European Commission

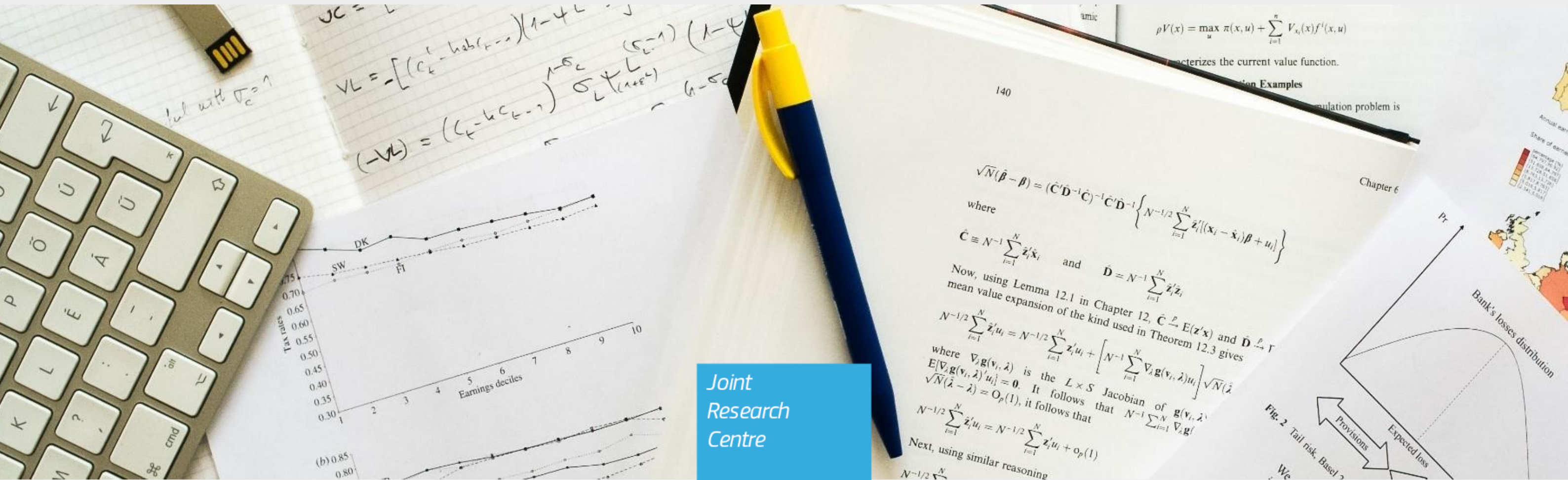
# New administrative geospatial data for agricultural policy evaluation

*An application to EU crop diversity obligations*

Brutti, Z., Freo, M., Serlenga, L.

2025

*JRC Working Papers in Economics and Finance, 2025/2*



Joint Research Centre

**EU Science Hub**
https://joint-research-centre.ec.europa.eu

JRC141131

Ispra: European Commission, 2025



How to cite this report: European Commission, Joint Research Centre, Brutti Z., Freo M. and Serlenga, L., *New Administrative Geospatial Data for Agricultural Policy Evaluation,* JRC Working Papers in Economics and Finance, 2025/2 European Commission, Ispra, 2025, JRC141131.

## Executive summary

This study showcases a new class of administrative, geo-spatial data sourced from agricultural subsidy registers as a powerful tool for agricultural policy evaluation. Beyond full national coverage and accurate identification of land use, the key novel feature of GeoSpatial Aid Application (GSAA) data consists in the ability to link agricultural parcels managed by the same farm, enabling causal analysis at the holding level. Using Spanish GSAA data, we evaluate an EU-wide environmental regulation, also unveiling the occurrence of strategic behaviour among a subgroup of farm holdings. We concisely discuss implications for future research endeavours in the agricultural policy domain.

***Policy context***

Assessing the impact of agricultural policies on various outcomes of interest, such as those in the environmental domain, implies substantial data demands. When obligations are set at the farm level, holding-level data is needed to produce a credible analysis of the causal impact of policy interventions through changes in farming practices. Many ex-post, causal studies are based on farm survey exercises or, when administrative parcel-level data is employed, are limited to specific sub-national areas. Most importantly, studies employing administrative data with universal coverage of the farm population, and relating individual land parcels to their farm holding, are extremely rare. This is a serious limitation for the evaluation of agricultural policies imposed at the holding level, especially if obligations depend on specific holding characteristics such as total size.

***Key conclusions***

- GeoSpatial Aid Applications (GSAA) data is a powerful policy evaluation tool, especially when individual farm holding identifiers are included in the dataset.
- GSAA data combines the features of i) universal coverage of common agricultural policy applicants, ii) parcel-level information on land use, iii) exact geo-location of land plots iv) ability to calculate holding-level characteristics, such as total size, specialisation and land-allocation choices.
- The 2014 crop diversification requirement had an impact on the cropping choices and land allocation of Spanish holdings, namely on average crop varieties grown, share of land dedicated to the main crop and compliance indicators at the policy thresholds
- Policy impacts for the diversification requirement are significantly stronger at the 10-ha threshold with respect to the 30-ha threshold, indicating a stronger policy pressure on smaller-sized holdings and higher levels of pre-policy compliance among larger holdings
- A minority of farmers has engaged in strategic behaviour aimed at avoiding the new diversification obligations, by reducing the size of their arable land so to remain below the size thresholds foreseen by the policy.

# New administrative geospatial data for agricultural policy evaluation: an application to EU crop diversity obligations

Zelda Brutti*[1], Marzia Freo[2], and Laura Serlenga[3]

[1,2]European Commission, Joint Research Centre (JRC)
[3]University of Bari

January 22, 2025

**Abstract**

This study showcases a new class of administrative, geo-spatial data sourced from agricultural subsidy registers as a powerful tool for agricultural policy evaluation. Beyond full national coverage and accurate identification of land use, the key novel feature of GeoSpatial Aid Application (GSAA) data consists in the ability to link agricultural parcels managed by the same farm, enabling causal analysis at the holding level. Using Spanish GSAA data, we evaluate an EU-wide environmental regulation, also unveiling the occurrence of strategic behaviour among a subgroup of farm holdings. We concisely discuss implications for future research endeavours in the agricultural policy domain.

Keywords: GSAA data; geospatial data; common agricultural policy; environmental policy evaluation.

JEL classifications: Q12, Q18, Q58

*Corresponding author. Email: zelda.brutti@ec.europa.eu. Address: JRC S.3 Science for Modelling, Monitoring and Evaluation, Competence Centre on Microeconomic Evaluation (CC-ME). TP 267 – 26b – Via E. Fermi 2749 - 21027 - Ispra (VA) - Italy. We are grateful to Francesco Gianola, Sophie Helaine and the members of the JRC S.3 CC-ME team for their support and inputs throughout, and to participants to the 2024 EEA/ESEM Congress, the 23rd Journées Louis-André Gérard-Varet conference, the XXXVI SIEP Conference, and members of the NYU Abu Dhabi Social Sciences Division for their helpful comments. 

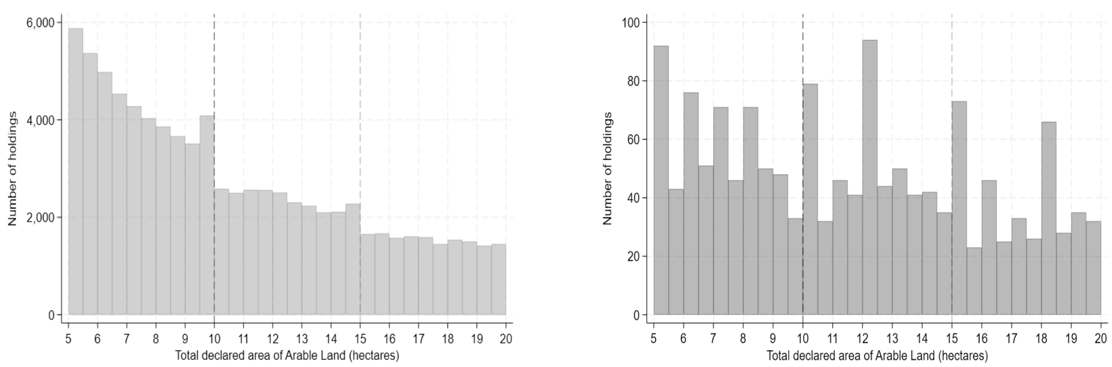
# 1 Introduction

The Common Agricultural Policy (CAP) is the European Union's largest expenditure program, accounting for roughly one third of the Union's total budget every year. The vast majority of financial resources executed under the CAP umbrella takes the form of direct payments to EU farm holdings, with the main purpose of supporting farm income and enhancing the competitiveness of the sector. Remarkably, nearly half of the agricultural income of the average EU farm derives from direct financial support (European Parliament 2023, Directorate-General for Agriculture and Rural Development 2017). Over the past two decades, the EU's policy agenda has witnessed a significant shift in focus towards environmental sustainability, and the CAP regulation has followed suit, with the promotion of environmentally friendly farming practices becoming a core objective. To receive CAP payments, farmers have to comply with rules that range from limits on the application of mineral fertilisers, to provisions that keep specific areas of nature free from agricultural use.

Assessing the impact of these eco-friendly practices on environmental outcomes implies substantial data demands, even leaving aside the issue of final environmental outcome measures. The fact that most environmental requirements are set at the farm level generates the need of holding-level data to produce a credible analysis of the causal impact of policy interventions through changes in farming practices. However, most of the existing evaluation studies draw conclusions based on simulations produced by models of the EU farming sector, operating either at the regional or the representative-farm level. The few ex-post, causal studies assessing changes actually occurred due to new environmental obligations are based on farm survey exercises or, when administrative parcel-level data is employed, are limited to specific sub-national areas. Most importantly, none of the existing evaluations employing administrative data is able to relate land parcels to their farm holding, which is a serious limitation for the evaluation of policies imposed at the holding level and often triggered by specific holding characteristics such as total size. This study pioneers a causal evaluation of a major EU environmental policy on farms, utilizing administrative data with universal coverage of a large EU country, with detailed information on land use at the land parcel level, and allowing for the identification of individual farm holdings.

The CAP policy component we focus on in this article is the so called 'Greening' package, a major initiative that was introduced with the 2014-2022 CAP and established new environmental requirements for farmers across the whole Union (Regulation EU No 1307/2013). Section 3 provides a thorough overview of the policy package and of the component we mainly focus on, the crop diversification (CD) obligation. The CD requirement, as well as the Greening package as a whole, received criticism for their low ambition and their weak potential for producing environmental impacts, despite the large financial commitment (European Court of Auditors 2017). However, rigorous ex-post evaluation studies on the diversification requirements are very hard to come by. For instance, the CD requirement has a size-threshold-based structure that lends itself to the implementation of discontinuity designs, enabling local impact estimation. However, discontinuity designs are demanding and draw on large amounts of data close to the thresholds of interest, further, it is crucial that data on key dimensions, such as the forcing variable on which thresholds are based, is precisely reported. Even the largest farm surveys such as as FADN (*Farm Accountancy Data Network*) struggle to comply with such requirements; in particular, the tendency to approximate land sizes to 'round' values is a common phenomenon and an insidious issue for RD studies, as illustrated in Figure 1.

Figure 1: Distribution of Spanish holdings by arable land size




(a) From GSAA administrative data (2022)

(b) From FADN survey data (2021)

This article is the first to build a counterfactual exercise evaluating the Greening crop diversification requirements using administrative data generated by the Spanish GeoSpatial Aid Application (GSAA) system. This is a component of the Integrated Administration and Control System (IACS), a set of computerised databases defined by Art. 68 of EU Regulation 1306/2013 towards the management and distribution of Common Agricultural Policy subsidies and whose structure is common to all EU countries. The universal coverage and precision of the administrative data generated by the GSAA system provide ideal circumstances for a rigorous application of discontinuity designs, as well as for the detection of previously undocumented phenomena, such as strategic behaviour by EU farmers in response to environmental policy obligations.

The first finding from this study is that the new environmental policy requirement had an impact on the cropping choices and land allocation of Spanish holdings, identifying sizable discontinuities in average crop varieties grown, share of land dedicated to the main crop and compliance indicators at the policy thresholds. The second finding is that policy impacts are significantly stronger at the 10 ha threshold with respect to the 30 ha threshold, indicating a stronger policy pressure on smaller-sized holdings and higher levels of pre-policy compliance among larger holdings. The third finding is that a share of farmers engaged in policy avoidance behaviour by reducing the size of their arable land so to remain below the size thresholds triggering the new diversification obligations; this is a phenomenon that is common in size-contingent policies (Garicano, Lelarge, and Van Reenen 2016) but that no previous evaluation of Greening has documented before. We provide a quantification of the avoidance behaviour and we estimate the effects of diversification obligations at the relevant policy thresholds after correcting for strategic selection.

Beyond the methodological contribution, the findings on the diversification requirement presented in this article are significant on their own. This environmental obligation has established standards across the EU over the past nine years (2014-2022) and its impact will continue to shape EU farming activities in the next future, as it has been built into the CAP strategic plans of several EU member states for the 2023-2027 programming period. Finally, our findings entail policy lessons that may be relevant for other size-contingent farm-level policy interventions implemented in the EU context.

The remainder of the article is structured as follows. Section 2 places the contribution of this study within the related literature; 3 provides background and detail about the environmental policy component that is the object of evaluation; Section 4 presents Spanish GSAA data; Section 5 illustrates novel findings on farmers' policy avoidance behaviour; Section 6 lays out the methodology employed for policy effect estimation after accounting for avoidance; Section 7 presents results and Section 8 concludes.

## 2 Related literature

Agricultural policy evaluation in the European features a wide range of methodological approaches and varying geographical scope. In most cases, the common objective of evaluation studies is to quantify the reaction of EU farmers to new policies, such as environmental requirements, in order to enable subsequent impact assessments on outcomes of interest in the economic, environmental or social dimension. The most frequently encountered difficulties concern the availability of good data for evaluation, ideally featuring a level of detail high enough to detect changes in farmers' behaviour, and a geographical coverage wide enough to draw conclusions valid beyond specific subnational areas. A good illustration is provided by the evaluation history of the the crop diversification obligation, which was introduced to EU farming in 2014 and which constitutes our empirical application. One wave of assessment studies was carried out in very early stages of the 2014-2020 CAP policy cycle, either before its implementation or in its first years, employing mostly predictive methodologies and structural modelling. These models receive as inputs pre-2014 data on holding characteristics and behaviour and simulate the expected reaction to the new policy. Examples from this literature are Gocht et al. (2017) and Van Zeijts et al. (2011), who make their predictions based on a large agro-economic model which operates with data considered representative of the farm-category (CAPRI); Louhichi et al. (2018), who mostly look at economic outcomes using an individual farm-level model based on data from the FADN survey (IFM-CAP); Mahy et al. (2015), whose individual-farm-level simulation model is calibrated to the region of Flanders.[1]

On the other hand, studies based on ex-post evaluation approaches are based on directly-observed farm behaviour and changes thereof. This eliminates the need to impose strong assumptions on policy take-up rates, and on the kind of reactions of EU farmers to the new policy; on the other hand, this sort of studies escalates the need for high-quality, micro-level data to leverage. As a consequence, the number of ex-post evaluations is considerably smaller than those employing predictive models, and the majority focuses on specific geographic areas, such as sub-national regions. Examples are Michalek (2022) studying multidimensional outcomes on a subset of Slovak regions and Kurdyś-Kujawska, Strzelecka, and Zawadzka (2021) looking at small-farm profits in Poland.

Closing in on the scope of our article, Bertoni et al. (2018) focus on one Italian region over the period 2011-2016 and use detailed, georeferenced data at the agricultural parcel-level to analyse changes in farmland allocation due to the new environmental obligations. This data stems from the Land Parcel Identification System (LPIS), which originates from the same monitoring system that produces GSAA

[1]The general agreement from these and similar studies is that only small proportions of land were to expected to be reallocated because of diversification requirements and that small subsequent impacts should be expected; in fact, several sources have criticized the low ambition of the CAP program in terms of environmental outcomes (among others, see European Court of Auditors 2017, Pe'er, Bonn, et al. 2020, Pe'er, Zinngrebe, Moreira, et al. 2019, The Economist 2017, Burrascano et al. 2016).

data, i.e. the *Integrated Administration and Control System - IACS*; however, connecting parcels to the holding managing them is not possible in LPIS sets. LPIS parcel-level data is also leveraged by a large ex-post study by Alliance Environnement and The Thünen Institute (2018), along with other sources such as the FADN survey and the 2010 Farm Structure Survey. This comprehensive study targets the effects of crop diversification obligations and of the Greening package more broadly, encompassing 10 EU member states. Notably, the authors of this study explicitly point to an important drawback due to data constraints: land parcels are not linked to data about the farm holding, preventing the analysis of how the cropping choices of individual farms changed due to the new policy measures (Alliance Environnement and The Thünen Institute (2018), page 23). Finally, there is a small and recent set of projects belonging to the earth observation and land-use monitoring literature, which raise awareness about the existence and policy-relevance of GSAA data, such as Baiamonte, Voican, and Loudjani 2023 and Van der Velde et al. 2025; again, these studies do not mention the farm holding identifiers as a relevant dimension of GSAA data, with their focus lying on the accurate quantification of crop- and land use diversification in the EU.

This paper contributes towards bridging methodological gaps previously detected in ex-post evaluation studies in the EU context. The setup we illustrate combines the advantages of having highly-detailed information on parcel-level land use, and those of knowing the total size and specialisation of the farm holding managing land plots, a feature which is key especially when policies are imposed at the farm-level and may be contingent on the characteristics of these, such as the 2014 Greening package.

## 3 The 'Greening' crop diversification policy

The introduction of the Greening package did not alter the overall budget of direct payments to EU farmers, but made approximately one third of the direct support received by individual farm holdings conditional on adopting or maintaining farming practices that contribute to the EU's environmental and climate goals (European Court of Auditors 2017). The financial resources earmarked for Greening between 2014 and 2022 represented about 23% of the whole CAP budget, or about EUR 12 billion per year; for the individual farmer, Greening payments represented on average 30% of all direct payments received, and non-compliance was penalized with a financial aid loss of up to 1.25 times the Greening portion, plus additional fines since 2017 (European Union 2013, Gocht et al. 2017). The three mandatory practices farm holdings had to start complying with in order to qualify for the portion of direct support earmarked for Greening are *Crop diversification*, to enhance the resilience of soil and ecosystems; *Maintenance of permanent grassland*, to support carbon sequestration and protect habitats; *Reservation of ecological focus areas*, to aid biodiversity.[2] In this article, we focus on the first Greening requirement, crop diversification, which was a complete novelty of the 2014-2021 CAP cycle and was applied uniformly across the Union's territory.[3] The crop diversification (CD) requirement introduced in 2014 is set according to the size of the farm's arable land (AL), that is land worked regularly and generally under temporary agricultural crops;[4] the conditions of obligation to the CD practice are summarized in Table 1 (European Parliament 2022).

Table 1: Crop diversification requirements to receive 'Greening' payments

| **Size of AL** | **Minimum requirement** | **Additional conditions** |
|---|---|---|
| Less than 10 hectares | -no requirement- | -no requirement- |
| Between 10 and 30 hectares | 2 crops | main crop $\leq$ 75% of AL |
| More than 30 hectares | 3 crops | main crop $\leq$ 75% of AL;<br>2 main crops $\leq$ 95% of AL |

*Source: Article 44 of Regulation (EU) No 1307/2013*

Note that the new diversification requirement boils down to banning the practice of monoculture for holdings below 30ha and the practice of mono- or biculture for holdings above 30ha, if one defines 'monoculture holdings' as those farms dedicating more than 75% of their AL to one single crop, and 'biculture holdings' as those growing dedicating more than 75% of their AL to one single crop, or more than 95% of AL to two crops.

The CD requirement, as well as the Greening package as a whole, received criticism for their low ambition and their weak potential for producing environmental impacts; One of the main points of objection to the new diversification requirement, raised in early-assessment studies, concerned the fact

[2]The new Greening rules were not applied to farmers falling under special circumstances, such as those included in the 'small farmers scheme', organic farmers and other specific exemptions applicable due to individual conditions. Individual EU countries were also allowed to introduce equivalent farming practices for farmers to choose as a way to comply with Greening requirements.

[3]Both the maintenance of permanent grassland and the preservation of ecological focus areas are concepts that display significant variation across Member States, due to differences in the interpretation and application of the EU regulations (European Commission 2017, Beaufoy et al. 2009, Pe'er, Zinngrebe, Hauck, et al. 2017). Moreover, maintenance of permanent grassland had been actively targeted by the CAP regulation already from 2003 (Regulation 1782/2003).

[4]It includes cereals, vegetables and a variety of other temporary crops, as well as temporary grassland and fallow land. It excludes permanent crops such as vineyards and fruit trees, permanent grassland, kitchen gardens and other special crops such as berries.

that most EU farmers were already compliant before the policy was enacted, so that very little reaction was expected, despite the large financial commitment (European Court of Auditors 2017).

A new CAP policy cycle started in January 2023, carrying along a number of modifications to the 'green architecture' overseeing the environmental and climate performance of EU farming. However, an examination of the national strategic plans approved at the time of writing shows that eight EU countries have transferred the previous crop diversification standards or slightly different versions of these, *including the contingency of obligations on farm size thresholds*, into their new National Plans, either as ad addition to new requirements (Austria, Spain, Ireland, Poland) or as an allowed alternative to these (BE-Wallonia, France, Netherlands, Portugal) (European Commission, Directorate-General for Agriculture and Rural Development et al. 2023). Overall, crop diversification standards as molded by the 2014-2020 CAP programming period are still financially supported through direct payments or other types of funding in 18 Member States (European Commission, Directorate-General for Agriculture and Rural Development et al. 2023, p.433).

# 4 Data

The main source of data employed in this study is the universe of land use declarations filed by Spanish farmers during the 2022 Common Agricultural Policy (CAP) campaign, with the purpose of applying for EU subsidies. CAP income support must be applied for by EU farmers on a yearly basis and in Spain, the procedure is condensed in the filling in of a 'Single Application' (*Solicitud única*) for CAP aid. The Single Application procedure has been established in Spain since 2003, simplifying previous formalities and allowing to apply simultaneously for all support types the farmer might be entitled to. Typically, the application period runs from February to April and is followed by verification procedures by the regional authorities managing CAP funds, under the coordination of the Spanish Agricultural Guarantee Fund (*FEGA - Fondo Español de Garantía Agraria*). All support items the farmer is entitled to are then aggregated into a 'single payment' (*Pago único*), which is transferred to the farmer between October and December. The Single Application must be addressed to the regional authority in which the exploitation or the majority of its surface is located.

Spain has implemented an infrastructure called Geographic Information System for Agricultural Parcels, SIGPAC (*Sistema de Información Geográfica de Parcelas Agrícolas*), which allows the geographical identification of the parcels declared by farmers, in any aid regime related to the area cultivated or used by livestock.[5] It was initially conceived with the purpose of making it easier for farmers to submit applications, with graphic support, as well as to facilitate administrative and on-the-spot controls; nowadays SIGPAC has become an enormously useful tool in fields other than agriculture (e.g. geology, infrastructures, urban planning). SIGPAC identifies land parcels as identified by the land registry, as well as their possible subdivisions, called ‘enclosures’: when there are different uses within a parcel (e.g. fruit trees, arable land crops, pastures), enclosures will delimit the precise area devoted to each of such uses. Enclosures are the basic unit for requesting and managing the different forms of CAP aid. SIGPAC also identifies graphic records as filed by the individual agricultural holdings in their annual Single Application: these graphic records are called "Graphic Declaration Line". These lines represent contiguous areas declared by farmers within SIGPAC enclosures. Each Declaration Line corresponds to a specific aid request and features a homogeneous use, product variety and exploitation system (e.g. dry or irrigated). We obtained geo-referenced SIGPAC data for the 2022 CAP campaign and were able to map the Declaration Line data from the universe of Single Applications filed across the whole country that year, creating a dataset whose most attractive features are the universal coverage, the high level of detail in land use and the possibility to group parcels under their farm holding - a feature extremely rare in the data sources prevailing in agricultural economics literature. Along with the exact geo-location and size of each land portion, the dataset records information about its declared use, including the specific crops grown on it, the declared exploitation system and a rough classification of the type of aid requested. The identity of the farm holding, however, remains anonymous, so that further linkage to external data sources such as income registries is not straightforward. Panel (a) of Figure 2 illustrates the data coverage and panel (b) illustrates the level of detail in the declaration data.

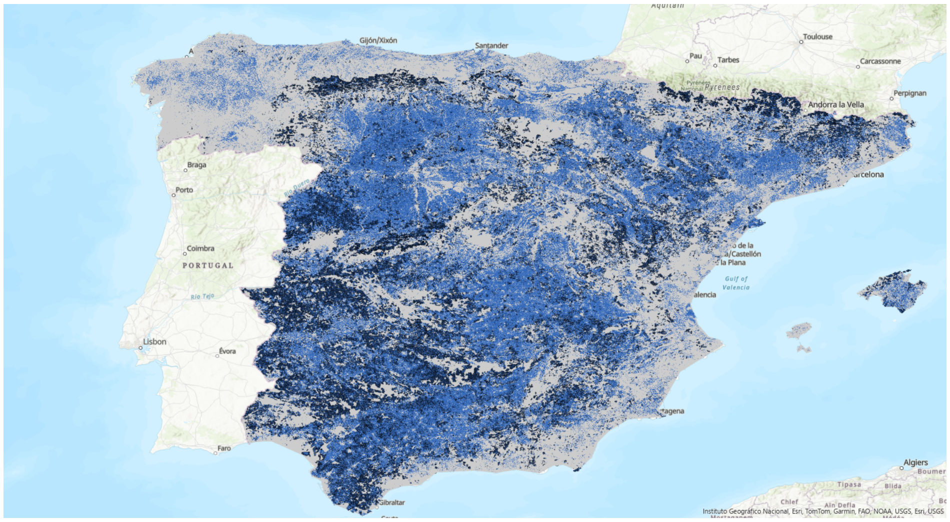
# 5 Descriptive findings and strategic behaviour

We have consolidated the declaration-level data into holding-level data, obtaining a dataset of 643,573 holdings, a figure consistent with the 648,691 Single Applications received according to a press state-

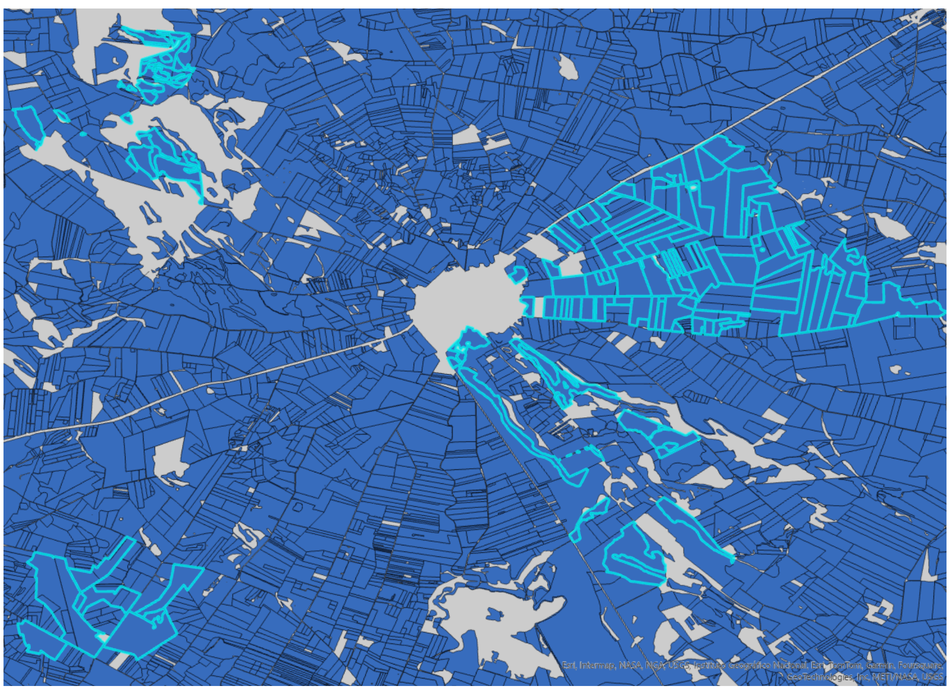
[5] In fact, Council Regulation (EC) No 1593/2000 of 17 July 2000 amending Regulation (EEC) No 3508/92 establishing an Integrated Management and Control System (IACS) for specific types of EU aid, requires the creation of a Digital Graphic System for the Identification of Agricultural Plots, using computer techniques for geographic information, also recommending the use of aerial or spatial orthoimages. This regulation also establishes that, as of January 1, 2005, each Member State must have a graphic database of all digitized crop plots, with a precision equivalent to at least a 1:10,000 scale cartography.

Figure 2: Illustration of the geocoded data from farmers' declarations (2022)

(a) National overview




(b) Local detail

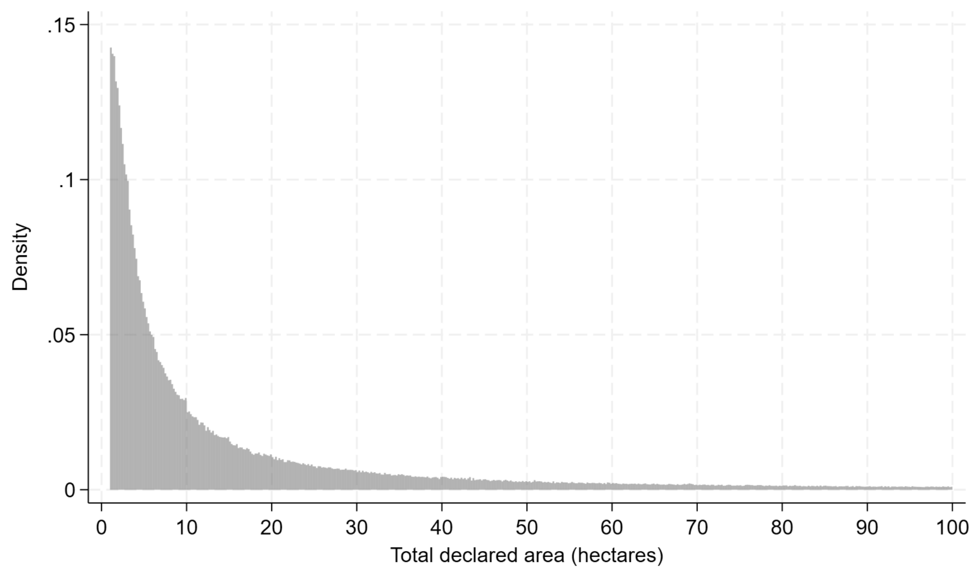


*Source: Authors' elaboration from SIGPAC/FEGA data (2022).* Agricultural perimeters declared by farmers are coloured dark blue. An illustration of a set of perimeters belonging to the same farmer are highlighted in panel (b).

ment by the Ministry of Agriculture back in June 2022 (Ministerio de Agricultura, Pesca y Alimentación 2022). For each farm holding, we have computed characteristics of interest under the Crop Diversification policy perspective, such as the total size of declared land, the total size of declared *arable land*, the number of arable land crops grown and the area of land occupied by each crop.

The size distribution of Spanish farms roughly follows a power law, featuring a very large number of small-sized holdings and very small numbers of large holdings; in fact, holding counts decrease as a power of their size. This sort of distribution is the most common one describing firm sizes as measured by turnover or number of employees, according to macroeconomics, trade, finance an industrial organization literature (Garicano, Lelarge, and Van Reenen 2016, Axtell 2001). Figure 3 illustrates the density of holdings by total size of declared agricultural area.

Figure 3: Distribution of Spanish farms by size of declared agricultural area (2022)



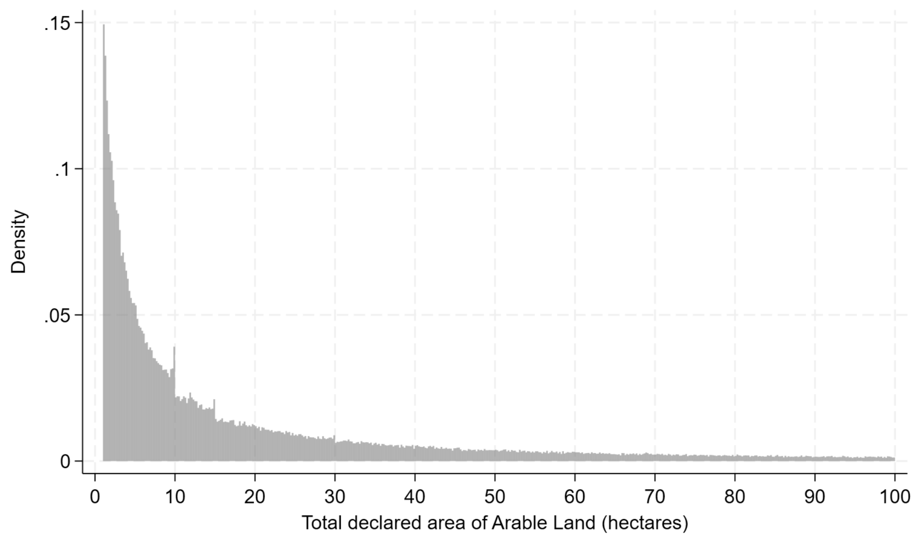

*Source: Authors' elaboration from SIGPAC/FEGA data (2022).* Holdings by total declared area (2,000 sqm bins), excluding holdings smaller than 1 hectare or larger than 100 hectares.

Importantly, the crop diversification (CD) requirement introduced in 2014 is not contingent on the total size of the farm, but only on the size of the farm's arable land (AL).[6] The AL size thresholds used to determine Greening obligations are policy-specific and do not have broader implications or applications, do not trigger fiscal incentives or any other subsidy scheme.

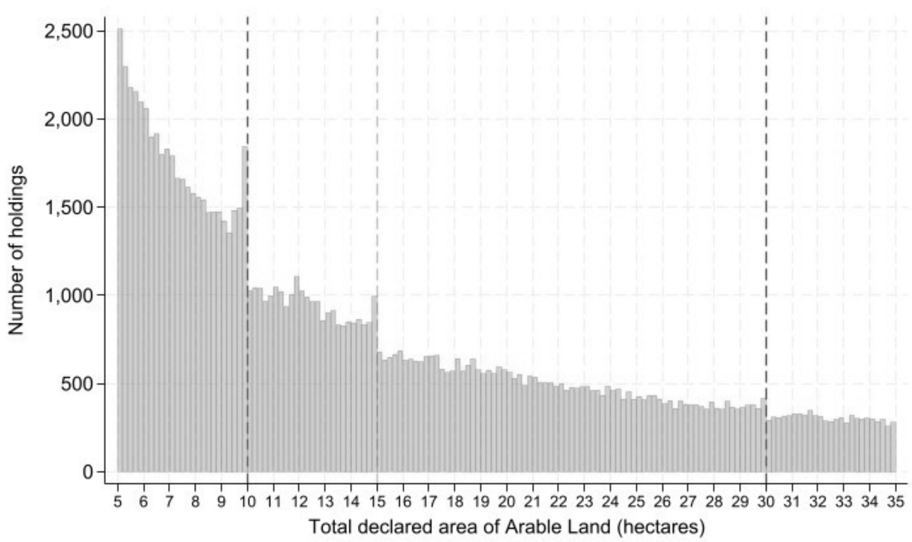
[6]See Footnote 4

Figure 4: Distribution of Spanish farms by declared size of arable land (2022)

(a) Full distribution between 1 and 100 hectares




(b) Detail around the 10 hectares crop diversification threshold




*Source: Authors' elaboration from SIGPAC/FEGA data (2022).*
Holdings by declared Arable Land (AL) area (2,000 sqm bins).

We have carried out a classification exercise, working through the roughly 370 codes identifying the possible uses of land portions declared by farmers and grouping them into crop categories, as well as distinguishing portions of land falling under the arable land class from those dedicated to uses excluded from the definition of arable land. Figure 4 illustrates the distribution of holdings by declared size of arable land. While the usual power law still characterises the distribution of Spanish arable land, there are departures from the pattern to be noted. Bulges of density are visible below the 10-hectares-threshold, below the 15-hectares-threshold and, to a much weaker extent, below the 30-hectares-threshold. As discussed in Section 1, the size thresholds of 10 ha and 30 ha are those corresponding to the Crop Diversification component of the Greening policy requirements; the 15 ha threshold is the one determining the obligation to comply with the Ecological Focus Area requirement, which goes beyond the scope of this article.

The occurrence of these size distortions is in line with what has been documented by previous literature on firm size distortions due to size-contingent policies and obligations (Garicano, Lelarge, and Van Reenen 2016). Economic theory predicts that there will be deviations from the 'natural' firm size distribution as a result of regulations that impose additional obligations or increases in costs for firms exceeding any specific size threshold. Firms are expected to bunch up below the policy threshold in order to avoid the obligation to comply with the regulation imposed on larger sizes, and this is precisely what we find empirical evidence for in the case of Spanish farm holdings.

In the case of agricultural holdings, sorting just below AL policy thresholds may be implemented via different strategies. One example is changing the ratio of permanent-to-AL crops grown on the farm, or of permanent grassland-to-AL crops, increasing the former and reducing the latter below the number of hectares triggering Greening obligations. Another possibility is to rent out a portion of AL to be managed by a different holding, so that the "excess" hectares enter the subsidy declaration of the renter rather than the one of the owner. The same way, farmers who, absent Greening, were renting arable land in excess of 10 / 15 / 30 hectares, may decide to rent fewer hectares as Greening is enforced. A third, more cumbersome way is to split the farm holding altogether, constructing "child" holdings each featuring fewer than 10 / 15 / 30 hectares of AL. These examples are just for illustrative purposes and do not constitute an exhaustive list of possibilities to obtain the result of farm holdings 'bunching' just below 10 hectares of AL size.

To the best of our knowledge, this is the first article that provides an empirical documentation of bunching behaviour of EU farmers below specific AL sizes as a result of the Greening policy provisions. In the Supplementary Material, Section A, we carry out a quantification of the magnitude of bunching at the 10-ha and 30-ha diversification policy thresholds and we use exempt farms as a placebo; Table 2 summarises results. We identify the magnitude of bunching to be in the ballpark of 11% excess holdings in the 2.5 ha just before the 10-ha policy threshold and 4% excess holdings in the 2.5 ha just before the 30-ha threshold. The number of farms missing in the 2.5 ha just after each of the two thresholds represents only about one quarter of the estimated 'bunchers', suggesting that the remaining part was originally spread in parts of the farm distribution further away from the threshold.

Table 2: Estimation of the magnitude of bunching around CD policy thesholds

| | Excess below | | Missing above |
|---|---|---|---|
| | N | % of total | N |
| 10-ha threshold | + 1,108 | + 11.7% | - 227 |
| 30-ha threshold | + 240 | + 4% | - 64 |

*Note:* Excess and missing farms are calculated in the 2.5ha below and above each threshold.

## 6 Methodology

Our outcomes of interest are measures of arable crop diversity among Spanish farms. Figure 5 plots the relationship between these outcomes and AL size, with flexible polynomials fitted to each side of the policy cutoffs; the left panels refer to the 10-ha threshold and the right panels to the 30-ha threshold. First, we focus on the share of monoculture farms and the share of mono- or biculture farms. We follow the Greening regulation in defining farms as 'monoculture holdings' if they grow only 1 crop on their AL, or dedicate more than 75% of their AL to that main crop. Similarly, 'biculture holdings' grow only 2 crops on their AL surface, or dedicate more than 75% of their AL to the first, or more than 95% to the first two crops; 'grass or fallow specialists' are never categorised as monoculture or biculture holdings, given their being exempt from diversification obligations. Panels (a) and (b) of Figure 5 plot the evolution of the share of monoculture or mono- and biculture farms along the AL size distribution. As a matter of fact, these shares can be read as the complements of policy compliance shares, if one recalls the fact that the diversification obligation comes down to banning the practice of monoculture (for holdings below 30ha) and biculture (for holdings above 30ha) defined as above. The plots show an overall negative relationship between the probability of being a mono- or biculture farm and AL size. In line with the findings of previous literature and ex-ante assessments of the Greening policy package (see reviews in Section 2), the data confirms that the majority of holdings appear to be 'naturally' compliant with the 2014 diversification requirements. In panel (a) one can observe that even below 10 ha of AL, that is, among the population which was exempted from diversifying, the share of monoculture holdings averages only between 0.3 and 0.4. Looking at larger farms in panel (b), the share of biculture holdings becomes lower than 0.2 even before reaching the 30-ha threshold triggering the ban.

Another point worth noting is the fact that the share of mono- or biculture holdings does not drop all the way to zero even for holdings owning above 10 or 30 ha of arable land. This is consistent with the fact that a minority of farm holdings is allowed to refrain from diversification due special schemes, equivalent eco-friendly practices such as organic farming (quantified at 8% of Spanish holdings in 2022)[7], or similar exceptions that we are not able to account for in the data.

Next, in Figure 5 we examine two additional outcomes of interest: the average variety of arable land crops grown on farms, in Panels (c) and (d), and the proportion of farmland allocated to the main crop or the first two crops, in panels (e) and (f). These variables have a dual interpretation: they serve as proxies for measuring crop diversity on farmland, and they represent the two key dimensions that farms can modify to comply with the new diversification requirements, namely increasing crop variety and adjusting land allocation. The four panels confirm the overall positive relationship between crop diversity and arable land size: farms owning more AL grow more varieties of crops on average and devote lower shares of their AL to the first or first two crops. However, it has to be noted that the slope of these relationships levels out as farms grow, so that after a certain size they become flat, implying that growth in size is accompanied by a less-than-proportional growth in crop diversity.

The outcome distributions in Figure 5 confirm the consequences of strategic sorting across the 10-ha threshold: in all three of the left panels one can notice evident changes in the relationship between outcome and the AL running variable, when approaching the 10-ha threshold from the left. As would be expected, farms engaging in avoidance behaviour have strong preferences for low diversification of their AL crops, so that they increase the share of monoculture holdings, decrease average crop counts and increase the average share dedicated to the main crop just before the 10-ha threshold. In line

[7]Data available on the Agri-food Data Portal, Organic Production, of the Directorate-General for Agriculture and Rural Development. Link https://agridata.ec.europa.eu/extensions/DashboardIndicators/OrganicProduction.html, accessed on 2024-11-26

with the previous evidence that bunching was substantially less at the 30-ha cutoff, the phenomenon is qualitatively similar but considerably more moderate around that size threshold as compared to the 10-ha threshold.

The strategic sorting behaviour invalidates the traditional Regression Discontinuity (RD) design as an identification strategy. In RD terminology, we are facing a case of violation of the no-manipulation assumption: sorting across the policy threshold introduces selection, so that and any discontinuities in outcomes measured at the threshold may be partially due to selection and not to the policy (Imbens and Lemieux 2008, McCrary 2008, Cattaneo and Titiunik 2022).

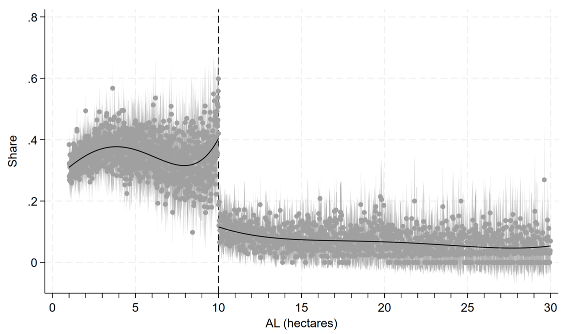
For this reason, we resort to the literature devoted to bunching (among others, see Diamond and Persson 2016, Saez 2010, Chetty et al. 2011, Kleven and Waseem 2013).[8] The core intuition behind the methodology is to predict how the manipulated distribution of farms and of outcomes would have looked, had there been no sorting across the threshold. This is achieved by using the 'safe' parts of the distribution to 'fill in' the shape inside the bunching regions (Diamond and Persson 2016). The complexity of the models and methods used to predict the shape inside the bunching regions varies depending on the application. Oftentimes, the prediction of the shape inside the manipulation area follows the construction of a theoretical model that explains the behaviour of the units of observation; the identification of the bunching region itself can be the object of estimation (Diamond and Persson 2016, Kleven and Waseem 2013). However, the key underlying assumption is that, in absence of manipulation, the distribution would have been smooth (Diamond and Persson 2016) and the standard approach boils down to fitting a flexible polynomial to the observed distribution, excluding data in a range around the policy threshold, and extrapolate the fitted distribution to the threshold (Kleven 2016).

In this analysis, the definition of bunching region is based on visual assessment of the data presented in Figure 4 and Figure 5, as well as on the quantification exercise presented in the Supplementary Material, Section A. We define the 'bunching' or 'manipulation' zone as the window including the last 2.5 ha of AL before and the first 2.5 ha of AL after the 10 and 30-ha policy threshold. Most of the excess density is concentrated in the last 2 ha before the threshold, so that our definition allows for imprecision in this assessment. After defining the bunching window, we approximate the relationship between the outcomes of interest and AL size with second-degree polynomials fitted to the data,

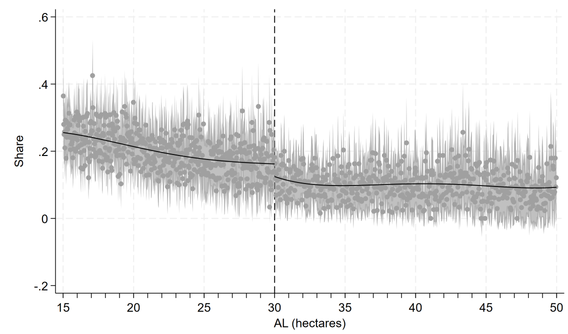
[8] Note that in this case, so-called "donut-RD" designs are not appropriate (Barreca, Lindo, and Waddell 2016). In our application we face excess bunching into a relatively wide area below the threshold (2 or 2.5 ha), rather than heaping at specific values of the running variable, e.g. due to rounding in reporting.

Figure 5: Relation between outcomes of interest and AL size, with fitted polynomials

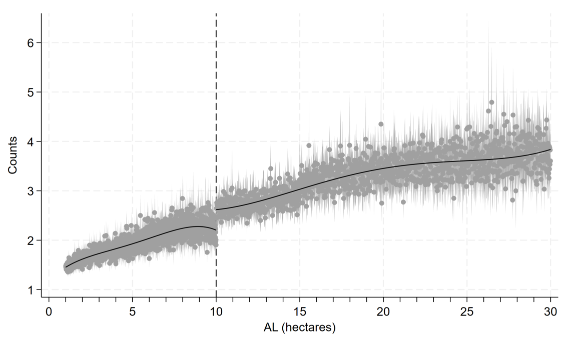
(a) Monoculture farms



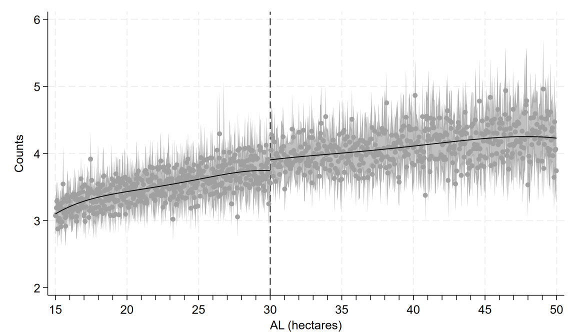

(b) Mono- or biculture farms



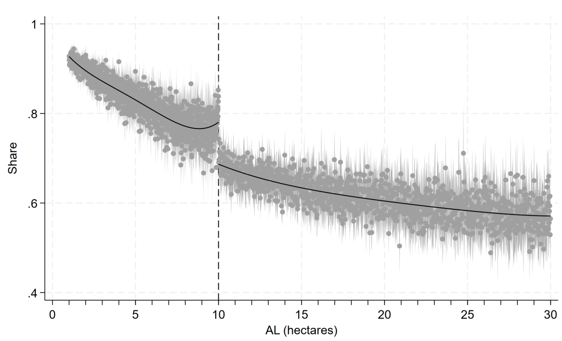

(c) AL crop varieties



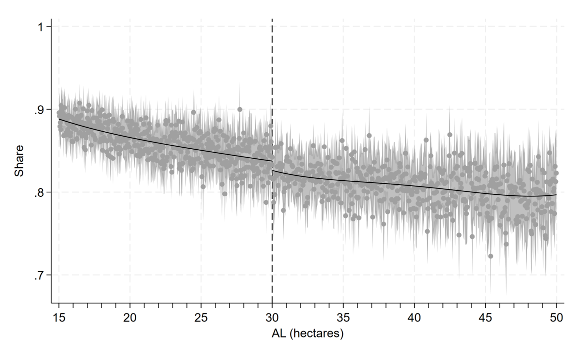

(d) AL crop varieties




(e) Share of AL devoted to main crop




(f) Share of AL devoted to first 2 crops




*Source: Authors' elaboration from SIGPAC/FEGA data (2022).*

separately on each side of the 10-ha threshold and 30-ha threshold, excluding the bunching window. Within the bunching window, the outcome distribution free of manipulation is approximated by the extrapolation of the polynomials from each side of the threshold.

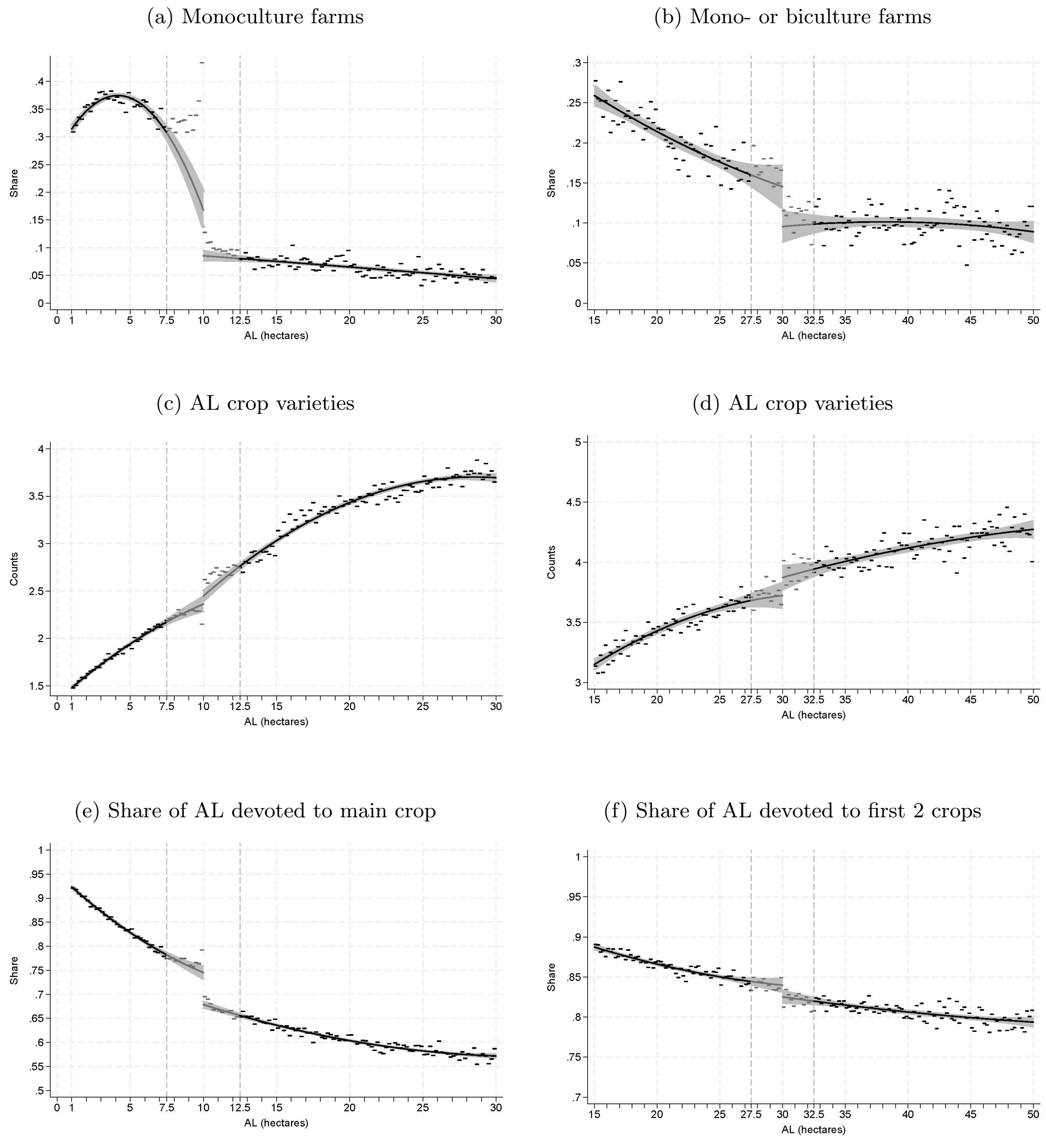
The discontinuity in outcomes at the threshold provides an estimate of the crop diversification policy effect. As noted previously, we may define the treatment as being a ban on the mono- or biculture practice, triggered at 10ha and 30ha of AL respectively. Further, we have noted around Figure 5 that compliance is significant even among holdings to the left of each policy threshold, and imperfect among holdings to the right of these. This leads to an interpretation of the threshold discontinuities as a local intention to treat (ITT) effects, i.e. the effects of treatment assignment rather than of treatment take-up (Hahn, Todd, and Van der Klaauw 2001, Cattaneo, Idrobo, and Titiunik 2023).

# 7 Results

Figure 6 illustrates main results graphically and Table 3 presents the associated numerical estimates. Policy effects are quantified by the magnitude of the discontinuities in polynomial intercepts at policy thresholds. Inference is based on conducting Wald tests on the difference in intercepts, allowing for arbitrary covariance in errors between each pair of polynomials (Hausman 1978, White 1982, White 1996).

We find that the introduction of the crop diversification obligation has significantly decreased the share of monoculture farms in the local neighbourhood of the 10-ha AL size threshold. Among farms approaching 10 ha of AL from the left, the proportion of monoculture holdings is estimated around 16%, and reduce to half when crossing the threshold, dropping by approximately 8 percentage points. The impact is significant at all conventional levels. At the same time, the local impact on average crop counts is estimated at approximately 0.08 crops, which represents a 3.5% rise from the pre-threshold average. However, the effect on crop counts is imprecisely estimated and statistical significance is only marginal. Finally, the decrease in the share of arable land dedicated to the first crop is highly significant and quantified at almost 7 percentage points, or a 9% drop across the policy threshold. The results suggest that the 2014 introduction of the crop diversification obligation has pushed away from the practice of monoculture a significant proportion of those holdings that would still be engaging in it in the counterfactual scenario. After adjusting for sorting, this proportion represents a minority of around 16% of the general population. A joint reading or the results of Table 3a suggests that the

Figure 6: Estimation of policy effects after correction for bunching




*Note:* Second-degree polynomials are fitted to each side of the 10-ha or 30-ha threshold. Dashed vertical lines indicate the start and the end of bunching zones, which are excluded from estimation; scattered dots indicate average outcomes in 0.2 ha bins; 95% confidence bands are shaded in grey.

shift away from monoculture may have been achieved through a combination of an introduction of new crops and an increase the land share dedicated to secondary crops, even though the latter margin is more precisely identified as compared to the former.

Table 3: Policy effect estimations

(a) 10-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Monoculture farms (share) | 106,413 | .168 | 48,310 | .085 | **-.083** | <0.01 |
| Crop varieties (counts) | 106,413 | 2.362 | 48,310 | 2.446 | **.084** | 0.094 |
| First crop (share) | 106,413 | .745 | 48,310 | .678 | **-.067** | <0.01 |

(b) 30-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Mono-/biculture farms (share) | 32,608 | .145 | 19,920 | .100 | **-.050** | <0.01 |
| Crop varieties (counts) | 32,608 | 3.723 | 19,920 | 3.874 | **.150** | 0.051 |
| First 2 crops (share) | 32,608 | .840 | 19,920 | .825 | **-.015** | .022 |

*Note:* Estimates of intercepts are obtained approximating the relationship between each outcome and second-order polynomials of AL size, from the left and from the right, excluding the bunching zone. Relationships are extrapolated into the bunching zone and up to the 10-ha or 30-ha threshold; the excluded bunching zone is set at 2.5ha before and after each threshold. Difference in intercepts estimate policy effects; p-values refer to Wald tests on the difference in intercepts and allows for covariance in errors between equations in each pair.

Turning to the results on the 30-ha policy threshold, in Panel (b) of Table 3b, all main takeaways are analogous to those on the 10-ha thresholds. However, in line with the finding of a weaker avoidance behaviour at the 30-ha threshold in Section 5, policy impacts are also weaker around this size class. The two results suggest an overall lower policy burden for larger holdings as compared to smaller ones, which is in line with the fact that diversification requirements grow less-than-proportionally with AL size. We estimate that the share of mono- or biculture farms dropped by approximately 5 percentage points, by 35% of their no-policy counterfactual around the threshold. The increase in crop varieties is quantified at 0.15 crops and is once again less precisely estimated with respect to the other two outcomes. We find the share dedicated to the first two crops to have decreased by approximately 1.5 percentage points.

Finally, Figure A.3 and Table A.1 in the Supplementary Material show that all results are robust to changing the bunching window to 2 ha on each side of the threshold. Further, Figure A.4 and Table A.2 show that increasing the order of polynomials to the third degree produces results that remain largely in line with previous specifications, but increases the noisiness of estimates and deteriorates inference, in line with Gelman and Imbens (2019).

## 8 Conclusion

The principal contribution of this article is demonstrating the enormous potential of administrative data from the Geospatial Aid Application (GSAA) information system to inform policy evaluation. The GSAA data provides a unique opportunity to analyze the impact of policy interventions on agricultural practices at a high level of spatial detail in land use, coupled with the ability to associate land parcels to the farm holding managing them. Especially when policies are targeted at the holding level and depend on holding characteristics such as total size or arable land size, this feature is key towards a credible identification of causal relationships between policies and their effects.

In this article, the first-time use of GSAA data for policy evaluation at national scale has yielded novel empirical evidence on policy avoidance and sorting behaviour of EU farmers at the size thresholds triggering crop diversification obligations. Essential to the identification of this result were the ability to measure the size of land plots and identify exact land use more precisely than what is commonly achieved through survey exercises such as the Farm Accountancy Data Network or agricultural census exercises, both of which are affected by rounding bias and coarse land use classifications.

The documentation and quantification of avoidance behavior in response to the 2014 Greening obligations is highly relevant in at least two respects. Firstly, it provides valuable insights for policymakers, shedding light on the perceived burden of the new requirements on farmers and offering a foundation for cost-benefit analyses. Secondly, it offers methodological guidance for future policy evaluation exercises, cautioning against the use of naive Regression Discontinuity (RD) designs, which are invalidated by threshold manipulation.

The analysis in this article also enriches the existing literature on the effects of the 2014 Greening package, which accounted for nearly a quarter of the total CAP budget during the 2014-2022 period and which continues to have a profound influence on farming practices in Europe, by contributing to shape the 2023-2027 CAP strategic plans.

Our innovative analysis based on GSAA data was able to confirm previous findings that only a minority of farm holdings were impacted by the introduction of crop diversification requirements, due to the fact that most are spontaneously compliant with the low level of ambition of the latter. We also identify heterogeneity in policy burden across holding size, in favour of larger holdings; this heterogeneity is a result of the less-than-proportional relation of Greening diversification requirements to AL size and it is empirically verified both in the intensity of avoidance behaviour and in the estimation of local policy effects on diversification levels.

We were able to provide numerical estimations of intention-to-treat effects of the crop diversification requirements around the 10-ha and 30-ha size thresholds. Our results demonstrate a local increase in diversity measures, which are sizeable considering that the share of holdings pushed to compliance by the introduction of the new requirements is small.

Finally, this study paves the ground for future research endeavours based on matching GeoSpatial Aid Application (GSAA) data from the yearly Common Agricultural Policy implementation campaigns to other geospatial data sources. Promising applications include precisely geolocated environmental outcome data such as soil, water or air quality measurements collected by national environmental agencies or EU-level initiatives such as the INSPIRE Geoportal. The ability to reconstruct the causal chain between agricultural policy initiatives targeted at individual farms, the resulting changes in farming practices at the farm level, and precisely geolocated environmental, social, or economic outcomes of interest represents a highly valuable perspective for future agricultural policy evaluation research and the evidence-based promotion of effective agricultural practices.

## References


Alliance Environnement and The Thünen Institute (2018). *Evaluation of the CAP Greening Measures.* Tech. rep. SWD(2018)478. European Commission.

Axtell, Robert L (2001). "Zipf distribution of US firm sizes". In: *science* 293.5536, pp. 1818–1820.

Baiamonte, Giorgio, Gilbert Voican, and Philippe Loudjani (2023). *Getting the most of Land Parcel Identification Systems (LPIS) and GeoSpatial Aid Application (GSAA) datasets.* Tech. rep. JRC Technical Report JRC133145. European Commission.

Barreca, Alan I, Jason M Lindo, and Glen R Waddell (2016). "Heaping-induced bias in regression-discontinuity designs". In: *Economic inquiry* 54.1, pp. 268–293.

Beaufoy, Guy et al. (2009). *Permanent pastures and meadows under the CAP: the situation in 6 countries.* Tech. rep. Derwentside (UK): European Forum on Nature Conservation and Pastoralism.

Bertoni, Danilo et al. (Dec. 2018). "Farmland Use Transitions After the CAP Greening: a Preliminary Analysis Using Markov Chains Approach". In: *Land Use Policy* 79, pp. 789–800.

Burrascano, Sabina et al. (2016). "Current European policies are unlikely to jointly foster carbon sequestration and protect biodiversity". In: *Biological Conservation* 201. Publisher: Elsevier, pp. 370–376.

Cattaneo, Matias D, Nicolás Idrobo, and Rocio Titiunik (2023). "A practical introduction to regression discontinuity designs: Extensions". In: *arXiv preprint arXiv:2301.08958.*

Cattaneo, Matias D and Rocio Titiunik (2022). "Regression discontinuity designs". In: *Annual Review of Economics* 14.1, pp. 821–851.

Chetty, Raj et al. (2011). "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records". In: *The quarterly journal of economics* 126.2, pp. 749–804.

Diamond, Rebecca and Petra Persson (2016). *The long-term consequences of teacher discretion in grading of high-stakes tests.* Tech. rep. National Bureau of Economic Research.

Directorate-General for Agriculture and Rural Development (2017). *CAP explained: direct payments for farmers 2015 2020.* eng. Luxembourg: Publications Office of the European Union. ISBN: 978-92-79-67876-9.

European Commission (Mar. 2017). *Report from the Commission to the European Parliament and the Council - on the implementation of the ecological focus area obligation under the green direct payment scheme.* Tech. rep. SWD(2017)121 final. Brussels.

European Commission, Directorate-General for Agriculture and Rural Development et al. (June 2023). *Mapping and analysis of CAP strategic plans - Publications Office of the EU.* Tech. rep. Publications Office of the European Union, 2023.

European Court of Auditors (2017). *Greening : a more complex income support scheme, not yet environmentally effective.* eng. Special report No 21. LU: Publications Office.

European Parliament (Apr. 2022). *First pillar of the common agricultural policy (CAP): II – Direct payments to farmers.* en.

— (Apr. 2023). *Financing of the CAP.* en.

European Union (2013). *Regulation No 1307/2013 of the European Parliament and the Council establishing rules for direct payments to farmers under support schemes within the framework of* *the common agricultural policy and repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009*.

Garicano, Luis, Claire Lelarge, and John Van Reenen (2016). “Firm size distortions and the productivity distribution: Evidence from France”. In: *American Economic Review* 106.11, pp. 3439–3479.

Gelman, Andrew and Guido Imbens (2019). “Why high-order polynomials should not be used in regression discontinuity designs”. In: *Journal of Business & Economic Statistics* 37.3, pp. 447–456.

Gocht, Alexander et al. (Sept. 2017). “EU-wide Economic and Environmental Impacts of CAP Greening with High Spatial and Farm-type Detail”. In: *Journal of Agricultural Economics* 68.3, pp. 651–681.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). “Identification and estimation of treatment effects with a regression-discontinuity design”. In: *Econometrica* 69.1, pp. 201–209.

Hausman, Jerry A (1978). “Specification tests in econometrics”. In: *Econometrica: Journal of the econometric society*, pp. 1251–1271.

Imbens, Guido and Thomas Lemieux (2008). “Regression discontinuity designs: A guide to practice”. In: *Journal of econometrics* 142.2, pp. 615–635.

Kleven, Henrik J (2016). “Bunching”. In: *Annual Review of Economics* 8, pp. 435–464.

Kleven, Henrik J and Mazhar Waseem (2013). “Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan”. In: *The quarterly journal of economics* 128.2, pp. 669–723.

Kurdyś-Kujawska, Agnieszka, Agnieszka Strzelecka, and Danuta Zawadzka (2021). “The impact of crop diversification on the economic efficiency of small farms in Poland”. In: *Agriculture* 11.3.

Louhichi, Kamel et al. (Apr. 2018). “Economic impacts of CAP greening: application of an EU-wide individual farm model for CAP analysis (IFM-CAP)”. In: *European Review of Agricultural Economics* 45.2, pp. 205–238.

Mahy, Louis et al. (2015). “Simulating farm level response to crop diversification policy”. In: *Land Use Policy* 45. Publisher: Elsevier, pp. 36–42.

McCrary, Justin (2008). “Manipulation of the running variable in the regression discontinuity design: A density test”. In: *Journal of econometrics* 142.2, pp. 698–714.

Michalek, Jerzy (2022). “Environmental and farm impacts of the EU RDP agri-environmental measures: Evidence from Slovak regions”. In: *Land Use Policy* 113. Publisher: Elsevier, p. 105924.

Ministerio de Agricultura, Pesca y Alimentación (2022). *Nota de Prensa: "Agricultores y ganaderos recibirán 4.856 millones de euros de las ayudas directas de la PAC 2022"*. Tech. rep. MAPA, FEGA.

Pe’er, Guy, Aletta Bonn, et al. (June 2020). “Action needed for the EU Common Agricultural Policy to address sustainability challenges”. en. In: *People and Nature* 2.2. Ed. by Kevin Gaston, pp. 305–316. ISSN: 2575-8314, 2575-8314. DOI: `10.1002/pan3.10080`.

Pe’er, Guy, Yves Zinngrebe, Jennifer Hauck, et al. (2017). “Adding some green to the Greening: Improving the EU’s Ecological Focus Areas for biodiversity and farmers”. In: *Conservation Letters* 10.5, pp. 517–530.

Pe’er, Guy, Yves Zinngrebe, Francisco Moreira, et al. (Aug. 2019). “A greener path for the EU Common Agricultural Policy”. eng. In: *Science* 365.6452, pp. 449–451. ISSN: 1095-9203. DOI: `10.1126/science.aax3146`.

Saez, Emmanuel (2010). “Do taxpayers bunch at kink points?” In: *American economic Journal: economic policy* 2.3, pp. 180–212.

The Economist (Nov. 2017). *Reforming the EU's Common Agricultural Policy in the context of the Sustainable Development Goals.*

Van der Velde, Marijn et al. (2025). "Policy relevant crop diversity monitoring based on earth observation and farmers declarations". In: *Environmental Research Letters*.

Van Zeijts, Henk et al. (2011). *Greening the Common Agricultural Policy. Impacts on farmland biodiversity on an EU scale.* Technical Report PBL-500136005. Den Haag, NL: Netherlands Environmental Assessment Agency PBL, p. 62.

White, Halbert (1982). "Maximum likelihood estimation of misspecified models". In: *Econometrica: Journal of the econometric society*, pp. 1–25.

— (1996). *Estimation, inference and specification analysis*. 22. Cambridge university press.

# Supplementary Material

New administrative geospatial data for agricultural policy evaluation: an application to EU crop diversity obligations

Zelda Brutti[1], Marzia Freo[2], and Laura Serlenga[3]

[1,2]European Commission, Joint Research Centre (JRC)
[3]University of Bari

January 22, 2025

## A Sorting across arable land policy thresholds

We expect strategic sorting across arable land (AL) thresholds to occur among those holdings that became subject to diversification obligations starting in 2014. The diversification obligation (and the Greening obligations in general) foresaw exemptions for special types of farms, mostly for those farms whose specialization is oriented to types of crops considered less straining for soil quality and biodiversity, such as grasses or other herbaceous forages. The exemptions applied in Spain for 2022 included holdings in which more than 75% of the arable land surface is used to produce legumes, grass or other crops herbaceous forages, or left fallow, or dedicated to a combination of both.[1] Holdings who are exempt from the Greening policy are not expected to show significant strategic sorting around policy thresholds, so that they are helpful towards the validation of the hypothesis that the bunching we observe at Greening thresholds are actually due to policy obligations.[2]

We have categorised farms into one of 12 categories on the basis of their land use declarations, depending on their *main* arable land use. For this analysis we consider the first four arable land use categories identified in the data: 83% of holdings declared either cereals or fallow land or temporary grassland or

[1]The "Grasses or other herbaceous forages" group the following species of grasses and legumes and their mixtures, which are traditionally either found in natural pastures, or are included in seed mixtures for pastures or meadows: Festuca, raygrass, Agrostis, Arrhenantherum, Dactyl, timothy, poa, clover (mixed with other grasses).

[2]Note that further exemptions are foreseen for those holdings applying mostly organic farming practices or practices considered 'equivalent' to Greening according to national definitions, as well as further exceptions based on the local agro-environmental conditions and specific needs; unfortunately these kind of holdings are not easily identifiable in our data, with the result that they will remain mingled among 'normal' farms in our analysis.

energy crops as their main arable land specialization. Despite the fact that our classification into "targeted" and "exempted" holdings is merely approximated based on 'main specialisation' and that we are missing some exemptions in the data, one would still expect holdings included in the 'cereals' and 'energy crops' categories to be significantly less likely to be exempted from diversification obligations with respect to holdings included in the 'fallow' and 'temporary grass' categories. We thus expect a significantly larger occurrence of sorting across the 10 ha and 30 ha thresholds for farms belonging to the first group with respect to those belonging to the second group.

## A.1 Magnitude of sorting

To estimate the magnitude of sorting or across the policy threshold, we follow one of the empirical approaches typical in the literature around "bunching" (Kleven 2016, Garicano, Lelarge, and Van Reenen 2016) of fitting a smooth function to the portion of interest of the empirical distribution, excluding the area in which bunching is suspected. Deviations between the predicted and the observed distribution within the bunching window provide a quantification of the sorting phenomenon. Based on the descriptive evidence on arable land size distribution (Figure 3) and crop diversity dynamics (Figure 5) collected in previous sections, we suspect sorting to happen approximately in the two hectares before and the two hectares after the Greening policy thresholds. This is in line with literature finding behavioural responses to be mostly very local (Kleven 2016).

When focusing on the 10-ha diversification threshold, our empirical scenario is complicated by the presence of *two different* policy threshold within a relatively limited size range: the sorting behavior due to the 10 ha Crop Diversification (CD) threshold is potentially in part overlapping with the sorting behavior due to the 15-ha Ecological Focus Area (EFA) threshold, which is not a primary object of interest in our current analysis. We therefore allow for a generous bunching window that overarches both thresholds and estimate our counterfactual predictions excluding farms whose declared arable land size is larger or equal to 7.5 ha but smaller or equal to 17.5 ha. Results are robust to variations to this window. In terms of the estimation window, we choose the empirical distribution of Spanish farms between 2.5 and 25 hectares of arable land; once again, results are not sensitive to variations in the chosen window. Figure A.1 shows the results we obtained by fitting an exponential function to the distribution of holdings belonging to estimation window, but excluding those in the bunching window. Panel (a) shows results for holdings specialising in 'cereals and energy crops', while panel (b) considers those identified as 'fallow and grass' specialists in our classification exercise.

Comparing the counterfactual density prediction to the observed distribution, one can observe the presence of an excess density of holding before the 10-ha CD threshold (as well as before the 15-ha EFA threshold), followed by missing density to the right of such threshold. The difference between the observed number of holdings and the smooth counterfactual, in the window between 7.5 and 10 ha, yields an estimated 1,100 holdings having strategically sorted to the left of the 10-ha threshold. This corresponds to an excess of approximately 11.7% of holdings in the AL size class betwen 7.5 and 10 ha. In the right-hand side of the bunching window, that is between 10 and 12.5 ha, the estimated number of missing holdings is only 227 (3% of the total), suggesting that the remaining 'bunchers' may have originated from further away from the 10 ha threshold.

On the other hand, as expected, the pattern of excess-followed-by-missing density can not be spotted in the results of Panel (b), who focus on the population of fallow and grass specialists, most of whom we expect to benefit from Greening exemptions.

When repeating the analysis focusing on the 30-ha diversification threshold, we had to widen our size bins from 0.2 to 0.5 ha, due to the lower number of holdings found around that AL size class. The parameters of the exponential function have been estimated to fit the distribution of holdings in the window between 17.5 and 50 ha of AL, excluding those in the bunching window between 27.5 and 32.5 ha. Figure A.2 shows the results obtained for 'cereal and energy crop' specialists in Panel (a) and those for 'fallow and grass' specialists in Panel (b). Beyond a general increase in the volatility of the empirical distribution, due to the lower farm counts in each bin, it is evident that the bunching phenomenon at the 30-ha CD threshold is more limited with respect to the one found at the 10-ha CD threshold.

Our quantification of the excess density to the left of the 30-ha threshold yields an estimate of around 240 holdings having strategically sorted to stay below 30 ha of AL, representing approximately 4% of the population of holdings between 27.5 and 30 ha of AL. To the immediate right of the threshold, between 30 and 32.5 ha, we estimate only around 64 holdings missing, which again suggests that the remaining 'bunchers' may have originated from further away from the 30-ha threshold.

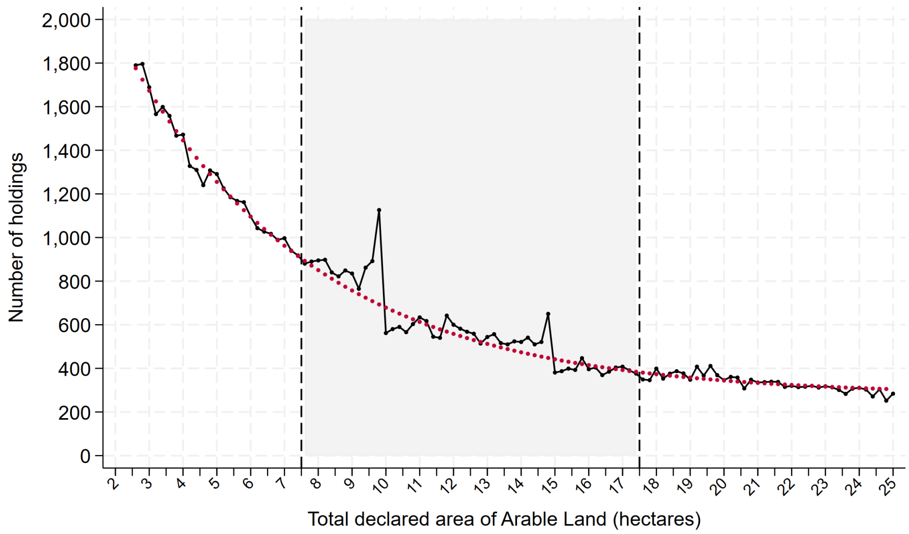
## References


Garicano, Luis, Claire Lelarge, and John Van Reenen (2016). "Firm size distortions and the productivity distribution: Evidence from France". In: *American Economic Review* 106.11, pp. 3439–3479.

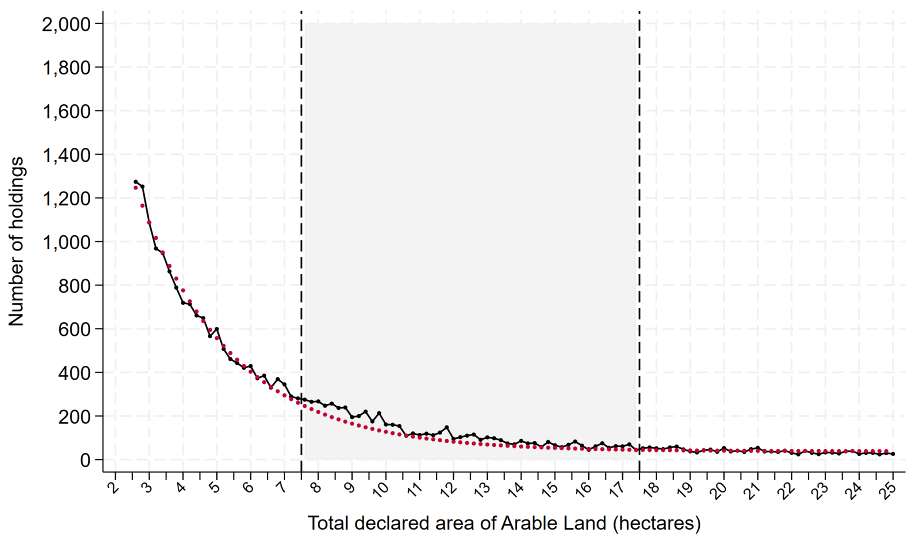
Kleven, Henrik J (2016). "Bunching". In: *Annual Review of Economics* 8, pp. 435–464.

Figure A.1: Actual distribution of holdings VS. exponential function fit - focus around 10ha

(a) Farms declaring cereals or energy crops as their main arable land crop



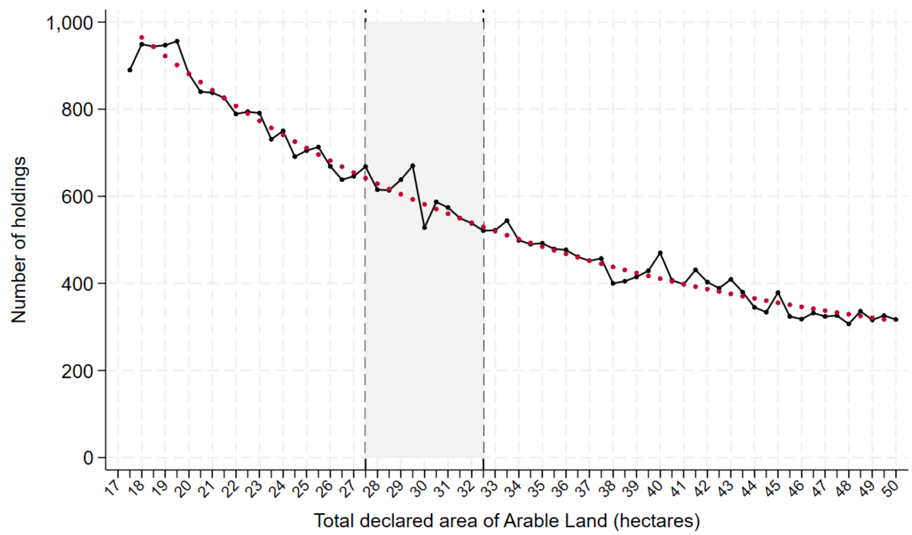

(b) Farm declaring fallow or temporary grassland for ≥ 75% of their AL



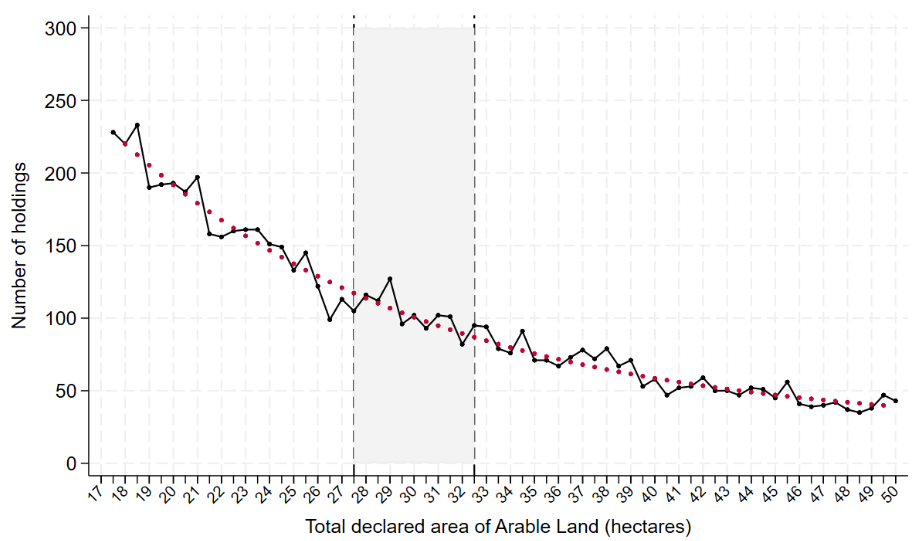

*Source: Authors' elaboration from SIGPAC/FEGA data (2022).*
Holdings by declared Arable Land (AL) area (2,000 sqm bins).
Areas shaded in gray have not been used to estimate the functional fit.

Figure A.2: Actual distribution of holdings VS. exponential function fit - focus around 30 ha

(a) Farms declaring cereals or energy crops as their main arable land crop




(b) Farm declaring fallow or temporary grassland for ≥ 75% of their AL

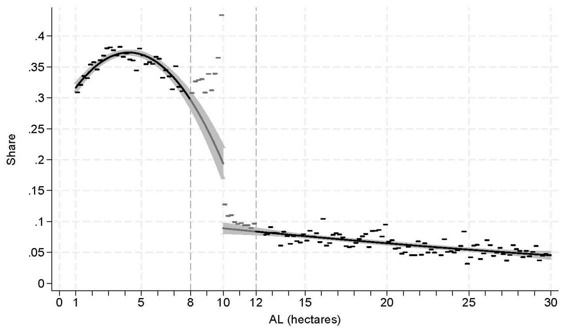



*Source: Authors' elaboration from SIGPAC/FEGA data (2022).*
Holdings by declared Arable Land (AL) area (5,000 sqm bins).
Areas shaded in gray have not been used to estimate the functional fit.

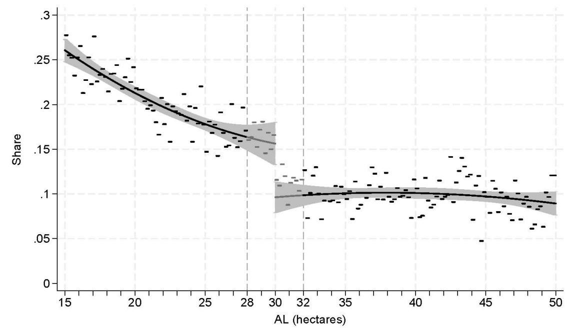
# B Supplementary figures

Figure A.3: Estimation of policy effects after correction for bunching (2-ha bunching window)

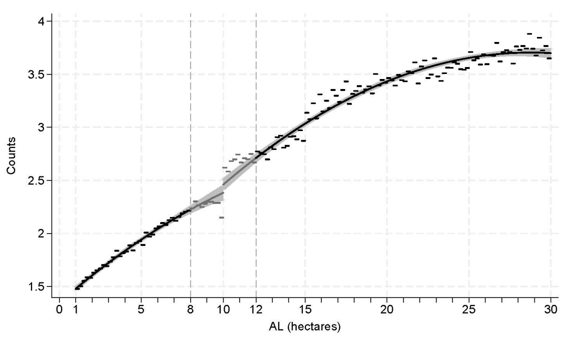
(a) Monoculture farms



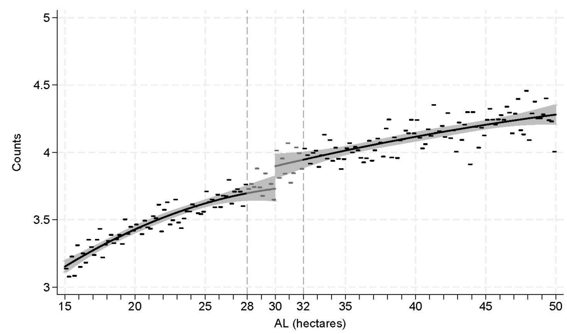

(b) Mono- or biculture farms



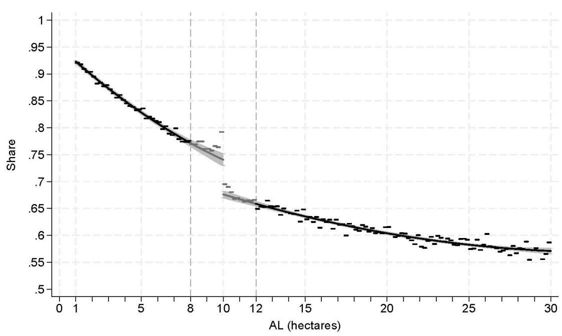

(c) AL crop varieties



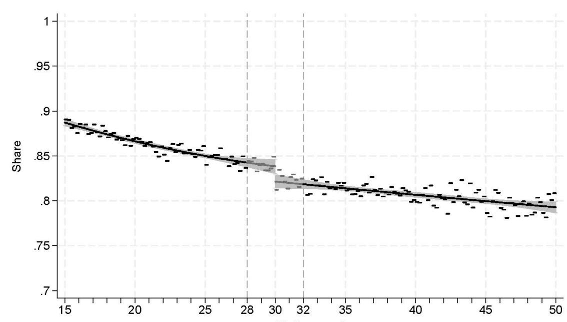

(d) AL crop varieties




(e) Share of AL devoted to main crop



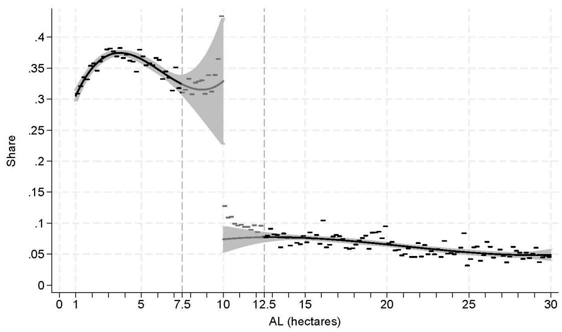

(f) Share of AL devoted to first 2 crops



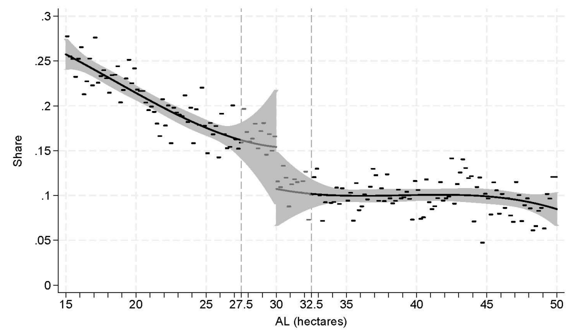

*Note:* Second-degree polynomials are fitted to each side of the 10-ha or 30-ha threshold. Dashed vertical lines indicate the start and the end of bunching zones, which are excluded from estimation; scattered dots indicate average outcomes in 0.2 ha bins; 95% confidence bands are shaded in grey.

Figure A.4: Estimation of policy effects after correction for bunching (3-rd order polynomials)

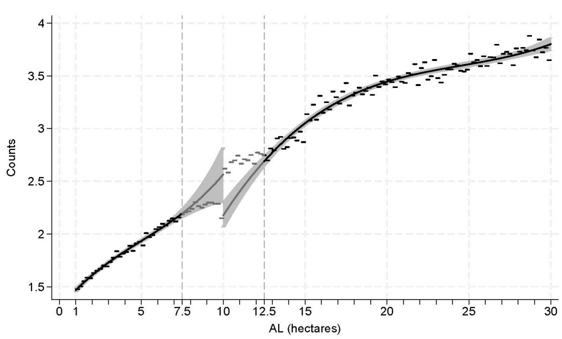
(a) Monoculture farms



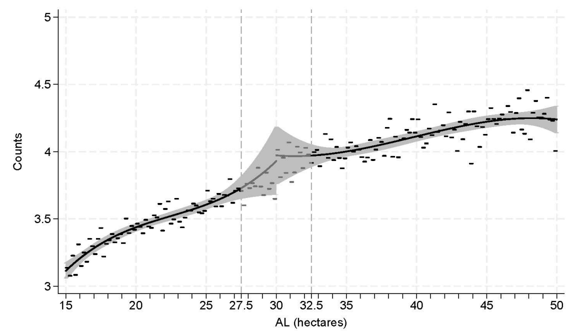

(b) Mono- or biculture farms



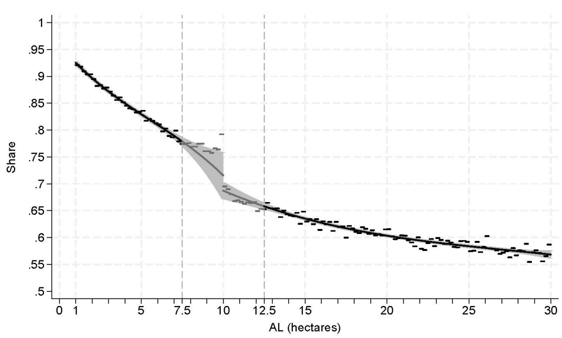

(c) AL crop varieties



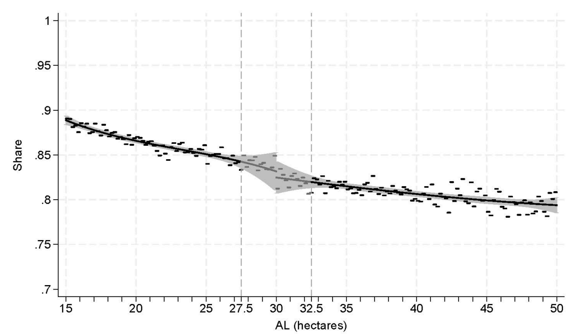

(d) AL crop varieties




(e) Share of AL devoted to main crop




(f) Share of AL devoted to first 2 crops




*Note:* Third-degree polynomials are fitted to each side of the 10-ha or 30-ha threshold. Dashed vertical lines indicate the start and the end of bunching zones, which are excluded from estimation; scattered dots indicate average outcomes in 0.2 ha bins; 95% confidence bands are shaded in grey.

# C Supplementary tables

Table A.1: Policy effect estimations - 2-ha bunching window

(a) 10-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Monoculture farms (share) | 110,444 | .194 | 50,815 | .089 | **-.105** | <0.01 |
| Crop varieties (counts) | 110,444 | 2.383 | 50,815 | 2.460 | **.077** | 0.061 |
| First crop (share) | 110,444 | .741 | 50,815 | .676 | **-.064** | <0.01 |

(b) 30-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Mono-/biculture farms (share) | 33,546 | .156 | 20,674 | .096 | **-.060** | <0.01 |
| Crop varieties (counts) | 33,546 | 3.731 | 20,674 | 3.900 | **.167** | 0.011 |
| First 2 crops (share) | 33,546 | .838 | 20,674 | .821 | **-.017** | <0.01 |

*Note:* Estimates of intercepts are obtained approximating the relationship between each outcome and second-order polynomials of AL size, from the left and from the right, excluding the bunching zone. The relationship is then interpolated up to the 10-ha threshold. The excluded bunching zone is set at 2ha before and after the threshold; p-values refer to Wald tests on the difference in intercepts and allows for covariance in errors between equations in each pair.

Table A.2: Policy effect estimations - 3rd degree polynomials

(a) 10-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Monoculture farms (share) | 106,413 | .329 | 48,310 | .074 | **-.255** | <0.01 |
| Crop varieties (counts) | 106,413 | 2.565 | 48,310 | 2.176 | **-.389** | <0.01 |
| First crop (share) | 106,413 | .716 | 48,310 | .687 | **-.028** | 0.220 |

(b) 30-ha threshold

| | Below threshold | | Above threshold | | | |
|---|---|---|---|---|---|---|
| Outcome | N | Intercept | N | Intercept | **Difference** | p-value |
| Mono-/biculture farms (share) | 32,608 | .154 | 19,920 | .107 | **-.047** | 0.226 |
| Crop varieties (counts) | 32,608 | 3.930 | 19,920 | 3.973 | **.043** | 0.800 |
| First 2 crops (share) | 32,608 | .716 | 19,920 | .687 | **-.028** | 0.627 |

*Note:* Estimates of intercepts are obtained approximating the relationship between each outcome and third-order polynomials of AL size, from the left and from the right, excluding the bunching zone. The relationship is then interpolated up to the 10-ha threshold. The excluded bunching zone is set at 2.5ha before and after the threshold; p-values refer to Wald tests on the difference in intercepts and allows for covariance in errors between equations in each pair.

## Getting in touch with the EU

### In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

### On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

## Finding information about the EU

### Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

### EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

### EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

### EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

**EU Science Hub**
Joint-research-centre.ec.europa.eu