

Biodiversity and the design of result-based payments: Evidence from Germany

Sergei Schaub^{a,*}, Alexander Pfaff^b, Petyo Bonev^{a,c}

^a Managerial Economics, Agroscope, Ettenhausen, Switzerland

^b Sanford School of Public Policy, Duke University, Durham, NC, USA

^c Swiss Institute for Empirical Economic Research, University of St. Gallen, St. Gallen, Switzerland

ARTICLE INFO

JEL classification:

Q15
Q18
Q57

Keywords:

Agricultural policy
Policy design
Agri-environmental payments
Result-based payments
Biodiversity
Birds

ABSTRACT

Paying farmers for measured outcomes—i.e., results, not actions—is promoted for raising the effectiveness and flexibility of efforts to address agriculture's environmental damages. One key design choice is how exactly to reward these measured results. Continuous rewards are possible, yet, in practice, observed species outcomes have been rewarded using a single threshold (compliant/not compliant) or, to move toward continuity, a few thresholds (e.g., low/medium/high). We assess whether more continuous rewards—specifically, multiple target thresholds for plant species—raise bird diversity. We study a pilot scheme in Germany's federal state of Lower Saxony, where an incentive with one threshold is the baseline. Using citizen-science bird data (offering over 6.7m entries across 16 years) and staggered difference-in-differences estimation, we find that the pilot scheme using multiple target thresholds for plant species raised bird diversity versus the single-threshold baseline (same lower threshold, but no further thresholds). Our results show the benefits of even small shifts in incentive design.

1. Introduction

Intensive agriculture threatens biodiversity (Foley et al., 2011; Pe'er et al., 2014; Leclère et al., 2020). In response, governments worldwide have implemented agri-environmental policies to reduce this pressure on nature (Pe'er et al., 2022; Baylis et al., 2022; Pannell and Rogers, 2022; Elmiger et al., 2023). Such efforts are significant in scale, including the 16.6b euros spent by the European Union (EU) in 2023 to support “green measures”¹ in the agriculture sector (European Commission, Directorate-General for Agriculture and Rural Development et al., 2023) to help achieve commitments under the global Kunming–Montreal Global Biodiversity Framework (UN, 2022) and the UN Decade on Ecosystem Restoration (2021–2030) (IUCN, 2022). One important tool to address these challenges is agri-environmental schemes. To date, such schemes mostly reward farmers for implementing desired practices, though some also include result-based payments. However, the effectiveness of payments is questioned, in general, and the gains from shifting design are poorly understood (Navarro and López-Bao, 2019; Pe'er et al., 2019, 2020, 2022).

Result-based payments are, themselves, a significant design shift that could increase impact relative to more common action-based payments (Burton and Schwarz, 2013; Elmiger et al., 2023; Kelemen et al., 2023; Sattler et al., 2023). In the latter, farmers are paid for

* Corresponding author.

E-mail addresses: sergei.schaub@agroscope.admin.ch (S. Schaub), alex.pfaff@duke.edu (A. Pfaff), petyo.bonev@agroscope.admin.ch (P. Bonev).

¹ “Green measures” comprise a broad set of interventions to achieve environmental and climate objectives. Beyond these, they include “Eco-schemes,” “Environment and Climate Measures,” and non-productive investments in “green infrastructure.”

what they do, regardless of outcomes. In the former, farmers are paid for achieving predefined outcomes.

To date, the relatively few result-based payments in agriculture often use vascular plants as indicators of biodiversity (Burton and Schwarz, 2013; Elmiger et al., 2023). Typically, a count of distinct plant species determines whether a farmer's field is eligible for result-based payments. For example, farmers might be paid if at least 4 of 40 desirable species (on a list) are present within a given field. To raise the chances of being paid, farmers might adjust a field's management to make it more "flora-friendly," which can, in turn, yield spillover benefits for different taxa, including birds (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012).

One critical choice within the design of incentives, in the context of result-based payments, is how exactly to reward measured results. One could reward the achievement of a single threshold (such as 4 out of 40 listed species, in the example above). Alternatively, an incentive scheme could, instead, offer different rewards for different thresholds, moving toward a continuous rewards function by paying more for more species (e.g., from a list of 40, paying more for 6 than for 4 species).

However, despite discussions about using payments for measured results, there are few such schemes in practice. Thus, little is known about their impact on biodiversity (see Schaub et al., 2025 on a Swiss policy) or about the effects of different designs. To our knowledge, the additional impact of using multiple thresholds in result-based payments has not been studied to date.

We offer the first empirical study on the impact of multiple thresholds in result-based payment schemes. Specifically, we compare a *multiple-thresholds design* with a *single-threshold design* that uses the same low payment threshold but no additional thresholds for rewards. These designs co-existed in Germany's federal state of Lower Saxony, the focus of our study. We highlight two key challenges for analyzing the impacts from such result-based agri-environmental payments. The first one is the availability of biodiversity data. Biodiversity is costly to monitor, and datasets with many spatial units covering many years are rare. The second challenge is the endogenous enrollment of fields: fields that feature higher biodiversity before such a scheme starts are more likely to be enrolled by farmers (e.g., Kleijn and Sutherland, 2003; Hart and Latacz-Lohmann, 2005; Gómez-Limón et al., 2019; Bertoni et al., 2020). We address both challenges in our paper, as described below.

We leverage a biodiversity dataset with over 6.7 million citizen-science entries for birds, across 16 years of observation, for Lower Saxony. Even though birds are typically not targeted by result-based schemes, as plants are, the effect of agri-environmental schemes on bird diversity is still relevant. First, increased plant diversity, in response to a result-based payment scheme, can increase bird diversity both directly and indirectly (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012; Krause and Culmsee, 2013; Allan et al., 2014; Müller et al., 2022). On the direct side, induced changes in field management to support plant diversity can positively affect bird diversity. On the indirect side, the policy can lead to an increase in plant diversity, which can, itself, positively impact (insect diversity and) bird diversity. The second reason for the relevance of birds is that they and their diversity can help to understand systems' ecological states, as birds are sensitive to land-use changes and being in the higher ranks of the food chain makes them sensitive to changes in other trophic levels, such as insects (Gregory et al., 2005; Fraixedas et al., 2020; Li et al., 2020; Lees et al., 2022). Thus, schemes' effects upon bird diversity are interesting in their own right.

To deal with endogenous plot enrollment, we leverage a policy reform in Lower Saxony that with "staggered adoption" introduced payments with multiple thresholds for measured plant species. Given a baseline single-threshold scheme, a multi-threshold scheme was piloted in one region in 2007/08. In 2012, the pilot was extended to additional regions, and by 2014, the multi-threshold scheme was implemented across all of Lower Saxony (except in nature protection areas). This staggered implementation generated a natural control group, specifically the regions that at a given point in time are not yet treated. We leverage this for our analysis, using a nonparametric, staggered difference-in-differences approach.

One major advantage of our setup is that payments were introduced in large regions. This allows us to assess impacts on mobile species, such as birds. In contrast, when treatment happens at the field level, the mobility of birds across fields can make such an assessment challenging. Thus, we can use information on bird populations which, compared to other taxa, are uniquely available across extensive temporal and spatial scales (Kamp et al., 2021; Lees et al., 2022).

We find that multiple thresholds within result-based payments can help to raise bird diversity. For the initial pilot area, adding an additional threshold raises bird diversity by ~ 0.28 standard deviations (90 % confidence interval = [0.00 to 0.55]; 95 % confidence interval = [-0.03 to 0.59]). This adds to any impacts from single-threshold result-based payments, so it is a lower bound on the impact of multi-threshold payments relative to a baseline with no payments. Our findings highlight the importance of the designs of result-based payments going beyond the binary choice between action- and result-based approaches.

Our study contributes to assessing the impacts of agri-environmental payments on biodiversity in agricultural landscapes. Based on the measure of biodiversity, this literature can be broadly divided into studies using *in situ* (i.e., "in-the-field") measured biodiversity (plant, insect, or bird populations) versus indirect proxies, such as management practices (e.g., fertilizer use) and policy take-up (e.g., area enrolled). Studies using measured biodiversity are most often correlational, not accounting for decisions to enroll only certain fields (e.g., Kleijn and Sutherland, 2003; Roth et al., 2008; Baker et al., 2012; Meichtry-Stier et al., 2014; Marja et al., 2018). Studies addressing the endogenous selection of fields do so by using matching to identify similar fields (e.g., Kleijn et al., 2001; Kleijn et al., 2006; Knop et al., 2006) or by comparing trends across similar fields (Kleijn and van Zuijlen, 2004). They have focused on action-based payment, however, and often find small to moderate effects.

An exception, in the sense of combining diversity outcomes with attention to identification, is Schaub et al. (2025). They examine the payment increase of a single-threshold result-based payment scheme in Switzerland based on measured plant diversity. Compared to fields already eligible for payments but facing no shift in incentives, they find that the payment increase raised plant diversity in fields whose pre-reform species counts were just below the threshold of 6 indicator plant species. These fields likely faced lower adjustment costs.

Several studies using proxies for biodiversity seek to account for the selection of fields to schemes (e.g., Chabé-Ferret and Subervie,

2013; Bertoni et al., 2020; Stetter et al., 2022; Wuepper and Huber, 2022; Coderoni et al., 2023; Zimmert et al., 2024). They often find small to no effects of agri-environmental payments. One major advantage of using proxies is the temporal and spatial availability of data easily linked at the plot or farm level. Yet, while these studies are important for understanding uptake and input decisions,² proxies cannot indicate the magnitudes of effects on biodiversity, as their links to biodiversity are strongly context specific (e.g., Dormann et al., 2008; Socher et al., 2012; Báldi et al., 2013; Graham et al., 2018; Montgomery et al., 2020).

Finally, we contribute to literature on citizen-science data and drivers of changes in biodiversity. Citizen science is an emerging valuable tool, one complementing “traditional” data sources, that can temporally and spatially extend ecological data (Kosmala et al., 2016; Fraisl et al., 2022). Such data have helped identify species’ temporal trends (e.g., Schultz et al., 2017; Neff et al., 2022) and spatial distributions (e.g., Tiago et al., 2017; Johnston et al., 2020), in addition to assessing the influences of farms’ sizes and protected areas on biodiversity (Noack et al., 2022; Wauchope et al., 2022). Yet, to our knowledge, they have not been used to assess the causal effects of agri-environmental payments on biodiversity.

The rest of this paper is set out as follows. Section 2 offers policy background, while Section 3 describes the data that we employ. Section 4 lays out our empirical approach, Section 5 provides our results, and, finally, Section 6 concludes.

2. Policy background

We study multi-threshold versus single-threshold result-based agri-environmental payments in the German federal state of Lower Saxony, which is Germany’s second-largest state (at ~48k km²) and larger than Switzerland or the Netherlands (State Office for Statistics of Lower Saxony, 2023). Over half of Lower Saxony’s land is used for agricultural production, often intensive (Lomba et al., 2017), with mixed environmental and production conditions hosting livestock, arable crop, and grassland production in various climates and soils (Lomba et al., 2017). Some regional patterns can be seen: poorer, sandy soils predominate in the northeast, while higher-yielding soils are more prevalent in the middle-east, southeast, and south (Schüler and Noack, 2019; IRENES, 2020). These patterns correspond to certain land uses, with more grassland—for livestock—in the northern coastal region in contrast with more extensive arable farming (including cereals, potatoes, and sugar beets) in the middle-east, southeast, and south (Schüler and Noack, 2019; IRENES, 2020). In the southwest of the state, production focuses on animal husbandry (including pig, poultry, and egg production; Schüler and Noack, 2019; Lower Saxony Ministry of Food, Agriculture and Consumer Protection, 2019). Studying agri-environmental payments in Lower Saxony is interesting in part because of this heterogeneity of conditions and production systems, which can be relevant for a range of settings common across Europe.

As Germany is in the European Union (EU), its agricultural policy is strongly regulated by the EU’s Common Agricultural Policy (CAP). The CAP has two pillars: one focuses on direct payments to support farmers’ income, and the other focuses on rural development (Détang-Dessendre et al., 2023). Programs under the second pillar are cofinanced by Member States, who have flexibility in terms of program design. In Germany, the states design and implement their respective state programs.

Since the early 2000s, Lower Saxony has financed agri-environmental programs via their Rural Development Policy (Federal Agricultural Research Center, 2005; Andersson et al., 2017). For the policy period from 2007 to 2013, the Rural Development Policy was reformed, which included the introduction of a set of new agri-environmental payment schemes (the rural development program for this period was called PROFIL; Tietz et al., 2006). These new schemes included a result-based payments scheme for species-rich grasslands. We call these “result-based payments”. They were—at least in part—initially only introduced in pilot areas (Fig. 1), providing a unique opportunity to study their effects on biodiversity.

The result-based payment scheme is based on plant species and has two “layers” (or target thresholds driving payments). The first layer (NAU/BAU B2) was introduced in the entire state in 2007. Farmers were eligible for payments of 110 euros ha⁻¹ year⁻¹ when four or more indicator plant species were present on grasslands (Dickel et al., 2010; Most et al., 2015). The list had 31 plant species, or species groups, referred to as “indicator plant species” because they are often linked with certain habitat compositions and biodiversity (Most and Keienburg, 2006).³

The second layer (KoopNat FM 411; Fig. 1) was introduced in 2007/08 in pilot regions.⁴ A field was eligible for an extra payment of 110 euros ha⁻¹ year⁻¹, adding to the payment from the first layer, if its number of indicator plant species was at least 6 (Dickel et al., 2010). This pilot was then spatially extended in 2012 (Fig. 1). Thus, some regions had only one payment layer (i.e., a single target threshold), while pilot regions had two layers (i.e., two target thresholds, here referred to as *multi-threshold* result-based payments). We focus upon the impacts of this pilot.

These initial pilot regions in Lower Saxony featured different grassland types. Stakeholders including academics, environmental planning officers, and farmers were invited to engage in selecting the pilot regions. Yet the announcement of which regions were eligible was only made shortly before farmers could apply. Thus, farmers could not manage for diversity in anticipation. In contrast to the initial pilot region, the later extensions of the scheme’s pilot region in 2012 included areas with higher-expected-species grasslands (outside nature-protection areas) and where stakeholders had signaled their high interest in such species schemes. Farmers in the 2012 extension might have shifted behaviors before payments schemes were implemented.

² While Zimmert et al. (2024) use a biodiversity score, a useful extension of other work in approximating biodiversity, the score is not an *in situ* measure of biodiversity; it is based upon management practices (Jeanneret et al., 2014).

³ For the detailed list, see Table S1 in the Supplementary Information.

⁴ We know the pilot boundaries in 2008 but not 2007, so we aggregated those regions and their treatment timing. Because most of the areas implemented the pilot scheme in 2008 (Most et al., 2015), we use 2008 as the reference timepoint for the treatment.

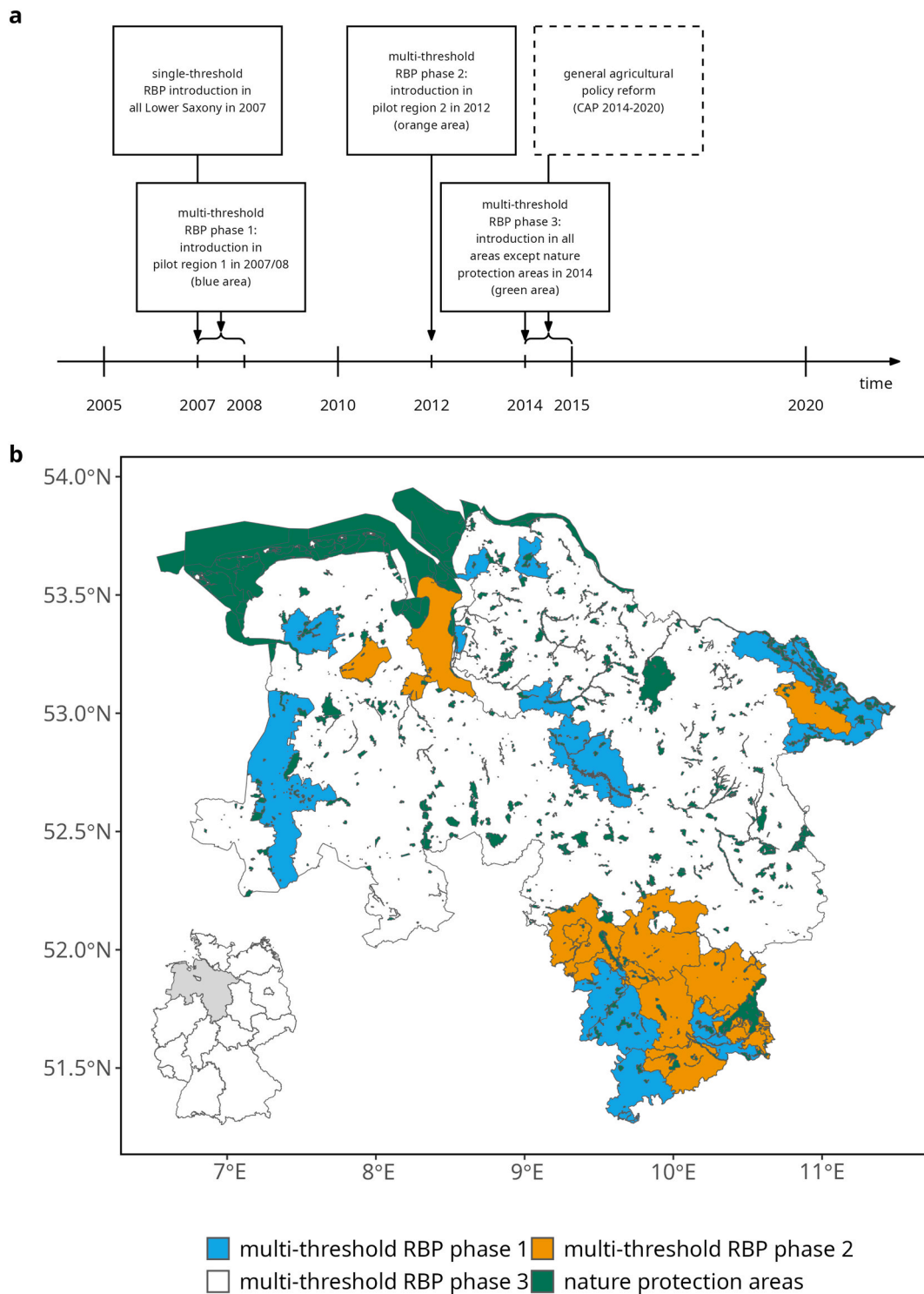


Fig. 1. Temporal (panel a) and spatial (panel b) development of the result-based payments in the federal state of Lower Saxony. RBP = result-based payments. Nature protection areas are defined here as those not eligible for multi-threshold result-based payments in phase 3, and include nature reserves, type C biosphere reserves, and national parks.

Since the agricultural policy underwent another reform, which started in 2014 and led to a large expansion of the area eligible for result-based payments (Text S1), we focus on the period before 2014. Further, we note that other spatially restricted agri-environmental payments in Lower Saxony existed (Text S1 for details), which we account for in the analyses.

3. Conceptual framework

In this section, we provide the conceptual framework that underlies our empirical analysis. Fig. 2's simplified causal graph depicts our key variables as nodes, linked by causal arrows. Node "A" depicts the eligibility criteria for result-based payments that should incentivize farmers to adjust field management (node "B") to sustain or increase plant diversity (node "C")—i.e., the policy's target outcome. As per our discussion of ecological direct and indirect effects in Section 1 (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012; Krause and Culmsee, 2013; Allan et al., 2014; Müller et al., 2022), field management and plant diversity can affect insect and bird diversity (nodes "D" and "E"). Thus, field management can have a direct effect on plant diversity (arrow 2) plus direct and/or indirect effects on insect and bird diversity. We aim to identify and estimate the total effect of the policy on bird diversity, which consists of the sum of the causal paths $(1 \rightarrow 4)$, $(1 \rightarrow 3 \rightarrow 6)$, $(1 \rightarrow 2 \rightarrow 5 \rightarrow 6)$, and $(1 \rightarrow 2 \rightarrow 7)$. In this graph, plant and insect diversity are mediators of the total effect of the policy on bird diversity. We do not observe these mediators but focus on the total effect of the policy on bird diversity. It can be identified without knowledge of mediation channels (Imai et al., 2010; Pearl, 2012).

Extending these ideas, we add a simple microeconomic model to convey cost–benefit tradeoffs and management incentives triggered by a single- and a multi-threshold result-based payment scheme for plant diversity. For a clear, simple model that includes all the ideas, here we abstract from the temporal details of the actual, staggered implementation of multi-threshold payments in Lower Saxony. Specifically, to focus on the incentives from introducing policies, we assume that a second, multi-threshold scheme is introduced at a single point in time and, further, that every farmer has only one field.⁵ We thus extend the model in Schaub et al. (2025), who focused on a single-threshold payment and plant species as a biodiversity outcome.

Let us start with a single-threshold payment, P_1 , for a specific number of indicator plant species, i.e., a threshold or a target T_1 . The cost of reaching any target number n of indicator species depends upon the pre-payment number, n_0 , as a positive linear function $c^*(n - n_0)$.⁶ Thus, a farmer's profit from trying to achieve at least the single target threshold is $R_1 - c^*(T_1 - n_0)$.⁷

This gain function implies that when maximizing profits, farmers will each choose their n_1^* , the endogenous number of indicator species optimal for them under a single-threshold payment.⁸ With low enough reward R_1 , no farmer will find it worthwhile to change management practices (e.g., reduce the number of cuts, reduce fertilizer application, or overseed) to achieve that single threshold T_1 . With high enough reward R_1 , all the farmers not above T_1 would be willing to adjust. For intermediate reward levels, only farmers with high pre-payment numbers of species (but not above T_1) and, thus, lower adjustment costs, would respond to payment by changing management. In sum, farmers below target T_1 and with adjustment costs below reward R_1 respond. (Farmers with pre-payment indicator species diversity above T_1 can get payment without adjusting management.)

Next, we consider a shift from a single-threshold payment scheme to a multi-threshold scheme, with a second payment R_2 at threshold T_2 . Decisions under the multi-threshold scheme follow the same logic as under the single-threshold scheme. Given that the multi-threshold comes after a single-threshold policy (versus the single-threshold being introduced after no policy at all), for new choices we can just replace the initial species number (n_0) with the optimal single-threshold species number given R_1 and T_1 (i.e., n_1^*). The cost to reach T_2 depends on the single-threshold optimum, n_1^* , again with a linear cost, i.e., $c^*(T_2 - n_1^*)$. Farmers with $T_2 > n_1^* \geq T_1$ face gains of reaching T_2 of $R_2 - c^*(T_2 - n_1^*)$, while farmers who optimally did not achieve T_1 under R_1 consider $R_1 + R_2 - c^*(T_2 - n_1^*)$. With a low enough reward R_2 for achieving T_2 —again, R_2 is added to payment R_1 for achieving T_1 —no farmer will adjust their management of their (single) field to achieve the second threshold.

With high enough R_2 , however, every farmer that did not yet reach T_2 would optimally respond. For this latter case, those farmers who optimally choose to manage to try to reach T_2 to get R_2 potentially include some farmers who, say for a low R_1 , did not even try to manage to reach T_1 . For intermediate R_2 , only farmers with higher post-single-threshold-payment species (n_1^*) will respond to the second-threshold reward R_2 . This may only include those farmers who got R_1 , via either high initial pre-payment indicator species ($n_0 \geq T_1$) or their management to reach T_1 ($n_1^* \geq T_1 > n_0$). When a farmer has an initial level of indicator species (n_0) above the threshold, now T_2 , they will receive R_2 , as well as R_1 , without having to make field management adjustments.

In sum, our conceptual model highlights that multi-compared to single-threshold payments can lead to both (i) more farmers adjusting their field management to raise plant diversity and (ii) farmers adjusting management more significantly to reach higher levels of plant diversity. As a consequence, this can raise bird diversity (see also Fig. 2).⁹ Thus, from here we move to empirical testing, having flagged policy-design issues critical for impact, such as the reward levels and targets. In addition, we note that impacts also

⁵ This model can be analogously applied to a setting with staggered implementation and farmers with several fields.

⁶ When it is about reaching at least n , then these linear costs are $c^*(n - n_0) \cdot 1[n_0 < n]$, where $1[\cdot]$ is an indicator function equal to one when its condition is satisfied. Further, one could also imagine reasons for non-linearities (either concavity or convexity) in such a cost function or, more generally, differences between farmers in the costs.

⁷ Farmers are finally eligible for payment when $n \geq T_1$.

⁸ This could be above T_1 , of course, if $n_0 \geq T_1$, which would imply that a farmer faces zero cost to get R_1 .

⁹ At the conceptual level, none of these predictions would change if the adjustment costs described above were non-linear (which they could be), provided that marginal costs are non-negative.

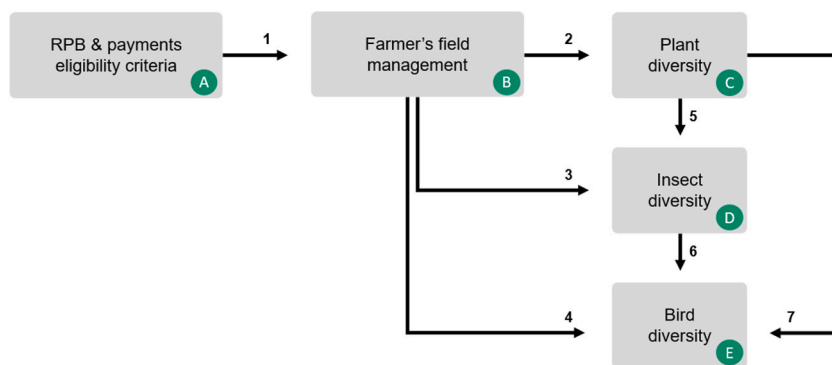


Fig. 2. Causal graph depicting the potential direct and indirect effects of field management on bird diversity. In the figure, each directed arrow denotes a causal impact and each node a variable. RBP = result-based payments.

depend upon the species numbers before such a policy. In the following, we empirically assess how the introduction of multi-versus single-threshold payments affected biodiversity beyond plant diversity, by focusing on bird diversity.

4. Data

4.1. Bird data

We utilize data from the Ornitho (<http://www.ornitho.de>) citizen-science project, within which volunteers report observations for birds across Germany. During 2005–20, ~6,000 volunteers recorded roughly 6.7 million entries in Lower Saxony.¹⁰ We assign each record to a 16 km² unit of observation, which aggregates upward from the original 1 × 1 km² sampling in the Ornitho data.¹¹ Birds are mobile, so effective management of any field will affect bird observations not only in that field but also more widely, implying spillovers from impactful adjustments (and policies). Larger observational units reduce the number of times a given new bird is counted as present. They also considerably reduce the share of observations in which zero birds are counted (Table S1). On the other hand, too large a unit—beyond spatial spillovers—reduces the power of our tests for no good reason and can blend unmanaged with managed fields. For robustness, we varied the size of our observational units from 1 km² up to 68 km² (Fig. S1).

We excluded observations on the East Frisian Islands, which are inherently different to the mainland. We also deleted 276 entries for which we did not know the coordinates (1 × 1 km² grid) of the observation but only the municipality. Further, we excluded all the regions (and their observations) in which, between 2008 and 2013, action-based payments schemes for birds and other animals in restricted areas were introduced (Section 2 and Text S1). More generally, this space-time dataset is unbalanced since not every observation was surveyed every year.

The data include whether a bird species was seen and the number of individual birds observed for that species. We compute the average number of individual birds seen per species, per observation, across different volunteer observers, and then we rescale each of the variables to have a mean of zero and standard deviation of one. The data also include an anonymous ID for any volunteer reporting a sighting. Using this, we can calculate the number of volunteers per spatial observational unit, which we later use in our estimation as a time-varying covariate. The data show that the number of volunteers generally increased over time (Fig. S2) and that volunteers were quite evenly distributed across space (Fig. S3).

4.2. Treatment data

To identify when an area was exposed to the “treatment” (i.e., when an area became eligible for multi-threshold result-based payments), we use data provided by the Lower Saxony State Department for Waterway, Coastal and Nature Conservation. We label an observation as “treated” when at least 50 % of it is in the pilot area. We also vary this treatment definition (i.e., the % in the pilot area) in our sensitivity analyses to examine “at least 70 %” and “at least 90 %.”

4.3. Bird species outcome definition

We focus on common bird species in studying the effects of shifting to multi-threshold payments. We expect that this is where agri-environmental result-based payments might have an influence, given that it has been shown that agriculture and grassland management can be (particularly) important for common bird species (Donald et al., 2001; Benton et al., 2002; Batáry et al., 2007; Kleijn

¹⁰ Note that in autumn of 2011 ornitho.de (the online platform) was launched. The data before 2011 was retrospectively entered.

¹¹ The sampling grid used by Ornitho (following the German “Halbminutenfelder”) is approximately 1 × 1 km (i.e., 1 km²).

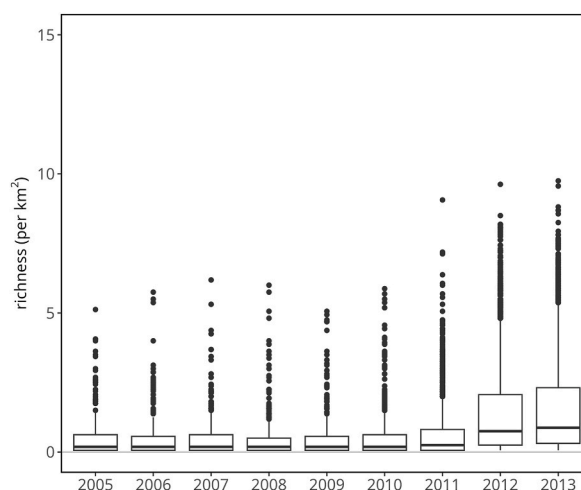


Fig. 3. Development of species richness over time of *common* bird species. The development for the Hill–Shannon index of *common* bird species, species richness over time of *all* bird species, and species richness over time of *farmland* bird species are shown in Fig. S4–6, respectively. The scale of the y-axis of this plot is selected to easily compare the trends with Figs. S5 and S6.

et al., 2011; Busch et al., 2020; Rigal et al., 2023). Moreover, volunteers might be better at detecting common birds compared to rare bird species (Cox et al., 2012; Kosmala et al., 2016). We define as common all those recorded during 2005–2020 in the German Common Bird Monitoring Survey (Mitschke et al., 2005; Table S3).¹²

In addition to common birds, we test the effects on all birds and farmland birds to better understand which birds are most affected. Using farmland species may help to reduce potential selection biases resulting from volunteers' decisions about what species to record (Noack et al., 2022). Our classification of farmland species is based on lists by Busch et al. (2020) and Noack et al. (2022) (Table S4). Species in those lists are seen as indicator bird species for farmland species overall as they occupy various ecological niches in agricultural landscapes (Busch et al., 2020).

4.4. Diversity measure

Our main measure of diversity is species richness, a common and simple measure of how many different species were observed (Roswell et al., 2021). We focus on this because citizen-science data are thought to offer more precision for presence, for any species, than for abundance, i.e., number of individual birds observed per species (Bird et al., 2014; Kosmala et al., 2016). In our sensitivity analyses, we also examine the effects on the Hill–Shannon index, which in addition takes into account the species' abundances (Roswell et al., 2021; Text S2).

Fig. 3 shows richness over time in our data, for common bird species, using 16km² units. An upward trend in total birds observed by citizen-science is expected, given the increasing popularity of participation in such citizen observation (Knape et al., 2022), distinct from actual population trends for the birds (Busch et al., 2020; Kamp et al., 2021; Lees et al., 2022; Rigal et al., 2023). Yet, these data nonetheless allow for a comparison of regions, as below we show that the trends before the implementation are not different between the regions (Section 6).

4.5. Additional data

We use several additional data sources in our analysis (see *estimation procedure* in Section 5.2 and *additional observed covariates* in Section 6.3). First, we use the share of nature protection areas (Fig. S7) from Lower Saxony State Department for Waterway, Coastal and Nature Conservation (2023) and NUMIS (2002). These areas suggest high perceived natural value and were excluded from result-based payments until 2021.

Second, we use the share of agricultural land and grassland in 2006, with a resolution of 100 × 100m, from the EU's Copernicus Land Monitoring Service (2020) to account for agricultural areas where farmers could enroll land to the payment scheme.

Third, we use the mean of grassland yield to account for opportunity costs, which can influence non-participation in any agri-environmental payment scheme (Schaub et al., 2023), from the State Office for Statistics of Lower Saxony (2024). These data cover 2002–2007 at the county level (Fig. S8 for the county distribution in Lower Saxony) and are matched to each grid based on the county in which the majority of the grid is located.

¹² These common bird species encompass 180 species in total.

Group	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
M	0	0	0	2	2	2	2	2	2	2
S	0	0	0	1	1	1	1	2	2	2
B	0	0	0	1	1	1	1	1	1	2

Fig. 4. Time change by region in the number of thresholds (2 = “multi-”, 1 = “single-”). Each scheme has a threshold of 4 indicator plant species. The multi-threshold scheme adds one threshold of 6 indicator plant species, with another payment “layer.” Note that here, we indicate the introduction in 2007/08 as if it were in 2008.

Fourth, we use topsoil organic carbon content (Jones et al., 2005) and soil pH (Reuter et al., 2008)¹³ at a $1 \times 1\text{km}$ and $5 \times 5\text{km}$ resolution, respectively, from the European Soil Data Centre (Panagos et al., 2022) to account further for differences in opportunity costs and yield potential. Fifth, we use nitrogen surplus (Batool et al., 2022) and cattle density (Du et al., 2025) in 2005 at an approximate resolution of $10 \times 10\text{km}$ and $5 \times 5\text{km}$, respectively, which provide information on management intensity.

Sixth, we use human population density, in a $1 \times 1\text{ km}$ grid, from the German 2011 census from the Statistical Offices of the Federal Government and the States (2015) to account for the likelihood of recording in citizen-science data (Hertzog et al., 2021). While 2011 is after the 2008 treatment, we assume this treatment does not affect human population.

Seventh, in our robustness tests, we consider yearly weather variables. These variables include, following Palmer et al. (2025), annual average min and max temperature and precipitation from 2005 to 14, have a resolution of $\sim 4 \times 4\text{km}$, and are taken from TerraClimate (Abatzoglou et al., 2018). For all variables, we compute the average value per unit then rescale so that variables have a mean of zero and a standard deviation of one.

5. Empirical strategy

We are interested in the impact of a multi-threshold compared to a single-threshold result-based payments scheme on bird diversity. In this section, we present notation and treatment definitions (in Section 5.1), then we outline our identification and estimation strategy (Section 5.2).

5.1. Notation and treatment effects of interest

Let t denote time, and let $D \in \{0, 1, 2\}$ be a treatment variable with three values: 0 (no result-based payments), 1 (result-based payments with one threshold or “single-threshold” payment), and 2 (for a result-based payments scheme with two thresholds or “multi-threshold” payment). Let $Y_{t,i}(d)$ denote the potential biodiversity (proxied by bird species richness within our main analysis) that would be observed for unit i in period t had the region received a treatment $d = 0, 1, 2$. The corresponding measured outcome is defined as Y_{it} (the index i is frequently omitted for convenience). With this notation, we can define the following average treatment effect:

$$\Delta_{t,d'}(d, d') = \mathbb{E}[Y_t(d) - Y_t(d') | D = d']. \quad (1)$$

Intuitively, $\Delta_{t,d'}(d, d')$ describes the average effect on biodiversity from switching from mode d' to d for those who are under treatment arm d' . When $d = d'$, this is a standard average treatment effect on the treated. Since we focus on treatment as the introduction of the payment in a region rather than the enrollment of individual fields in a payment scheme, the effects should be interpreted as Intention-to-Treat (ITT) effects.

It is helpful to introduce the following additional notation. Let B , S , and M be three different groups of regions reflecting phase 3, 2, and 1 when multi-threshold payments were introduced (compare with Fig. 1). We illustrate the temporal development of the number of thresholds in each of these groups in Fig. 4. Group M gets the multi-threshold payment $d = 2$ in period 2008, group S in 2012, and group B in 2014, the point in time at which the multi-threshold payment scheme is introduced for every region.

5.2. Identification and estimation

Identifying strategy. We first focus on the identification of $\Delta_{2012,S}(2, 1) = \mathbb{E}[Y_{2012}(2) - Y_{2012}(1) | S]$. This is the one-period effect of introducing a multi-threshold payment relative to a single-threshold payment for the group S . Thus, we interpret a system with one threshold as the baseline or control condition, while a system with two conditions is the treatment. The effect is identified under the assumption

¹³ While the soil pH data were published in 2008 (i.e., in the first treatment year), all input data comes from preceding years; hence, the data (as well as the data on organic carbon content published in 2005) represent pre-treatment conditions.

$$(PT1) \mathbb{E}[Y_{2012}(1) - Y_{2012}(1)|S] = \mathbb{E}[Y_{2012}(1) - Y_{2012}(1)|B] \quad (2)$$

which is the standard parallel trends assumption.¹⁴ Below, we provide evidence for the validity of *PT1*. Estimation of $\Delta_{2012,S}(2, 1)$ can be performed as in the canonic two-period difference-in-difference setup using only observations from 2011 to 2012 for groups *B* and *S*. To increase the efficiency of this estimator, we additionally include the pre-treatment periods 2008–10 as well as the post-treatment period 2013. Note that our analysis uses only post-treatment periods up to 2013, as in 2014 the agricultural policy in Lower Saxony underwent another reform, leading to the widescale adoption of multi-threshold payments (Fig. 1). This restriction poses a limitation for our analysis, since it only allows us to assess the change in biodiversity in 2 periods after the reform. This is problematic if biodiversity needs more time to adjust, as has been documented (e.g., Chamberlain et al., 2000; Ernoul et al., 2006; Uezu and Metzger, 2016), or when farmers' management adjustments in response to the payment introduction take time (e.g., Hynes and Garvey, 2009; Moret-Bailly and Muro, 2024).

To deal with this problem, we additionally use the following approach. An estimator that uses all periods until 2013 and all groups can be constructed when the multi-threshold payment ($d = 2$) is considered to be the treatment and all other regimes ($d = 0, 1$) as the control. Since $d = 2$ is introduced in 2008 (for group *M*), 2012 (for group *S*), and 2014 (for group *B*), we obtain a setting with a staggered implementation of the treatment. We use the nonparametric estimator of Callaway and Sant'Anna (2021). The major advantage of this approach is that it utilizes all available information and provides effect estimates for different periods after the treatment (and, thus, the unfolding of the effect over time can be studied).

The potential problem with this estimator is that by pooling 0 and 1, the interpretability of the estimates becomes more complicated. In the following, we show that this is not an issue and that the estimates can be interpreted as lower bounds for the true effects. Specifically, the effects suggested by Callaway and Sant'Anna (2021) depend on the quantity

$$ATT(g, t) = E[Y_{it}(d) - Y_{it}(0)|G_g = 1] \quad (3)$$

where g is the point in time of first treatment, and G_g is a binary variable indicating whether treatment takes place in period g . This is the so-called group-time effect. Based on $ATT(g, t)$, we can compute the group effect as

$$ATT^{group} = \frac{1}{\tau - g + 1} \sum_{t=g}^{\tau} ATT(g, t) \quad (4)$$

which is a weighted aggregation of the group-time effect and where τ is the number of periods we observe (Callaway and Sant'Anna, 2021). $ATT(g, t)$ is identified via (Marcus and Sant'Anna, 2021):

$$ATT(g, t) = E[Y_{it} - Y_{it-1}|G_g = 1] - E[Y_{it} - Y_{it-1}|D_t = 0, G_g = 0] \quad (5)$$

These quantities are then estimated with sample averages. In our case, since before 2008 all individuals have a treatment status $D = 0$, and since after 2008 no individual is under $D = 0$, this approach amounts to estimating quantities of the type

$$Q := E[Y_{POST}(2) - Y_{PRE}(0)|M] - E[Y_{POST}(2) - Y_{PRE}(0)|B_2], \quad (6)$$

where *POST* refers to the period between 2008 and 2014 and *PRE* represents a period before 2008, and $B_2 = B \cup S$ means the union of the groups *B* and *S*. The quantity Q can be consistently estimated from the observed data by substituting the expectations with their finite sample averages. Assume also the parallel trends assumption for the case of no treatment,

$$(PT0) \mathbb{E}[Y_{POST}(0) - Y_{PRE}(0)|M] = \mathbb{E}[Y_{POST}(0) - Y_{PRE}(0)|B_2] \quad (7)$$

Below, we provide compelling evidence that this assumption is satisfied. Moreover, it is straightforward to show that under *PT0*, the quantity Q can be represented as a difference between two different one-period treatment effects on the treated. Specifically, it holds that

$$Q = E[Y_{POST}(2) - Y_{POST}(0)|M] - E[Y_{POST}(1) - Y_{POST}(0)|B_2]. \quad (8)$$

The proof of the equality is provided in Text S3 of the Supplementary Information. The first component of the right-hand side of the above equality is the treatment effect of introducing multi-threshold payments relative to no result-based payments for the group *M*, while the second component represents the treatment effect of introducing single-threshold payments (relative to no result-based payments) for the group B_2 . Thus, this difference in treatment effects is difficult to interpret. However, under the mild assumption that $E[Y_{POST}(1) - Y_{PRE}(0)|B_2]$ is non-negative, we can derive the following inequality¹⁵

$$Q \leq E[Y_{POST}(2) - Y_{POST}(0)|M]. \quad (9)$$

¹⁴ In our multi-period setup, $\Delta_{2012,S}(2, 1)$ is more generally identified under the assumption $(PT1) \mathbb{E}[Y_{2012}(1) - Y_t|S] = \mathbb{E}[Y_{2012}(1) - Y_t|B]$, where $t < 2012$.

¹⁵ This assumption is supported by recent findings of Schaub et al. (2025) and Wuepper and Huber (2022). Both studies consider the effect of single-threshold result-based payments on biodiversity, which is precisely the setup envisioned by the assumption. While this literature finds that for some farms, the payments come as a windfall gain, there is no evidence for a boomerang (i.e., negative) effect in these cases, and the average effect of the policy is positive.

Thus, the estimable terms Q can be interpreted as *lower bounds* for the quantities $ATT(g, t)$.

Estimation procedure. We use the nonparametric estimator of Callaway and Sant'Anna (2021). The estimator builds on a doubly robust estimator, which leverages propensity scores (Sant'Anna and Zhao, 2020), allowing flexible consideration of observed pre-treatment characteristics and the use of unbalanced data. We include two types of covariates in the estimation. First, we consider time-invariant and pre-treatment covariates, denoted by X_i , such as the share of nature protection areas, the share of agricultural land, the mean grassland yield, and human population density (see Section 4.5). Second, as the only time-varying covariate in our main analysis, we consider the number of volunteers that record birds, Z_{it} . This can be an important variable, as it links directly to the likelihood of number of birds observed and the “effort” spend to observe. Following, for example, Wauchope et al. (2022), we model the log of it in our analysis. The main identification assumption here is that these covariates are not affected by the treatment. We provide empirical evidence that this is not the case in Section 6.3.

Remark 1. To assess the robustness of our results, we follow the recent literature on covariates in staggered difference-in-differences models (Caetano and Callaway, 2024; Caetano et al., 2022; Karim and Webb, 2024) and estimate a variety of different specifications with covariates. Specifically, we consider separate specifications with only time-constant pre-treatment covariates, specifications with time-constant and pre-treatment values of time-varying covariates, or specifications that include all sorts of covariates, including time-varying covariates. The validity of the estimates of these different specifications depend on different conditional parallel trends assumptions and, in the case of time-varying covariates, on assumptions regarding the relationship between the treatment and the covariates (Caetano et al., 2022). In this latter case, identification relies on the specific exogeneity and confoundedness assumptions (assumptions Cov-exogeneity and Cov-unconfoundedness in Caetano et al. (2022)). Violation of these assumptions leads to conditioning on colliders, or, put differently, bad controls (Karim and Webb, 2024).

There are two important points to be made here. First, unlike the selection on observables approaches, such as matching or augmented inverse probability estimation, the (staggered) difference-in-differences approach does not require that all factors of the outcome are accounted for (and, therefore, observed). Instead, observed covariates are used to relax the parallel trends assumption and only impose it within groups with similar observed characteristics (Heckman et al., 1997, 1998). Selection upon unobservables is allowed for, even conditionally on observed covariates. Second, the only diagnostic tool for assessing the validity of the different parallel trends assumption is estimating placebo effects in pre-treatment periods (see Remark 2 in Caetano et al., 2022, and also Heckman et al., 1997, 1998). As discussed below, our estimates provide strong evidence for the validity of the different parallel trends assumption. In addition, we also provide evidence for the Cov-exogeneity and Cov-unconfoundedness of our main time-varying covariate (i.e., the number of volunteers).

6. Results

6.1. Main results – common bird species

Our main results examine common bird species. Consider first our estimates for $\Delta_{2012,S}(2,1)$, which is the impact of changing from a single- to a multi-threshold payment for the 2012 pilot area. The group-time effects and group effects are displayed in Fig. 5, where the group effects are the weighted aggregation over group-time effects (see Equation (4)). As discussed above, we are only able to estimate the group-time effects for the first two periods after the treatment, which may not reveal time-lagged effects given that farm management and biodiversity may need time to adjust (e.g., Chamberlain et al., 2000; Ernout et al., 2006; Hynes and Garvey, 2009; Uezu and Metzger, 2016; Moret-Bailly and Muro, 2024). Both estimates are very close to zero in magnitude. Yet, it may take time for any effect to unfold. All pre-treatment estimates are also very close to zero, and the corresponding confidence bands contain zero almost at their centers, indicating that the parallel trends assumption is plausible.

To assess the possibility that effects arise over time, we pool control groups $d = 0$ and $d = 1$ then use a staggered difference-in-differences approach (see Section 5.2). This approach allows us to obtain effect estimates for 6 post-treatment periods (Fig. 6). Consistent with the previous approach, these estimates also are practically zero for the first two post-treatment periods (Fig. 6a). However, from period three, effects are positive and increasing and, as shown, can be interpreted as *lower bounds*. The point estimate of the 2008 group effect (ATT^{2008}) is 0.28 standard deviations (Fig. 6b).

Since the related literature predicted non-negative effects without providing a specific prior, we use confidence intervals, following Imbens (2021), rather than p-values. In other words, our objective is to estimate a policy effect, not test a given Null hypothesis (see, also, Cox, 2020; Wasserstein and Lazar, 2016). The 90 % and 95 % confidence intervals for this ATT^{2008} estimate are [0.00 to 0.55] and [−0.03 to 0.59], respectively. The lower bound of the 95 % confidence interval just includes zero, while that of the 90 % interval does not.^{16,17}

Finally, all pre-treatment estimates displayed in Fig. 6a are again practically equal to zero, and the tight confidence bounds imply

¹⁶ That said, of course plausibility is not uniform along the confidence intervals. Under mild regularity conditions, an outcome at the point estimate is 4 times as plausible as an outcome at the extremes of a 90 % confidence interval, and that ratio becomes about 7 times more plausible than at the extremes for a 95 % confidence (Romer, 2020).

¹⁷ In our multi-period setup, $\Delta_{2012,S}(2,1)$ is more generally identified under the assumption $(PT\ 1) \ E[Y_{2012}(1) - Y_t|S] = E[Y_{2012}(1) - Y_t|B]$, where $t < 2012$.

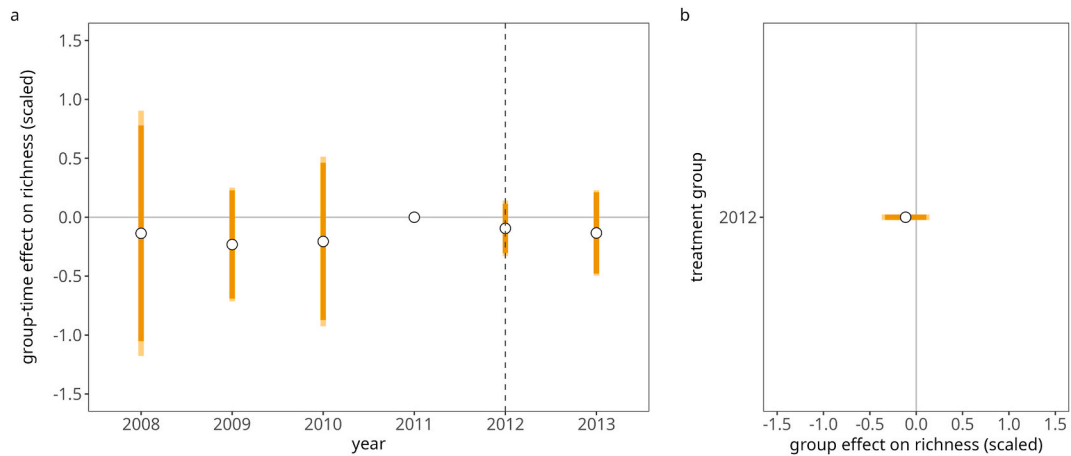


Fig. 5. Effects of treatment on common bird species richness, considering the 2012 pilot area. Panel a shows the group-time effect, and panel b shows the group effect. The light and dark bars indicate the 95 % and 90 % confidence intervals, respectively, and standard errors are clustered at the county level.

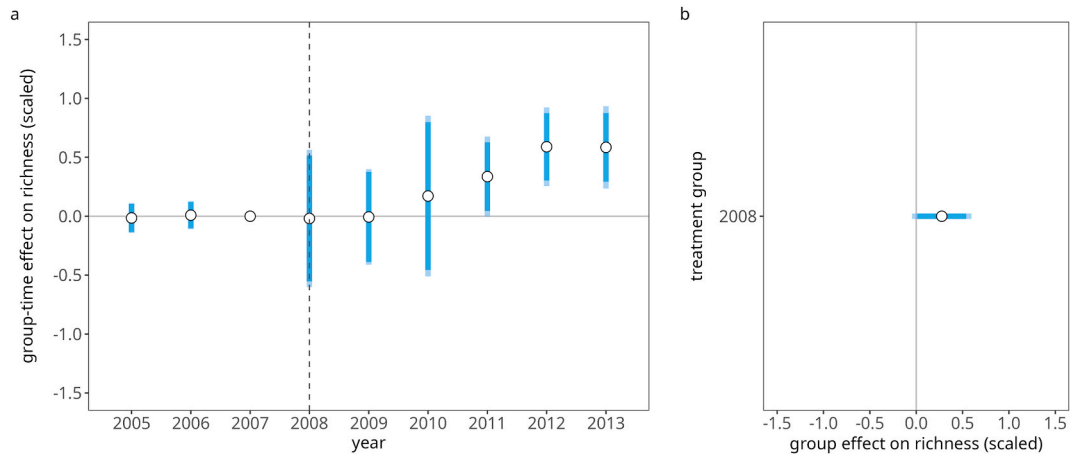


Fig. 6. Effects of treatment on common bird species richness, considering the 2008 pilot area. Panel a shows the group-time effect, and panel b shows the group effect. The light and dark bars indicate the 95 % and 90 % confidence intervals, respectively, and standard errors are clustered at the county level.

that these measurements are very precise. This provides further evidence for the validity of the parallel trends assumption.

6.2. Alternative bird population focus

We now present estimates using data on *all* (Fig. 7a) and *farmland* bird species (Fig. 7b) (instead of *common*). To simplify, we focus in this section on the group effect. This group effect include, for the 2008 pilot area, 6 post treatment periods, which allows us to capture when effects occur with a longer time lag, while only two post treatment periods are included for the 2012 pilot area (see Figs. S9 and S10 for group-time treatment effects).

These estimates reveal two insights for the 2008 pilot area (we do not observe effects for the 2012 pilot area). First, the point estimate for all birds, 0.27, is very similar to that for common birds (0.28), even if more uncertain. Second, the multi-threshold result-based payments had an even clearer positive group effect, again beyond a single-threshold effect, on farmland birds for the 2008 pilot area (Fig. 7b). The point estimate is 0.41, with 90 % and 95 % confidence intervals of [0.21 to 0.61] and [0.19 to 0.63], respectively. Finally, Figs. S9 and S10 show that the parallel trends assumption is supported.

6.3. Further results and robustness checks

Number of volunteers as a covariate. To evaluate whether the number of volunteers can be included in the regression as a time-varying covariate, we follow Caetano et al. (2022). Specifically, they adopt a parallel trends assumption conditional on pre-treatment

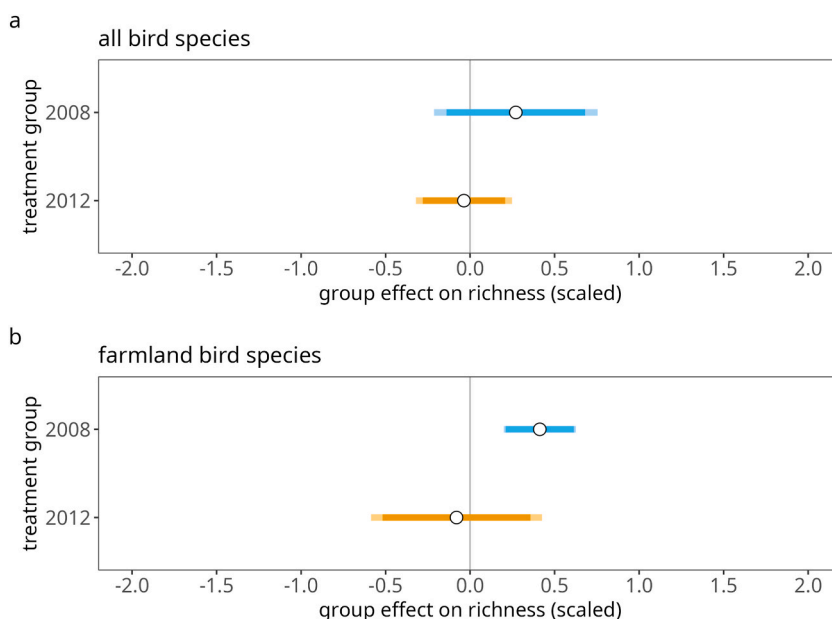


Fig. 7. Group effects of treatment on all (panel a) and farmland (panel b) bird species richness. The 2008 pilot area is indicated in blue, and the 2012 pilot area is indicated in orange. The model specification follows our main estimation specification. The light and dark bars indicate the 95 % and 90 % confidence intervals, respectively, and standard errors are clustered at the county level. All results of the group-time effects on all bird species and farmland bird species can be found in [Figs. S9 and S10](#), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

covariates and show that identification and standard procedures hold if, in addition, the pre-treatment covariates are exogenous (where “exogeneity” follows a weaker definition, see their assumption 3a) and unconfounded. To evaluate these assumptions, we follow the two approaches described in [Remark 2](#) in [Caetano et al. \(2022\)](#). The “direct” approach is to provide the placebo effect of bird diversity estimates for pre-treatment periods. As discussed in Section 6.1, our placebo effect estimates support the conditional parallel trends assumption. As an “indirect” approach, we estimate a staggered difference-in-differences model with the number of observers as a dependent variable ([Fig. S11](#)). First, the results show that all post-treatment effect estimates are close to zero, which makes it plausible to assume that the treatment has no effect on the number of observers and supports the exogeneity assumption of [Caetano et al. \(2022\)](#). Second, the unconfoundedness assumption is related to the untreated potential covariate outcome (i.e., the number of volunteers). This assumption is supported by the pre-treatment effect estimates being also very close to zero. Thus, the number of observers is a valid covariate (in the sense of [Caetano et al., \(2022\)](#)).

In addition to these robustness checks, we also run a model without accounting for the number of volunteers (i.e., we drop the time-variant number of observers variable Z_{it} from the specification). The results are displayed in [Fig. S12](#) and are very similar to those in our main analysis. Importantly, the pre-treatment trends are shown to be parallel.

Avidity bias.¹⁸ One potential pitfall of our empirical strategy to use birds as a proxy for biodiversity is the so-called avidity bias ([Moeltner and Shonkwiler, 2005](#)). Specifically, observations of birds are potentially collected in locations known for their high bird diversity, which are at the same potential locations as species-rich grasslands.¹⁹ This type of sampling is known as on-site sampling, which is a form of stratified sampling ([Shaw, 1988](#)). The potential problem arises “if the probability of inclusion in a given sample is commensurate to the size of the variable to be recorded” ([Moeltner and Shonkwiler, 2005](#)). A rich literature has discussed this bias and related ones in the context of on-site sampling and, more generally, choice models ([Englin and Shonkwiler, 1995](#); [Moeltner and Shonkwiler, 2005](#); [Hindsley et al., 2011](#); [Kuriyama et al., 2013](#)).

To address the possibility of such a bias, we rely on several alternative strategies. First, we follow the most common method in the literature and account for the endogenous stratification by adjusting the observations using their (estimated) inverse probability weights ([Hindsley et al., 2011](#); [Wooldridge, 2002](#)). The intuition behind weighting the observations in the context of choice-based sampling, which was first suggested by [Manski and Lerman \(1977\)](#) and then further developed by a variety of studies ([Imbens, 1992](#); [Hindsley et al., 2011](#)), is to oversample underrepresented regions. The probability of choice of location (the one that determines which region is underrepresented) is estimated as a first step, and then each observation is multiplied by its inverse probability weight. In the context of nonparametric difference-in-differences, [Sant’Anna and Zhao \(2020\)](#) and others develop the properties of a doubly robust estimator based on the propensity score. An analogous principle is used in the staggered version of the estimators that we use in

¹⁸ For an additional discussion on the “quality of volunteers,” see Text S4.

¹⁹ We thank an anonymous referee for pointing this out.

this paper. To account for differences in bird diversity across regions, we use a rich set of predictors, such as local population density, share of agricultural land, regional grassland yield, and conservation area.

Second, note that in contrast to choice models, our difference-in-differences estimates are only affected by avidity bias if the treatment interacts with (or is related to) the on-site intensity of observation in a dynamic way. Specifically, a major threat here is that the treatment affects the number of observers because it makes the observation activity less costly or more attractive. To account for this possibility, we utilize a major property of our dataset—namely, we observe the number of observers of each grid (i.e., unit of observation). This allows us to estimate the effect of the treatment on the number of observers. Above, we show the absence of any such effect.

Finally, if the local intensity of observation does not interact with the treatment, the avidity bias boils down to estimating heterogeneous (location-specific) effects. In this case, the treatment effect estimate for each location is unbiased; however, it may differ significantly from the overall average treatment effect. This possibility does not represent a threat to our results for the following reasons. Even if we estimate a treatment effect only for bird-rich and, thus, indirectly for grassland-rich locations, it can be argued that this effect is the ultimate effect of interest and not an effect over all possible regions. In particular, since flora diversity hotspots are found mainly in grassland-rich areas, precisely those areas are the target of the agri-environmental scheme.

Measurement error. One potential pitfall of our dataset is that it is collected by ordinary citizens and not by professionals. Thus, even though in our main results we focus on common species, misreporting may exist. Therefore, we perform an additional estimation with an alternative dataset. Specifically, we use the German Common Bird Monitoring Survey (CBS) as a check on the validity of the bird sightings in the citizen-science data (Mitschke et al., 2005).²⁰ In the CBS, all common birds are systematically recorded based on stratified sampling areas and independent of the species interests of the observer (Text S5). The results show that Ornitho data would indeed appear to be valid for depicting changes in bird populations (Table S5; Fig. S13), despite sighting locations and times not being identical in the Ornitho and CBS data.

Furthermore, we note that using citizen-science data—i.e., data collected through unstructured methods—has often been shown to be a reliable source of information,²¹ and the conclusion of studies that use both citizen-science and structured datasets often align (Hoyer et al., 2012; Aceves-Bueno et al., 2017; Callaghan et al., 2018; Henckel et al., 2020; Noack et al., 2022; Binley et al., 2025), supporting its validity in applications such as ours.

Measurement and definition choices. We run several checks to understand the sensitivity of our main findings with respect to the measurement and definition choices. First, we redefine “treated” from having at least 50 % of area eligible for multi-threshold result-based payments to thresholds of 70 % and 90 %. Our results remain robust to these changes (Fig. S14).

Second, in our main analysis, the size of the observations is 16km². Our results are robust to altering the observation size to 1km², 4km², and 64km² (Fig. S15).

Third, we account for potential earlier anticipation effects by farmers in the 2008 pilot regions, which could arise if most of the treatment effect of the 2008 pilot area already happened in 2007 (note that we merged the 2007 and 2008 areas; see Section 2). To this end, we set 2006 instead of 2007 as the reference year to which the treatment effect is compared. The estimates remain similar to those in our main specification (Fig. S16).

Fourth, instead of our current species richness measure, we make use of the Hill–Shannon index (which considers both abundance and richness) as the measure of diversity. We find that the results are consistent with our main results (Fig. S17).

Additional observed covariates. We re-estimate our model including different sets of observed covariates. First, we consider the possibility that our results are influenced by weather. Specifically, there are two ways in which weather could impact our estimates. On the one hand, weather can impact the intensity of observation. For example, bad weather could deter some observers from going out and, thus, decrease the number of observations in the sample. On the other hand, trends in weather caused by climate change might affect the size and location of bird populations. With these considerations, we re-estimated the main model with common bird diversity as well as the number of volunteers and extended it by including weather. Specifically, following Palmer et al. (2025), we controlled for the time-varying covariates of annual average min and max temperature as well as precipitation, which are by definition exogenous. The results of the estimations with common bird diversity and the number of volunteers as the outcome variables support our main findings (Figs. S18 and S19).

Second, we control for the share of other spatially restricted programs introduced in earlier reforms than the reform that included the introduction of result-based payments. These programs include the “Wild herbs, spatial biotope types scheme” and the “Nordic visiting birds scheme” (Text S1). The results support our main findings (Fig. S20).

Third, so far we have considered nature protection area status as being an indication of high perceived natural value, including all areas established by 2022. In this robustness check, we re-run our estimation considering only those areas that were established in 2008 or earlier. The results support our main findings (Fig. S21).

Fourth, we use the share of grassland instead of agricultural land in 2006 as a pre-treatment covariate to consider that grassland rather than agricultural land more generally might affect our estimates. This change supports our main analysis (Fig. S22).

Fifth, next to pre-treatment grassland yields we include additional variables that may be linked to spatial differences in opportunity costs and grassland yield potential, as well as management intensity information in the area. Specifically, we include pre-treatment

²⁰ We use the CBS data only for the purpose of validation, since while it is systematically selected based on random sampling, it is not suitable for our main analysis due to very low numbers of sampling areas in the pilot areas.

²¹ The reliability might be higher when comparing the differences between spatially closely related regions rather than general changes over time, given the increasing popularity of participation in such citizen observation over time (Knape et al., 2022; Kamp et al., 2016).

information on organic carbon content, soil pH, nitrogen surplus, and cattle density. Our results are robust against the inclusion of these variables (Fig. S23).²²

Assessment of the overlap assumption. The overlap assumption (Callaway and Sant'Anna, 2021) requires that the characteristics of an observation do not deterministically define the treatment status. Specifically, it is necessary to ensure that conditioning on covariates in our conditional difference-in-differences approach is indeed feasible. A violation of the overlap assumption would lead to very large (close to 1) or very small (close to 0) estimates of the propensity score, leading to very large or even infinite standard errors. In order to assess the overlap assumption, we estimate the propensity score (including only the time-invariant covariates of our main estimation) using a logit model. The results shows that this assumption is satisfied (Fig. S26).

Remark 2. One advantage of our empirical approach is that a consistent estimation of the propensity score does not invalidate our estimates. As in the cross-section doubly robust estimation (Knaus, 2022) and in the standard two period-two groups difference-in-differences estimation (Chang, 2020), both the propensity score and the outcome equation are nuisance parameters, and misspecifying one of them is allowed as long as the other one is consistently estimated (the standard double robustness property; see Robins et al., 1994; Scharfstein et al., 1999).

Remark 3. In terms of identification, we do not match on observed covariates as in selection-on-observables methods, such as propensity score matching. We use covariates only to relax the parallel trends assumption. As in the case of unconditional difference-in-differences estimators, the plausibility of this assumption can only be verified with the inspection of pre-treatment placebo effects (Caetano et al., 2022; Heckman et al., 1997, 1998). As discussed above, we provide robust evidence that these pre-treatment trends are parallel.

7. Concluding remarks

Governments across the globe use agri-environmental payments to reduce the loss of biodiversity in agricultural systems, with several objectives including those set under the Kunming–Montreal Global Biodiversity Framework (UN, 2022) and the UN Decade on Ecosystem Restoration (2021–2030) (IUCN, 2022). To increase the impact, result-based payments might be a useful innovation (Burton and Schwarz, 2013; Elmiger et al., 2023; Kelemen et al., 2023; Sattler et al., 2023). However, there is a lack of knowledge about the impact of such result-based approaches in general, and, in particular, we highlight the lack of prior consideration of different payment designs.

We help close that gap by estimating the effects of multi-threshold payments compared to those of single-threshold payments on biodiversity. Thus, we assess the effect of moving the assessment of biodiversity in result-based payments toward a continuous matrix. Bird diversity is used as a measure of biodiversity, which can be seen as a spillover benefit from increasing plant diversity and the associated changes in management required to achieve that increase. We exploit the staggered rollout of a multi-threshold result-based payment scheme in the German federal state of Lower Saxony and citizen-science bird data collected over 16 years, with over 6.7 million data entries.

Our results highlight that multi-threshold result-based payments can increase bird diversity when compared to single-threshold payments and that this effect is only established with a time lag. For the initial pilot area, the average effect over time is about 0.28 standard deviations (90 % confidence interval = [0.00 to 0.55]; 95 % confidence interval = [−0.03 to 0.59]). Related to the post-treatment average of bird diversity, our point estimate indicates an increase of 35 %²³ over a period of six years. However, as discussed in Section 6.1, it is important to account for uncertainty (Imbens, 2021; Romer, 2020). Our lower bounds for the 95 % and 90 % confidence intervals—representing the “statistical worst-case scenario”—are close to zero, yet they are only approximately one-seventh and one-quarter as likely, respectively. Therefore, our findings predict, on average, large gains for biodiversity from a multi-threshold result-based payment scheme and underscore the importance not only of choosing between result- and action-based payments but also of carefully designing result-based payments to increase valued impacts.

Our findings contribute to the existing literature on the effects of results-based payments by combining “in-the-field” measured biodiversity with a causal identification strategy (i.e., accounting for the endogenous selection of fields into schemes). Most previous studies have focused on either “in-the-field” measured biodiversity or on causal identification (e.g., Kleijn and Sutherland, 2003; Roth et al., 2008; Baker et al., 2012; Chabé-Ferret and Subervie, 2013; Meichtry-Stier et al., 2014; Marja et al., 2018; Bertoni et al., 2020; Stetter et al., 2022; Wuepper and Huber, 2022; Zimmert et al., 2024). An exception is Schaub et al. (2025), who investigated a payment increase within a single-threshold scheme in Switzerland and found positive effects on plant diversity—primarily for fields that already nearly met the eligibility criteria prior to the payment change. Wuepper and Huber (2022), analyzing the same policy change as Schaub et al. (2025), showed that a 1 % increase in payment led to a 1 % increase in enrolled area. In contrast to Schaub et al. (2025) and Wuepper and Huber (2022), we examine a newly introduced multi-threshold payment scheme and its effects on bird diversity. Our analysis reveals relatively large biodiversity effects, suggesting potentially greater effectiveness compared to single-threshold schemes.

²² We also use time-varying nitrogen surplus and cattle density information in another model estimation, the results of which support our main findings (Fig. S24). Similarly, we estimate our main model specification using annual instead of pre-treatment grassland yields (see Text S6 for additional data information); this adjustment also supports our results (Fig. S25). However, we note that the inclusion of these time-varying variables could potentially bias our estimates. The reason for this is that nitrogen surplus, cattle density, and grassland yields might be impacted by the treatment themselves and are then “bad controls” (Karim and Webb, 2024).

²³ Or, in absolute terms, 0.4 species per square km.

More generally, the relatively large effect sizes observed in our study stand in contrast to the typically smaller impacts (or associations) found for action-based payments (e.g., Kleijn and Sutherland, 2003; Roth et al., 2008; Baker et al., 2012; Chabé-Ferret and Subervie, 2013; Bertoni et al., 2020; Stetter et al., 2022; Zimmert et al., 2024), which may indicate that multi-threshold results-based payments are more effective in promoting biodiversity outcomes than conventional action-based approaches.

An important question for policymakers is whether the multi-threshold result-based payment scheme is desirable overall. This is a complex question that has several dimensions. The first is the cost-effectiveness of the program in the narrow sense of the ratio of biodiversity output to subsidies. Yet, our theoretical model and empirical evidence suggest that a two-threshold system reduces the number of fields that receive windfall gains and, thus, might be superior to a one-threshold system in terms of cost-effectiveness. However, the overall evaluation depends on the absolute value of windfall gains (i.e., the number of fields receiving windfall gains multiplied by the payment) and, therefore, on the payment levels and the number of fields with indicator species exceeding the different thresholds in a two-threshold system. On the other hand, considering the extra utility of farmers generated by windfall gains complicates the cost-effectiveness evaluation. An overall welfare calculation would require a comparison of this utility and the societal utility generated by higher biodiversity. Such calculations require measuring the societal “biodiversity reservation wage” (see, e.g., Wuepper and Huber, 2022, for an approach to measure willingness to pay for biodiversity). While it is not in the scope of our paper to conduct these calculations, such tradeoffs need to be explicitly accounted for by policymakers in considering whether such programs should be implemented.

Our findings suggest important future research, next to estimating the societal biodiversity reservation wage. First, future studies could identify the costs of multi-compared to single-threshold payments. Second, it would be valuable to investigate how farmers have adjusted their management practices in response to single-versus multi-threshold payment schemes and to examine the roles of landscape features and composition in shaping the impacts of result-based payments. Third, future research could extend our study by looking at other taxa (e.g., plants, insects, and soil microbiomes) because, while birds can provide important insights about the ecosystem state (Gregory et al., 2005; Fraixedas et al., 2020; Li et al., 2020; Lees et al., 2022), they cannot alone convey the complexity of environmental change (e.g., Siddig et al., 2016).

CRediT authorship contribution statement

Sergei Schaub: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexander Pfaff:** Writing – review & editing, Writing – original draft, Methodology. **Petyo Bonev:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

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

Acknowledgement

We thank ornitho.de, the Lower Saxony State Department for Waterway, Coastal and Nature Conservation, and Federation of German Avifaunists for sharing the data. Moreover, we thank Annette Most and Jürgen Ludwig for their generous sharing of time and expertise. We also thank Nadja El Benni and the participants at the Conference of the Swiss Society for Agricultural Economics and Rural Sociology for fruitful discussions and comments. We gratefully acknowledge the support by the Agroscope Research Program “Indicate—Measuring and Optimizing Farm Environmental Impacts.”

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jeem.2025.103247>.

Data availability

The data that has been used is confidential.

References

- Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* 5 (1), 1–12.
- Aceves-Bueno, E., Adeleye, A.S., Feraud, M., Huang, Y., Tao, M., Yang, Y., Anderson, S.E., 2017. The accuracy of citizen science data: a quantitative review. *Bull. Ecol. Soc. Am.* 98 (4), 278–290.
- Allan, E., Bossdorf, O., Dormann, C.F., Prati, D., Gossner, M.M., Tscharntke, T., Fischer, M., 2014. Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl. Acad. Sci.* 111 (1), 308–313.
- Andersson, A., Högård, S., Rabinowicz, E., 2017. Evaluation of results and adaptation of EU rural development programmes. *Land Use Policy* 67, 298–314.

- Baker, D.J., Freeman, S.N., Grice, P.V., Siriwardena, G.M., 2012. Landscape-scale responses of birds to agri-environment management: a test of the English environmental stewardship scheme. *J. Appl. Ecol.* 49 (4), 871–882.
- Báldi, A., Batáry, P., Kleijn, D., 2013. Effects of grazing and biogeographic regions on grassland biodiversity in Hungary—analysing assemblages of 1200 species. *Agric. Ecosyst. Environ.* 166, 28–34.
- Batáry, P., Báldi, A., Erdős, S., 2007. Grassland versus non-grassland bird abundance and diversity in managed grasslands: local, landscape and regional scale effects. *Biodivers. Conserv.* 16, 871–881.
- Batool, M., Sarrazin, F.J., Attinger, S., Basu, N.B., Van Meter, K., Kumar, R., 2022. Long-term annual soil nitrogen surplus across Europe (1850–2019). *Sci. Data* 9 (1), 612.
- Baylis, K., Coppess, J., Gramig, B.M., Sachdeva, P., 2022. Agri-environmental programs in the United States and Canada. *Rev. Environ. Econ. Pol.* 16 (1), 83–104.
- Benton, T.G., Bryant, D.M., Cole, L., Crick, H.Q., 2002. Linking agricultural practice to insect and bird populations: a historical study over three decades. *J. Appl. Ecol.* 39 (4), 673–687.
- Bertoni, D., Curzi, D., Aletti, G., Olper, A., 2020. Estimating the effects of agri-environmental measures using difference-in-difference coarsened exact matching. *Food Policy* 90, 101790.
- Binley, A.D., Hanson, J.O., Robinson, O.J., Golet, G.H., Bennett, J.R., 2025. Quantifying the value of participatory science data for conservation decision-making. *J. Appl. Ecol.*
- Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Frusher, S., 2014. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* 173, 144–154.
- Birrer, S., Spiess, M., Herzog, F., Jenny, M., Kohli, L., Lugin, B., 2007. The Swiss agri-environment scheme promotes farmland birds: but only moderately. *J. Ornithol.* 148, 295–303.
- Burton, R.J., Schwarz, G., 2013. Result-oriented agri-environmental schemes in Europe and their potential for promoting behavioural change. *Land Use Policy* 30 (1), 628–641.
- Busch, M., Katzenberger, J., Trautmann, S., Gerlach, B., Droschmeister, R., Sudfeldt, C., 2020. Drivers of population change in common farmland birds in Germany. *Bird. Conserv. Int.* 30 (3), 335–354.
- Caetano, C., Callaway, B., Payne, S., Rodrigues, H.S.A., 2022. Difference in differences with time-varying covariates. *arXiv preprint arXiv:2202.02903*.
- Caetano, C., Callaway, B., 2024. Difference-in-Differences with time-varying Covariates in the Parallel Trends Assumption. No. 2406.15288.
- Callaghan, C.T., Martin, J.M., Major, R.E., Kingsford, R.T., 2018. Avian monitoring—comparing structured and unstructured citizen science. *Wildl. Res.* 45 (2), 176–184.
- Callaway, B., Sant’Anna, P.H., 2021. Difference-in-differences with multiple time periods. *J. Econom.* 225 (2), 200–230.
- Chabé-Ferret, S., Subervie, J., 2013. How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-Matching. *J. Environ. Econ. Manag.* 65 (1), 12–27.
- Chamberlain, D.E., Fuller, R.J., Bunce, R.G., Duckworth, J.C., Shrubbs, M., 2000. Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *J. Appl. Ecol.* 37 (5), 771–788.
- Chang, N.C., 2020. Double/debiased machine learning for difference-in-differences models. *Econom. J.* 23 (2), 177–191.
- Coderoni, S., Esposti, R., Varacca, A., 2023. How differently do farms respond to agri-environmental policies? A probabilistic machine-learning approach. *Land Econ.* 100 (2), 370–397.
- Copernicus Land Monitoring Service, 2020. CORINE Land Cover 2006 (vector/raster 100 M), pp. 6–yearly. Brussels, Belgium.
- Cox, T.E., Philippoff, J., Baumgartner, E., Smith, C.M., 2012. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecol. Appl.* 22 (4), 1201–1212.
- Cox, D.R., 2020. Statistical significance. *Annual Review of Statistics and Its Application* 7 (1), 1–10.
- Détang-Dessendre, C., Guyomard, H., Mauguin, P., 2023. Evolving the Common Agricultural Policy for Tomorrow’s Challenges. Quae, Versailles Cedex, France.
- Dickel, R., Reiter, K., Roggendorf, W., Sander, A., 2010. Halbzeitbewertung Von PROFIL - Zahlungen Für Agrarumweltmaßnahmen. Braunschweig/Hannover, Germany.
- Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural intensification and the collapse of Europe’s farmland bird populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268 (1462), 25–29.
- Dormann, C.F., Schweiger, O., Arens, P., Augenstein, I., Aviron, S.T., Bailey, D., Zobel, M., 2008. Prediction uncertainty of environmental change effects on temperate European biodiversity. *Ecol. Lett.* 11 (3), 235–244.
- Du, Z., Yu, L., Zhao, Y., Li, X., Liu, X., Li, X., Wang, H., 2025. Annual global gridded livestock mapping from 1961 to 2021. *Earth Syst. Sci. Data Discuss.* 2025, 1–20.
- Elmiger, N., Finger, R., Ghazoul, J., Schaub, S., 2023. Biodiversity indicators for result-based agri-environmental schemes—current state and future prospects. *Agric. Syst.* 204, 103538.
- Englin, J., Shonkwiler, J.S., 1995. Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation. *Rev. Econ. Stat.* 104–112.
- Ernault, A., Tremauville, Y., Cellier, D., Margerie, P., Langlois, E., Alard, D., 2006. Potential landscape drivers of biodiversity components in a flood plain: past or present patterns? *Biol. Conserv.* 127 (1), 1–17.
- European Commission, Directorate-General for Agriculture and Rural Development, Folkeson Lillo, C., Chartier, O., Valli, C., Jongeneel, R., Selten, M., van Asseldonk, M., Avis, K., Rouillard, J., Underwood, E., Parissaki, M., Bertolozzi, D., Devot, A., 2023. Mapping and Analysis of CAP Strategic Plans: Assessment of Joint Efforts for 2023–2027. Publications Office of the European Union, Brussels, Belgium.
- Federal Agricultural Research Center, 2005. Aktualisierung Der Halbzeitbewertung Von ProLand Niedersachsen. Braunschweig, Germany.
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Zaks, D.P., 2011. Solutions for a cultivated planet. *Nature* 478 (7369), 337–342.
- Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.Y., Danielsen, F., Haklay, M., 2022. Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers* 2 (1), 64.
- Fraixedas, S., Lindén, A., Piha, M., Cabeza, M., Gregory, R., Lehtikoinen, A., 2020. A state-of-the-art review on birds as indicators of biodiversity: advances, challenges, and future directions. *Ecol. Indic.* 118, 106728.
- Gómez-Limón, J.A., Gutiérrez-Martín, C., Villanueva, A.J., 2019. Optimal design of agri-environmental schemes under asymmetric information for improving farmland biodiversity. *J. Agric. Econ.* 70 (1), 153–177.
- Graham, L., Gaulton, R., Gerard, F., Staley, J.T., 2018. The influence of hedgerow structural condition on wildlife habitat provision in farmed landscapes. *Biol. Conserv.* 220, 122–131.
- Gregory, R.D., Van Strien, A., Vorisek, P., Gmelig Meyling, A.W., Noble, D.G., Foppen, R.P., Gibbons, D.W., 2005. Developing indicators for European birds. *Phil. Trans. Biol. Sci.* 360 (1454), 269–288.
- Hart, R., Latacz-Lohmann, U., 2005. Combating moral hazard in agri-environmental schemes: a multiple-agent approach. *Eur. Rev. Agric. Econ.* 32 (1), 75–91.
- Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Rev. Econ. Stud.* 64 (4), 605–654.
- Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. *Econometrica* 66 (5), 1017–1098.
- Henckel, L., Bradter, U., Jönsson, M., Isaac, N.J., Snäll, T., 2020. Assessing the usefulness of citizen science data for habitat suitability modelling: opportunistic reporting versus sampling based on a systematic protocol. *Divers. Distrib.* 26 (10), 1276–1290.
- Hertzog, L.R., Frank, C., Klimke, S., Röder, N., Böhner, H.G., Kamp, J., 2021. Model-based integration of citizen science data from disparate sources increases the precision of bird population trends. *Divers. Distrib.* 27 (6), 1106–1119.
- Hindsley, P., Landry, C.E., Gentner, B., 2011. Addressing onsite sampling in recreation site choice models. *J. Environ. Econ. Manag.* 62 (1), 95–110.
- Hoyer, M.V., Wellendorf, N., Frydenborg, R., Bartlett, D., Canfield Jr, D.E., 2012. A comparison between professionally (florida department of environmental protection) and volunteer (Florida LAKEWATCH) collected trophic state chemistry data in Florida. *Lake Reserv. Manag.* 28 (4), 277–281.

- Hynes, S., Garvey, E., 2009. Modelling farmers' participation in an agri-environmental scheme using panel data: an application to the rural environment protection scheme in Ireland. *J. Agric. Econ.* 60 (3), 546–562.
- Imai, K., Keele, L., Yamamoto, T., 2010. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* 25 (1), 51–71.
- Imbens, G.W., 1992. An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica: J. Econom. Soc.* 1187–1214.
- Imbens, G.W., 2021. Statistical significance, p-values, and the reporting of uncertainty. *J. Econ. Perspect.* 35 (3), 157–174.
- IRENES, 2020. IRENES Case Study Factsheet: Lower Saxony, Germany. Interreg Europe. Available at https://projects2014-2020.interregurope.eu/fileadmin/user_upload/tx_tevprojects/library/file_1608202776.pdf.
- IUCN, 2022. Decade on Ecosystem Restoration 2021–2030. Gland, Switzerland.
- Jeanneret, P., Baumgartner, D.U., Knuchel, R.F., Koch, B., Gaillard, G., 2014. An expert system for integrating biodiversity into agricultural life-cycle assessment. *Ecol. Indic.* 46, 224–231.
- Johnston, A., Moran, N., Musgrove, A., Fink, D., Baillie, S.R., 2020. Estimating species distributions from spatially biased citizen science data. *Ecol. Model.* 422, 108927.
- Jones, R.J., Hiederer, R., Rusco, E., Montanarella, L., 2005. Estimating organic carbon in the soils of Europe for policy support. *Eur. J. Soil Sci.* 56 (5), 655–671.
- Kamp, J., Frank, C., Trautmann, S., Busch, M., Dröschmeister, R., Flade, M., Sudfeldt, C., 2021. Population trends of common breeding birds in Germany 1990–2018. *J. Ornithol.* 162 (1), 1–15.
- Karim, S., Webb, M.D., 2024. Good controls gone bad: difference-in-differences with covariates. *arXiv preprint arXiv:2412.14447*.
- Kelemen, E., Megyesi, B., Matzdorf, B., Andersen, E., van Bussel, L.G., Dumortier, M., Yacamán-Ochoa, C., 2023. The prospects of innovative agri-environmental contracts in the European policy context: results from a Delphi study. *Land Use Policy* 131, 106706.
- Kleijn, D., Berendse, F., Smit, R., Gilissen, N., 2001. Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. *Nature* 413 (6857), 723–725.
- Kleijn, D., Sutherland, W.J., 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? *J. Appl. Ecol.* 40 (6), 947–969.
- Kleijn, D., van Zuijlen, G.J., 2004. The conservation effects of meadow bird agreements on farmland in Zeeland, the Netherlands, in the period 1989–1995. *Biol. Conserv.* 117 (4), 443–451.
- Kleijn, D., Baquero, R.A., Clough, Y., Díaz, M., De Esteban, J., Fernández, F., et al., 2006. Mixed biodiversity benefits of agri-environment schemes in five European countries. *Ecol. Lett.* 9 (3), 243–254.
- Kleijn, D., Rundlöf, M., Scheper, J., Smith, H.G., Tschernitz, T., 2011. Does conservation on farmland contribute to halting the biodiversity decline? *Trends Ecol. Evol.* 26 (9), 474–481.
- Knappe, J., Coulson, S.J., van der Wal, R., Arlt, D., 2022. Temporal trends in opportunistic citizen science reports across multiple taxa. *Ambio* 1–16.
- Knaus, M.C., 2022. Double machine learning-based programme evaluation under unconfoundedness. *Econom. J.* 25 (3), 602–627.
- Knop, E.V.A., Kleijn, D., Herzog, F., Schmid, B., 2006. Effectiveness of the Swiss agri-environment scheme in promoting biodiversity. *J. Appl. Ecol.* 43 (1), 120–127.
- Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. *Front. Ecol. Environ.* 14 (10), 551–560.
- Krause, B., Culmes, H., 2013. The significance of habitat continuity and current management on the compositional and functional diversity of grasslands in the uplands of Lower Saxony, Germany. *Flora-Morphology, Distribution, Functional Ecology of Plants* 208 (5–6), 299–311.
- Kuriyama, K., Hilger, J., Hanemann, M., 2013. A random parameter model with onsite sampling for recreation site choice: an application to Southern California shoreline sportfishing. *Environ. Resour. Econ.* 56 (4), 481–497.
- Leclère, D., Obersteiner, M., Barrett, M., Butchart, S.H., Chaudhary, A., De Palma, A., DeClerck, F.A., et al., 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* 585 (7826), 551–555.
- Lees, A.C., Haskell, L., Allinson, T., Bezeng, S.B., Burfield, I.J., Renjifo, L.M., Butchart, S.H., 2022. State of the world's birds. *Annu. Rev. Environ. Resour.* 47, 231–260.
- Li, Y., Miao, R., Khanna, M., 2020. Neonicotinoids and decline in bird biodiversity in the United States. *Nat. Sustain.* 3 (12), 1027–1035.
- Lomba, A., Strohbach, M., Jerrentrup, J.S., Dauber, J., Klimek, S., McCracken, D.I., 2017. Making the best of both worlds: can high-resolution agricultural administrative data support the assessment of high nature value farmlands across Europe? *Ecol. Indic.* 72, 118–130.
- Lower Saxony Ministry of Food, Agriculture and Consumer Protection, 2019. The agriculture and food industry in Lower Saxony. https://www.mg-niedersachsen.de/fileadmin/Downloads/D_Weltweit_Maerkte_erschliessen/mgn_broschuere_international_EN_web.pdf.
- Lower Saxony State Department for Waterway, Coastal and Nature Conservation, 2023. Naturschutzrechtlich besonders geschützte Teile von Natur und Landschaft. https://www.umwelt.niedersachsen.de/startseite/service/umweltkarten/natur_amp_landschaft/besonders_geschuetzte_teile_von_natur_und_landschaft/naturschutzrechtlich-besonders-geschuetzte-teile-von-natur-und-landschaft-9065.html.
- Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. *Econometrica: J. Econom. Soc.* 1977–1988.
- Marcus, M., Sant'Anna, P.H., 2021. The role of parallel trends in event study settings: an application to environmental economics. *Journal of the Association of Environmental and Resource Economists* 8 (2), 235–275.
- Marja, R., Viik, E., Mänd, M., Phillips, J., Klein, A.M., Batáry, P., 2018. Crop rotation and agri-environment schemes determine bumblebee communities via flower resources. *J. Appl. Ecol.* 55 (4), 1714–1724.
- Meichtry-Stier, K.S., Jenny, M., Zellweger-Fischer, J., Birrer, S., 2014. Impact of landscape improvement by agri-environment scheme options on densities of characteristic farmland bird species and brown hare (*Lepus europaeus*). *Agric. Ecosyst. Environ.* 189, 101–109.
- Mitschke, A., Sudfeldt, C., Heidrich-Riske, H., Dröschmeister, R., 2005. Das neue Brutvogelmonitoring in der Normallandschaft Deutschlands—Untersuchungsgebiete, Erfassungsmethode und erste Ergebnisse. *Vogelwelt* 126, 127–140.
- Moeltner, K., Shonkwiler, J.S., 2005. Correcting for on-site sampling in random utility models. *Am. J. Agric. Econ.* 87 (2), 327–339.
- Montgomery, I., Caruso, T., Reid, N., 2020. Hedgerows as ecosystems: service delivery, management, and restoration. *Annual Review of Ecology, Evolution, and Systematics* 51, 81–102.
- Moret-Bailly, S., Muro, M., 2024. The Costs and Benefits of Transitioning to Sustainable Agriculture in the EU – a Synthesis of Existing Knowledge. Institute for European Environmental Policy. <https://ieep.eu/publications/the-costs-and-benefits-of-transitioning-to-sustainable-agriculture-in-the-eu-a-synthesis-of-existing-knowledge/>.
- Most, A., Keenburger, T., 2006. Entwicklung und Erprobung von Methoden für die ergebnisorientierte Honorierung ökologischer Leistungen im Grünland Nordwestdeutschlands. *Ökonomie Der Honorierung Ökologischer Leistungen*. Norddeutsche Naturschutzakademie. Schneverdingen, Germany.
- Most, A., Heuser, R., Jacob, A., Jaschinski, T., Ludwig, J., Mexner, A., Peerenboom, C., Pelzer, A., Rasper, M., Sandkühler, K., Schacherer, A., Schader, P.G., Thorwest, A., 2015. Wirkungen Des Kooperationsprogramms Naturschutz Und Weiterer Niedersächsischer Und Bremer Agrarumweltmaßnahmen Auf Die Biodiversität. Norden, Germany.
- Müller, S., Gossner, M.M., Penone, C., Jung, K., Renner, S.C., Farina, A., Scherer-Lorenzen, M., 2022. Land-use intensity and landscape structure drive the acoustic composition of grasslands. *Agric. Ecosyst. Environ.* 328, 107845.
- Navarro, A., López-Bao, J.V., 2019. EU agricultural policy still not green. *Nat. Sustain.* 2 (11), 990.
- Neff, F., Korner-Nievergelt, F., Rey, E., Albrecht, M., Bollmann, K., Cahenzli, F., Knop, E., 2022. Different roles of concurring climate and regional land-use changes in past 40 years' insect trends. *Nat. Commun.* 13 (1), 7611.
- Noack, F., Larsen, A., Kamp, J., Levers, C., 2022. A bird's eye view of farm size and biodiversity: the ecological legacy of the iron curtain. *Am. J. Agric. Econ.* 104 (4), 1460–1484.
- NUMIS, 2002. Zonierung Elbtalau. Hannover, Germany.
- Palmer, C., Groom, B., Silici, L., Langton, S., 2025. Biodiversity–food trade-offs when agricultural land is spared from production. *Am. J. Agric. Econ.*
- Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Montanarella, L., 2022. European Soil data centre 2.0: Soil data and knowledge in support of the EU policies. *Eur. J. Soil Sci.* 73 (6), e13315.
- Pannell, D., Rogers, A., 2022. Agriculture and the environment: policy approaches in Australia and New Zealand. *Rev. Environ. Econ. Pol.* 16 (1), 126–145.
- Pearl, J., 2012. Interpretable conditions for identifying direct and indirect effects. UCLA Cognitive Systems Laboratory Technical Report.

- Pe'er, G., Dicks, L.V., Visconti, P., Arlettaz, R., Báldi, A., Benton, T.G., et al., 2014. EU agricultural reform fails on biodiversity. *Science* 344 (6188), 1090–1092.
- Pe'er, G., Zinngrebe, Y., Moreira, F., Sirami, C., Schindler, S., Müller, R., Lakner, S., 2019. A greener path for the EU common agricultural policy. *Science* 365 (6452), 449–451.
- Pe'er, G., Bonn, A., Bruelheide, H., Dieker, P., Eisenhauer, N., Feindt, P.H., Lakner, S., 2020. Action needed for the EU common agricultural policy to address sustainability challenges. *People and Nature* 2 (2), 305–316.
- Pe'er, G., Finn, J.A., Díaz, M., Birkenstock, M., Lakner, S., Röder, N., Guyomard, H., 2022. How can the European common agricultural policy help halt biodiversity loss? Recommendations by over 300 experts. *Conserv. Lett.* 15 (6), e12901.
- Reuter, H.I., Lado, L.R., Hengl, T., Montanarella, L., 2008. Continental-scale digital soil mapping using European soil profile data: soil pH. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie* 19 (1), 91–102.
- Rigal, S., Dakos, V., Alonso, H., Aunin, A., Benkő, Z., Brotons, L., Devictor, V., 2023. Farmland practices are driving bird population decline across Europe. *Proc. Natl. Acad. Sci.* 120 (21), e2216573120.
- Robins, J.M., Rotnitzky, A., Zhao, L.P., 1994. Estimation of regression coefficients when some regressors are not always observed. *J. Am. Stat. Assoc.* 89 (427), 846–866.
- Romer, D., 2020. In praise of confidence intervals. *AEA Papers and Proceedings* 110, 55–60.
- Roswell, M., Dushoff, J., Winfree, R., 2021. A conceptual guide to measuring species diversity. *Oikos* 130 (3), 321–338.
- Roth, T., Amrhein, V., Peter, B., Weber, D., 2008. A Swiss agri-environment scheme effectively enhances species richness for some taxa over time. *Agric. Ecosyst. Environ.* 125 (1–4), 167–172.
- Sant'Anna, P.H., Zhao, J., 2020. Doubly robust difference-in-differences estimators. *J. Econom.* 219 (1), 101–122.
- Sattler, C., Barghusen, R., Bredemeier, B., Dutilly, C., Prager, K., 2023. Institutional analysis of actors involved in the governance of innovative contracts for agri-environmental and climate schemes. *Glob. Environ. Change* 80, 102668.
- Scharfstein, D.O., Rotnitzky, A., Robins, J.M., 1999. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Am. Stat. Assoc.* 94 (448), 1096–1120.
- Schaub, S., Ghazoul, J., Huber, R., Zhang, W., Sander, A., Rees, C., Finger, R., 2023. The role of behavioural factors and opportunity costs in farmers' participation in voluntary agri-environmental schemes: a systematic review. *J. Agric. Econ.* 74 (3), 617–660.
- Schaub, S., Roth, T., Bonev, P., 2025. The effect of result-based agri-environmental payments on biodiversity: evidence from Switzerland. *Am. J. Agric. Econ.*
- Schüler, S., Noack, E.M., 2019. Does the CAP reflect the population's concerns about agricultural landscapes? A qualitative study in Lower Saxony, Germany. *Land Use Policy* 83, 240–255.
- Schultz, C.B., Brown, L.M., Pelton, E., Crone, E.E., 2017. Citizen science monitoring demonstrates dramatic declines of monarch butterflies in western North America. *Biol. Conserv.* 214, 343–346.
- Shaw, D., 1988. On-site samples' regression: problems of non-negative integers, truncation, and endogenous stratification. *J. Econom.* 37 (2), 211–223.
- Siddig, A.A., Ellison, A.M., Ochs, A., Villar-Leeman, C., Lau, M.K., 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in ecological indicators. *Ecol. Indic.* 60, 223–230.
- Socher, S.A., Prati, D., Boch, S., Müller, J., Klaus, V.H., Hölzel, N., Fischer, M., 2012. Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *J. Ecol.* 100 (6), 1391–1399.
- State Office for Statistics of Lower Saxony (Landesamt für Statistik Niedersachsen), 2023. Bevölkerungsstand: Einwohnerzahl Niedersachsens - Tabellen. Hannover, Germany.
- State Office for Statistics of Lower Saxony (Landesamt für Statistik Niedersachsen), 2024. Erntestatistik Online - Ernteergebnisse in Niedersachsen Seit 1991. Hannover, Germany.
- Statistical Offices of the Federal Government and the States, 2015. Deutschlandweite Ergebnisse Des Zensus 2011 Im 1km-Gitter. Wiesbaden, Germany.
- Stetter, C., Mennig, P., Sauer, J., 2022. Using machine learning to identify heterogeneous impacts of agri-environment schemes in the EU: a case study. *Eur. Rev. Agric. Econ.* 49 (4), 723–759.
- Tiago, P., Pereira, H.M., Capinha, C., 2017. Using citizen science data to estimate climatic niches and species distributions. *Basic Appl. Ecol.* 20, 75–85.
- Tietz, A., Grajewski, R., Anter, J., Bathke, M., Bergschmidt, A., Dickel, R., Eberhardt, W., Ebers, H., Fahrman, B., Fengler, B., Forstner, B., Franz, K., Moser, A., Pufahl, A., Raue, P., Reiter, K., Roggendorf, W., Sander, A., Schwarz, G., Spengler, M., 2006. Ex-post-Bewertung PROFIL – Programm Zur Förderung Im Ländlichen Raum Niedersachsen Und Bremen 2007 Bis 2013.
- Uezu, A., Metzger, J.P., 2016. Time-lag in responses of birds to Atlantic forest fragmentation: restoration opportunity and urgency. *PLoS One* 11 (1), e0147909.
- UN, 2022. Adoption of the "Kunming-Montreal Global Biodiversity Framework. Canada, Montreal. <https://www.unep.org/news-and-stories/statements/adoption-kunming-montreal-global-biodiversity-framework-gbf>.
- Vickery, J.A., Bradbury, R.B., Henderson, I.G., Eaton, M.A., Grice, P.V., 2004. The role of agri-environment schemes and farm management practices in reversing the decline of farmland birds in England. *Biol. Conserv.* 119 (1), 19–39.
- Wasserstein, R.L., Lazar, N.A., 2016. The ASA statement on p-values: context, process, and purpose. *Am. Statistician* 70 (2), 129–133.
- Wauchope, H.S., Jones, J.P., Geldmann, J., Simmons, B.I., Amano, T., Blanco, D.E., Sutherland, W.J., 2022. Protected areas have a mixed impact on waterbirds, but management helps. *Nature* 605 (7908), 103–107.
- Wooldridge, J.M., 2002. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Port. Econ. J.* 1 (2), 117–139.
- Wuepper, D., Huber, R., 2022. Comparing effectiveness and return on investment of action-and results-based agri-environmental payments in Switzerland. *Am. J. Agric. Econ.* 104 (5), 1585–1604.
- Zimmert, F., Jan, P., Bonev, P., 2024. Participation in biodiversity schemes and environmental performance: overall farm-level impact and spillover effects on non-enrolled land. *Eur. Rev. Agric. Econ.* 51 (3), 690–724.

Reference:

- Schaub, S., Pfaff, A., Bonev, P., 2025. Biodiversity and the Design of Result-based Payments: Evidence from Germany. University of St. Gallen, School of Economics and Political Science. Discussion Paper no. 2025-02. <https://ux-tauri.unisg.ch/RePEc/usc/econwp/EWP-2502.pdf>.