Universität St.Gallen

Biodiversity and the Design of Result-based Payments: Evidence from Germany

Sergei Schaub, Alexander Pfaff, and Petyo Bonev

January 2025 Discussion Paper no. 2025-02

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor:	Mirela Keuschnigg University of St.Gallen School of Economics and Political Science Department of Economics Müller-Friedberg-Strasse 6/8 CH-9000 St.Gallen Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Müller-Friedberg-Strasse 6/8 CH-9000 St.Gallen
Electronic Publication:	http://www.seps.unisg.ch

Biodiversity and the Design of Result-based Payments: Evidence from Germany[1]

Sergei Schaub, Alexander Pfaff, and Petyo Bonev

Author's address: Dr. Sergei Schaub
Managerial Economics, Agroscope
Tänikon 1
8356 Ettenhausen
Email sergei.schaub@agroscope.admin.ch

[1] 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."

Affiliations of Sergei Schaub (corresponding author): Managerial Economics, Agroscope, Switzerland; Alexander Pfaff: Sanford School of Public Policy, Duke University, USA; Petyo Bonev: Managerial Economics, Agroscope, Switzerland and Swiss Institute for Empirical Economic Research, University of St.Gallen, Switzerland.

Abstract

Paying farmers for measured outcomes – i.e., results, not actions − is promoted for reducing risk and raising flexibility in addressing agriculture's environmental damages. One key design choice is how exactly to reward those measured results. Continuous rewards are possible yet, in practice, observed species outcomes have been rewarded using a single threshold (compliant/not) 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 − raised bird diversity. We study a pilot scheme in Germany's Lower Saxony, where an incentive with one threshold is the baseline. Using citizen-science bird data (offering over 6.7m entries across 16 years), 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 findings show potential for benefits from even small shifts in incentive designs.

Keywords

Agricultural policy, Policy design, Agri-environmental payments, Results-based payments, Biodiversity, Birds.

JEL Classification

Q15, Q18, Q57

1. Introduction

Intensive agriculture threatens biodiversity (Foley et al., 2011; Pe'er et al., 2014; Leclère et al., 2020). In response, governments have implemented agri-environmental policies worldwide 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, globally, including for instance 15.4b Euros spent by the European Union (EU) in 2023 to support "green measures" in the agriculture sector (European Commission et al., 2023) to move toward achieving international commitments under the Kunming-Montreal Global Biodiversity Framework (UN, 2022) and the UN Decade on Ecosystem Restoration (2021–2030) (IUCN, 2022). An important such tool is agri-environmental schemes. To date, these schemes mostly reward farmers for the implementation of desired practices, though some also include result-based payments. However, the payments' effectiveness is questioned, in general, and potential gains from shifting design have been 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 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 any farmer's field is eligible for result-based payments. For example, farmers are paid if at least 4 of 40 listed, desired species are present in a given field. To increase the chance of being paid, farmers might adjust a field's management to make it more "flora-friendly", which can also lead to spillover benefits for different taxa, including birds (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012).

A key policy-design choice, in the context of result-based payments, is how exactly to reward results. Schemes can reward simply the achievement of a single threshold (like 4 out of 40 listed species). Yet they could instead offer different rewards for different targets, i.e., move toward continuity in rewards by paying more for higher numbers of species (e.g., out of 40 listed species, more for 6 than for 4).

Yet, despite the push to use payments based on measured desired results, there are not many such schemes and little is known about their impact on biodiversity (see Schaub et al. 2025 on an early Swiss policy) or about the effects of different designs. To our knowledge, additional impact from using multiple thresholds in result-based payments to date has not been studied.

We offer the first empirical study on the impact of multiple thresholds in result-based payments. Specifically, we compare a *multiple-thresholds design* with a *single-threshold design* that uses the same low payment hurdle but no additional hurdles or additional rewards. These designs co-existed in the German federal state of Lower Saxony, the focus of our study. We highlight two key challenges present when analyzing 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 for 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 policy, can increase bird diversity directly and indirectly (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012). On the direct side, the induced changes in field management can positively affect bird diversity. On the indirect side, the policy could lead to an increase in plant diversity, which can 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 on higher-ranks of food chains makes them sensitive to changes in other tropic levels, such as insects (Gregory et al., 2005; Fraixedas et al., 2020; Li et al., 2020; Lees et al., 2022). Thus, bird diversity is interesting in its own right.

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

One major advantage of our setup is that payments were introduced in specific large regions. This allows us to assess impact 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]). As this effect adds to any impacts from single-threshold result-based payments, this is a lower bound on the overall impact of multi-threshold payments relative to a baseline with no payments. Our findings highlight the importance of the design of result-based payments, going beyond the binary choice between action-based 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 some fields in schemes (e.g., Kleijn and Sutherland, 2003; Roth et al., 2008; Baker et al., 2012; Meichtry-Stier et al., 2014; Marja et al., 2018). Studies that take into account the endogenous selection of fields do so by matching 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). These have focused on action-based payments 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 a single-threshold payment scheme in Switzerland, based upon measured results for plant diversity. Comparing with fields already eligible for payments, i.e., facing no change in incentives, they find that the payment rise triggered by policy reform raised plant diversity for fields whose measured species before the reform were just below the threshold (of 6 indicator plant species). Those fields likely faced lower costs of adjustment.

Several studies using proxies for biodiversity seek to account for 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 temporal and spatial availability of data easily linked at plot or farm level. Yet while these studies are important for understanding uptake, and input decisions[1], 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).

[1] While Zimmert et al. (2024) use a biodiversity score, which is a useful extension of most other related studies when it comes to approximating biodiversity, the score is not an *in-situ* measure of biodiversity but is based on a set of management practices (Jeanneret et al., 2014).

Finally, we contribute to literature on citizen-science data and drivers of changes in biodiversity. Citizen science is an emerging and valuable tool, complementing “traditional” data sources, that can extend temporal and spatial dimensions of ecological datasets (Kosmala et al., 2016; Fraisl et al., 2022). Such data have been, for example, successfully used to 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 the influences of farms' sizes and protected areas on biodiversity (Noack et al., 2022; Wauchope et al., 2022). To our knowledge, however, these data have not been used to assess causal effects of agri-environmental payments on biodiversity.

Below, Section 2 offers policy background, while Section 3 describes the data that we employ. Section 4 lays out our empirical approach, Section 5 provides results, and Section 6 discusses.

2. Policy background

We study multi-threshold versus single-threshold result-based agri-environmental payments in the German state of Lower Saxony, which is Germany's second-largest state (at ~48k km²), larger than say Switzerland or the Netherlands (State Office for Statistics of Lower Saxony, 2023). Over half of its 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 varied climatic and soil settings (Lomba et al., 2017). Studying agri-environmental payments in Lower Saxony is interesting precisely 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 co-financed 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 the 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 (Figure 1.), providing a unique opportunity to study their effects on biodiversity.

The result-based payment scheme had two "layers" (or target thresholds driving payments). The first was introduced in the entire state in 2007 (that is NAU/BAU B2). Farmers were eligible for payments of 110 Euro 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, which are referred to as indicator plant species because they are considered to be linked to certain habitat compositions and biodiversity (Most and Keienburg, 2006).[2]

The second layer was introduced 2007/08, in pilot regions (KoopNat FM 411; Figure 1.).[3] A field was eligible for an extra payment of 110 Euro ha^{-1} $year^{-1}$, added 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 (Figure 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.

[2] For the detailed list, see Table S1 in the Supplementary Information.

[3] We know the pilot boundaries in 2008 but not 2007, so we aggregated those regions and their treatment timing. Because most of the area started in 2008 (Most et al., 2015), we use 2008 as reference time for the treatment.

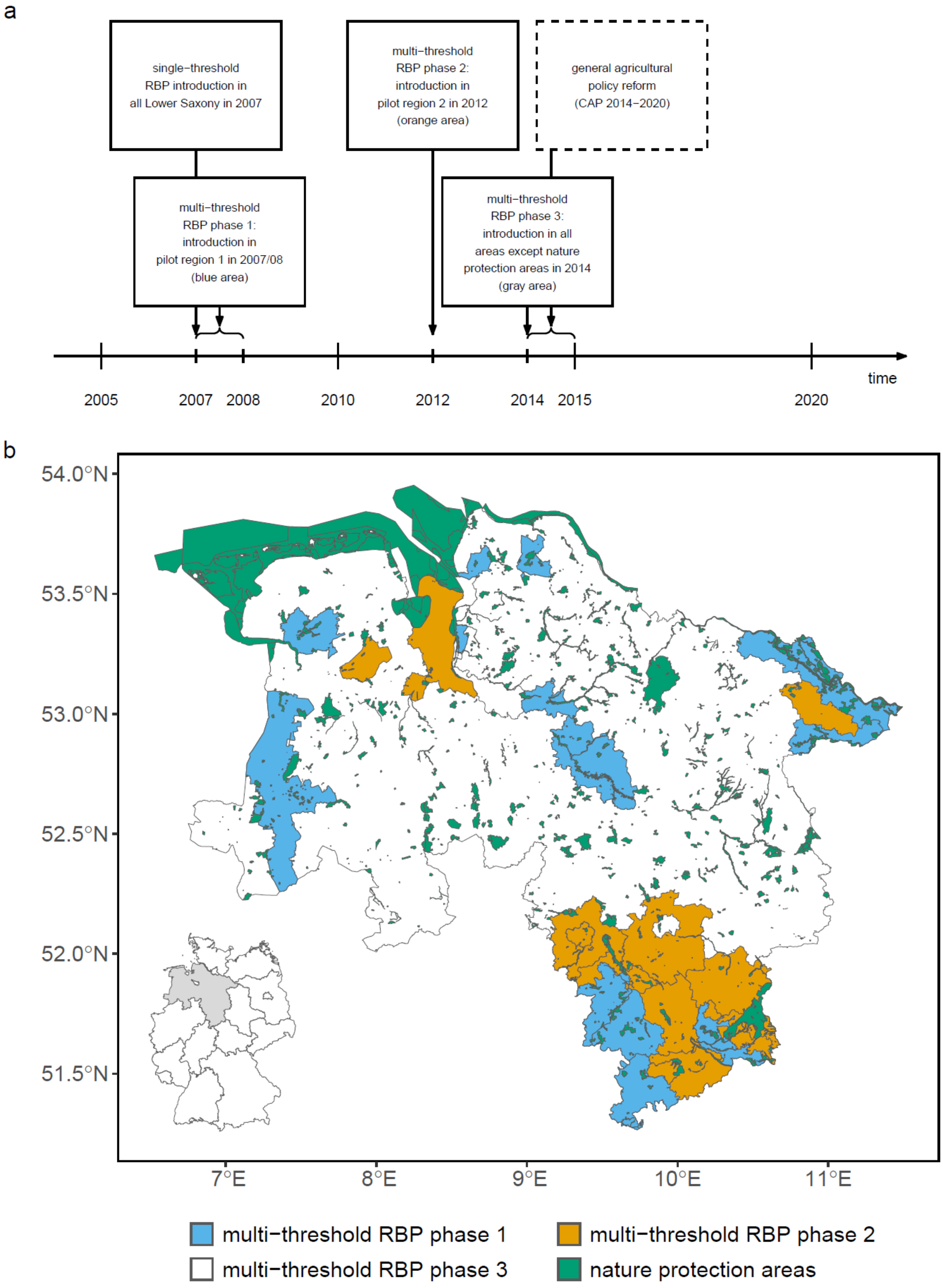

Figure 1. Temporal (panel a) and spatial (panel b) development of the result-based payment in Lower Saxony. RBP = results-based payments. Nature protection areas are here defined as those not eligible for multi-threshold result-based payments in phase 3 and include nature reserves, biosphere reserves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

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 pilot regions. Yet the announcement of which regions were eligible was only 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.

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 of time before 2014. Further, we note that other spatially restricted agri-environmental payments in Lower Saxony existed (Text S1 for details), for which we account in the analyses.

3. Conceptual framework

We present in this section our conception of the cost-benefit tradeoffs involved in farmers' responses to a single-threshold, initially, and then to a shift to a multi-threshold result-based payment scheme. Abstracting from the staggered implementation of multi-threshold payments in Lower Saxony, to keep this model simple we assume a second scheme is introduced at a single point in time and, further, that every farmer has only one field.[4] We 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 $c * (n - n_0)$.[5] Thus,

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

[5] When it is about reaching at least n, then these linear costs are $c * (n - n_0) * \mathbb{1}[n_0 < n]$, where $\mathbb{1}[\cdot]$ 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.

a farmer's profit from trying to achieve at least the single target threshold is $R_1 - c * (T_1 - n_0)$.[6]

This gain function implies that when maximizing profits farmers each will choose their $n^*_{1,}$, the endogenous number of indicator species optimal for them under a single-threshold payment.[7] With low enough R_1, no farmer will change management practices (e.g., reduce number of cuts, reduce fertilizer application, or overseed) to achieve that single threshold. Yet, with a high enough R_1, any farmer who had not yet reached the target T_1 would make changes. In between, only farmers with higher pre-payments numbers of species (but not above T_1) and, thus, lower adjustment costs will respond to the introduction of the payment by changing management. In sum, farmers below the target and with adjustment costs smaller than payment gains respond. (Farmers with pre-payment indicator species diversity above T_1 can get the payment without adjusting management.)

Next, we consider a shift from a single-threshold payment scheme to a multi-threshold scheme that adds a second payment R_2 at threshold or target T_2. As the multi-threshold scheme follows a single-threshold scheme, further decisions are analogous to those for the initial management choice, simply replacing the initial species (n_0) with the optimal species defined under 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., $c * \left(T_2 - n^*_{1,}\right)$. 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 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, these farmers who optimally choose to manage to try to reach T_2 to get R_2

[6] Farmers are finally eligible for payment when $n \geq T_1$.

[7] This could be above T_1, of course, if $n_0 \geq T_1$, which would imply a farmer face's zero cost to get R_1.

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. Perhaps it will be only the farmers who got R_1, via either high initial pre-payments 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 do field-management adjustments.

We highlight here that, as discussed in *Section 1*, we assume that any increase in plant diversity, in response to payments, can increase bird diversity (Kleijn and Sutherland, 2003; Vickery et al., 2004; Birrer et al., 2007; Baker et al., 2012).

In sum, this conceptual model highlights that multi- compared to single-threshold payments can lead to: (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. In doing so, we flag policy-design issues critical for scheme impact, such as the reward levels and targets. We note too that impacts also depend upon the species numbers pre-policy. Below, we empirically assess how the introduction of multi- versus single-threshold payments affected biodiversity beyond the plant diversity, 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, ~ 6000 volunteers recorded roughly 6.7 million entries in Lower Saxony.[8] We assign each record to a 16 km^2 unit of observation, which aggregates upward from the original 1x1km^2 sampling in ornitho data.[9] Birds are mobile, so effective management of any field will affect bird observations not only on

---

[8] We note that in autumn of 2011 ornitho.de (the online platform) was launched. The data before 2011 was retrospectively entered.
[9] The sampling grid by ornitho (following the German "Halbminutenfelder") is approximately 1 × 1 km, i.e., 1 km^2.

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 were 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 km^2 up to 68 km^2 (Figure S1).

We excluded the observations on the East Frisian Islands, inherently different to the mainland. We also deleted 276 entries for which we do not know the coordinates (1x1km^2 grid) of the observation but only 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 with a sighting. Using that, we can calculate the number of volunteers per spatial observational unit, which we later use as in our estimation as time-varying covariate.

4.2 Treatment data

To identify when an area was exposed to the 'treatment', i.e., when an area become 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 then also vary this treatment definition (i.e., % in pilot area) in sensitivity analyses to also examine "at least 70%" and "at least 90%".

4.3 Bird species outcome definition

We focus on common bird species in studying effects of shifting to multi-threshold payments. We expect this is where agri-environmental result-based payments might have an influence. Moreover, volunteers might be better in 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 to 2020 in the German Common Bird Monitoring Survey (Mitschke et al., 2005; Table S3).

In addition to *common* birds, we test the effects on *all* 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 effects on the Hill–Shannon index, which in addition takes into account the species' abundances (Roswell et al., 2021; Text S2).

Figure 2. shows richness over time in our data, for common bird species, using 16km^2 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 comparison of regions, as below we show that the trends before the implementation are not different between the regions (*Section 6*).

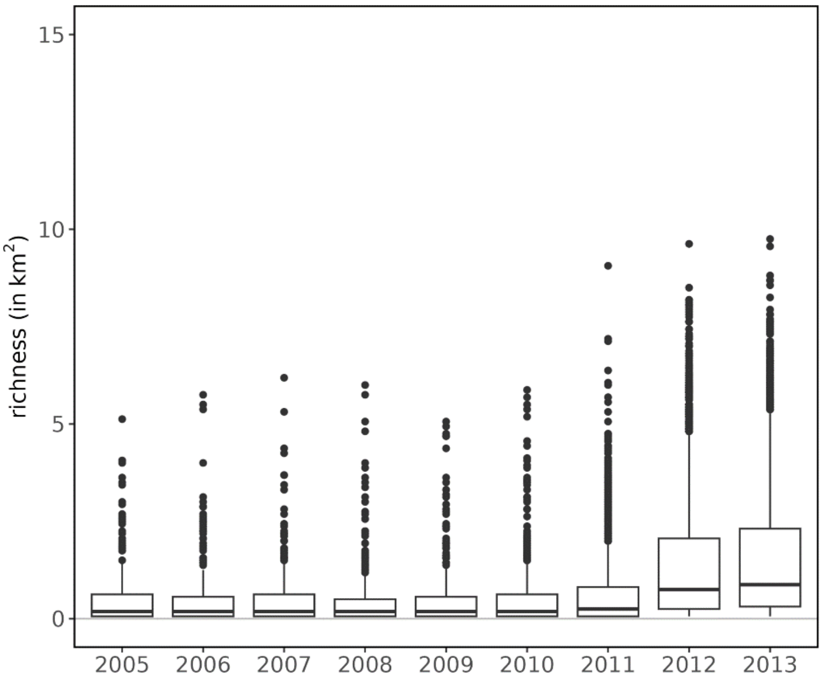

Figure 2. Development of species richness over time of *common* bird species. The development for Hill-Shannon index of *common* bird species, number of voluteers per unit, species richness over time of *all* bird species, and species richness over time of *farmland* bird species over time are shown in Figure S2, Figure S3, Figure S4, and Figure S5, respectivly. The scales of the y-axis of this plot is selected to easily compare the trends with Figure S4 and Figure S5.

4.5 Additional data

We use four additional data sources in our analysis representing time invariant information (see *Estimation procedure* in *Section 5.2*). First, the share of nature protection areas (Figure S6), from Lower Saxony State Department for Waterway, Coastal and Nature Conservation (2023) and NUMIS (2002). They suggest high perceived natural value but were excluded from result-based payments until 2021. Second, the share of agricultural land, to account for agricultural areas where farmers could enroll land to the payment scheme, from the EU's Copernicus Land Monitoring Service (2020) with a resolution of 100x100m. Third, 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 State Office for Statistics of Lower Saxony (2024). They cover 2002-07 at the county level (Figure S7 for the county distribution in Lower Saxony). Fourth, human population density, to account for likelihood of recording in citizen-science data (Hertzog et al., 2021), from the German 2011 census from the Statistical Offices of the Federal Government and the States (2015) in a 1x1km grid. While 2011 is after the 2008 treatment, we assume this treatment does not affect human population. For all variables we compute the average value per unit then rescale so variables have a mean of zero and standard deviation of one.

5. Empirical strategy

We are interested in 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 definition (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_{ti} (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.

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 Figure 1). We illustrate the temporal development of the number of thresholds in each of these groups in Figure 3. Group M gets the multi-threshold payment $d = 2$ in period

2008, group S in 2012, while B in 2014, the point in time in which the multi-threshold payment scheme is introduced for every region.

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

Figure 3. Time change by region in the number of thresholds (2 = "multi-", 1 = "single-"). Each scheme has a threshold at 4 indicator plant species. The multi-threshold scheme *adds* one threshold at 6 indicator plant species, with another payment "layer". Note that here we indicated the introduction in 2007/08 as if it were in 2008.

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. The effect is identified under the assumption

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

which is the standard parallel trends assumption.[10] 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 and 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 (Figure 1). This restriction poses a limitation for our analysis, since it only allows to assess the change in

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

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; Ernoult et al., 2006; Uezu and Metzger, 2016).

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 it 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

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

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 time-group effect. Based on $ATT(g,t)$ we can compute the group effect as:

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

where τ are the number of periods we observe (Callaway and Sant'Anna, 2021). $ATT(g,t)$ is identified via (Marcus and Sant'Anna, 2021):

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

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

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

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,

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

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

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

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 relatively 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 of treatment effects is difficult to interpret. However, under the mild assumption that $\mathbb{E}[Y_{POST}(1) - Y_{PRE}(0)|B_2]$ is non-negative, we can derive the following inequality:

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

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). It allows the use of unbalanced data and flexible consideration of observed pretreatment characteristics. We include two types of covariates in the estimation. First, we consider as the time-invariant covariates, denoted by X_i, such as share of nature protection areas, share of agricultural land, mean grassland yield, and human population density, see *Section 4.5*. Second, as the only time-varying covariate, we consider the number of volunteers that record birds, Z_{it}. This can be an important variable as it directly links 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*.

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 multi-threshold payment for the 2012 pilot area. The time-group effects and group effects are displayed in Figure 4. As discussed above, we are only able to estimate the time-group effects for the first two periods after the treatment. Both estimates are very close to zero in magnitude. It may take time for any effect to unfold. All pre-treatment estimates are also very close to zero and corresponding confidence bands contain zero almost at their centers, indicating that the parallel trends assumption is plausible.

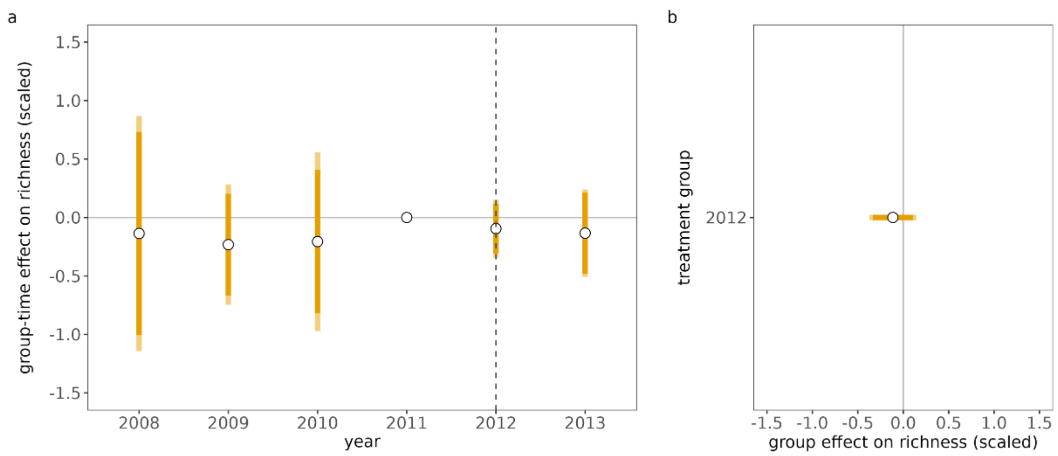

Figure 4. Effects of treatment on common bird species richness considering the 2012 pilot area. Panel a shows the time-group effect and panel b the group effect. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

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 (Figure 5). Consistent with the previous approach, these estimates also are practically zero for the first two post-treatment periods (Figure 5a). Yet, 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 (Figure 5b).

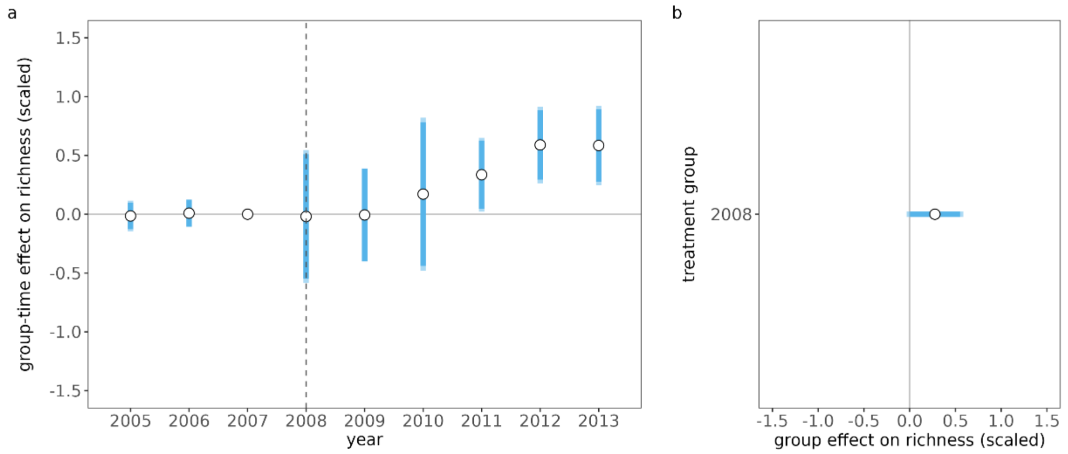

Figure 5. Effects of treatment on common bird species richness considering the 2008 pilot area. Panel a shows the time-group effect and panel b the group effect. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

Since 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 similar, respectively [0.00 to 0.55] and [-0.03 to 0.59]. The lower bound of the 95% confidence interval just includes zero, while that of the 90% interval does not.[11] We note the upper bounds imply strong policy effects.[12]

[11] 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).

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

Finally, all pre-treatment estimates displayed in Figure 5a are again practically equal to zero and the tight confidence bounds imply 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* and *farmland* bird species (instead of *common*). To simplify, we discuss here only the group effects (Figure S8 and Figure S9 for group-time treatment effects). These reveal two findings 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 (i.e., 0.28), even if more uncertain. Thus, it seems that *common*, compared to *rare* birds (which are defined as *all* minus *common* birds), were more affected by the multi-threshold result-based payments. However, volunteers may be worse at detecting *rare* birds (Cox et al., 2012; Kosmala et al., 2016). 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 (Figure 6b). The point estimate is 0.41, with a 90% and 95% confidence interval of [0.21 to 0.61] and [0.19 to 0.63], respectively. Finally, Figure S8 and Figure S9 show that the parallel trends assumption is supported.

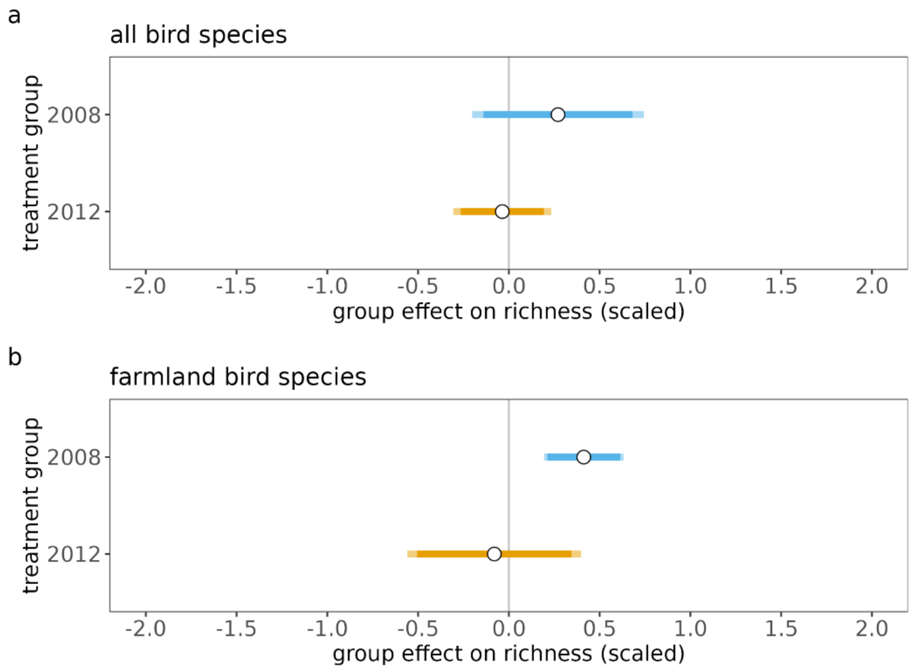

Figure 6. Group effects of treatment on all (panel a) and farmland (panel b) bird species richness. The 2008 pilot area is indicated in blue and 2012 pilot area is indicated in orange. The model specification follows our main estimation specification. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level. All results of the group-time effects on all bird species and farmland bird species can be found in Figure S8 and Figure S9, respectively.

6.3 Further results and robustness checks

Time varying covariates. To evaluate whether the number of observers is a valid covariate, the main assumption here is that time-varying covariates are not affected by treatments (if so, they are "bad controls"; Caetano et al., 2022; Caetano and Callaway, 2023). In order to assess, we follow Caetano et al., 2022 and re-estimate our main model but using Z_{it} instead of Y_{it} as dependent variable (Figure S10). Reassuringly, we find no evidence that treatment affected the number of observers. Thus, we can use Z_{it} as a control in our staggered difference-in-differences.

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. Though literatures commonly have found that data collected by experts and volunteers are comparable (Bird et al., 2014), 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 citizen-science data (Mitschke et al., 2005).[13] In the CBS, systematically all common birds are recorded based on stratified sampling areas and independent of species interests of the observer (Text S4). The results show that ornitho data would indeed appear to be valid for depicting changes in bird populations (Table S5; Figure S11).

Alternative specifications. We run several checks to understand the sensitivity of our main findings with respect to the measurement and definition choices. First, we run a model without accounting for the number of volunteers (i.e., drop the time-variant variable, Z_{it}). These results are similar to those in our main analysis (Figure S12).

Second, 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 (Figure S13).

Third, in our main analysis we focus the size of observations is 16km^2. Results are robust to altering the observations' size to 1km^2, 4km^2, and 64km^2 (Figure S14).

Fourth, we control for other spatially restricted programs introduced in earlier reforms than the reform that included the introduction of result-based payment. These programs include the "Wild herbs, spatial biotope types scheme", and "Nordic visiting birds scheme" (Text S1). This adjustment does not change the interpretation of our results (Figure S15).

Fifth, 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 (Figure S16).

Assessment of the overlap assumption. The overlap assumption requires that the characteristics of an observation does not deterministically define the treatment status. We checked the overlap assumption by estimating the propensity score (considering only time-

[13] 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.

invariant covariates of our main estimation) using a logit model and check if the scores are not bounded to one or zero. The results shows that this assumption is satisfied (Figure S17).

7. Concluding remarks

Governments across the globe use agri-environmental payments to lower loss of biodiversity in agricultural systems, with several objectives including those set under the Kunming-Montreal Global Biodiversity Framework (UN, 2022) and UN Decade on Ecosystem Restoration (2021–2030) (IUCN, 2022). To increase 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, we lack 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 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 payment towards 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 a staggered rollout of a multi-threshold result-based payments in the German state of Lower Saxony and citizen-science bird data over 16 years with over 6.7 million data entries.

Our results highlight that multi-threshold result-based payments can increase bird diversity as compared to single-threshold payments and that this effect 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]). We show that this can be interpreted as the lower bound effect of the multi-threshold result-based payments. Our findings 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 suggest important future research. First, future studies could identify 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).

8. References

Andersson, A., Höjgård, 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. *Journal of Applied Ecology*, 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. *Agriculture, Ecosystems & Environment*, 166, 28-34.

Baylis, K., Coppess, J., Gramig, B. M., & Sachdeva, P. (2022). Agri-environmental programs in the United States and Canada. *Review of Environmental Economics and Policy*, 16(1), 83-104.

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.

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. *Biological Conservation*, 173, 144-154.

Birrer, S., Spiess, M., Herzog, F., Jenny, M., Kohli, L., & Lugrin, B. (2007). The Swiss agri-environment scheme promotes farmland birds: but only moderately. *Journal of Ornithology*, 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., Droeschmeister, R., & Sudfeldt, C. (2020). Drivers of population change in common farmland birds in Germany. *Bird Conservation International*, 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., (2023). Difference-in-Differences with Time-Varying Covariates in the Parallel Trends Assumption. arXiv preprint arXiv:2202.02903.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 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. *Journal of Environmental Economics and Management*, 65(1), 12-27.

Chamberlain, D. E., Fuller, R. J., Bunce, R. G., Duckworth, J. C., & Shrubb, M. (2000). Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *Journal of Applied Ecology*, 37(5), 771-788.

Coderoni, S., Esposti, R., & Varacca, A. (2023). How differently do farms respond to agri-environmental policies? A probabilistic machine-learning approach. *Land Economics*, 100(2), 370-397.

Copernicus Land Monitoring Service (2020). CORINE Land Cover 2006 (vector/raster 100 m), Europe, 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. *Ecological Applications*, 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.

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. *Ecology Letters*, 11(3), 235-244.

Elmiger, N., Finger, R., Ghazoul, J., & Schaub, S. (2023). Biodiversity indicators for result-based agri-environmental schemes–Current state and future prospects. *Agricultural Systems*, 204, 103538.

Ernoult, 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?. *Biological Conservation*, 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., & Lehikoinen, A. (2020). A state-of-the-art review on birds as indicators of biodiversity: Advances, challenges, and future directions. *Ecological Indicators*, 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. *Journal of Agricultural Economics*, 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. *Biological Conservation*, 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. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 269-288.

Hart, R., & Latacz-Lohmann, U. (2005). Combating moral hazard in agri-environmental schemes: a multiple-agent approach. *European Review of Agricultural Economics*, 32(1), 75-91.

Hertzog, L. R., Frank, C., Klimek, 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. *Diversity and Distributions*, 27(6), 1106-1119.

Imbens, G. W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives*, 35(3), 157-174.

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. *Ecological Indicators*, 46, 224-231.

Johnston, A., Moran, N., Musgrove, A., Fink, D., & Baillie, S. R. (2020). Estimating species distributions from spatially biased citizen science data. *Ecological Modelling*, 422, 108927.

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. *Journal of Ornithology*, 162(1), 1-15.

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?. *Journal of Applied Ecology*, 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. *Biological Conservation*, 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. *Ecology Letters*, 9(3), 243-254.

Knape, J., Coulson, S. J., van der Wal, R., & Arlt, D. (2022). Temporal trends in opportunistic citizen science reports across multiple taxa. *Ambio*, 1-16.

Knop, E. V. A., Kleijn, D., Herzog, F., & Schmid, B. (2006). Effectiveness of the Swiss agri-environment scheme in promoting biodiversity. *Journal of Applied Ecology*, 43(1), 120-127.

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10), 551-560.

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. *Annual Review of Environment and Resources*, 47, 231-260.

Li, Y., Miao, R., & Khanna, M. (2020). Neonicotinoids and decline in bird biodiversity in the United States. *Nature Sustainability*, 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?. *Ecological Indicators*, 72, 118-130.

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_landsch aft/besonders_geschutzte_teile_von_natur_und_landschaft/naturschutzrechtlich-besonders-geschuetzte-teile-von-natur-und-landschaft-9065.html

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. *Journal of Applied Ecology*, 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). *Agriculture, Ecosystems & Environment*, 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.

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.

Most, A., & Keienburg, 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.

Navarro, A., & López-Bao, J. V. (2019). EU agricultural policy still not green. *Nature Sustainability*, 2(11), 990-990.

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. *Nature Communications*, 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. *American Journal of Agricultural Economics*, 104(4), 1460-1484.

NUMIS (2002). Zonierung Elbtalaue. Hannover, Germany.

Pannell, D., & Rogers, A. (2022). Agriculture and the environment: Policy approaches in Australia and New Zealand. *Review of Environmental Economics and Policy*, 16(1), 126-145.

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, Birkenstock M, Lakner S, Röder N (2021) The Common Agricultural Policy post-2020: Views and recommendations from scientists to improve performance for biodiversity. Thünen Working Paper No. 175, Johann Heinrich von Thünen-Institut.

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. *Conservation Letters*, 15(6), e12901.

Rigal, S., Dakos, V., Alonso, H., Auniņš, A., Benkő, Z., Brotons, L., ... & Devictor, V. (2023). Farmland practices are driving bird population decline across Europe. *Proceedings of the National Academy of Sciences*, 120(21), e2216573120.

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. *Agriculture, Ecosystems & Environment*, 125(1-4), 167-172.

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. *Global Environmental Change*, 80, 102668.

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. *Journal of Agricultural Economics*, 74(3), 617-660.

Schaub, S., Roth, T., & Bonev, P. (2025). The effect of result-based agri-environmental payments on biodiversity: Evidence from Switzerland. *American Journal of Agricultural Economics*. *In press*.

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. *Biological Conservation*, 214, 343-346.

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. *Ecological Indicators*, 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. *Journal of Ecology*, 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. *European Review of Agricultural Economics*, 49(4), 723-759.

Tiago, P., Pereira, H. M., & Capinha, C. (2017). Using citizen science data to estimate climatic niches and species distributions. *Basic and Applied Ecology*, 20, 75-85.

Tietz, A., Grajewski, R., Anter, J., Bathke, M., Bergschmidt, A., Dickel, R., Eberhardt, W., Ebers, H., Fährmann, 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. Montreal, Canada. 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. *Biological Conservation*, 119(1), 19-39.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American 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.

Wuepper, D., & Huber, R. (2022). Comparing effectiveness and return on investment of action- and results-based agri-environmental payments in Switzerland. *American Journal of Agricultural Economics*, 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. *European Review of Agricultural Economics*, 51(3), 690-724.

Supplementary information

Figures

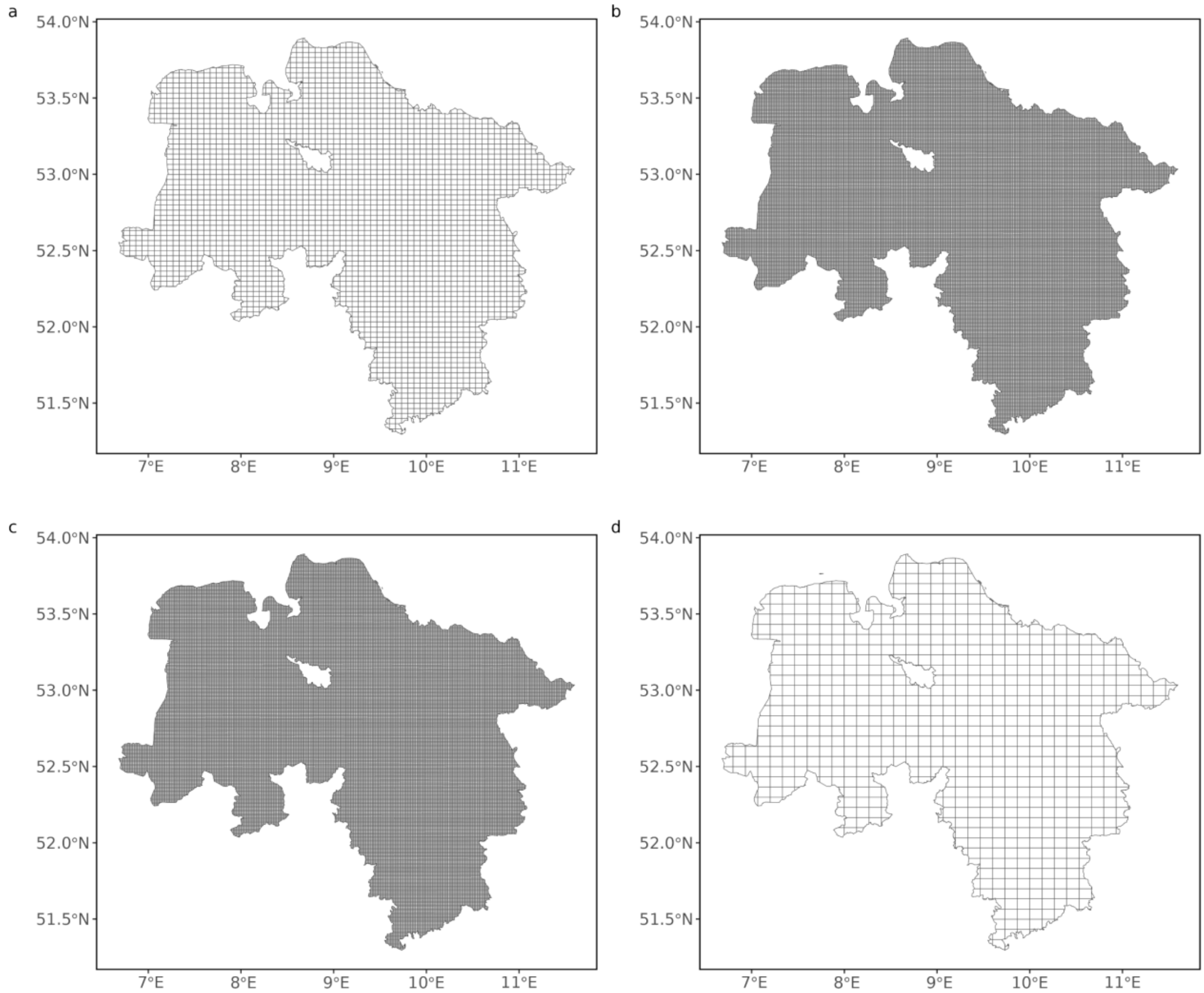

Figure S1. Overview of sizes of observation: 4x4 (main definition; panel a), 1x1 (panel b), 2x2 (panel c), and 8x8 (panel d).

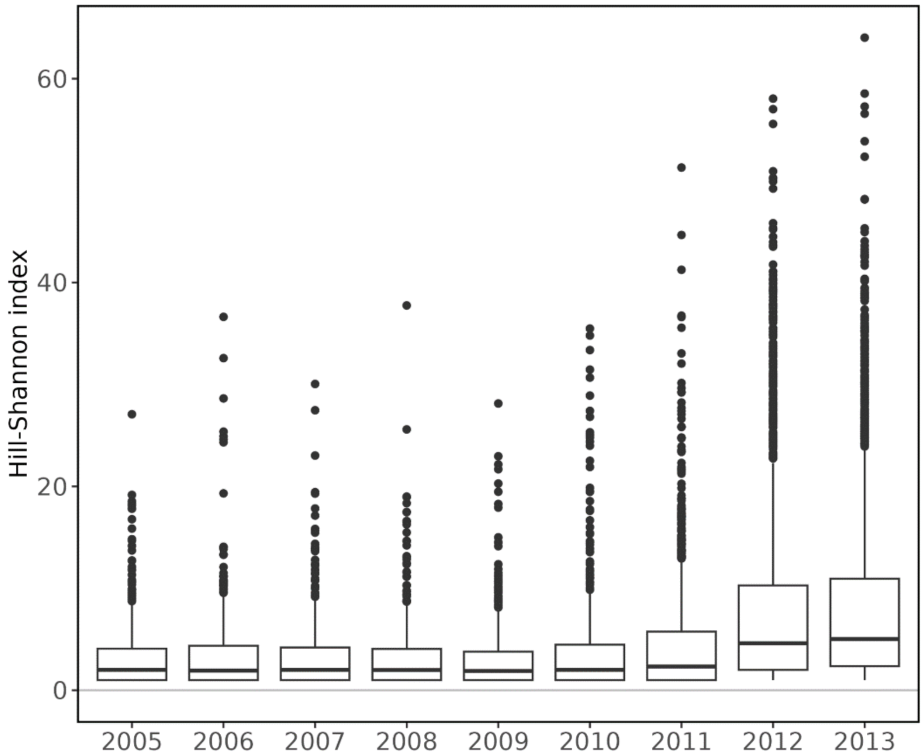

Figure S2. Development of Hill-Shannon index of common bird species over time.

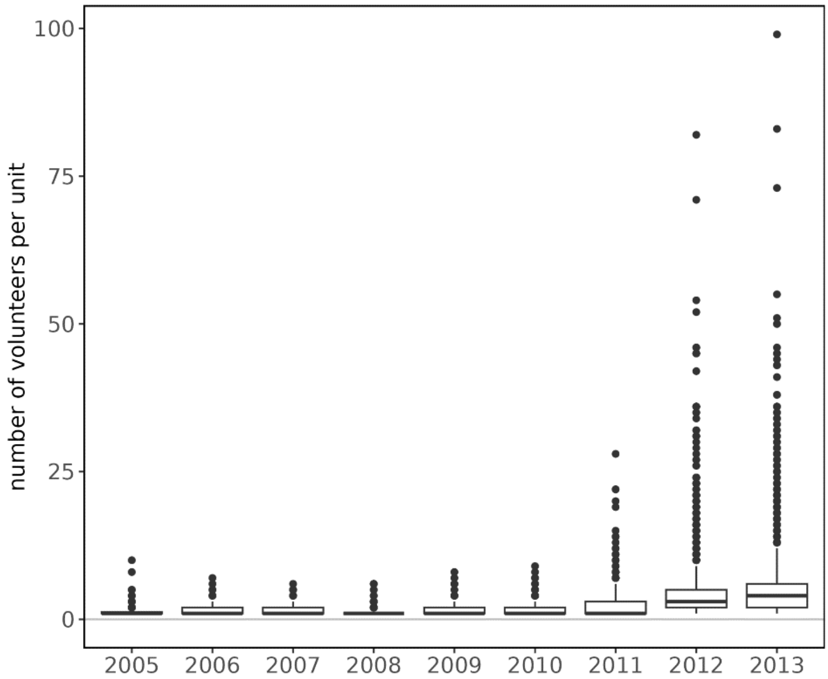

Figure S3. Development of number of volunteers per site over time.

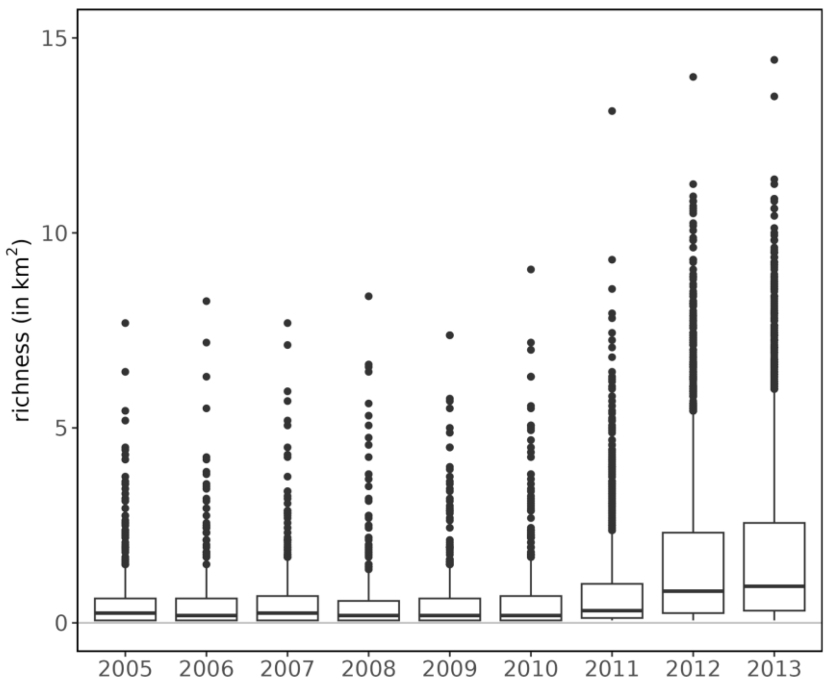

Figure S4. Development of species richness over time of all bird species.

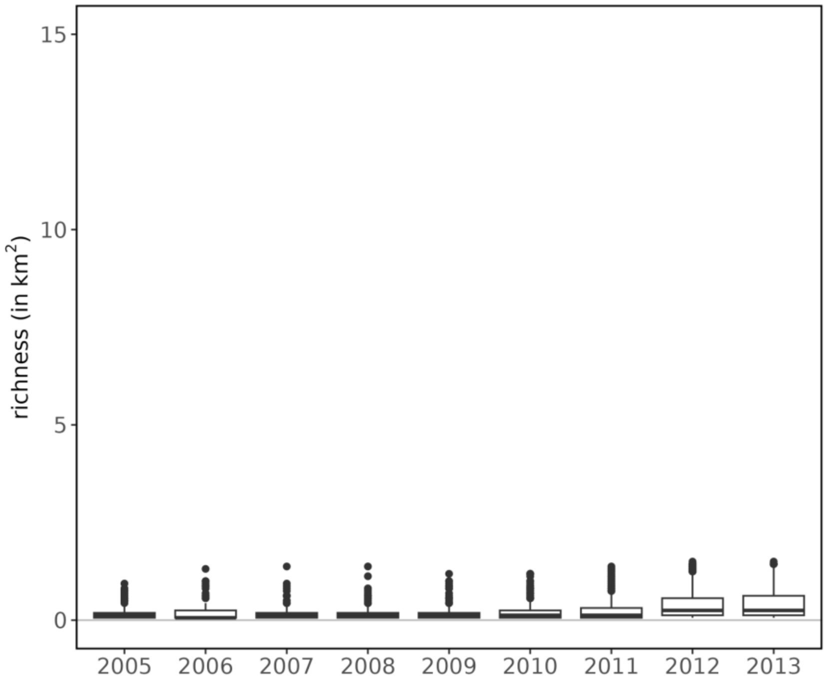

Figure S5. Development of species richness over time of farmland bird species.

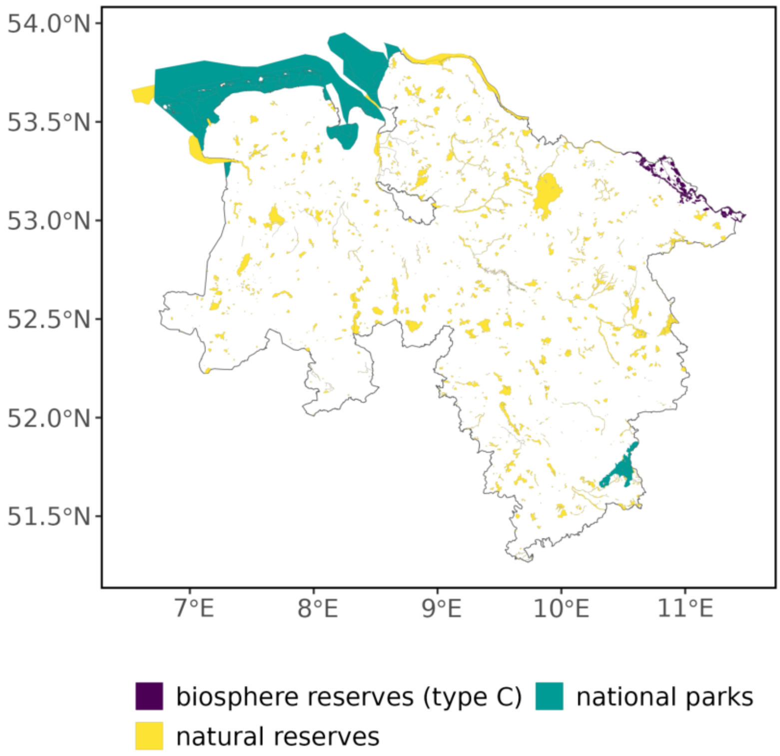

Figure S6. Overview of nature protection areas. We define those as nature protection areas that were excluded in the study period from receiving multi-threshold result-based payments. These include biosphere reserves (type C), national parks, and nature reserves.

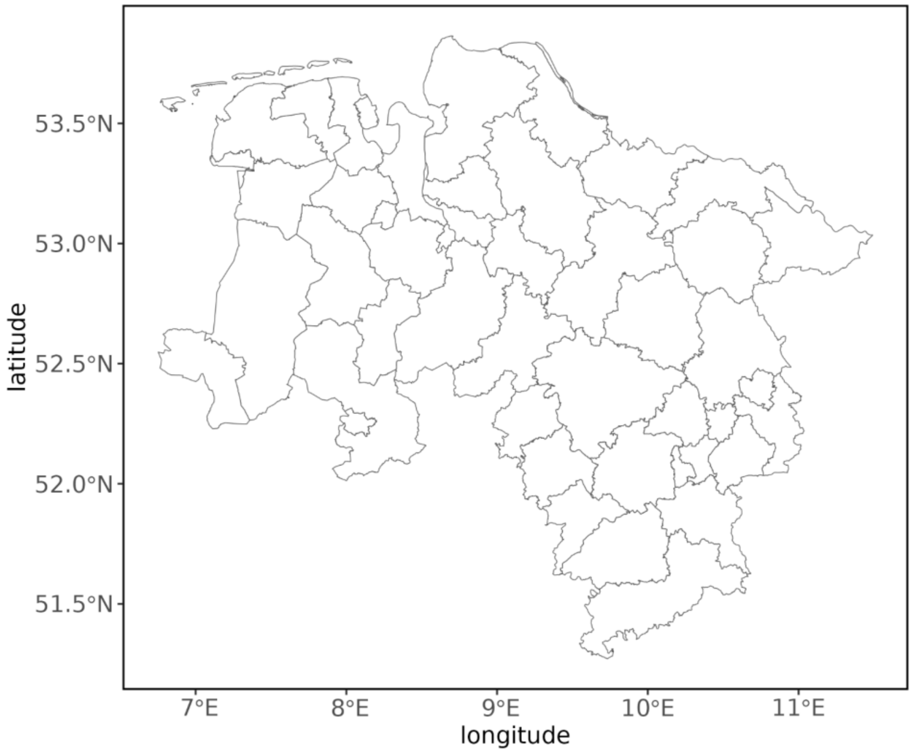

Figure S7. Overview of counties of Lower Saxony.

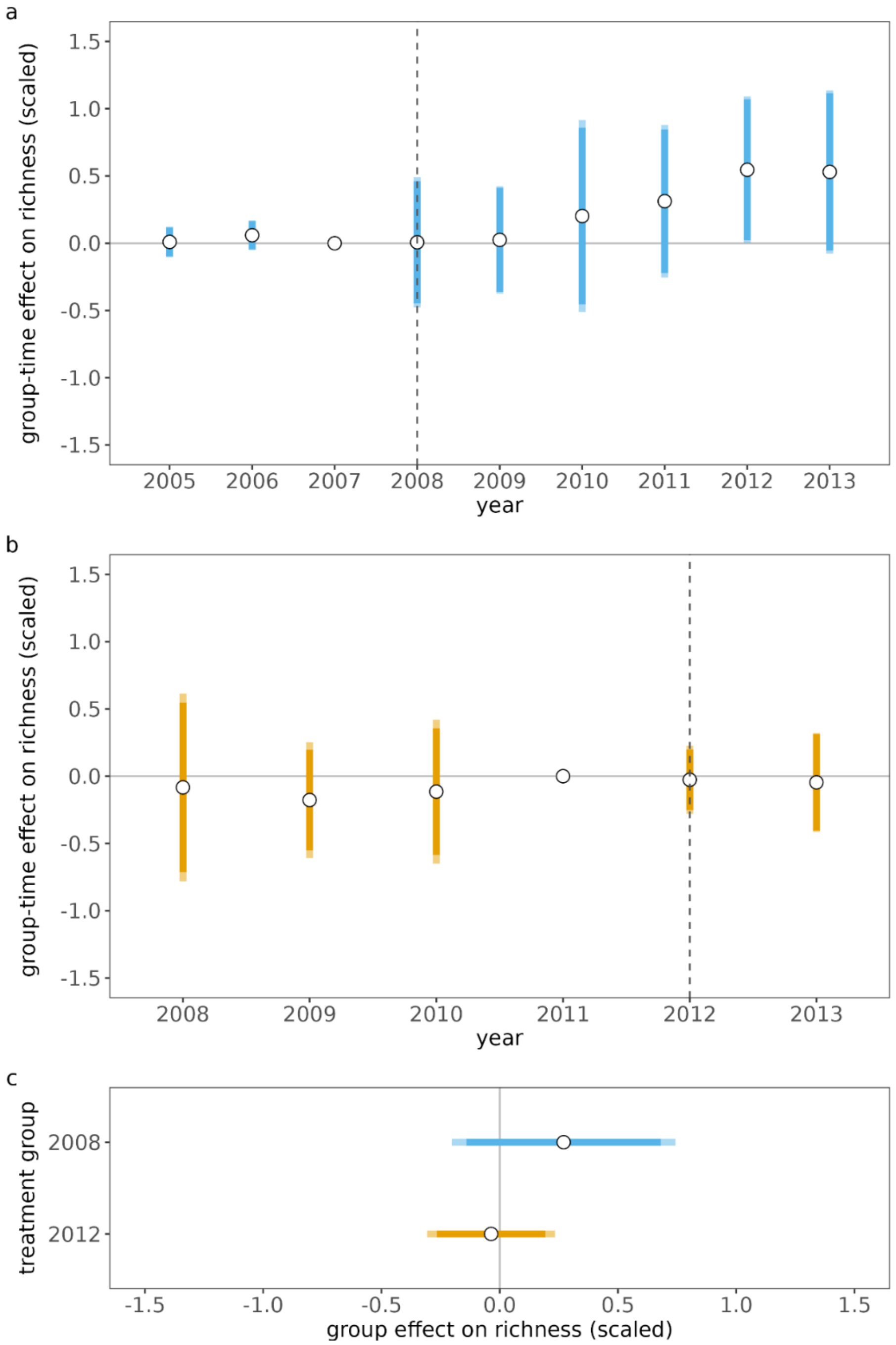

Figure S8. Effects of treatment on bird species richness, considering observations of all bird species. Panel a shows the time-group effect of the 2008 pilot area, panel d the time-group effect of the 2012 pilot area, and panel c the group effects. The 2008 pilot area is indicated in blue and 2012 pilot 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 a clustered at the county level. The associated group and group-time effects on the number of volunteers are shown in Figure S18.

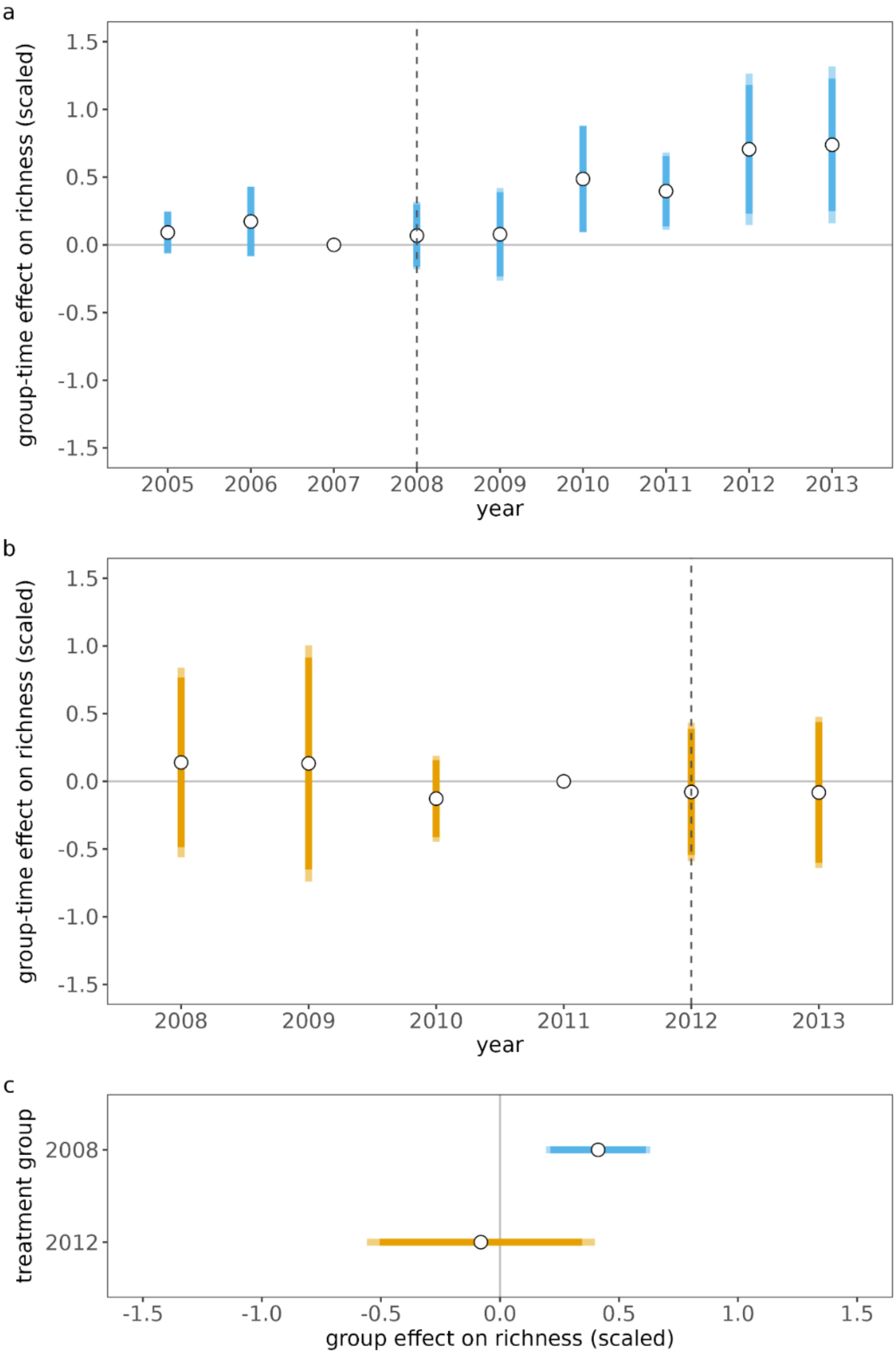

Figure S9. Effects of treatment on bird species richness, considering observations of farmland bird species. Panel a shows the time-group effect of the 2008 pilot area, panel d the time-group effect of the 2012 pilot area, and panel c the group effects. The 2008 pilot area is indicated in blue and 2012 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 a clustered at the county level. The associated group and group-time effects on the number of volunteers are shown in Figure S19.

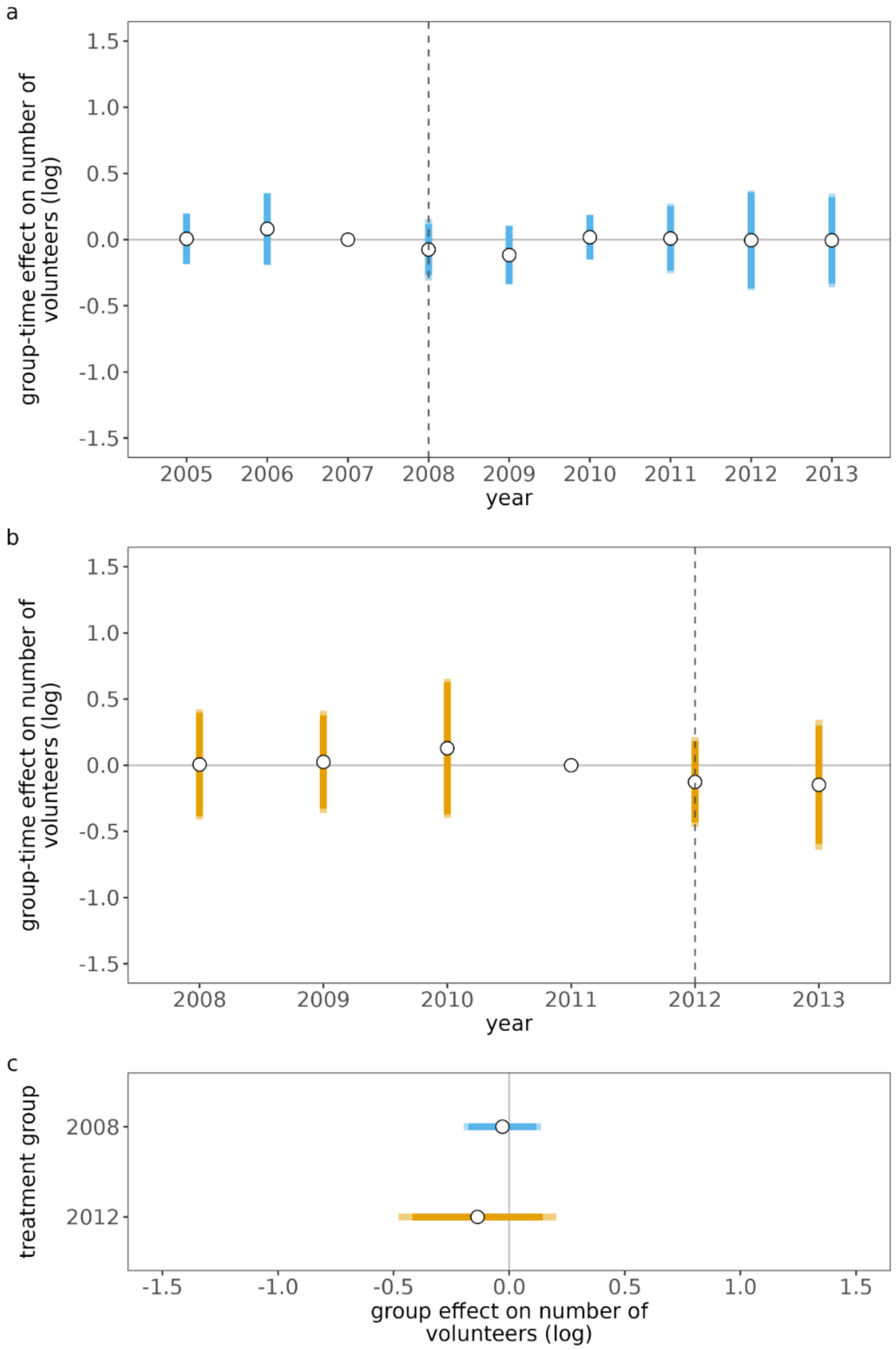

Figure S10. Effects of treatment on number of volunteers. Panel a shows the time-group effect of the 2008 pilot area, panel d the time-group effect of the 2012 pilot area, and panel c the group effects. The 2008 pilot area is indicated in blue and 2012 pilot area is indicated in orange. The model specification follows our main estimation specification. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

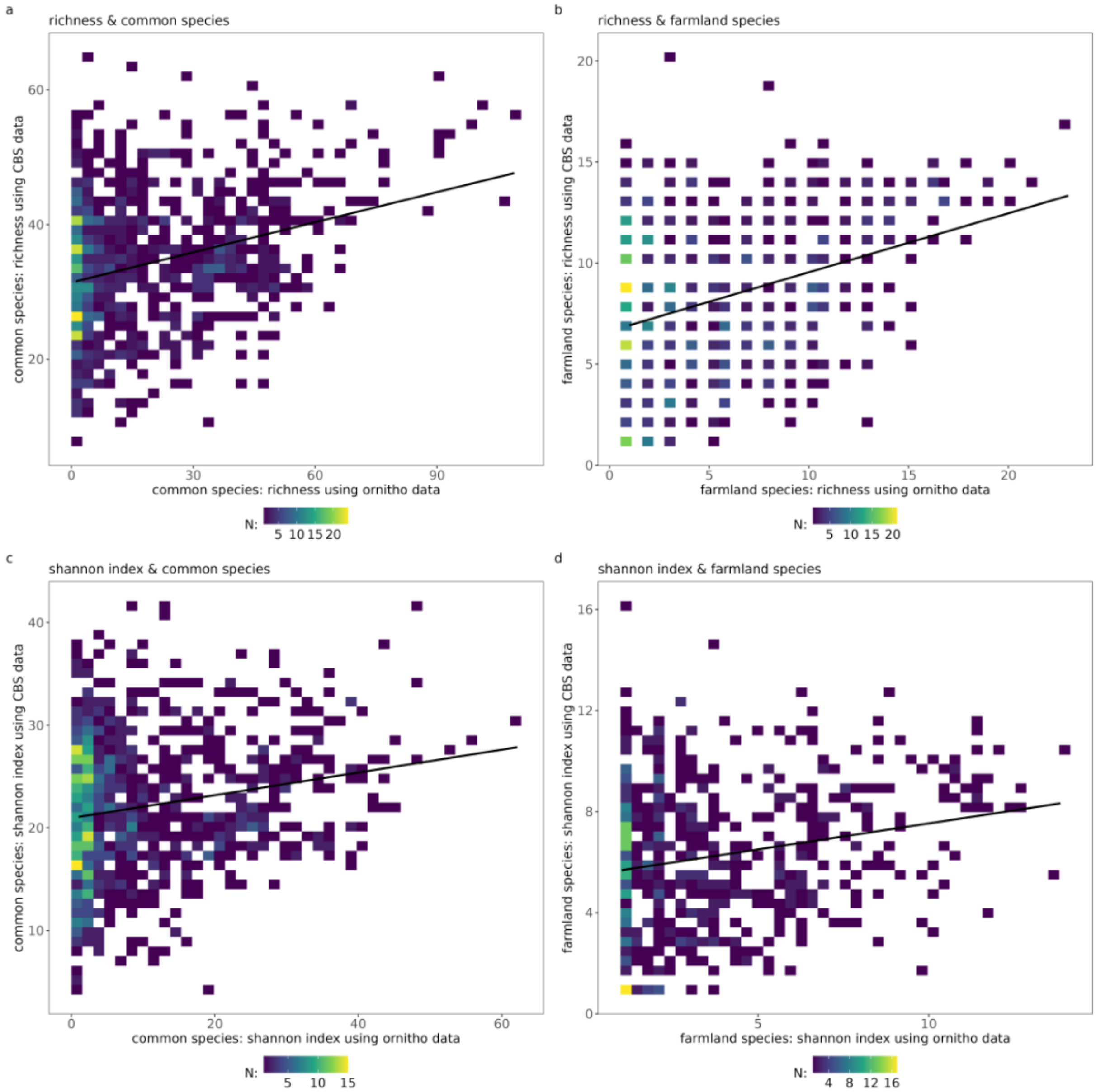

Figure S11. Relationship between diversity measures of the CBS data (y-axes) and ornito data (x-axes) on the CBS sampling areas.

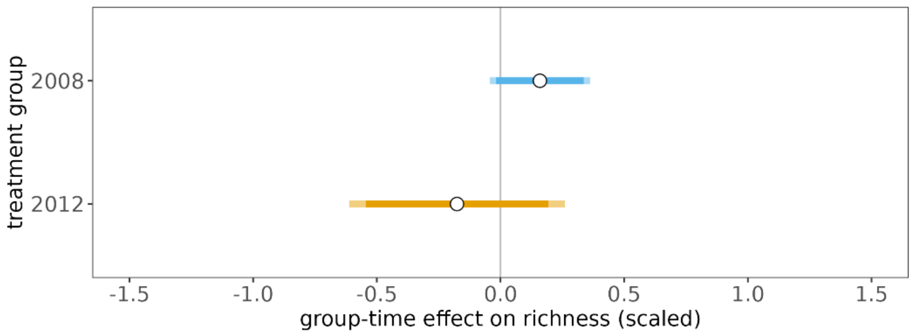

Figure S12. Group effects of treatment on common bird species richness excluding number of volunteers as covariate. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

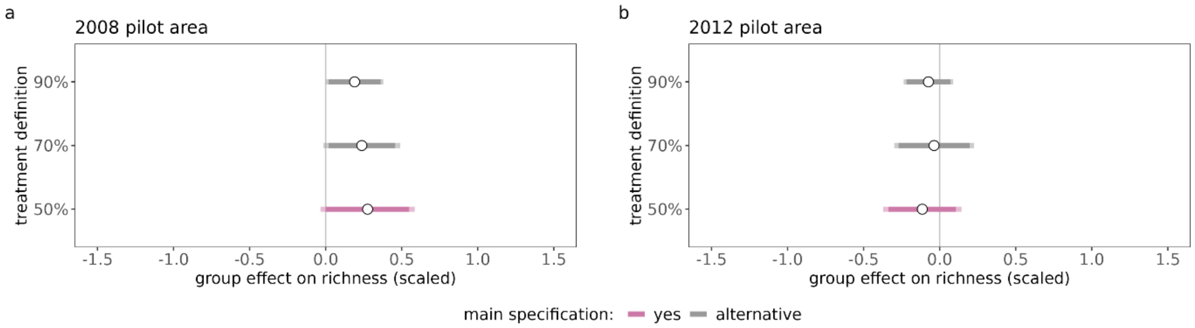

Figure S13. Group effects of treatment on common bird species richness over a range of treatment threshold definition. Panel a shows the results for the 2008 pilot area and panel b shows the results for the 2012 pilot area. The main estimation results are shown in red and alternative specification The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

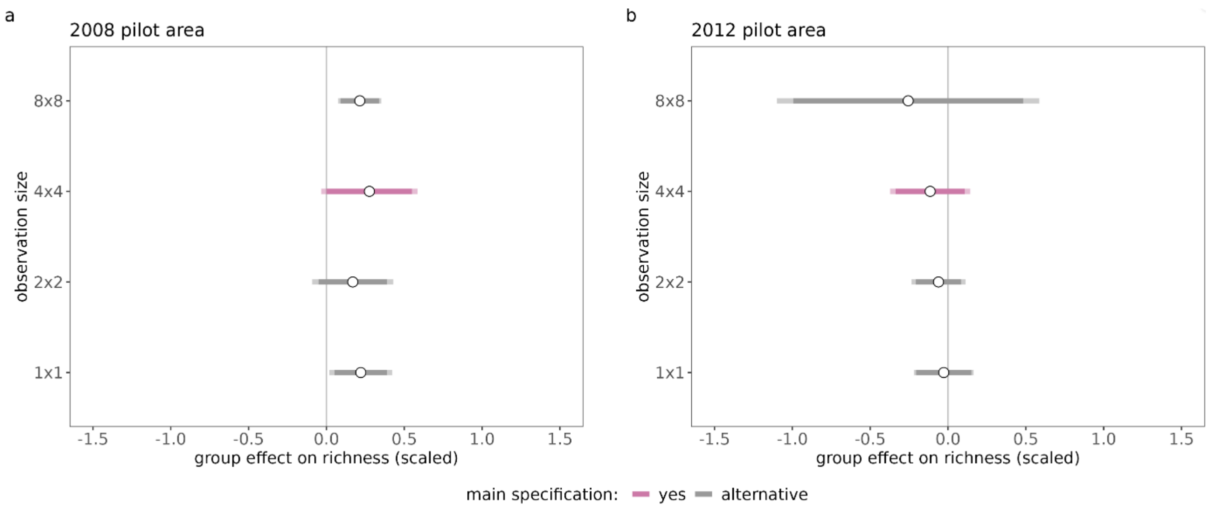

Figure S14. Group effects of treatment on common bird species richness over a range of observation sizes. Panel a shows the results for the 2008 pilot area and panel b shows the results for the 2012 pilot area. The main estimation results are shown in red and alternative specification in gray. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

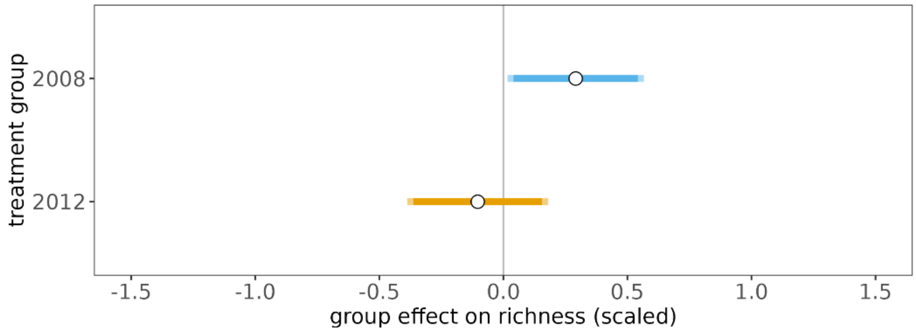

Figure S15. Group effects of treatment on common bird species richness, considering other spatially restricted programs established in earlier reforms. The 2008 pilot area is indicated in blue and 2012 pilot area is indicated in orange. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

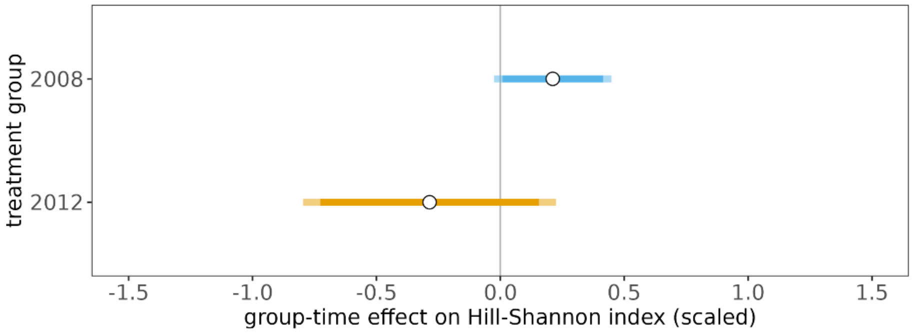

Figure S16. Group effects of treatment on Hill-Shannon index, considering common bird species. The 2008 pilot area is indicated in blue and 2012 pilot area is indicated in orange. The light and dark bars indicate the 95% and 90% confidence intervals, respectively, and standard errors a clustered at the county level.

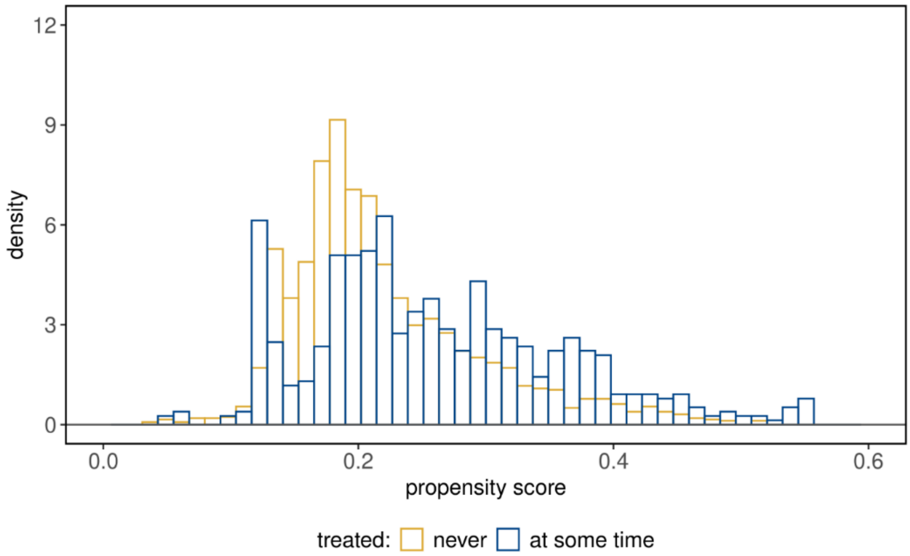

Figure S17. Propensity score of being treated depending on the time-invariant covariates.

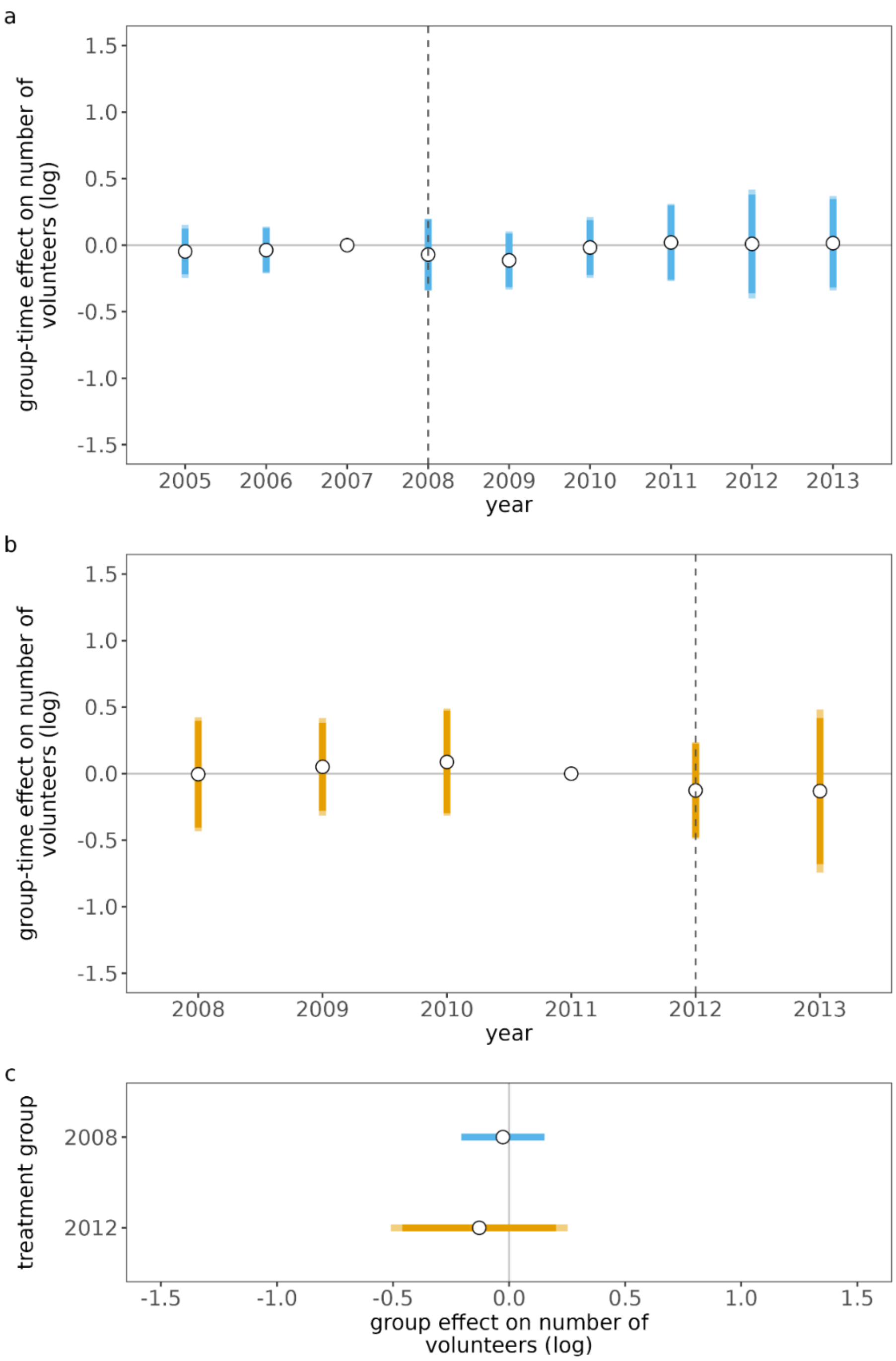

Figure S18. Effects of treatment on number of volunteers, considering observations of all bird species. Panel a shows the time-group effect of the 2008 pilot area, panel d the time-group effect of the 2012 pilot area, and panel c the group effects. The 2008 pilot area is indicated in blue and 2012 pilot 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 a clustered at the county level.

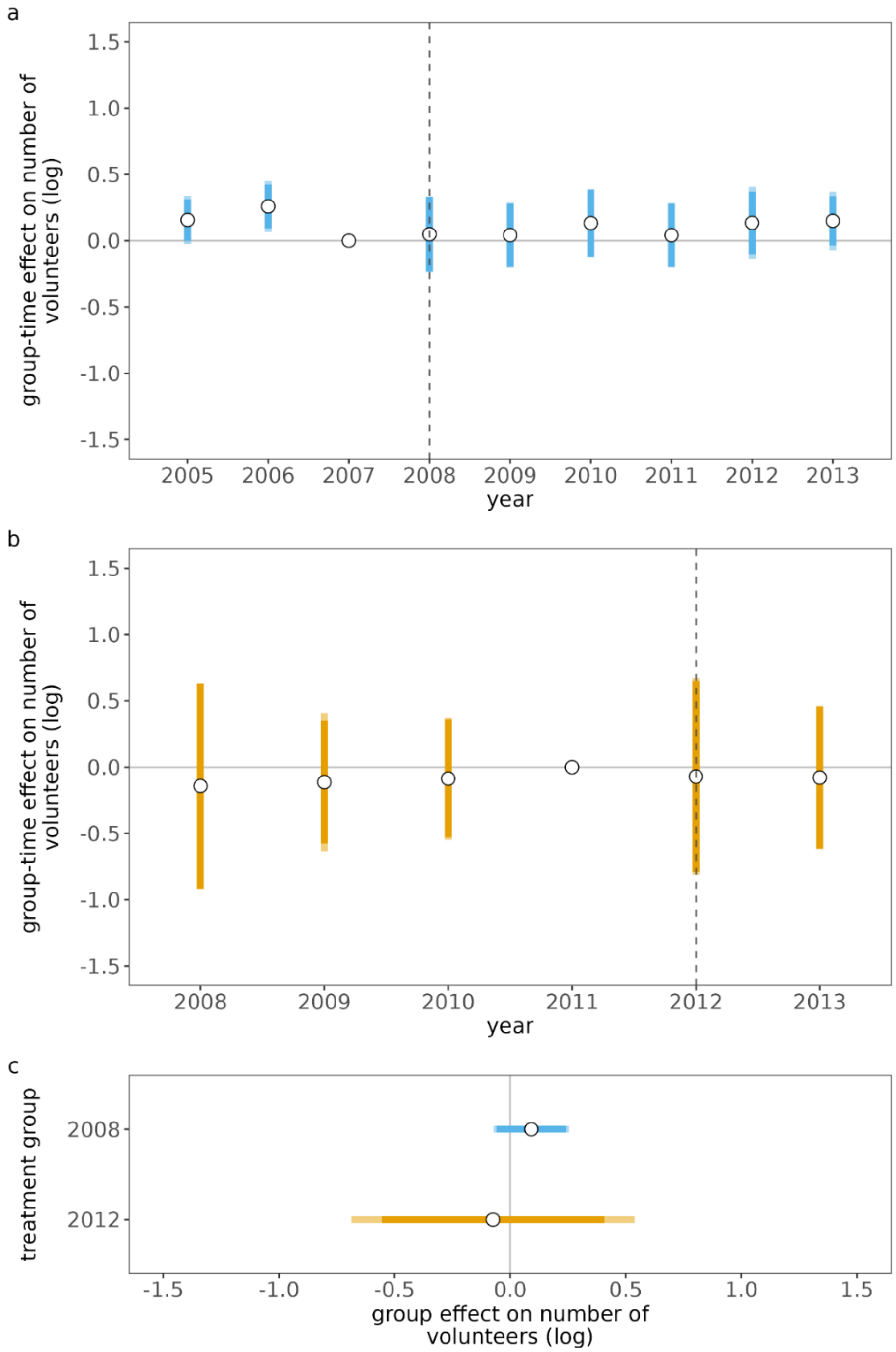

Figure S19. Effects of treatment on number of volunteers, considering observations of farmland bird species. Panel a shows the time-group effect of the 2008 pilot area, panel d the time-group effect of the 2012 pilot area, and panel c the group effects. The 2008 pilot area is indicated in blue and 2012 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 a clustered at the county level.

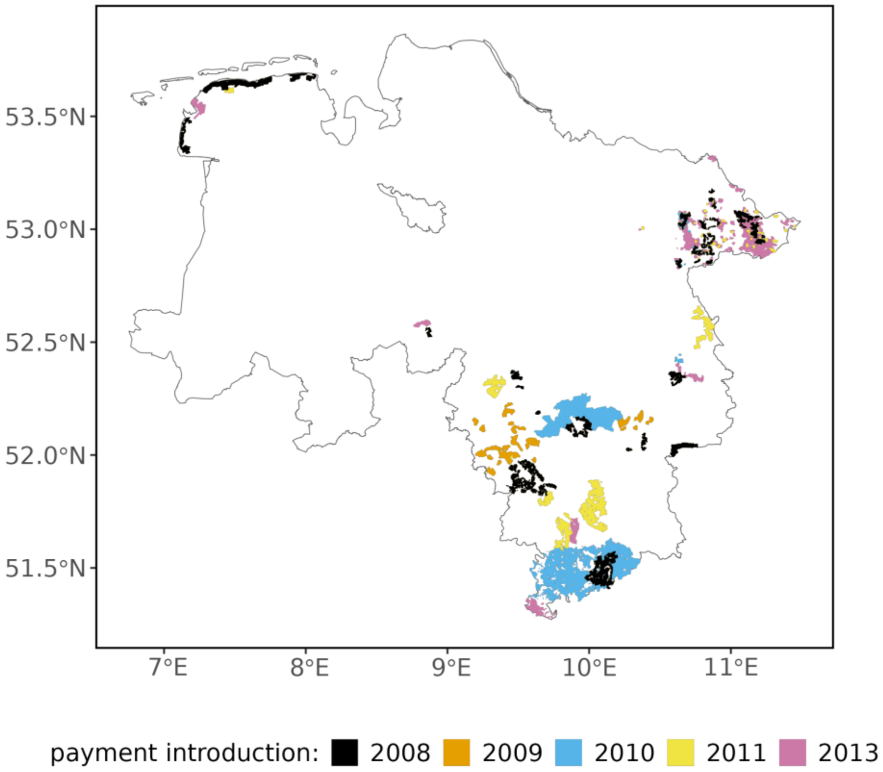

Figure S20. Temporal and spatial development of the action-based payment scheme for reducing the management intensity of arable land using field edge strips and to support birds and other animals in restricted areas.

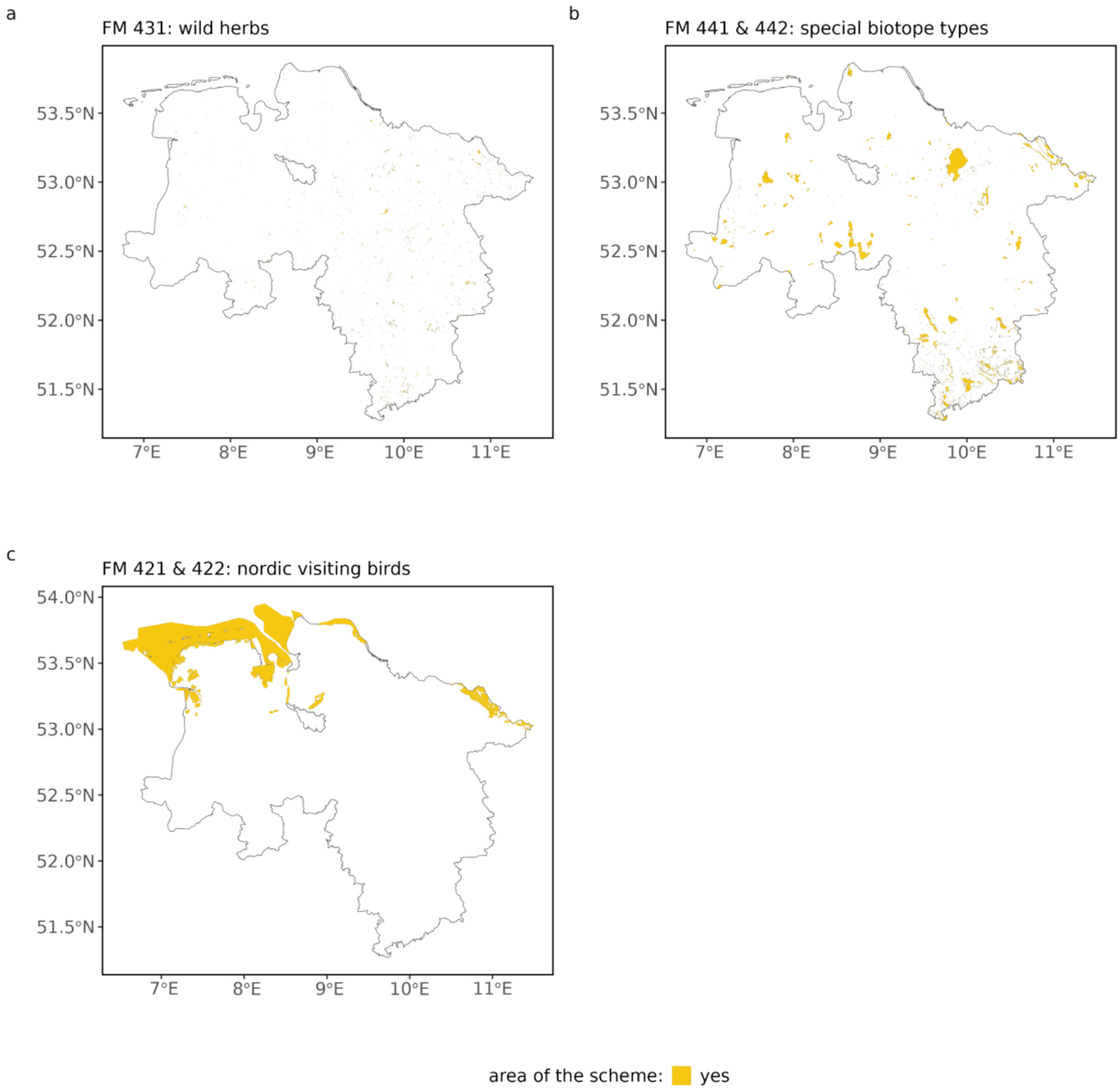

Figure S21. Overview of the area of schemes protecting wild herbs, spatial biotope types, and Nordic visiting birds. Note that the information of those areas are taken from Lower Saxony Ministry for Environment, Energy, and Climate Protection (2023) and represent the policy areas of 2023. Earlier information is not available as GIS information but the information aligns with Most et al. (2015). We only adopted the data of panel c, as we could visually better fit the area to the policy condition in 2013 using Most et al. (2015).

Tables

Table S1. Name of indicator plant species of result-based payments.

	Scientific name
1.	*Achillea millefolium*
2.	*Achillea ptarmica*
3.	*Ajuga reptans*
4.	*Alchemilla spec.*
5.	*Anthoxanthum odoratum*
6.	*Apiaceae* (excluding *Anthriscus sylvestris*)
7.	*Bistorta officinalis*
8.	*Caltha palustris*
9.	*Cardamine pratensis*
10.	*Carex spec*. (including *Scirpus spec.* and *Bolboschoenus spec.*)
11.	*Centaurea spec.*
12.	*Cirsium oleraceum*
13.	*Galium spec.* (white flowering, excluding *Galium aparine*)
14.	*Galium verum*
15.	*Knautia spec., Scabiosa spec.,* and *Succisa spec.*
16.	*Lathyrus pratensis*
17.	*Leucanthemum spec.*
18.	*Lotus spec.*
19.	*Luzula spec.*
20.	*Medicago lupulina, Trifolium dubium,* and *Trifolium campestre*
21.	*Plantago lanceolata*
22.	*Prunella vulgaris*
23.	*Ranunculus acris*
24.	*Ranunculus flammula*
25.	*Rhinanthus spec.*
26.	*Rumex acetosa* and *Rumex thyrsiflorus*
27.	*Silene flos-cuculi*
28.	*Stellaria graminea* and *Stellaria palustris*
29.	*Trifolium pratense*
30.	*Veronica chamaedrys*
31.	*Vicia cracca*

Source: Most and Keienburg (2006).

Table S2. Share of units with bird diversity information.

Year	Observation size	Number of observations	Share of units with bird diversity information of all grids within Lower Saxony[1]	Share of units with bird diversity information of all grids that had at least one time a species recording within Lower Saxony[1]
2005	4x4	372	11.93%	13.62%
2006	4x4	373	11.96%	13.66%
2007	4x4	369	11.83%	13.51%
2008	4x4	444	14.24%	16.26%
2009	4x4	460	14.75%	16.84%
2010	4x4	594	19.04%	21.75%
2011	4x4	1383	44.34%	50.64%
2012	4x4	2476	79.38%	90.66%
2013	4x4	2540	81.44%	93.01%
2005	1x1	848	1.81%	4.17%
2006	1x1	905	1.93%	4.45%
2007	1x1	852	1.82%	4.19%
2008	1x1	943	2.01%	4.64%
2009	1x1	966	2.06%	4.75%
2010	1x1	1286	2.75%	6.33%
2011	1x1	3975	8.49%	19.55%
2012	1x1	13064	27.91%	64.26%
2013	1x1	14715	31.44%	72.38%
2005	2x2	590	4.92%	7.01%
2006	2x2	615	5.13%	7.30%
2007	2x2	591	4.93%	7.02%
2008	2x2	671	5.60%	7.97%
2009	2x2	695	5.80%	8.26%
2010	2x2	928	7.74%	11.02%
2011	2x2	2508	20.93%	29.79%
2012	2x2	6425	53.62%	76.32%
2013	2x2	6908	57.65%	82.05%
2005	8x8	203	24.40%	26.54%
2006	8x8	219	26.32%	28.63%
2007	8x8	233	28%	30.46%
2008	8x8	267	32.09%	34.90%
2009	8x8	279	33.53%	36.47%
2010	8x8	340	40.87%	44.44%
2011	8x8	612	73.56%	80%
2012	8x8	751	90.26%	98.17%
2013	8x8	753	90.50%	98.43%

[1]The number of all units within Lower Saxony is 3119, 46810, 11982, and 832 for the observation size $16km^2$, $1km^2$, $4km^2$, and $64km^2$, respectively (Figure S1). The number of all units that had at least one time a species recording within Lower Saxony is 2731, 20329, 8419, and 765 for the observation size $16km^2$, $1km^2$, $4km^2$, and $64km^2$, respectively.

Table S3. Name of common bird species.

Scientific name (column 1)	Scientific name (column 2)	Scientific name (column 3)	Scientific name (column 4)
Acanthis cabaret	*Coccothraustes coccothraustes*	*Larus fuscus*	*Poecile palustris*
Accipiter gentilis	*Coloeus monedula*	*Limosa limosa*	*Porzana porzana*
Accipiter nisus	*Columba livia f. domestica*	*Linaria cannabina*	*Prunella modularis*
Acrocephalus arundinaceus	*Columba oenas*	*Locustella fluviatilis*	*Pyrrhula pyrrhula*
Acrocephalus palustris	*Columba palumbus*	*Locustella luscinioides*	*Rallus aquaticus*
Acrocephalus schoenobaenus	*Corvus corax*	*Locustella naevia*	*Recurvirostra avosetta*
Acrocephalus scirpaceus	*Corvus cornix*	*Lophophanes cristatus*	*Regulus ignicapilla*
Aegithalos caudatus	*Corvus corone*	*Loxia curvirostra*	*Regulus regulus*
Aegolius funereus	*Corvus frugilegus*	*Lullula arborea*	*Remiz pendulinus*
Alauda arvensis	*Coturnix coturnix*	*Luscinia luscinia*	*Riparia riparia*
Alcedo atthis	*Crex crex*	*Luscinia megarhynchos*	*Saxicola rubetra*
Alopochen aegyptiaca	*Cuculus canorus*	*Luscinia svecica*	*Saxicola rubicola*
Anas crecca	*Cyanistes caeruleus*	*Lyrurus tetrix*	*Scolopax rusticola*
Anas platyrhynchos	*Cygnus olor*	*Mareca strepera*	*Serinus serinus*
Anser anser	*Delichon urbicum*	*Milvus migrans*	*Sitta europaea*
Anthus pratensis	*Dendrocopos major*	*Milvus milvus*	*Somateria mollissima*
Anthus trivialis	*Dendrocoptes medius*	*Motacilla alba*	*Spatula clypeata*
Apus apus	*Dryobates minor*	*Motacilla cinerea*	*Spatula querquedula*
Ardea cinerea	*Dryocopus martius*	*Motacilla flava*	*Spinus spinus*
Asio flammeus	*Emberiza calandra*	*Muscicapa striata*	*Sterna hirundo*
Asio otus	*Emberiza citrinella*	*Netta rufina*	*Sterna paradisaea*
Athene noctua	*Emberiza hortulana*	*Nucifraga caryocatactes*	*Sternula albifrons*
Aythya fuligula	*Emberiza schoeniclus*	*Numenius arquata*	*Streptopelia decaocto*
Botaurus stellaris	*Erithacus rubecula*	*Oenanthe oenanthe*	*Streptopelia turtur*
Branta canadensis	*Falco subbuteo*	*Oriolus oriolus*	*Strix aluco*
Bubo bubo	*Falco tinnunculus*	*Panurus biarmicus*	*Sturnus vulgaris*
Bucephala clangula	*Ficedula hypoleuca*	*Parus major*	*Sylvia atricapilla*
Buteo buteo	*Ficedula parva*	*Passer domesticus*	*Sylvia borin*
Calidris pugnax	*Fringilla coelebs*	*Passer montanus*	*Sylvia communis*
Caprimulgus europaeus	*Fringilla montifringilla*	*Perdix perdix*	*Sylvia curruca*
Carduelis carduelis	*Fulica atra*	*Periparus ater*	*Sylvia nisoria*
Carpodacus erythrinus	*Gallinago gallinago*	*Pernis apivorus*	*Tachybaptus ruficollis*
Certhia brachydactyla	*Gallinula chloropus*	*Phasianus colchicus*	*Tadorna tadorna*
Certhia familiaris	*Garrulus glandarius*	*Phoenicurus ochruros*	*Tringa ochropus*
Charadrius alexandrinus	*Glaucidium passerinum*	*Phoenicurus phoenicurus*	*Tringa totanus*
Charadrius dubius	*Grus grus*	*Phylloscopus collybita*	*Troglodytes troglodytes*

Charadrius hiaticula	*Haematopus ostralegus*	*Phylloscopus sibilatrix*	*Turdus iliacus*
Chloris chloris	*Haliaeetus albicilla*	*Phylloscopus trochiloides*	*Turdus merula*
Chroicocephalus ridibundus	*Hippolais icterina*	*Phylloscopus trochilus*	*Turdus philomelos*
Ciconia ciconia	*Hirundo rustica*	*Pica pica*	*Turdus pilaris*
Ciconia nigra	*Jynx torquilla*	*Picoides tridactylus*	*Turdus torquatus*
Cinclus cinclus	*Lanius collurio*	*Picus canus*	*Turdus viscivorus*
Circus aeruginosus	*Lanius excubitor*	*Picus viridis*	*Tyto alba*
Circus cyaneus	*Larus argentatus agg.*	*Podiceps cristatus*	*Upupa epops*
Circus pygargus	*Larus canus*	*Poecile montanus*	*Vanellus vanellus*

Table S4. Name of farmland bird species.

Scientific name
Acrocephalus palustris
Alauda arvensis
Carduelis carduelis
Emberiza calandra
Emberiza citrinella
Falco tinnunculus
Hippolais icterina
Hirundo rustica
Lanius collurio
Limosa limosa
Linaria cannabina
Locustella naevia
Luscinia megarhynchos
Motacilla alba
Motacilla flava
Passer montanus
Saxicola rubetra
Streptopelia turtur
Sturnus vulgaris
Sylvia borin
Sylvia communis
Sylvia curruca
Turdus pilaris
Turdus viscivorus
Falco tinnunculus

Source: Busch et al. (2020) and Noack et al. (2022).

Table S5. Correlation between CBS diversity indices (columns) and ornitho indices (rows).

	Richness – common species	Richness – farmland species	Shannon index – common species	Shannon index – farmland species
Richness – common species	0.307			
	(<0.001)			
Richness – farmland species		0.364		
		(<0.001)		
Shannon index – common species			0.192	
			(<0.001)	
Shannon index – farmland species				0.217
				(<0.001)
N	940	613	940	613

Numbers in parentheses indicate p-values.

Texts

Text S1. Additional policy background information.

Follow up policy periods and other payments

The 2014-2020 Common Agricultural Policy (CAP) of the EU, was decided in December 2013 and was mainly implemented in Lower Saxony in 2014 and 2015. Amongst others the CAP 2014-2020 led to an expansion area of the multi-threshold result-based payments to entire state except nature protection areas (i.e., nature reserves, biosphere reserves of type C, and national parks) and from 2015 onwards the two-threshold design (with target threshold of 4 and 6 indicator plant species) changed to with three-threshold design (with target threshold of 4, 6, and 8 indicator plant species). However, other aspects, as the list of indicator plant species on which payments depend did not change (Keienburg et al., 2006; Lower Saxony Ministry of Food, Agriculture and Consumer Protection, 2023).

Other spatially restricted payment schemes

Lower Saxony also introduced in 2008 action-based payment scheme for reducing the management intensity of arable land using field edge strips and to support birds and other animals in restricted areas (Figure S20; Dickel et al., 2010). The area of the action-based payments was extended each year based on county-recommendations given the regulatory restrictions of the policy, except in 2012 due to limited financial resources. We consider the introduction of these payments in our analysis as the areas partly overlap with the pilot area of the multi-threshold result-based payments. In other words, we exclude the areas were both these and multi-threshold result-based payments were introduced.

Moreover, other spatially restricted payments were introduced in earlier reforms in Lower Saxony, including the “Wild herbs, spatial biotope types scheme”, and “Nordic visiting birds scheme” (Figure S21). We consider those areas in a robustness check.

Text S2. Calculation of Hill–Shannon index.

The Hill–Shannon index is calculated as follows:

$$Hill\ Shannon\ Index = e^{-\sum_{j=1}^{S} p_j \ln(p_j)} \quad \text{(S.1)}$$

where S is the number of different species, j indicates a species, p_j is the proportion of all individuals of a species, n_j, of the total number of individuals on an area, N, i.e., $p_j = n_j/N$.

Text S3. Proof of $Q = \mathbb{E}[Y_{POST}(2) - Y_{POST}(0)|M] - \mathbb{E}[Y_{POST}(1) - Y_{POST}(0)|B_2]$.

For proofing our stated results in Equation (5) first note that

$$Q = \mathbb{E}[Y_1(2) - Y_0(0)|M] - \mathbb{E}[Y_1(1) - Y_0(0)|B_2] \quad \text{(S.2)}$$

$$= \mathbb{E}[Y_1(2) - Y_1(0) + Y_1(0) - Y_0(0)|M] - \mathbb{E}[Y_1(1) - Y_1(0) + Y_1(0) - Y_0(0)|B_2] \quad \text{(S.3)}$$

$$\overset{rearranging}{=} \mathbb{E}[Y_1(2) - Y_1(0)|M] + \mathbb{E}[Y_1(0) - Y_0(0)|M] - \mathbb{E}[Y_1(1) - Y_1(0)|B_2] -$$

$$\mathbb{E}[Y_1(0) - Y_0(0)|B_2] \quad \text{(S.4)}$$

$$\overset{PT0}{=} \mathbb{E}[Y_1(2) - Y_1(0)|M] - \mathbb{E}[Y_1(1) - Y_1(0)|B_2] \quad \text{(S.5)}$$

$$= \Delta_{1,M}(2,0) - \Delta_{1,B_2}(1,0). \quad \text{(S.6)}$$

Thus, under $PT0$, it holds

$$Q = \Delta_{1,M}(2,0) - \Delta_{1,B_2}(1,0). \quad \text{(S.7)}$$

Text S4. Common Bird Monitoring Survey dataset.

We use the German Common Bird Monitoring Survey (CBS) from 2005 to 2020 (Mitschke et al., 2005; Kamp et al., 2021) to validate the ornitho data. Initially, we planned also using the CBS dataset for the main analysis, however, due to a very low overlap of the sampling areas with the treatment area[14] we now only use it for validating the ornitho data. The results show that ornitho data would indeed appear to be valid for depicting changes in bird populations (Table S5; Figure S11).

For the CBS, volunteers systematically collect information about birds in a 1x1 km large sampling area four times a year (between March 10 and June 20). Each area was sampled four times record to ensure that different breeding birds are covered by the data. The sampling areas were defined in two points in time. In 2003 based on a state(s)-wide initiative, a random sample of 100 areas were drawn in the region of Lower Saxony and Bremen (Mitschke et al., 2005). This was followed by a national-wide initiative, in which another 198 areas were drawn based on a two-layered stratified random sampling approach (Mitschke et al., 2005; Mitschke, 2008). The first stratum is based on Germany's "environmental regions" (which comprise information on soil, climate and vegetation) and the second stratum on land cover types (e.g., grassland, arable land, and forest) (Mitschke et al., 2005). In our study, we only focus on Lower Saxony comprising in total 273 relevant sampling areas. Moreover, we only consider those areas that have both a recording in the CBS and ornitho data, resulting in 194 sampling areas.

[14] Especially, when we consider the overlapping area between result-based payments and action-based payment scheme to support birds and other animals (see *Section 2*). The average number of observations in either the 2008 or 2012 pilot region per year would be 25.4.

References of Supplementary Information

Busch, M., Katzenberger, J., Trautmann, S., Gerlach, B., Droeschmeister, R., & Sudfeldt, C. (2020). Drivers of population change in common farmland birds in Germany. *Bird Conservation International*, 30(3), 335-354.

Dickel, R., Reiter, K., Roggendorf, W., & Sander, A. (2010). Halbzeitbewertung von PROFIL - Zahlungen für Agrarumweltmaßnahmen. Braunschweig/Hannover, Germany.

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. *Journal of Ornithology*, 162(1), 1-15.

Keienburg, T., Most, A., Prüter, J. (2006): Entwicklung und Erprobung von Methoden für die ergebnisorientierte Honorierung ökologischer Leistungen im Grünland Nordwestdeutschlands. Norddeutsche Naturschutzakademie. Schneverdingen, Germany.

Lower Saxony Ministry for Environment, Energy, and Climate Protection (Niedersächsisches Ministerium für Umwelt, Energie und Klimaschutz) (2023). Agrarumwelt- und Klimamaßnahmen zur Verbesserung der Biodiversität (AUKM) des Niedersächsischen Ministeriums für Umwelt, Energie und Klimaschutz (MU). Hannover, Germany.

Lower Saxony Ministry of Food, Agriculture and Consumer Protection (2023). GL 5 - Artenreiches Grünland (GL51/GL52/GL53). Hannover, Germany.

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.

Mitschke, A. (2008). Amsel, Drossel, Fink und Star–Erste Ergebnisse aus fünf Jahren Monitoring häufiger Brutvögel in Niedersachsen und Bremen. *Vogelkundliche Berichte Niedersachs*, 40, 163-180.

Most, A., & Keienburg, 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.

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. *American Journal of Agricultural Economics*, 104(4), 1460-1484.