic Control Method, Stationarity and Pointwise Statistical Inference. ⟨http://www.uone-tech.cn/ChinaStataConference2021/pdf/lujiaxuan.pdf⟩.
Mennig, P. (2024). A never-ending story of reforms: on the wicked nature of the Common Agricultural Policy. *npj Sustainable Agriculture, 2*, 20. https://doi.org/10.1038/s44264-024-00027-z
OECD. (2019). *Evaluating the Impact of Agricultural Policies*. OECD Publishing.
OECD. (2024). *Agricultural Policy Monitoring and Evaluation 2024: Innovation for Sustainable Productivity Growth*. Paris: OECD Publishing.
OECD. (2025). *Agri-Environmental Indicators – 2025 Update*. Paris: OECD Publishing.
Popescu, A., Tindeche, C., Mărcuță, A., Mărcuță, L., & Honțuș, A. (2021). Labor productivity in Romania's agriculture in the period 2011–2020 and its forecast for 2021–2025 horizon. *Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development, 21*(3), 673–678.
Prigoreanu, I., Radu, G., Grigore Sava, A., Costuleanu, C. L., Ungureanu, G., & Ignat, G. (2025). Assessing the economic sustainability of the EU and Romanian farming sectors. *Sustainability, 17*(10), 4440.
Schnake-Mahl, A., Diez Roux, A. V., Bilal, U., Schwartz, G. L., & Burris, S. (2025). Rigorous policy measurement: causal inference challenges and opportunities. *American Journal of Epidemiology*. kwae468.
Sotte F. (a cura di). (2009). La politica di sviluppo rurale 2007-2013. Un primo bilancio per l'Italia, Quaderni del Gruppo 2013- Edizioni Tellus, Roma.
Sotte, F., (2012). La Politica di Sviluppo Rurale. In: De Filippis, F. (a cura di). La nuova PAC 2014-2020. Un'analisi delle proposte della Commissione. Edizioni Tellus, Roma: 89-105.
Sotte, F., & Brunori, G. (2025). *European Agricultural Policy: History and Analysis*. Switzerland, Basel: Springer Nature.
Stetter, C., Mennig, P., & Sauer, J. (2022). Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study. *European Review of Agricultural Economics, 49*(4), 723–739.
Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199.
Swales, V. (2007). The Role of Pillar 2 in Delivering Environmental Outcomes. A paper prepared by for the LUPG/BfN conference 'Future Policies for Rural Europe 2013 and beyond', Brussels, 19 – 20 September.
Targetti, S., Niedermayr, A., Hafner, K., & Schaller, L. (2024). New pathways for improved delivery of public goods from agriculture and forestry. *Bio-Based and Applied Economics, 13*(1), 3–11.
UCL Institute for Innovation and Public Purpose. (2020). Alternative policy evaluation frameworks and tools: Exploratory study. BEIS Research Paper Number 2020/044. *Department for Business, Energy Industrial Strategy*. ⟨https://www.gov.uk/government/publications/alternative-policy-evaluation-frameworks-and-tools-exploratory-study⟩.
VanderWeele, T. J., & Hernán, M. A. (2013). Causal Inference Under Multiple Versions of Treatment. *Journal of Causal Inference, 1*(1), 1–20.
Wageningen Economic Research. (2024). *Farm income reports*. Wageningen (Netherlands): Wageningen Economic Research. ⟨https://www.wur.nl/en/research-results/farm-income⟩.
Wooldridge, J. M. (2023). Simple Approaches to Nonlinear Difference-in-Differences with Panel Data. *The Econometrics Journal, 26*(3), C31–C66.

Roberto Esposti is a full professor of "Economic Policy" at the Department of Economics and Social Sciences at the Università Politecnica delle Marche. He also teaches courses of "Environmental Economics", "Ecological Economics" and "European Agricultural Policy", as well as of "Panel Data Econometrics" in the PhD program in Economics. His main areas of research include policy evaluation in the field of environmental and agricultural policies, technological progress and innovation, structural change and local development.