# Agricultural Productivity and Biodiversity Effects: Theory and Evidence*

Chloé Antoine† Jean-Marc Bourgeon‡ José De Sousa§

July 3, 2026

**Abstract**

Agricultural specialization maximizes land use efficiency but it also amplifies pest pressure, thereby requiring heavy pesticide use. This paper evaluates the economic benefits of reducing specialization to leverage natural pest control. We develop a model in which the resilience of a farmer's field is endogenous to the crop and pesticide choices of neighboring farmers. We account for both within-crop and between-crop cross-field externalities. Using global data on 40 crops and 39,116 counties, we estimate the impact of these externalities on crop productivity with instrumental variables constructed from agro-climatic suitability measures. Our results show that a one-percent increase in crop diversity raises crop yields by 1.3 to 1.8%, while a one-percentage-point increase in a crop's own farmland share lowers that crop's yield by 9 to 13%. Counterfactual simulations indicate that revenue-maximizing reallocation would roughly double the diversity of crops grown and that even a marginal reduction in specialization raises production.

*Keywords:* Crop yield, biodiversity, crop diversity, pest control, ecosystem services
*JEL codes:* Q15, Q51, Q57

*For fruitful discussions and valuable comments, we are grateful to Alex Bao, François Bareille, Geoffrey Barrows, Stefan Baumgärtner, Gregory Corcos, Christophe Gouel, Ben Groom, Nikolai Jäger, Chanrithy Lao, as well as conference and seminar participants at BioEcon 26th Annual Conference, Bordeaux School of Economics, CREST, EAERE 30th Annual Conference, EAERE Winter School 2025, FAERE 12th Annual Conference, INRAE SMART, Paris-Saclay Applied Economics, Pau University, and USSEE 13th Biennial Conference. All errors are our own.

†CREST, École Polytechnique, Institut Polytechnique de Paris, France. Email: chloe.antoine@ensae.fr.

‡INRAE, UMR Paris-Saclay Applied Economics, and CREST, École Polytechnique, Institut Polytechnique de Paris, France. Email: bourgeon@agroparistech.fr

§LEMMA, Université Paris Panthéon-Assas and LIEPP, Sciences Po, France. Email: jose.de-sousa@assas-universite.fr.

# 1 Introduction

This paper evaluates the economic benefits of reducing agricultural specialization to leverage natural pest control. Farmers face a complex dilemma, pondering two factors critical to crop survival: abiotic factors, related to weather and soil conditions, and biotic factors, related to the proliferation of pests. On one hand, specializing in the most suitable crop enables to exploit the abiotic characteristics of the land, increasing yield returns. On the other, specialization provides an ideal environment for pests to spread, decreasing yield returns. The result is a dilemma between maximizing individual yields through monoculture and minimizing the risk of pest damages (Weitzman, 2000). Uncontrolled, pest damage can be severe: yield losses for staple crops like maize, rice, and wheat could range between 50% and 70% (Oerke, 2006).

The standard response to pest threats is the use of pesticides, which comes at a high ecological cost (Tilman, 1999). Pesticides contaminate groundwater, degrade soils, and drive biodiversity loss.[1] Pesticide use also imposes a social cost as biodiversity loss undermines key ecosystem services like pollination and natural pest control.[2] Additional costs arise from the negative effects of degraded ecosystem services on human health (Frank, 2024; Frank and Sudarshan, 2024).

To limit the detrimental effects of pesticides, several countries have banned the most dangerous substances and intend to limit the use of remaining chemicals through successive regulations. However, such restrictions may reduce yields, while global food demand is projected to rise by 46% over the next three decades (Gouel and Guimbard, 2017). In this context, reducing specialization and increasing crop diversity emerges as a promising solution to balance the need for both environmental sustainability and high productivity. Experimental studies have demonstrated that diverse ecosystems are generally more productive (Tilman, Polasky, and Lehman, 2005). Regarding agriculture, increasing crop diversity has the potential to naturally reduce pest pressure. The presence of multiple crops dilutes the availability of a pest's preferred food source, making it harder for the pest to proliferate (Pianka, 2011). In addition, crop diversity fosters a more complex ecosystem by attracting pollinators and natural pest predators (Bianchi, Booij, and Tscharntke, 2006; Redlich, Mar-

[1] Insecticides are identified as a major factor in the global 1% annual decline of insect populations (Van Klink et al., 2020), with losses reaching 76% near agricultural zones. In France, farmland bird populations fell by 25% between 1989 and 2009, the largest drop across habitats (Jiguet et al., 2012).

[2] In the US, the insect-mediated services are valued at $57 billion annually (Losey and Vaughan, 2006).

tin, and Steffan-Dewenter, 2018; Noack et al., 2022; Strobl, 2022). However, evidence on the economic viability of increasing crop diversity remains limited.

In this paper, we quantify global productivity gains from diversification. We combine a probabilistic model for crop production with geo-spatial data on land use, using a novel identification strategy. Our analysis thus contributes to the literature in two main ways. The first contribution is to propose a new model of crop production with endogenous biotic risk, stemming from pests. Building on the ecological approach of Weitzman (2000), Brock and Xepapadeas (2003), and Bellora and Bourgeon (2019), we assume that crop production is stochastic due to abiotic and biotic factors. Farmers choose which crop to grow and how much pesticide to use to maximize expected profits. Crucially, biotic risk depends not only on a farmer's own decisions but also on the crop and pesticide choices of neighboring farmers, generating *cross-field* biotic externalities. Accordingly, we distinguish two theoretical effects: a within-crop externality following a change in a crop's land share, and a between-crop externality following a change in the mix of other crops.

Specifically, a field's survival probability decreases with the number of neighboring fields growing the same crop and increases with the diversity of other crops species grown on neighboring fields and pesticide use. Using this model, we derive the equilibrium land distribution among crops based on individual farmers' decisions and a testable equation for crop productivity.

Our second contribution is to estimate the within- and between-crop effects on productivity. We operationalize these effects at the county level. Own-crop farmland shares capture within-crop effects, whereas the Shannon index, which accounts for both crop richness and relative abundance, captures between-crop effects. We address endogeneity, including the possibility that current productivity responds to prior crop diversity, by implementing an identification strategy designed to recover causal effects.

Using this strategy, which we discuss below, we find that both within- and between-crop effects significantly impact productivity. These effects are heterogeneous. Grouping the 40 crops into agronomic categories, we find them concentrated in cereals, a category characterized by large-scale specialization and substantial pest pressure. We also show that the between-crop effect is significantly larger for crops exposed to more pest species.

We then run two counterfactual exercises. First, we redistribute 1% of the dominant crop's area equally to all other cultivated crops and show that the production of the dominant crop *increases* by 13% on average. The reason is that the reduction in production due to

the decrease in cultivated area is more than offset by the increase in productivity. Second, we simulate the rent-maximizing allocation of farmland and find that revenue maximization calls for substantially greater crop diversity.

Prior studies on the productivity-biodiversity relationship in agriculture focused on wheat *genetic* diversity: planting fewer wheat varieties lowers yields in Canada (Carew, Smith, and Grant, 2009), and a 1% increase in genetic diversity raised mean wheat yields by 0.11% in southern Italy (Di Falco and Chavas, 2006). Smale et al. (1998) find more nuanced results: greater wheat genetic diversity increases yields in Pakistan's rainfed areas but has a small negative effect in irrigated regions. Expanding the scope, Bellora et al. (2018) use satellite data from South Africa to show that higher crop diversity enhances resilience and productivity across various crops. Focusing on northwestern France, Bareille and Dupraz (2020) show that wheat and milk yields increase with farm-level crop diversity. Our paper also complements empirical studies on the benefits of crop diversity on alternative agricultural outcomes than crop yields.[3] Overall, the literature indicates a positive and significant productive value of crop diversity. Spanning 52 studies, Bareille and Largier (2026) find that a 1% increase in crop biodiversity increases economic outcomes (including crop yields) by 0.75% on average.

A large ecological literature provides micro-foundations for the relationship between crop diversity, pest pressure, and agricultural productivity. A central mechanism is the *resource concentration hypothesis* (Root, 1973), whereby specialized systems increase the spatial concentration of host plants, facilitating pest colonization and reproduction. In contrast, more diverse cropping systems dilute host availability and reduce pest proliferation. A complementary mechanism is the *natural enemies hypothesis*, according to which greater plant diversity supports a richer community of predators and parasitoids, thereby enhancing biological pest control (Root, 1973; Andow, 1991; Altieri, 1999).

These mechanisms operate at broader spatial scales through landscape spillovers. Crop diversity and landscape complexity increase the abundance of natural enemies and reduce pest pressure in focal fields (Thies and Tscharntke, 1999; Tscharntke et al., 2005). Meta-analyses confirm that biodiversity and landscape heterogeneity are associated with stronger natural pest control (Letourneau et al., 2011; Chaplin-Kramer et al., 2011; Rusch et al., 2016; Karp et al., 2018). Our framework provides a tractable representation of these bi-

[3]For instance, Di Falco and Chavas (2009) and Groom and Pereira Fontes (2021) on cereal *production* in Ethiopia, Auffhammer and Carleton (2018) on *revenue* resilience to droughts in India, and Renard and Tilman (2019) on the stability of national *harvests* across 91 countries.

ological processes. By modeling cross-field biotic externalities, we capture how both the spatial concentration of crops and the diversity of neighboring crops shape pest pressure and, ultimately, productivity.

Our work distinguishes itself in several ways. First, we move beyond localized perspectives by assembling a global dataset covering 39,116 counties. To do so, we combine geospatial data from FAO-GAEZ, on the potential suitability and productivity of land, and Earthstat, on actual yields and harvested areas. Second, our analysis spans a larger crop coverage that allows measuring *between-crop* diversity more consistently. Third, we document how these productivity gains vary across agronomic crop groups and with pest pressure. Eventually, we propose an identification strategy that leverages exogenous variations in cropland patterns driven by local climate and soil conditions. Specifically, we use FAO-GAEZ suitability scores to construct *potential* counterparts to actual farmland shares and crop diversity, which serve as instrumental variables (IV). While related suitability-based instruments have been used in previous work (Fiszbein, 2022), our contribution is to apply this logic to estimate within-crop and between-crop effects on agricultural productivity in a global, crop-level setting.

The remainder of the paper is organized as follows. Section 2 develops the model of crop production with cross-field biotic externalities. Section 3 describes the global data and the identification strategy. Section 4 reports the structural estimates, and a range of robustness checks. Section 5 then uses the estimates to simulate the productivity and revenue gains from reducing specialization. Section 6 concludes.

# 2 Theoretical Framework

## 2.1 General Setup

We consider a world of $N$ countries, where each country $n$ is divided into $J_n$ spatial units. For convenience, these units are referred to as counties, though the model can be applied at other scales. Each county $j \in \{1, \ldots, J_n\}$ is subdivided into fields of equal size. Each field, or unit of land, is cultivated by a farmer. The farmer grows a single crop using one unit of labor and pesticides. Each crop $k \in \{1, \ldots, K\}$ is treated as a homogeneous good.[4]

[4]This framework abstracts from farm-management margins such as labor allocation, machinery constraints, or risk diversification that may arise in diversified farming systems. These margins would affect crop allocation incentives but not the biotic mechanisms through which specialization increases pest pressure and diversification reduces it.

### 2.1.1 Productivity

Productivity varies across crops and locations. The *actual* output of crop $k$ per unit of land and labor in area $j$ will always be lower than or equal to its *potential* output. Factors reducing production fall into two categories: abiotic (weather and soil conditions such as temperature, light and precipitation) and biotic (pests and interactions among plants, insects, and bacteria).

Farmers use pesticides to mitigate production losses due to biotic factors, but at a cost to human health and the environment. Governments limit this externality by taxing pesticide use with the environmental tax $\tau_j$. To maintain neutrality in government budget and farmers' income, we suppose that the tax policy is offset by a subsidy $T_j^k$, equal to the average tax payment in county $j$ for crop $k$.[5]

Farmers choose pesticide intensity $\rho$ to maximize profits:

$$\pi_j^k \equiv \max_{\rho} \mathbb{E}[\tilde{y}_j^k] p_j^k - \tau_j \rho + T_j^k - c_j \tag{1}$$

with $\tilde{y}_j^k$ the farmer's stochastic production level, $p_j^k$ the price of crop $k$, and $c_j$ the cost of labor and land in county $j$.

### 2.1.2 Crop Resilience

Adverse abiotic and biotic shocks occur independently across fields and result in the destruction of the entire production on affected fields. Biotic risks depend not only on a farmer's own decisions but also on the crop and pesticide choices of neighboring farmers due to *cross-field* biotic externalities.

The resilience of a field growing crop $k$ in county $j$ is thus a function of the farmer's pesticide intensity, $\rho$, as well as the cross-field biotic effects within the county. Let $\psi_j^k(\rho; \boldsymbol{S_j}, \boldsymbol{\rho_j})$ denote the resilience function, with $\boldsymbol{S_j} \equiv \{S_j^k : k = 1, \ldots, K\}$ the distribution of fields between crops, and $\boldsymbol{\rho_j} \equiv \{\bar{\rho}_j^k : k = 1, \ldots, K\}$ the vector of average pesticide use. $S_j^k$ is the share of fields dedicated to crop $k$ in county $j$, with $S_j^k \geq 0$ and $\sum_k S_j^k = 1$, while $\bar{\rho}_j^k$ denotes

[5] For simplicity, we abstract from the production and market dynamics of agrochemicals and assume that farmers are endowed with a large stock of agrochemicals that is not depleted by farming activities. In this setting, what matters for farmers' decisions is the effective cost of pesticide use, which is captured through the tax on pesticide application. The subsidy ensures that pesticide taxation does not mechanically affect farmers' income, allowing us to isolate its effect on pests and productivity. None of these assumptions are central to the biotic mechanisms that our model describes.

the average treatment applied to all fields growing crop $k$ in $j$. This resilience function corresponds to the probability that crop $k$ survives on a particular field and, at equilibrium, to the fraction of fields in county $j$ that survive among the fields that grow crop $k$.[6] Using this notation, (1) becomes

$$\pi_j^k = \max_{\rho} \bar{a}_j^k \psi_j^k(\rho; \boldsymbol{S_j}, \boldsymbol{\rho_j}) p_j^k - \tau_j \rho + T_j^k - c_j, \tag{2}$$

where $\bar{a}_j^k$ is the potential productivity of crop $k$ in county $j$. Hence, the expected productivity $z$ is given by[7]

$$z_j^k = \psi_j^k(\rho_j^{k\star}; \boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star}) \bar{a}_j^k. \tag{3}$$

Assuming that adverse weather shocks and pest attacks occur independently, the resilience function is the product of three stochastic components, i.e.[8]

$$\psi_j^k(\rho; \boldsymbol{S_j}, \boldsymbol{\rho_j}) \equiv A_j^k \mu_j^k(\rho) B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j}), \tag{4}$$

where $A_j^k < 1$ is the probability that a field survives abiotic shocks, $\mu_j^k(\rho)$ is the within-field biotic probability, and $B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j})$ captures cross-field biotic effects. We specify the within-field biotic probability as

$$\mu_j^k(\rho) \equiv \exp\left[-(\hat{\rho}_j^k - \rho)^2 / 2\lambda_j^k\right], \tag{5}$$

with $\hat{\rho}_j^k$ the unrestricted treatment level of pesticides and $\lambda_j^k > 0$ a measure of pest resistance, such that a larger $\lambda_j^k$ implies more effective pesticides. By choosing pesticide intensity $\rho = \hat{\rho}_j^k$, the farmer ensures that this probability equals one. However, this is not the farmer's optimal

[6] The resilience function $\psi_j^k(\rho; S_j, \rho_j)$ should be interpreted as a reduced-form representation of underlying pest population dynamics evolving over the growing season. Rather than modeling these dynamics explicitly, we assume that pest pressure reaches its steady-state level conditional on crop allocation and pesticide use. This is natural given that biological dynamics operate at a higher frequency than land-use decisions. Appendix B shows that such a steady-state approximation leads to a reduced-form specification consistent with our model.

[7] There is a difference between $S_j^k$, the share of land allocated to crop $k$, and $\psi_j^k$, the fraction of output that survives pest pressure. Pest dynamics depend on the availability of host plants (i.e., planted area), whereas realized output reflects ex-post biological losses.

[8] Abiotic risk distinguishes productivity differences driven by environmental conditions from those driven by biotic interactions, which helps link the model to the data. A richer framework could model pest dynamics as dependent on climatic conditions. However, implementing it empirically would require detailed, high-frequency data on pest incidence and its interaction with weather variables, which are not available at the global scale of our analysis. We therefore adopt a parsimonious specification in which abiotic and biotic risks are separable, while controlling empirically for agro-climatic conditions through attainable yields.

strategy when $\tau_j > 0$.

Denoting by $\mathcal{K}_j$ the subset of crops grown in the county, the cross-field biotic effects are defined as:

$$B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j}) \equiv \prod_{\ell \in \mathcal{K}_j} \mu_j^\ell(\bar{\rho}_j^\ell) \exp\left[-\hat{\kappa}_k(S_j^\ell)\right]. \tag{6}$$

The first term, $\mu_j^\ell(\bar{\rho}_j^\ell)$, measures the fraction of fields sown with crop $\ell$ that survive direct biotic factors. The second term, $\exp\left[-\hat{\kappa}_k(S_j^\ell)\right]$, captures the biotic externalities from crop $\ell$ to crop $k$, conditional on crop $\ell$ surviving. The sign and magnitude of these externalities depend on $\ell$ and on its share in county $j$ through function $\hat{\kappa}_k(\cdot)$, specified below.

## 2.2 Market Equilibrium

We assume perfect competition with atomistic farmers. As shown in Appendix A, equilibrium pesticide use for crop $k$ is the dominant strategy, $\rho_j^{k\star} = \hat{\rho}_j^k - \lambda_j^k \tau_j / c_j$, which is independent of the equilibrium crop allocation $\boldsymbol{S}_j^\star$ and pesticide applications $\boldsymbol{\rho}_j^\star$.[9] Moreover, since at the land and labor market equilibrium farmers' expected profit is zero for any crop choice, expected gross revenue of land is equalized across crops grown in the county: $p_j^k z_j^k = c_j$ for all $k \in \mathcal{K}_j \equiv \{k = 1, \ldots, K | S_j^{k\star} > 0\}$. As a result, farmers play mixed strategies over crop choices, yielding the equilibrium land allocation $\boldsymbol{S}_j^\star$.[10] We also obtain the following result:

**Proposition 1**. *In equilibrium, actual productivity for crop $k \in \mathcal{K}_j$ is given by*

$$z_j^{k\star} = \frac{a_j^k \exp\left[-\sum_{\ell \in \mathcal{K}_j} \hat{\kappa}_k(S_j^{l\star})\right]}{t_j^k \prod_{\ell \in \mathcal{K}_j} t_j^\ell}, \tag{7}$$

*with $a_j^k \equiv \bar{a}_j^k A_j^k$ the attainable yield of crop $k$ in county $j$ and $t_j^k \equiv \exp\left[\lambda_j^k (\tau_j / c_j)^2 / 2\right]$.*

**Proof**. See Appendix A.1.

Equation (7), together with the zero-profit condition $p_j^k z_j^{k\star} = c_j$, characterizes the equilibrium land allocation. Equilibrium requires $p_j^k z_j^{k\star} \leq c_j$ for all $k \in \{1, \ldots, K\}$, with equality holding for crops $k \in \mathcal{K}_j$, crops that are actually grown, i.e. for all $k$ such that $S_j^{k\star} > 0$. This condition arises from the endogenous land allocation at equilibrium and the presence of cross-field biotic externalities captured by (6). If a given crop $k$ were to generate strictly

[9] Superscript $\star$ denotes equilibrium values.

[10] Hence, $S_j^{k\star}$ refers both to the probability that a farmer in $j$ chooses crop $k$ and to the equilibrium share of that crop in $j$.

higher expected revenue than other crops, farmers would reallocate their fields toward $k$, increasing its equilibrium share $S_j^{k\star}$. Through the within-crop (specialization) externality, this expansion raises pest pressure and reduces productivity $z_j^{k\star}$ until the zero-profit condition is restored. Conversely, crops with lower expected revenues lose fields, which relaxes pest pressure. As a result, equilibrium land allocation reflects a balance between agro-climatic comparative advantage, captured by $a_j^k$, and biotic externalities, captured by the dependence of $z_j^{k\star}$ on $S_j^{k\star}$. Crops that do not satisfy $p_j^k z_j^{k\star} \geq c_j$ even at arbitrarily small shares are not cultivated in equilibrium. This explains both observed specialization patterns and the absence of certain crops in some locations.

Actual productivity, as defined in (7), depends on the attainable productivity, accounting for abiotic risk, $a_j^k$, cross-field effects, $\exp\left[-\sum_{\ell\in\mathcal{K}_j}\hat{\kappa}_k(S_j^{\ell\star})\right]$, and pest policy effects $t_j^k$ and $\prod_{\ell\in\mathcal{K}_j} t_j^\ell$. The first pest policy effect, $t_j^k = 1/\mu_j^k(\bar{\rho}_j^{k\star})$, corresponds to the inverse of the survival probability of crops to direct biotic factors, while the second policy effect arises from cross-field externalities. Thus, an increase in the environmental tax $\tau_j$ reduces pesticide use in *all* fields, thereby diminishing the positive cross-field externalities of pesticides on $k$'s survival. This is reflected by an increase in $t_j^k$ and all $t_j^\ell$ in the denominator of (7).

### 2.3 Cross-Field Biotic Effects

The following assumption specifies the structure of the cross-field biotic externalities.

**Assumption 1** (Decomposition of cross-field biotic externalities)**.** *For each crop $k \in \mathcal{K}_j$*

$$\hat{\kappa}_k(S_j^k) = \kappa_{0k} S_j^k \quad \textit{(within-crop effects)},$$

*and, if $K_j \equiv |\mathcal{K}_j| > 1$, for all $\ell \in \mathcal{K}_j, \ell \neq k$*

$$\hat{\kappa}_k(S_j^\ell) = -\kappa_k \left[\frac{1}{K_j - 1}\ln\left(\frac{1}{K_j - 1}\right) - \frac{S_j^\ell}{1 - S_j^k}\ln\left(\frac{S_j^\ell}{1 - S_j^k}\right)\right] \quad \textit{(between-crop effects)}.$$

Under Assumption 1, when more than one crop is grown in county $j$, we have

$$B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j}) = \prod_{\ell\in\mathcal{K}_j} \mu_j^\ell(\bar{\rho}_j^\ell) \exp\left[-\kappa_{0k} S_j^k - \kappa_k\left(\bar{H}_j - H_j^k\right)\right],$$

with

$$H_j^k \equiv -\sum_{\ell \in \mathcal{K}_j \setminus k} \frac{S_j^\ell}{1 - S_j^k} \ln\left(\frac{S_j^\ell}{1 - S_j^k}\right) \tag{8}$$

the *leave-one-out* Shannon index, which measures the diversity of all crops $\ell \neq k$ grown in the county, and $\bar{H}_j \equiv \ln(K_j - 1)$ its maximum. The difference $\bar{H}_j - H_j^k$ measures the gap between *maximum* diversity (when all other existing crops are equally represented) and *actual* diversity. This specification aligns with standard ecological diversity measures and ensures that $B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j}) \leq 1$.

The intuition behind Assumption 1 is that pest pressure depends on both the abundance of the host crop, that is, the crop on which the pest can feed and reproduce, and the composition of surrounding crops. When the share of crop $k$ increases, host crops become more spatially concentrated, which raises the carrying capacity of pests specialized in $k$ and increases pest pressure on that crop. By contrast, greater crop diversity in the surrounding fields lowers pest pressure through two complementary channels. First, it dilutes the availability of suitable hosts. Second, it can enhance biological control by supporting a richer community of natural enemies. In that sense, the within-crop term captures a host concentration effect, while the between-crop term captures dilution and natural-enemy effects. This asymmetry between within-crop and between-crop effects is motivated by host specificity in pest dynamics: the share of crop $k$ directly affects the population of pests specialized in $k$, whereas the shares of other crops matter indirectly through dilution and biological-control mechanisms.

Using this specification, when more than one crop is grown in the county, (7) becomes[11]

$$z_j^{k\star} = \frac{a_j^k \exp\left[-\kappa_{0k} S_j^{k\star} - \kappa_k(\bar{H}_j - H_j^{k\star})\right]}{t_j^k \prod_{\ell \in \mathcal{K}_j} t_j^\ell} \tag{9}$$

Two distinct biotic externalities affect crop $k$'s productivity. The first externality is a *specialization* effect, captured by the within-crop coefficient $\kappa_{0k}$, which is linked to an increase in the crop's land share. The second externality is a diversity effect, captured by the between-crop coefficient $\kappa_k$, which is linked to a decrease in the diversity of non-$k$ crops relative to the maximum level. Both effects are expected to reduce productivity: higher specialization increases pest pressure, while lower diversity weakens area $j$'s ecosystem, reducing natural

[11] For monoculture, actual productivity is given by $z_j^{k\star} = a_j^k e^{-\kappa_{0k}}/(t_j^k)^2$.

pest control.[12] Appendix B shows that this reduced-form specification can be derived from a simple pest population model. In this model, the carrying capacity increases with the share of the host crop and decreases with crop diversity. Additionally, pesticide use raises pest mortality.

### 2.4 Rent-Maximizing Equilibrium

So far, we analyzed allocation choices made by atomistic farmers who do not coordinate. We now consider a benevolent social planner who aims to determine the land allocation $\boldsymbol{S_j} \equiv \{S_j^k\}_{k=1}^K$ that maximizes total agricultural revenues in county $j$, taking into account cross-field externalities. The planner solves

$$\max_{\boldsymbol{S_j}} \left\{ \sum_{k=1}^{K} p_j^k z_j^k S_j^k : \sum_{k=1}^{K} S_j^k = 1 \right\}.$$

Our second proposition follows from this maximization problem:[13]

**Proposition 2**. *To maximize total agricultural revenues, the share of crop $k$ in area $j$ must satisfy*

$$S_j^k = \frac{1}{\sum_{\ell \in \mathcal{K}} \frac{\kappa_{0k}}{\kappa_{0\ell}} \frac{p_j^k z_j^k}{p_j^\ell z_j^\ell}} \left[ 1 + \sum_{\ell \in \mathcal{K}} \frac{1}{\kappa_{0\ell}} \left( \frac{p_j^k z_j^k}{p_j^\ell z_j^\ell} - 1 \right) + \sum_{\ell \in \mathcal{K}} \sum_{m \in \mathcal{K}} \frac{p_j^\ell z_j^\ell}{p_j^m z_j^m} \frac{\kappa_\ell S_j^\ell}{\kappa_{0m}} \left( \frac{\partial H_j^\ell}{\partial S_j^k} - \frac{\partial H_j^\ell}{\partial S_j^m} \right) \right], \tag{10}$$

where $\mathcal{K} \equiv \{k = 1, \ldots, K | S_j^k > 0\}$.

**Proof**. See Appendix A.3.

[12] The corresponding equilibrium farmland shares are derived in Appendix A.2.

[13] Appendix A.3 derives Proposition 2 and shows how biodiversity-induced changes in effective yields translate into changes in crop revenues and aggregate productivity.

# 3 From Model to Data: Empirical Implementation

This section outlines our empirical strategy for estimating the model predictions regarding within-crop and between-crop effects. We first describe the data sources and variable construction, including measures of actual and attainable productivity. We then link the model theoretical predictions to the data and discuss the identification strategy.

## 3.1 Data

We use two types of data: (i) *potential* measures of land suitability and attainable productivity, and (ii) *actual* measures of yields and harvested areas. All measures are crop-specific and at a 5-arc-minute resolution (grid cells equivalent to 10km×10km at the equator). We map cells to countries using the Global Administrative Areas Database (GADM v4.0), and then aggregate 5-arc-minute grid-cell measures to the *county* level $j$ (so-called 'admin2'). This aggregation reflects a trade-off between ecological relevance and measurement error.

Regarding ecological relevance, the literature shows that pest and natural-enemy dynamics depend on surrounding landscape composition at multiple spatial scales rather than only within the focal plot or grid cell (O'Rourke, Rienzo-Stack, and Power, 2011). In our setting, the county provides a tractable proxy for that surrounding landscape.

Regarding measurement error, grid cell-level EarthStat and GAEZ variables are partly constructed by allocating or predicting agricultural outcomes from coarser administrative statistics and agronomic information. Measurement error is likely to be large in cells with little cropland. Aggregation to the county level therefore reduces measurement error in yields, crop shares, and diversity measures, while preserving substantial within-country variation. If spillovers operate over a smaller radius than the county, this aggregation should attenuate estimated effects toward zero, so our estimates can be interpreted as conservative. Overall, our sample covers 39,116 counties, 148 countries and 40 crops. Appendix Tables C1 and C2 list the crops and countries. We present the main data sources below and use alternative datasets as robustness checks in Section 4.3.

### 3.1.1 Potential Measures

We source *potential* measures from version 4.0 of the Global Agro-Ecological Zoning (GAEZ) project (Fischer et al., 2021). GAEZ has the advantage of estimating yields from an agronomic model, which assesses the suitability of each cell for each crop based on edaphic and

climatic resources, matched to crop-specific requirements. As a result, we obtain *exogenous* productivity measures for each cell and crop, independent of the actual land use (e.g., the cell being cultivated with another crop).

**Suitability.** The version 4 of the FAO-GAEZ database reports a continuous suitability index at the 5-arc-minute grid-cell level. The layer we use is the average suitability index of the total grid cell. Each 5-arc-minute cell is decomposed into underlying 30 arc-second land units, which differ in soil type and slope class combinations. GAEZ evaluates each land unit separately and assigns a suitability rating and attainable yield after applying crop-specific yield-reduction factors linked to thermal conditions, water stress, agro-climatic constraints, and soil and terrain limitations. The grid-cell suitability index is then obtained by aggregating these land-unit results over the full cell, as an area-weighted average over its soil and slope composition. The index can therefore be interpreted as a rescaled measure of attainable yield relative to the crop's maximum attainable yield under the corresponding crop, input, and water-supply configuration; for example, a value of 8,000 corresponds roughly to 80% of that benchmark. Cells corresponding to urban areas, protected areas and areas of high biodiversity are identified and assigned a suitability index of 0 for all crops.

**Productivity.** Based on the suitability predictions, attainable productivity (t/ha) is assessed for various combinations of input levels (*high* or *low*)[14] and sources of water supply (*rainfed* or *irrigated*). Among the available combinations, we select the attainable production density variable for a high level of inputs[15] and rainfed water source.[16] We use this measure as a globally comparable and exogenous benchmark for agro-climatic productivity, not as a proxy for realized management conditions. We do not combine rainfed and irrigated potentials using observed irrigation shares, because such shares would depend on realized agricultural choices and would introduce endogeneity into a control variable. This measure averages attainable production across the entire cell, regardless of the actual percentage of cultivated area. Following Costinot, Donaldson, and Smith (2016), when FAO-GAEZ distinguishes between species (e.g., *dryland rice* vs. *wetland rice*), but only aggregate information

[14] A high level represents advanced land management and market-oriented agriculture, and a low level represents traditional land management and subsistence agriculture.

[15] As attainable yields are by definition greater than or equal to actual yields in our model, we prefer to select the highest values available for attainable yields.

[16] Data for irrigated attainable yields is limited, excluding major crops like rice and maize.

is available in the actual measures (e.g., *rice*), we use the maximum value between the two species.

### 3.1.2 Actual Measures

We use year-2000 measures of *actual* production (metric tons, t), harvested area (hectares, ha), and yields (t/ha) from the EarthStat database (Monfreda, Ramankutty, and Foley, 2008). These variables combine agricultural census and survey statistics, averaged over 1997-2003, with a satellite-derived cropland map, referred to as 'cropland extent'. Sub-national census crop areas are mapped to the 5-arc-minute grid by distributing each crop *within the observed cropland extent*, in proportion to the cell's satellite-derived cropland fraction (Ramankutty et al., 2008). The area of a crop in a cell equals its cropland fraction times the crop's share of total cropland in the surrounding administrative unit.

Importantly for our identification strategy, this allocation uses observed cropland extent alone and makes *no* use of agro-climatic suitability, so the realized crop diversity and farmland shares are not constructed from the FAO-GAEZ suitability information from which our instruments are built (Section 3.2). To match weight units between *actual* and *potential* measures, we apply crop-specific wet-to-dry conversion factors provided by FAO-GAEZ. Because EarthStat and FAO-GAEZ combine observed information with modeled or downscaled components, and because their underlying inputs are not perfectly contemporaneous, some measurement error is unavoidable. In particular, the FAO-GAEZ potential measures from which we build our instruments can contain non-classical measurement error (Araujo and Possebom, 2025) and need not predict realized outcomes well at fine spatial scales (Morando, 2024). We view this primarily as a source of noise that may reduce precision or attenuate estimated relationships, rather than as a mechanism likely to generate spurious positive effects.

The agricultural sector shows a high degree of specialization (see Appendix Tables C1 and C2). Ten crops account for about 76% of the global harvested area, and 50 countries represent over 91% of total harvested area. In some cases, actual yields exceed attainable yields. This discrepancy may arise from reporting errors in EarthStat or inaccuracies in FAO-GAEZ's agronomic predictions.[17] We address these issues in the estimation below. As an independent robustness check (see Section 4.3) we use observed national agricultural statistics that rely

[17] One limitation of the attainable productivity measure is that it does not account for genetically modified organisms and irrigation.

on no agro-climatic model: the U.S. Census of Agriculture (USDA National Agricultural Statistics Service, 2017), Brazil's Municipal Agricultural Production survey (IBGE, 2017), and India's district-level agricultural statistics (ICRISAT, 2017).

#### 3.1.3 Controls

The estimation conditions on county-level agro-climatic and geographic controls. The agro-climatic and edaphic variables include growing-period length, mean temperature, precipitation, potential evapotranspiration, terrain altitude, soil root conditions, and soil workability. These variables come from FAO-GAEZ layers (Fischer et al., 2021), aggregated to the county level. County area is derived from the GADM administrative boundaries. We additionally control for each county's average distance to the nearest surface water body, as a proxy for surface-water access. This measure is built from the HydroSHEDS global river network (Lehner, Verdin, and Jarvis, 2008) and the HydroLAKES lake database (Messager et al., 2016). Regional heterogeneity analyses use the World Bank country-region classification. For the robustness analysis (see Section 4.3), we assemble additional agro-ecological controls: county-level fertilizer (N + P + K) application from the EarthStat nutrient layers (West et al., 2014), area equipped for irrigation from the FAO Global Map of Irrigation Areas (Siebert et al., 2013), and human-induced soil-degradation severity from ISRIC GLASOD (Oldeman, Hakkeling, and Sombroek, 1991).

### 3.2 Estimation and Identification Strategy

#### 3.2.1 Estimating Equation

Building on the model, we estimate the log form of (9):

$$\ln z_j^k = \ln a_j^k - \kappa_{0k} S_j^k - \kappa_k(\bar{H} - H_j^k) - \ln t_j^k - \sum_{\ell \in \mathcal{K}_j} \ln t_j^\ell,$$

where county $j$ is the spatial unit of observation. Because information on pesticide regulations and pest resistance is imperfect, we proxy $t_j^k$ with country×crop fixed effects, which absorb country–crop determinants of pesticide use. This rich set of fixed effects captures factors such as nationally set pesticide taxes, crop-specific pest resistance, and any country–crop co-movement between climate and pest pressure. Eventually, we estimate the following equation:

$$\ln z_j^k = \alpha_k \ln a_j^k + \beta_{0k} S_j^k + \beta_k H_j^k + \gamma_i^k + \epsilon_j^k. \tag{11}$$

For each crop $k$, we regress log actual productivity, $\ln z_j^k$, on: (i) log attainable productivity, $\ln a_j^k$, (ii) the crop $k$'s *farmland share*, $S_j^k$, which represents the share of county $j$'s cultivated area devoted to crop $k$ (*within-crop* effects), (iii) the county crop diversity index, $H_j^k$ (*between-crop effects*),[18] and (iv) country $i$ fixed effects, $\gamma_i^k$. $\epsilon$ is the usual error term. As attainable yields $a_j^k$ are estimated by GAEZ under the assumption of advanced land management everywhere, we expect $\alpha_k < 1$. Consistent with the model, we also expect $\beta_{0k} < 0$ and $\beta_k > 0$ for all $k$, reflecting negative within-crop effects and positive between-crop effects.

### 3.2.2 Measuring Crop Diversity

**Mean Proportional Abundance.** As defined in the theoretical section, we measure crop diversity using the (leave-one-out) Shannon index $H_j^k$ (see Eq. 8). The standard index is widely used in the ecological literature as a measure of the mean proportional abundance of crops in a location.[19] It captures both crop richness and relative abundance (Shannon, 1948). Higher values indicate greater farmland heterogeneity, with a maximum value equal to $\ln K$ when all crops are grown in equal proportions.

**Effective Diversity.** To provide a more intuitive measure of diversity of crops than the mean proportional abundance, we use Hill's numbers, known as the effective number of crops:

$$D_j^q \equiv \left[ \sum_{\ell \in \mathcal{K}_j} (S_j^\ell)^q \right]^{1/(1-q)} \tag{12}$$

where $q \geq 0$ determines the sensitivity of the diversity measure to the relative abundance of crops. The index $D_j^q$ represents diversity by considering both the total number of crops and their relative abundance. In other words, $D_j^q$ represents the number of equally abundant crops needed to match the observed mean proportional abundance in county $j$, where crops may not be equally distributed. As $q \to 1$, Hill's numbers corresponds to the exponential of the Shannon index. Consequently, $\beta_k$ in our estimated equation (11) can be interpreted

[18] As $\bar{H}$ is constant across all observations, we omit it from the regression.
[19] The standard index is defined as $-\sum_{\ell \in \mathcal{K}_j} S_j^\ell \ln(S_j^\ell)$.

as the elasticity of the actual yield of crop $k$ with respect to effective diversity in the other crops, derived from the leave-one-out Shannon index $H_j^k$.

We compute the effective number of crops (Eq. 12) for all counties in our sample. Figure 1 shows that counties around the world are highly specialized. On average, the actual effective number of crops is only about 4.5 (Panel A), far below the potential effective number of about 22 (Panel B), reflecting the dominance of a few crops in observed allocations.

### 3.2.3 Accounting for Endogeneity

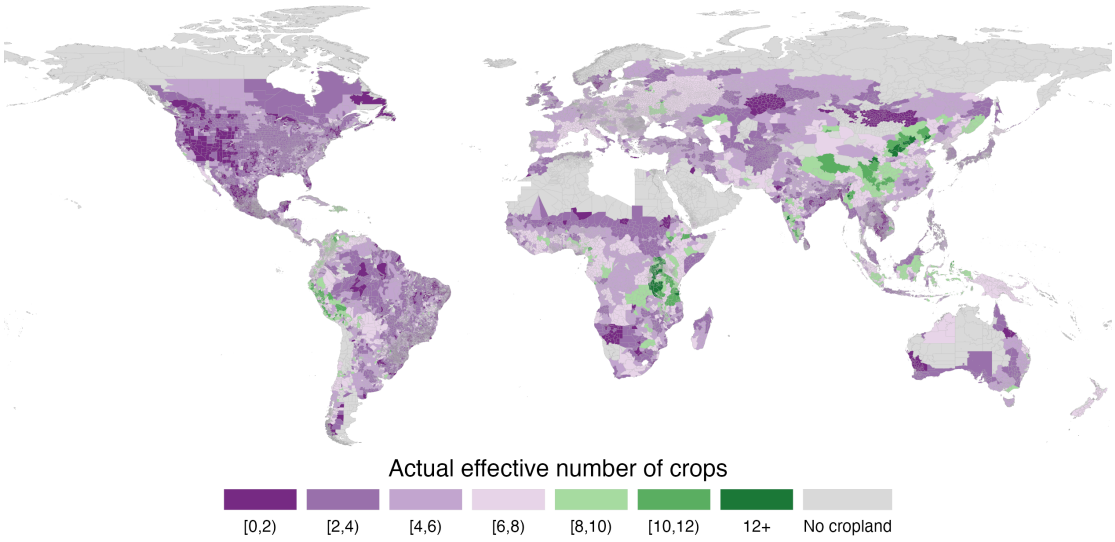
In practice, crop diversity and farmland shares may be endogenous to actual yields. First, since actual yields in our data are averaged over a six-year period, productivity in a given year may influence crop diversity in subsequent years. For example, improved wheat yields could encourage specialization in wheat, increasing its farmland share and reducing overall diversity. In this case, OLS estimates are likely downward-biased in absolute terms, for both within-crop specialization and between-crop diversity effects. Second, omitted factors unrelated to soil and climate conditions may influence both agricultural productivity and diversity. For instance, counties with better access to pesticide markets might exhibit higher yields and greater specialization, leading to upward-biased OLS estimates.

Our identification strategy leverages FAO-GAEZ suitability data to construct instrumental variables for farmland shares and crop diversity. Attainable yields absorb the direct effect of agro-climatic productivity on output, while the instruments exploit the *relative* suitability of land across crops within a county.

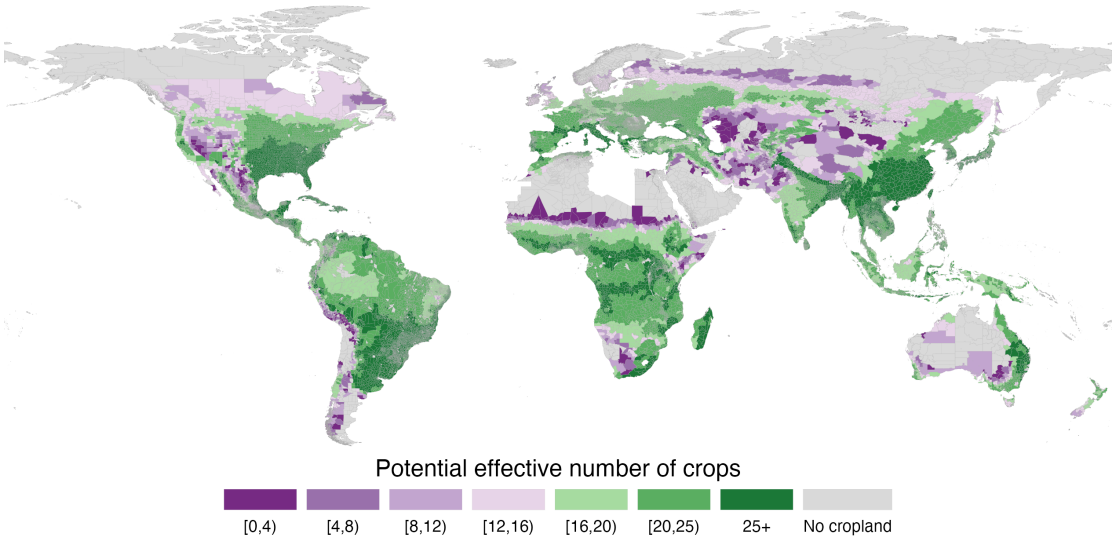
**Instruments for Farmland Shares.** We consider two instruments for $S_j^k$, the share of county $j$'s cultivated area devoted to crop $k$. The first instrument is the average share of land assessed by FAO-GAEZ as very suitable (when farmers can reach 80 to 100% of the maximum attainable yield for crop $k$) or suitable (60 to 80%). This variable is provided by FAO-GAEZ.

The second instrument is built from our estimation of *potential* farmland share. In each county $j$, we compute the potential farmland share of crop $k$, $\tilde{S}_j^k$, as the ratio of the suitability of crop $k$, $\text{SI}_j^k$ to the sum of the suitability of all crops:[20]

[20]Such index of relative suitability is also used in Berman et al. (2023) to construct price indexes.




**(a)** Actual




**(b)** Potential

**Figure 1:** Effective crop diversity (Hill's numbers)

*Notes*: Actual effective number of crops (1a) and potential effective number of crops (1b) at the county level. Computed based on FAO-GAEZ, Earthstat.

$$\tilde{S}_j^k = \frac{\mathrm{SI}_j^k}{\sum_{\ell \in \mathcal{K}_j} \mathrm{SI}_j^\ell}.$$

If land allocation prioritizes crops whose growth requirements align most effectively with local climate and soil characteristics, the two instruments should satisfy the relevance condition. Reassuringly, regressing actual shares on potential shares across all crops and counties yields a positive, significant coefficient of 0.770. We also compare our measure of potential against actual farmland shares for six major crops and observe positive correlations.

**Instrument for Crop Diversity.** We construct an index of potential crop diversity using our estimated potential shares $\tilde{S}_j^k$:

$$\tilde{H}_j = -\sum_{\ell \in \mathcal{K}_j} \tilde{S}_j^\ell \ln(\tilde{S}_j^\ell).$$

Exponentiating gives the corresponding potential Hill's numbers ($\tilde{D}_j^1$), displayed in Figure 1b. We also compute the leave-one-out version of the potential Shannon index for each crop $k$, which serves as an instrument for its actual counterpart:

$$\tilde{H}_j^k = -\sum_{\ell \in \mathcal{K}_j \setminus k} \frac{\tilde{S}_j^\ell}{1 - \tilde{S}_j^k} \ln\left(\frac{\tilde{S}_j^\ell}{1 - \tilde{S}_j^k}\right).$$

The relevance of the instrument is straightforward: counties whose natural conditions are suitable to a very limited number of crops are more likely to specialize in these crops, resulting in lower diversity indices.

**Identifying Assumptions.** We require three assumptions to identify within- and between-crop effects. Conditional on attainable yields and country fixed effects: (1) there must be sufficient within-country variation in farmland shares and diversity; (2) natural conditions must explain variations in farmland shares and diversity (relevance assumption); and (3) the remaining variation in potential farmland shares and diversity must not affect realized yields other than through actual crop patterns (exclusion restriction). This assumption is plausible in our setting because the instruments are based on modeled suitability rather than observed crop choices. The instruments are constructed from relative suitability patterns across crops

### 3.3 Identification and Inference

Equation (11) includes two endogenous regressors: the leave-one-out Shannon index (log effective diversity) and the own-crop farmland share. We instrument them with their FAO-GAEZ potential counterparts. Because errors may be heteroskedastic and spatially correlated, we assess instrument strength using heteroskedasticity-robust diagnostics. We report the Kleibergen-Paap rank Wald (F)-statistic (Kleibergen and Paap, 2006), the robust analogue of the Cragg-Donald statistic with multiple endogenous regressors. We also report Sanderson-Windmeijer (SW) conditional $F$-statistics (Sanderson and Windmeijer, 2016) for each endogenous regressor separately.

Following Andrews, Stock, and Sun (2019), Keane and Neal (2024), and Henningsen et al. (2025), we do not select the specification, the instruments, or the clustering scheme on the basis of the first-stage $F$-statistic. Conventional first-stage $F$ thresholds can be unreliable. Although refinements such as the effective $F$ (Montiel Olea and Pflueger, 2013) and the $tF$ adjustment (Lee et al., 2022) have been proposed, screening on $F$ can still make conventional inference reject too often in the presence of weak instruments. We report two weak-IV-robust inference procedures. First, we compute the Anderson-Rubin (AR) test (Anderson and Rubin, 1949), which remains valid under weak identification. Second, because our specification is overidentified, we also report the conditional likelihood ratio (CLR) test (Kleibergen, 2005). Following the practical recommendations in Keane and Neal (2024), we compute the CLR test using the continuously updated GMM (CUE) estimator, which is designed to provide weak-IV-robust inference under heteroskedasticity.[21] To probe the exclusion restriction more directly, we also report plausibly exogenous bounds following Conley, Hansen, and Rossi (2012). This robustness test allows for violations of the exclusion restriction. Specifically, each instrument is allowed to have a direct effect on yields of up to half of its reduced-form association with yields. We then report the resulting range of 2SLS confidence intervals. Finally, our main inference clusters standard errors on a 2.5° spatial grid (∼275 km), following Bester, Conley, and Hansen (2011), to allow for spatial correlation.

[21] In practice, we implement this procedure in Stata by using the `cue` option in `ivreg2.ado`, together with robust standard errors. To accommodate multiple high-dimensional fixed effects, we built the command `weakiv_hdfe.ado`, which extends `weakiv.ado` to our setting.

**Table 1:** First-stage Estimates

| | IV without controls | | IV with controls | |
|---|---|---|---|---|
| | Diversity | Farmland share | Diversity | Farmland share |
| Potential effective diversity | 0.144*** | -0.011*** | 0.077*** | -0.011*** |
| | (0.030) | (0.002) | (0.027) | (0.002) |
| Potential crop share | 0.292** | 0.153*** | -0.015 | 0.135*** |
| | (0.127) | (0.022) | (0.126) | (0.021) |
| Suitability share | 0.063*** | 0.004*** | 0.077*** | 0.005*** |
| | (0.013) | (0.002) | (0.013) | (0.002) |
| Kleibergen-Paap rk Wald F | 26.11 | 26.11 | 26.60 | 26.60 |
| Sanderson-Windmeijer conditional F | 37.59 | 38.60 | 27.35 | 44.94 |
| Area, agro-climatic + water controls | No | No | Yes | Yes |
| Country × crop FE | Yes | Yes | Yes | Yes |
| R-squared | 0.553 | 0.672 | 0.560 | 0.672 |
| Observations | 586,355 | 586,355 | 586,355 | 586,355 |

*Notes*: First-stage estimates of each endogenous regressor, effective diversity and farmland share, on the three excluded instruments (potential effective diversity, potential farmland share, and suitability share), *without* and *with* controls: log county area, agro-climatic variables, and distance to surface water, corresponding to columns (3) and (4) of Table 2. All columns include country×crop fixed effects and are estimated on the same baseline sample. The table also reports the Kleibergen-Paap rk Wald $F$-statistic for joint instrument relevance and the Sanderson-Windmeijer conditional $F$-statistic for the separate identification of each regressor. Standard errors are clustered on a 2.5° spatial grid in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# 4 Empirical Results

## 4.1 Baseline Estimates

We first estimate Eq. 11 under the assumption that cross-field biotic externalities are common across crops, so that $\beta_{0k} = \beta_0$ and $\beta_k = \beta$ for all $k$. Table 2 reports the estimated yield elasticity ($\alpha$), the farmland-share effect ($\beta_0$)), and the diversity effect, measured by the leave-one-out Shannon index ($\beta$). Columns (1)-(2) report OLS estimates, whereas columns (3)-(5) report IV estimates. Table 1 reports the corresponding first-stage estimates.

The IV specifications move from the baseline specification to specifications adding agro-climatic and water controls, with column (5) reporting the continuously updated GMM (CUE) estimator. All specifications include country×crop fixed effects, and standard errors are clustered on a (2.5°) spatial grid.

We begin with the first stage. Both endogenous regressors are strongly and significantly predicted by their instruments, with the expected signs. In column 1, the estimate of the potential effective diversity, which captures the diversity implied by the county's agro-climatic

**Table 2:** Baseline Estimates

| Dependent Variable | Log Actual Yield | | | | |
|---|---|---|---|---|---|
| | OLS | | IV | | |
| | (1) | (2) | (3) | (4) | (5) |
| Log attainable yield | 0.099*** | 0.080*** | 0.112*** | 0.116*** | 0.116*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Log effective diversity | 0.105*** | 0.137*** | 1.300*** | 1.655*** | 1.734*** |
| | (0.02) | (0.01) | (0.26) | (0.35) | (0.35) |
| Farmland share | 0.376*** | 0.347*** | -9.529*** | -14.787*** | -13.837*** |
| | (0.04) | (0.04) | (1.54) | (1.89) | (1.69) |
| Estimator | OLS | OLS | 2SLS | 2SLS | CUE |
| Area, agro-climatic + water controls | No | Yes | No | Yes | Yes |
| Country × crop FE | Yes | Yes | Yes | Yes | Yes |
| Clustering | 2.5° grid | 2.5° grid | 2.5° grid | 2.5° grid | 2.5° grid |
| Anderson-Rubin p-value | | | 0.000 | 0.000 | 0.000 |
| Hansen J p-value | | | 0.011 | 0.198 | 0.198 |
| Observations | 586,355 | 586,355 | 586,355 | 586,355 | 586,355 |

*Notes*: Estimations of Eq. (11): (1) OLS, baseline; (2) OLS, with controls; (3) 2SLS, baseline; (4) 2SLS, with controls; (5) Continuously-updated GMM (CUE), with controls. "Controls" denotes log county area, agro-climatic variables, and distance to surface water. All specifications include country×crop fixed effects. First-stage estimates are reported in Table 1, and the full set of coefficients, including controls, in Online Appendix Table O2. Column (5) is the continuously-updated GMM (CUE) estimator of Keane and Neal (2024); because `ivreg2` cannot absorb the high-dimensional fixed effects, the country×crop fixed effects are partialled out by the Frisch-Waugh-Lovell method while the controls are kept inside the GMM, with the control coefficients omitted from the table. Standard errors are clustered on a 2.5° spatial grid in parentheses with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

suitability for *other* crops, is positive. The estimate shows that farmers grow a more diverse crop mix in counties whose environment favors a broader set of crops. In column 2, both estimates of the potential farmland share and the suitability share are positive. Counties that are more suitable for crop $k$ devote a larger share of farmland to that crop.[22]

The instruments are jointly relevant. The Kleibergen-Paap rk Wald $F$-statistics exceed 20, and the Sanderson-Windmeijer conditional $F$-statistics show that each endogenous regressor is separately identified. Appendix Figures D1 and D2 depict these first-stage relationships. After partialling out country×crop fixed effects, each endogenous regressor displays a clear upward relationship with its excluded instrument. The corresponding reduced-

[22] In IV terms, what matters is that the instruments generate independent first-stage variation for the two endogenous regressors, so that the first-stage matrix has full rank; the instruments need not each load only on their own endogenous regressor.

form relationships, log yield against each instrument, are shown in Appendix Figures D3 and D4, and the multivariate reduced-form estimates in Appendix Table D1. Each instrument can affect yields through both endogenous regressors, so its reduced-form coefficient takes the sign of whichever channel dominates. The potential effective diversity and the suitability share mainly predict effective diversity in the first stage, and because diversity raises yields, their reduced-form coefficients are positive. The potential crop share mainly predicts the own-crop farmland share, and because specialization lowers yields, its reduced-form coefficient is negative. The reduced-form signs are therefore those the model predicts. The diversity and farmland-share coefficients, estimated crop by crop under each of the three candidate instrument sets, are reported in the Online Appendix (Table O1).

Turning to the second-stage estimates, actual yields respond only weakly to attainable yields (the elasticity $\alpha$ is around 0.1, well below one). This echoes Costinot and Donaldson (2012), who likewise find that realized output tracks FAO-GAEZ potential only weakly.[23] Agronomic potential nonetheless retains significant predictive power for realized yields.

The OLS coefficient on farmland share is positive in columns 1 and 2, whereas the IV estimates are negative, as expected, in columns 3 to 5. The OLS coefficient on effective diversity has the expected positive sign but is much smaller than its IV counterpart. The instruments thus correct substantial bias in the OLS estimates. This pattern is what one expects if crop allocation is endogenous to unobserved local conditions: counties with unobserved advantages for their dominant crop, or with more intensive, input-heavy management, tend to be both higher-yielding and more specialized, and hence less diverse. This omitted-variable correlation biases the OLS farmland-share coefficient upward and the OLS effective-diversity coefficient toward zero. Instrumenting with exogenous agro-climatic potential removes this correlation. The IV diversity coefficient is therefore larger, while the farmland-share coefficient becomes negative and larger in magnitude.[24]

Effective diversity captures *between*-crop effects. Since the regressor is the log effective number of crops, the coefficient is an elasticity. A 1% increase in the effective number of crops, defined as the number of equally abundant crops that would reproduce the county's observed diversity (Eq. (12)), is associated with a 1.3-1.8% increase in actual yields (columns 3 to 5).

[23] Their analysis differs from ours. They estimate a simple output-on-predicted-output relationship, whereas we estimate a yield equation with fixed effects and instruments.

[24] By comparison, the OLS elasticity (without county fixed effects) is 0.1, closer to Di Falco and Chavas (2006) on wheat genetic diversity: a 1% increase in genetic diversity raised mean wheat yields by 0.11% in southern Italy.

This result is consistent with Auffhammer and Carleton (2018).[25]

The own-crop farmland share captures specialization, or *within*-crop externalities. Its magnitude must be interpreted with care because farmland shares are strongly right-skewed. Most crops occupy only a small fraction of county cropland, while a few dominant crops occupy much larger shares. The coefficient is therefore a local effect: a one-percentage-point increase in a crop's own farmland share lowers that crop's yield by about 9 to 13% across the IV specifications. This effect is economically large and consistent with biotic pressure being a first-order determinant of yields (Oerke, 2006).

Adding agro-climatic and water controls in column 4 leaves both IV coefficients unchanged: highly significant, positive for effective diversity and negative for farmland share. The magnitudes are, if anything, somewhat larger after adding these controls, which suggests that the instruments are not operating primarily through the agro-climatic controls included in the regression. Column 4 is our preferred specification.

Finally, the continuously updated GMM estimates in column 5 are close to the 2SLS estimates in column 4: 1.78 versus 1.66 for effective diversity, and $-13.1$ versus $-14.2$ for farmland share. This suggests that weak-instrument bias is limited in the pooled specification. The Anderson-Rubin test reported in Table 2, which remains valid under weak instruments, rejects the joint null that effective diversity and farmland share have no effect on yield ($p < 0.001$). The Conley, Hansen, and Rossi (2012) plausibly exogenous bounds lead to the same conclusion. Even when each instrument is allowed to have a moderate direct effect on yield, the 95% union-of-confidence-intervals set for the effective-diversity effect is $[0.40, 3.48]$, bounded well away from zero (Appendix Table D2).

## 4.2 Heterogeneity

**Heterogeneity across crops.** We now examine how cross-field externalities vary across crops. We first group crops into four agronomic categories: cereals; grain legumes; other annual crops (oilseeds, fibre, roots, tubers, sugar beet, vegetables); and perennial (non-rotated multi-year crops: coffee, cocoa, tea, rubber, oil palm, coconut, olive, banana, sugarcane). We then estimate Eq. (11) separately for each group with country×crop fixed effects.

[25] By comparison, the OLS elasticity, estimated without county fixed effects, is 0.1, closer to Di Falco and Chavas (2006) on wheat genetic diversity: a 1% increase in genetic diversity raised mean wheat yields by 0.11% in southern Italy.

**Table 3:** Heterogeneity by Crop Type

| Dependent Variable | Log Actual Yield | | | |
|---|---|---|---|---|
| | Cereals | Legumes | Other annuals | Perennial |
| Log effective diversity | 2.67*** | 2.70*** | 0.15 | 3.64* |
| | (0.87) | (1.00) | (0.27) | (1.86) |
| Farmland share | -23.99*** | -4.69 | -13.02*** | -4.72 |
| | (5.69) | (4.17) | (1.46) | (4.09) |
| Kleibergen-Paap rk Wald F | 6.7 | 2.3 | 12.5 | 1.4 |
| SW conditional F (diversity) | 25.4 | 3.4 | 17.4 | 2.0 |
| SW conditional F (share) | 10.1 | 5.7 | 33.2 | 2.9 |
| Anderson-Rubin p-value | 0.000 | 0.000 | 0.000 | 0.000 |
| Hansen J p-value | 0.163 | 0.406 | 0.759 | 0.225 |
| Observations | 166,972 | 84,289 | 259,142 | 75,952 |

*Notes*: 2SLS estimates of Eq. (11) within each agronomic crop group: *Cereals* (wheat, maize, rice, barley, oat, rye, sorghum, millet, buckwheat); *Legumes* (soybean, groundnut, chickpea, cowpea, dry pea, pigeon pea, common bean, alfalfa); *Other annuals* (oilseeds, fibre, roots, tubers, sugar beet, vegetables); *Perennial* (non-rotated multi-year crops: coffee, cocoa, tea, rubber, oil palm, coconut, olive, banana, sugarcane). All specifications include country×crop fixed effects, controls (log county area, agro-climatic variables, and distance to surface water), and the same instruments (potential effective diversity; potential and suitability shares). Standard errors are clustered on a 2.5° spatial grid in parentheses with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The corresponding first-stage estimates for each crop group are reported in Online Appendix Table O4.

Table 3 reports the results. In all specifications, the Anderson-Rubin and Hansen $J$ $p$-values support the IV strategy. However, the second-stage estimates differ across crop types. Cereals show the clearest pattern: both effects are large and significant, positive for the between-crop effect (log effective diversity) and negative for the within-crop specialization effect (farmland share). For legumes, the diversity effect remains significant and positive, whereas the specialization effect is imprecisely estimated. For the broad "other annuals" group, we observe the opposite: the specialization effect is significant while the diversity effect is close to zero. Finally, for perennial crops, the estimates are mostly not statistically significant.

Overall, if the effects are not uniform across groups of crops, the heterogeneity supports our biodiversity mechanism that crop diversity affects productivity. The effects are indeed strongest for annual crops, especially cereals, where specialization and pest pressure are likely to be most relevant. Moreover, the weak estimates for perennial crops are consistent with the biodiversity mechanism, since rotation-based spillovers should be less relevant for non-rotated multi-year crops.

We next split crops by pest pressure rather than by botanical type. If diversity raises

yields by diluting pest pressure, the diversity effect should be stronger for crops exposed to many pest species and weaker for crops exposed to fewer pests. We measure pest pressure by the number of pest species recorded for each crop in the EPPO Global Database (EPPO, 2026) (see Appendix Table C3). We then split crops at the median of this measure and re-estimate equation (11) within each subsample.

Table 4 reports the results. The pattern is the one the biodiversity mechanism predicts. The between-crop diversity effect is concentrated in high-pest crops. A 1% increase in effective diversity raises yields by 2.7% for high-pest crops, compared with 0.8% for low-pest crops. The within-crop specialization effect, by contrast, is larger for low-pest crops. This suggests that the negative relationship between a crop's own farmland share and its yield may also operate through channels not captured by the number of pest species recorded in EPPO, such as generalist pests, soil-borne pathogens, or declining land quality as a crop's area expands. It may also reflect more intensive pest management for crops facing the highest pest pressure.

A fully interacted IV specification estimated on the combined high- and low-pest sample confirms that the diversity effect differs across the two groups. Equality of the high- and low-pest diversity coefficients is rejected ($p = 0.025$), while the difference in the farmland share coefficients is weaker ($p = 0.081$; see last two rows of Table 4). The high-pest first stage is weaker (Kleibergen-Paap $F = 9.5$), but the Anderson-Rubin test rejects the null of no effect ($p < 0.001$), so the biodiversity mechanism is robust to weak instruments in high-pest crops.

**Heterogeneity across regions.** As a final dimension of heterogeneity, we re-estimate our preferred specification separately within each World Bank region. Because splitting the sample reduces the within-region identifying variation, the region-specific first stages are weaker. We therefore rely on Anderson-Rubin weak-instrument-robust $p$-value, which rejects the joint null of no effect in all seven regions. The results are aligned with our between-crop diversity effect: the productivity gain from effective diversity is largest in the lower-input, less chemically-intensive regions, especially Sub-Saharan Africa and Latin America. By contrast, the effect is small, absent, or negative in more input-intensive systems, where chemical control substitutes for the biotic regulation that diversity provides. The within-crop specialization effect is more pervasive, with the farmland-share coefficient negative and significant in most regions (Appendix Table E1).

**Table 4:** Heterogeneity by Crop Pest Pressure

| Dependent Variable | Log Actual Yield | |
|---|---|---|
| | High pest | Low pest |
| Log effective diversity | 2.67*** | 0.76*** |
| | (0.85) | (0.24) |
| Farmland share | -12.25*** | -18.19*** |
| | (2.12) | (2.88) |
| Kleibergen-Paap rk Wald F | 9.5 | 14.8 |
| SW conditional F (diversity) | 9.3 | 36.4 |
| SW conditional F (share) | 25.5 | 23.4 |
| Anderson-Rubin p-value | 0.000 | 0.000 |
| Hansen J p-value | 0.585 | 0.976 |
| Observations | 339,958 | 246,397 |
| High = low, diversity ($p$) | 0.025 | |
| High = low, share ($p$) | 0.081 | |

*Notes*: 2SLS estimates of equation (11) for the high- and low-pest subsamples. Crops are split at the median number of pest species recorded per crop in the EPPO Global Database (EPPO, 2026); "High pest" refers to crops with an above-median count. The instrument set, controls, and clustering follow Table 3. The last two rows report the $p$-value of a test that the diversity (respectively, farmland-share) coefficient is equal across the two groups, from a fully-interacted IV that pools both halves under common clustered standard errors. Standard errors are clustered on a 2.5° spatial grid in parentheses with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The corresponding first-stage estimates for both groups are reported in Online Appendix Table O5.

### 4.3 Robustness

We assess the robustness of the baseline estimates of Table 2 using alternative data, clustering schemes, and FAO-GAEZ suitability thresholds, as well as additional controls and a permutation test. Table 5 reports the main robustness results.

**Alternative data.** We first re-estimate the model on independently-measured land use from a set of national agricultural statistics: the U.S. Census of Agriculture (USDA National Agricultural Statistics Service, 2017), Brazil's Municipal Agricultural Production survey (IBGE, 2017), and India's district-level agricultural statistics (ICRISAT, 2017). These sources report crop areas and yields by census or survey.[26] On this independent, multi-

[26] We also verified that the results are robust when using the SPAM (You et al., 2014) and CROPGRIDS (Tang et al., 2024) gridded crop data. We do not report these estimates, but they are available upon request. The reason is that both sources allocate crops to grid cells using FAO-GAEZ agro-climatic suitability (Fischer et al., 2021), the measure from which we build our instruments. A SPAM- or CROPGRIDS-based estimate would therefore have a partly mechanical first stage and, more importantly, a weakened exclusion restriction, because the instrument would be related to the constructed outcome through the allocation algorithm.

country sample the effective-diversity and farmland-share coefficients keep their signs and remain jointly significant. The first stage is weaker (Kleibergen-Paap $F \approx 5$), but the Anderson-Rubin test rejects the null of no effect.

**Identifying assumptions and inference.** First, we propose a permutation test to assess whether the reduced form reflects local agro-ecological variation rather than a spurious or mechanical correlation. We run a randomization-inference placebo that assigns to each county the potential diversity of a randomly drawn county from the same country. This permutation preserves the country-specific distribution of the instrument and the fixed-effect structure, but breaks the link between a county's own potential diversity and its yield. If the reduced form were driven by broad country-level patterns or by the construction of the instrument itself, the reassigned instrument should predict yield with similar strength. It does not. Across 500 permutations, the placebo coefficients are centered on zero, with a 5th to 95th percentile range from $-0.006$ to 0.005. By contrast, the original reduced-form coefficient, estimated using the actual county-level potential diversity measure, is 0.336, far outside the placebo distribution. None of the 500 permutations reaches this value, implying a permutation $p$-value below 0.002 (Appendix Table D3 and Appendix Figure D5).

Second, we examine possible violations of the exclusion restriction. In our setting, this restriction requires that our instruments affect yields only through effective diversity and own-crop concentration, and not through other determinants of productivity. Although this assumption cannot be tested directly, column 3 adds fertilizer use, irrigation, and land degradation as direct controls, addressing the concern that the estimates may be driven by differences in input intensity or land quality. Third, we verify that the estimates are not driven by the baseline construction of potential shares. We thus rebuild the instrument using a 40% rather than a 60% FAO-GAEZ suitability threshold (column 4). Fourth, we check that the results are not an artifact of the high-input GAEZ scenario by re-estimating the baseline specification with the *low-input* GAEZ suitability and potential-yield layers, together with alternative suitability thresholds (column 5). Across these three robustness checks, the effective-diversity and farmland-share estimates are virtually unchanged relative to the baseline specification in column 1 of Table 5.

Finally, we assess the robustness of inference. While the F-statistic is sensitive to the level at which standard errors are clustered, the Anderson-Rubin test is not. Re-clustering the base estimates at progressively coarser units, namely county (admin-2), 0.75°, 1°, and

**Table 5:** Robustness of the Baseline Estimates

| Dependent Variable | Log Actual Yield | | | | |
|---|---|---|---|---|---|
| | Baseline | Census | Additional controls | 40% threshold | Low-input GAEZ |
| Log effective diversity | 1.655*** | 2.929*** | 1.645*** | 2.162*** | 2.793*** |
| | (0.353) | (1.050) | (0.314) | (0.463) | (0.532) |
| Farmland share | -14.787*** | -1.432 | -14.079*** | -13.466*** | -9.211*** |
| | (1.891) | (0.984) | (1.883) | (2.020) | (1.701) |
| Kleibergen-Paap rk Wald F | 26.6 | 5.1 | 26.5 | 29.7 | 15.2 |
| SW conditional F (diversity) | 27.3 | 8.0 | 32.1 | 17.8 | 18.0 |
| SW conditional F (share) | 44.9 | 11.6 | 43.7 | 48.8 | 26.9 |
| Anderson-Rubin p-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Hansen J p-value | 0.198 | 0.038 | 0.277 | 0.590 | 0.042 |
| Observations | 586,355 | 43,802 | 586,082 | 586,355 | 584,939 |

*Notes*: Column 1 reproduces our preferred 2SLS estimate (Table 2, column 4). Each remaining column re-estimates it under the indicated variation. "Census" pools observed national agricultural statistics (the United States, Brazil and India); "Additional Controls" adds county-level fertilizer (N + P + K) application, irrigation extent and soil-degradation severity as controls; "40% Threshold" rebuilds the potential shares from a 40% GAEZ suitability cutoff; "Low-input GAEZ" rebuilds the suitability-based instruments (potential effective diversity, potential and suitability shares) and the attainable yield from GAEZ's low-input (rather than high-input) assumption. Standard errors are clustered on a 2.5° spatial grid in parentheses with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The corresponding first-stage estimates for each column are reported in Online Appendix Table O3.

2° spatial grids, and region (admin-1), leaves the result statistically significant throughout (see Appendix Table F1).

## 5 Simulations

Using our baseline 2SLS estimates from column 4 of Table 2, we simulate two alternative land allocations. First, we examine the effect of reducing marginally the share of the dominant crop in each county. We simulate a 1% reduction in the dominant crop's area, redistributed equally across all other crops already grown, and assess its effect on production. Second, we solve Eq. (10) numerically for the rent-maximizing farmland distribution.

**Table 6:** Mean Effects of Redistributing 1% of the Dominant Crop's Area

| Crop | Counties where crop dominates | | | | All counties | | | |
|---|---|---|---|---|---|---|---|---|
| | $\hat{P}$ | $\hat{S}$ | $\hat{z}_S$ | $\hat{z}_H$ | $\hat{P}$ | $\hat{S}$ | $\hat{z}_S$ | $\hat{z}_H$ |
| Alfalfa | 1.38 | 0.99 | 1.08 | 1.28 | 3.58 | 3.18 | 1.00 | 1.08 |
| Banana | 1.18 | 0.99 | 1.06 | 1.13 | 8.73 | 8.12 | 1.00 | 1.07 |
| Barley | 1.09 | 0.99 | 1.08 | 1.02 | 4.35 | 4.25 | 1.00 | 1.02 |
| Buckwheat | 1.04 | 0.99 | 1.05 | 1.00 | 4.08 | 4.02 | 0.99 | 1.02 |
| Cabbage | 1.07 | 0.99 | 1.08 | 1.01 | 4.96 | 4.86 | 0.99 | 1.02 |
| Carrot | 1.08 | 0.99 | 1.07 | 1.01 | 96.77 | 95.29 | 0.99 | 1.02 |
| Cassava | 1.10 | 0.99 | 1.08 | 1.02 | 22.53 | 22.19 | 1.00 | 1.02 |
| Chickpea | 1.07 | 0.99 | 1.07 | 1.01 | 168.49 | 165.86 | 0.99 | 1.02 |
| Cocoa | 1.17 | 0.99 | 1.06 | 1.12 | 5.92 | 5.53 | 1.00 | 1.06 |
| Coconut | 1.13 | 0.99 | 1.06 | 1.08 | 5.54 | 5.04 | 1.00 | 1.07 |
| Coffee | 1.17 | 0.99 | 1.06 | 1.12 | 5.94 | 5.36 | 1.00 | 1.07 |
| Cotton | 1.17 | 0.99 | 1.06 | 1.12 | 34.71 | 31.72 | 1.00 | 1.07 |
| Cow pea | 1.13 | 0.99 | 1.07 | 1.07 | 4.12 | 3.83 | 1.00 | 1.06 |
| Dry pea | 1.04 | 0.99 | 1.04 | 1.00 | 6.05 | 5.94 | 0.99 | 1.02 |
| Flax | . | . | . | . | 126.29 | 124.93 | 1.00 | 1.01 |
| Groundnut | 1.13 | 0.99 | 1.06 | 1.08 | 51.57 | 46.76 | 1.00 | 1.07 |
| Maize | 1.27 | 0.99 | 1.07 | 1.20 | 2.16 | 1.95 | 1.02 | 1.11 |
| Millet | 1.14 | 0.99 | 1.10 | 1.05 | 9.79 | 9.62 | 1.00 | 1.02 |
| Oat | 1.18 | 0.99 | 1.11 | 1.07 | 2.30 | 2.27 | 1.00 | 1.02 |
| Oil palm | 1.09 | 0.99 | 1.08 | 1.02 | 7.47 | 7.34 | 1.00 | 1.02 |
| Olive | 1.10 | 0.99 | 1.09 | 1.03 | 1241.36 | 1221.55 | 1.00 | 1.02 |
| Onion | 1.08 | 0.99 | 1.07 | 1.02 | 37.75 | 37.29 | 0.99 | 1.02 |
| Phaseolus bean | 1.14 | 0.99 | 1.06 | 1.09 | 409.81 | 383.97 | 1.00 | 1.08 |
| Pigeon pea | . | . | . | . | 3.82 | 3.76 | 1.00 | 1.02 |
| Rapeseed | 1.08 | 0.99 | 1.07 | 1.02 | 8.30 | 8.16 | 0.99 | 1.02 |
| Rice | 1.23 | 0.99 | 1.08 | 1.15 | 185.26 | 173.73 | 1.02 | 1.09 |
| Rubber | 1.09 | 0.99 | 1.08 | 1.01 | 8.91 | 8.81 | 1.00 | 1.02 |
| Rye | 1.09 | 0.99 | 1.08 | 1.02 | 10.03 | 9.85 | 1.00 | 1.02 |
| Sorghum | 1.16 | 0.99 | 1.06 | 1.11 | 4.40 | 4.09 | 1.00 | 1.06 |
| Soybean | 1.20 | 0.99 | 1.06 | 1.14 | 12.49 | 11.33 | 1.00 | 1.07 |
| Sugar beet | 1.05 | 0.99 | 1.06 | 1.01 | 38.25 | 37.80 | 1.00 | 1.02 |
| Sugar cane | 1.26 | 0.99 | 1.07 | 1.19 | 751.42 | 712.81 | 1.00 | 1.08 |
| Sunflower | 1.08 | 0.99 | 1.04 | 1.05 | 17.87 | 16.50 | 1.00 | 1.07 |
| Sweet potato | 1.04 | 0.99 | 1.03 | 1.02 | 176.34 | 162.66 | 1.00 | 1.07 |
| Tea | 1.09 | 0.99 | 1.08 | 1.02 | 21.04 | 20.71 | 1.00 | 1.02 |
| Tobacco | 1.14 | 0.99 | 1.06 | 1.09 | 96.81 | 91.19 | 1.00 | 1.06 |
| Tomato | 1.18 | 0.99 | 1.06 | 1.12 | 48.97 | 46.18 | 1.00 | 1.06 |
| Wheat | 1.16 | 0.99 | 1.06 | 1.10 | 2.26 | 2.05 | 1.01 | 1.08 |
| White potato | 1.09 | 0.99 | 1.05 | 1.05 | 26.01 | 24.99 | 1.00 | 1.06 |
| Yam | 1.04 | 0.99 | 1.05 | 1.00 | 2.33 | 2.29 | 0.99 | 1.02 |
| Total | 1.13 | 0.99 | 1.07 | 1.07 | 91.97 | 88.45 | 1.00 | 1.05 |

*Notes*: $\hat{x} = x_{new}/x$ is the ratio of post- to pre-redistribution values of production ($\hat{P}$), farmland share ($\hat{S}$), specialization-driven productivity ($\hat{z}_S$) and diversity-driven productivity ($\hat{z}_H$), with $\hat{P} = \hat{z}_S\,\hat{z}_H\,\hat{S}$. The first four columns average over counties where the crop dominates, the last four over all counties. Values are unavailable for flax and pigeon pea in the "Counties where crop dominates" columns, as no county is dominated by these crops.

### 5.1 Marginal Farmland Redistribution

**Decomposition of the production effect.** For each crop $k$, let $S^k_{new}$ denote the new farmland share and $H^k_{new}$ the new leave-one-out Shannon index after redistribution.[27] Holding input prices and pest resistance constant, the model predicts that the new expected yield $z^k_{new}$ is given by

$$z^k_{new} = z^k \exp\left[-\kappa_{0k}(S^k_{new} - S^k) + \kappa_k(H^k_{new} - H^k)\right],$$

for all $k \in \mathcal{K}$, where $\mathcal{K}$ is the set of crops with strictly positive production under market equilibrium. We thus only allow for changes at the intensive margin. The total variation in expected productivity can therefore be broken down into the product of variations due to changes in farmland share, $\hat{z}^k_S = \exp\left[-\kappa_{0k}(S^k_{new} - S^k)\right]$, and changes in crop diversity, $\hat{z}^k_H = \exp\left[\kappa_k(H^k_{new} - H^k)\right]$. Similarly, the variation in production breaks down as

$$\hat{P}^k = \hat{z}^k_S \times \hat{z}^k_H \times \hat{S}^k$$

where $\hat{P}^k = P^k_{new}/P^k$ and $\hat{S}^k = S^k_{new}/S^k$. Theoretically, the effect of redistribution on $\hat{P}^k$ is ambiguous and depends on the parameters $\kappa_{0k}$ and $\kappa_k$. We calibrate these using our pest-pressure-group IV estimates of Eq. (11), assigning each crop its group's coefficients. By construction, $\beta_k = \kappa_k$ and $\beta_{0k} = -\kappa_{0k}$. When a group estimate is not statistically significant, we set the corresponding parameter to zero. The resulting simulations should therefore be interpreted as illustrative counterfactuals rather than point forecasts, with greater weight placed on the pest-pressure groups—the high- and low-pest crop classes, split at the median number of recorded pest species—whose underlying coefficients are estimated more precisely.

**Predictions.** Table 6 reports mean production changes and their decomposition. Restricting to counties where one crop dominates, i.e., where it has a larger farmland share than all other crops, average production increases by 13% after redistribution. This implies that the drop in production associated with the reduction in harvested area (-1%) is more than offset by an increase in productivity (+14%). The decomposition indicates that this productivity gain splits roughly equally between increased diversity (+7%) and reduced specialization (+7%). Among the major staple crops, maize benefits the most, with production rising by

[27] For clarity, county subscripts are omitted; the unit of observation is unchanged.

27% in counties where it dominates, mostly through the diversity channel. Across all counties, the mean production change is dominated by area reallocation toward initially minor crops, whose harvested area multiplies under the redistribution, and is therefore uninformative on its own.

The main economic result is that, even in the average county, the between-crop diversity channel raises productivity by about 5%, while the within-crop specialization channel is roughly neutral. The redistribution shifts own-crop share, and hence specialization, away from the dominant crop and toward the other crops already grown in the county. By contrast, it raises the leave-one-out diversity faced by all crops. The aggregate productivity gain comes from diversification rather than from a reallocation of specialization across crops. Overall, these simulations indicate substantial scope for productivity gains from reallocating farmland toward more diversified patterns, especially in counties where production is concentrated in a dominant crop.

### 5.2 Rent-Maximizing Farmland Allocation

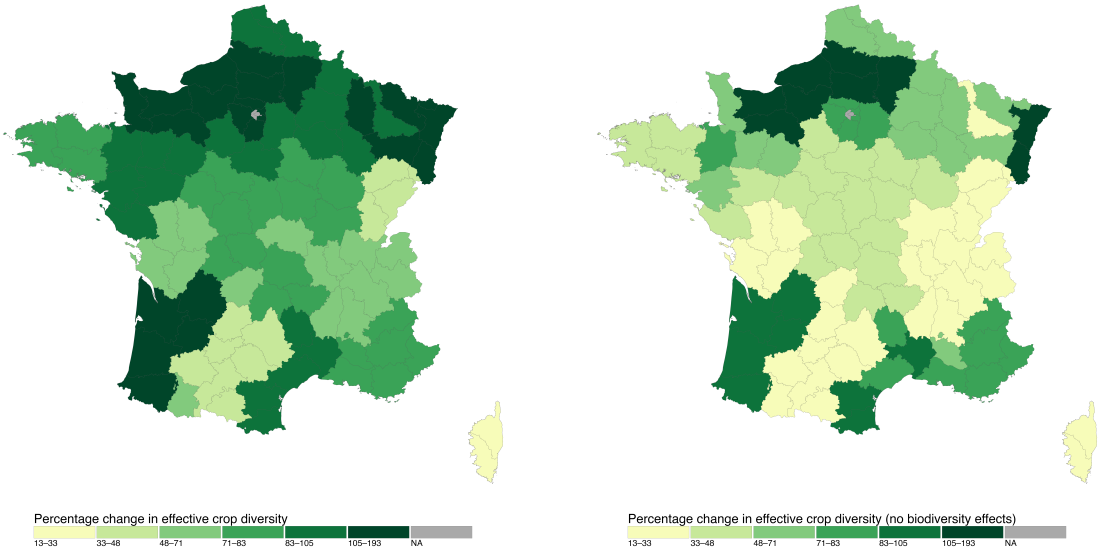
As a second simulation exercise, we consider a benevolent social planner who maximizes agricultural revenue, as described in Subsection 2.4. For each county and crop, we solve numerically for the rent-maximizing farmland shares from Eq. (10), parameterized with our **pest-pressure-group** IV estimates of Eq. (11).

We run this counterfactual for the 94 French counties in our sample. Each county grows at least four of the eight major cereal and oil crops considered: barley, maize, oat, rapeseed, sorghum, soybean, sunflower, and wheat. Each of these crops accounts for at least 1% of the county's cereal area. We reallocate land only among these eight crops and hold all other crop areas constant.

The rent-maximization treats output prices as fixed (Section 2).[28] After solving for the rent-maximizing farmland shares in each county,[29] we compute the resulting effective diversity and compare the planner's allocation with the observed one, in which the biotic externalities are not internalized.

[28] Cereals and oilseeds are homogeneous, globally-traded commodities for which a single region is a price-taker, so reallocating its land leaves relative prices unchanged; a large shift in, for example, local vegetable acreage would move prices and violate the fixed-price assumption.

[29] Because the specialization externality reflects a crop's prevalence in the whole landscape, the penalty is applied to each cereal's share of *total* county cropland, the share on which $\kappa_{0k}$ is estimated, rather than its share within the cereal group.




**(a)** Considering all cross-field externalities **(b)** Considering within-crop externalities only

**Figure 2:** Crop Diversity under Rent-Maximizing Farmland Allocation

*Notes*: The figures display the simulated percentage change in effective crop diversity (the Hill number, i.e. the exponential of the leave-one-out Shannon index) after the rent-maximizing distribution is implemented for all 94 French counties in our sample. We reallocate farmland within a group of eight cereal and oil crops (barley, maize, oat, rapeseed, sorghum, soybean, sunflower and wheat), keeping other crops' area constant. Every county grows at least four of these eight crops on at least 1% of its cereal area; grey areas are units with no cropland data. In Fig. 2b, we solve for rent-maximizing shares considering that $\kappa_k = 0$ for all crops $k$, i.e., there are no between-crop externalities.

Figure 2a maps the simulated percentage change in effective crop diversity at the county level. On average, the effective number of crops grown rises from about 6 to about 11, a mean increase of roughly 85% across counties. This result indicates that the planner selects substantially more diversified allocations than those observed in the data.[30] The increase varies across space, mostly reflecting initial differences in diversity. Because the planner reoptimizes only the cereal margin and holds all other crop fixed, these gains constitute a conservative, partial-equilibrium benchmark. A planner able to reallocate the entire landscape would diversify further.

To assess the role of between-crop externalities in the planner's allocation, we set $\kappa_k = 0$ for all crops in the subsample and solve for the rent-maximizing shares. Results appear

[30] We report effective diversity (Hill numbers, the exponential of the leave-one-out Shannon index) for consistency with the rest of the paper. In terms of the underlying Shannon index, the average increase is 35% in the benchmark and 24% in the no-biodiversity case below.

in Figure 2b. The effective number of crops rises only to about 9, against about 11 in the benchmark. Roughly two-thirds of the planner's diversification gain thus derives from internalizing the within-crop specialization externality, and the remaining third from the between-crop diversity externality. Both channels are economically substantial, and every county diversifies under both, consistent with the planner internalizing externalities that the decentralized equilibrium leaves unaddressed.

# 6 Conclusion

Absent any external control, conventional ecological wisdom holds that more specialized fields are subject to a higher risk of crop failure. Concentrating production in the most suitable crops, whether at the farm level to maximize returns or globally to exploit abiotic comparative advantage, should reduce agricultural productivity. Yet, the economic literature provides limited insight into the underlying mechanisms.

In this paper, we develop a probabilistic model of agricultural production that captures two cross-field biotic externalities: specialization (within-crop effects, associated with changes in a crop's land share) and diversity (between-crop effects, associated with changes in the mix of other crops). We derive farmers' equilibrium choices of crops and pesticide intensity and obtain a testable equation for crop productivity that we take to the data. We assemble a large county-level dataset combining potential and actual land use. We also propose a new instrumental-variables strategy that exploits variation in land suitability for particular crops and for supporting greater crop diversity. We find that both *within*- and *between*-crop externalities significantly affect productivity. We also find heterogeneity across crop types in our sample. A key finding is that reducing a crop's own farmland share and raising county-level crop diversity would raise yields for most crops, with the largest gains concentrated among crops facing high pest pressure, consistent with the pest-control mechanism at the core of the model. We illustrate these benefits with two simulation exercises. These exercises quantify counterfactual changes in productivity rather than changes in net private profitability. They do not incorporate management costs, learning requirements, labor needs, machinery constraints, or other adjustment costs associated with diversification. Such margins may partly explain why farmers specialize in practice despite positive yield effects from greater crop diversity.

## References


Altieri, Miguel A. 1999. "The Ecological Role of Biodiversity in Agroecosystems." *Agriculture, Ecosystems & Environment* 74 (1-3):19–31.

Anderson, T. W. and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics* 20 (1):46–63.

Andow, David A. 1991. "Vegetational Diversity and Arthropod Population Response." *Annual Review of Entomology* 36 (1):561–586.

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11:727–753.

Araujo, Rafael and Vítor Possebom. 2025. "Potato Potahto in The FAO-GAEZ Productivity Measures? Nonclassical Measurement Error with Multiple Proxies." *arXiv preprint arXiv:2502.12141* .

Auffhammer, M. and T. A. Carleton. 2018. "Regional Crop Diversity and Weather Shocks in India." *Asian Development Review* 35 (2):113–130.

Bareille, F. and P. Dupraz. 2020. "Productive Capacity Of Biodiversity: Crop Diversity And Permanent Grasslands In Northwestern France." *Environmental and Resource Economics* 77 (2):365–399.

Bareille, F. and B. Largier. 2026. "A Meta-Analysis on the Productive Value of Crop Biodiversity." *American Journal of Agricultural Economics* Forthcoming.

Bellora, C., E. Blanc, J.-M. Bourgeon, and E. Strobl. 2018. "Estimating the Impact of Crop Diversity on Agricultural Productivity in South Africa." In *Agricultural Productivity and Producer Behavior*, edited by W. Schlenker. University of Chicago Press, 185–215.

Bellora, C. and J.-M. Bourgeon. 2019. "Food Trade and Biodiversity Effects." *International Economic Review* 60 (4):1957–1999.

Berman, N., M. Couttenier, A. Leblois, and R. Soubeyran. 2023. "Crop Prices and Deforestation in the Tropics." *Journal of Environmental Economics and Management* 119:102819.

Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen. 2011. "Inference with Dependent Data Using Cluster Covariance Estimators." *Journal of Econometrics* 165 (2):137–151.

Bianchi, F. J. J. A., C. J. H. Booij, and T. Tscharntke. 2006. "Sustainable Pest Regulation in Agricultural Landscapes: A Review on Landscape Composition, Biodiversity and Natural Pest Control." *Proceedings of the Royal Society B: Biological Sciences* 273 (1595):1715–1727.

Brock, W. A. and A. Xepapadeas. 2003. "Valuing Biodiversity from an Economic Perspective: A Unified Economic, Ecological, and Genetic Approach." *American Economic Review* 93 (5):1597–1614.

Carew, R., E. Smith, and C. Grant. 2009. "Factors Influencing Wheat Yield and Variability: Evidence from Manitoba, Canada." *Journal of Agricultural and Applied Economics* 41 (3):1–15.

Chaplin-Kramer, Rebecca, Megan E. O'Rourke, Eleanor J. Blitzer, and Claire Kremen. 2011. "A Meta-Analysis of Crop Pest and Natural Enemy Response to Landscape Complexity." *Ecology Letters* 14 (9):922–932.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1):260–272.

Costinot, A. and D. Donaldson. 2012. "Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence." *American Economic Review* 102 (3):453–458.

Costinot, A., D. Donaldson, and C. Smith. 2016. "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields Around the World." *Journal of Political Economy* 124 (1):205–248.

Di Falco, S. and J.-P. Chavas. 2006. "Crop Genetic Diversity, Farm Productivity and the Management of Environmental Risk in Rainfed Agriculture." *European Review of Agricultural Economics* 33 (3):289–314.

———. 2009. "On Crop Biodiversity, Risk Exposure, and Food Security in the Highlands of Ethiopia." *American Journal of Agricultural Economics* 91 (3):599–611.

EPPO. 2026. "EPPO Global Database." European and Mediterranean Plant Protection Organization. URL https://gd.eppo.int/. Accessed 2026-06-04.

Fischer, G., F. O. Nachtergaele, H. T. van Velthuizen, F. Chiozza, G. Franceschini, M. Henry, D. Muchoney, and S. Tramberend. 2021. *Global Agro-Ecological Zones V4 – Model Documentation*. Rome: Rome, FAO.

Fiszbein, Martin. 2022. "Agricultural Diversity, Structural Change, and Long-Run Development: Evidence from the United States." *American Economic Journal: Macroeconomics* 14 (2):1–43.

Frank, E. 2024. "The Economic Impacts of Ecosystem Disruptions: Costs from Substituting Biological Pest Control." *Science* 385 (6713):eadg0344.

Frank, E. and A. Sudarshan. 2024. "The Social Costs of Keystone Species Collapse: Evidence from the Decline of Vultures in India." *American Economic Review* 114 (10):3007–3040.

Gouel, C. and H. Guimbard. 2017. "Nutrition Transition and the Structure of Global Food Demand." Working Papers 2017-05, CEPII research center.

Groom, B. and F. Pereira Fontes. 2021. "Revisiting the Link Between Cereal Diversity and Production In Ethiopia." *Q Open* 1 (2).

Hassell, Michael P. 1978. *The Dynamics of Arthropod Predator-Prey Systems*. Princeton, NJ: Princeton University Press.

Henningsen, Arne, Guy Low, David Wuepper, Tobias Dalhaus, Hugo Storm, Dagim Belay, and Stefan Hirsch. 2025. "Estimating Causal Effects with Observational Data: Guidelines for Agricultural and Applied Economists." *Journal of Agricultural Economics* .

IBGE. 2017. "Municipal Agricultural Production (PAM)." Instituto Brasileiro de Geografia e Estatística. URL https://sidra.ibge.gov.br/.

ICRISAT. 2017. "District Level Database For India." International Crops Research Institute for the Semi-Arid Tropics. URL http://data.icrisat.org/dld/.

Jiguet, F., V. Devictor, R. Julliard, and D. Couvet. 2012. "French Citizens Monitoring Ordinary Birds Provide Tools for Conservation and Ecological Sciences." *Acta Oecologica* 44:58–66.

Karp, Daniel S., Rebecca Chaplin-Kramer, Timothy D. Meehan, Emily A. Martin, Fabrice DeClerck, Heather Grab, Claudio Gratton, Lorin Hunt, Ashley E. Larsen, Adriana Martinez-Salinas et al. 2018. "Crop Pests and Predators Exhibit Inconsistent Responses to Surrounding Landscape Composition." *Science Advances* 4 (10):eaau7144.

Keane, Michael P. and Timothy Neal. 2024. "A Practical Guide to Weak Instruments." *Annual Review of Economics* 16:185–212.

Kleibergen, Frank. 2005. "Testing Parameters in GMM without Assuming That They Are Identified." *Econometrica* 73 (4):1103–1123.

Kleibergen, Frank and Richard Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133 (1):97–126.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter. 2022. "Valid T-Ratio Inference For IV." *American Economic Review* 112 (10):3260–3290.

Lehner, Bernhard, Kristine Verdin, and Andy Jarvis. 2008. "New Global Hydrography Derived from Spaceborne Elevation Data." *Eos, Transactions American Geophysical Union* 89 (10):93–94.

Letourneau, Deborah K., Inge Armbrecht, Bradley S. Rivera, Juan Manuel Lerma, Erika J. Carmona, Maria C. Daza, Sandra Escobar, Violeta Galindo, Carlos Gutierrez, Sergio D. Lopez et al. 2011. "Does Plant Diversity Benefit Agroecosystems? A Synthetic Review." *Ecological Applications* 21 (1):9–21.

Losey, J. E. and M. Vaughan. 2006. "The Economic Value of Ecological Services Provided by Insects." *BioScience* 56:311–323.

May, Robert M. 1976. "Simple Mathematical Models with Very Complicated Dynamics." *Nature* 261:459–467.

Messager, Mathis Loïc, Bernhard Lehner, Günther Grill, Irena Nedeva, and Oliver Schmitt. 2016. "Estimating the Volume and Age of Water Stored in Global Lakes Using a Geo-Statistical Approach." *Nature Communications* 7:13603.

Monfreda, C., N. Ramankutty, and J. A. Foley. 2008. "Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000." *Global Biogeochemical Cycles* 22 (1):GB1022.

Montiel Olea, José Luis and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3):358–369.

Morando, Bruno. 2024. "Testing The GAEZ Agronomic Model in the Fields: Evidence From Uganda." Working paper, Department of Economics, Finance & Accounting, Maynooth University.

Murdoch, William W., Cheryl J. Briggs, and Roger M. Nisbet. 2003. *Consumer-Resource Dynamics*. Princeton University Press.

Noack, F., A. Larsen, J. Kamp, and C. Levers. 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.

Oerke, E.-C. 2006. "Crop Losses to Pests." *The Journal of Agricultural Science* 144 (1):31–43.

Oldeman, L. R., R. T. A. Hakkeling, and W. G. Sombroek. 1991. "World Map of the Status of Human-Induced Soil Degradation: An Explanatory Note (GLASOD)." Tech. rep., ISRIC, Wageningen and UNEP, Nairobi.

O'Rourke, M. E., K. Rienzo-Stack, and A. G. Power. 2011. "A Multi-Scale, Landscape Approach to Predicting Insect Populations in Agroecosystems." *Ecological Applications* 21 (5):1782–1791.

Pianka, E. R. 2011. "Evolutionary Ecology." URL http://www.zo.utexas.edu/courses/bio373/erp-evolecol.html. 7th edition, e-book.

Ramankutty, Navin, Amato T. Evan, Chad Monfreda, and Jonathan A. Foley. 2008. "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000." *Global Biogeochemical Cycles* 22 (1):GB1003.

Redlich, S., E. A. Martin, and I. Steffan-Dewenter. 2018. "Landscape-Level Crop Diversity Benefits Biological Pest Control." *Journal of Applied Ecology* 55 (5):2419–2428.

Renard, D. and D. Tilman. 2019. "National Food Production Stabilized by Crop Diversity." *Nature* 571 (7764):257–260.

Root, Richard B. 1973. "Organization of a Plant-Arthropod Association in Simple and Diverse Habitats: The Fauna of Collards (Brassica Oleracea)." *Ecological Monographs* 43 (1):95–124.

Rusch, Adrien, Riccardo Bommarco, Mattias Jonsson, Henrik G. Smith, and Björn Ekbom. 2016. "Agricultural Landscape Simplification Reduces Natural Pest Control: A Quantitative Synthesis." *Biological Reviews* 91 (2):525–544.

Sanderson, Eleanor and Frank Windmeijer. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190 (2):212–221.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3):379–423.

Siebert, Stefan, Verena Henrich, Karen Frenken, and Jacob Burke. 2013. "Update of the Digital Global Map of Irrigation Areas to Version 5." Tech. rep., Rheinische Friedrich-Wilhelms-Universität Bonn and FAO. Global Map of Irrigation Areas (GMIA) v5.

Smale, M., J. Hartell, P. Heisey, and B. Senauer. 1998. "The Contribution of Genetic Resources and Diversity to Wheat Production in the Punjab of Pakistan." *American Journal of Agricultural Economics* 80:482–493.

Strobl, E. 2022. "Preserving Local Biodiversity through Crop Diversification." *American Journal of Agricultural Economics* 104 (3):1140–1174.

Tang, Fiona H. M., Thu Ha Nguyen, Giulia Conchedda, Leon Casse, Francesco N. Tubiello, and Federico Maggi. 2024. "CROPGRIDS: A Global Geo-Referenced Dataset of 173 Crops." *Scientific Data* 11:413.

Thies, Carsten and Teja Tscharntke. 1999. "Landscape Structure and Biological Control in Agroecosystems." *Science* 285 (5429):893–895.

Tilman, D. 1999. "Global Environmental Impacts of Agricultural Expansion: The Need for Sustainable and Efficient Practices." *Proceedings of the National Academy of Sciences* 96 (11):5995–6000.

Tilman, David, Stephen Polasky, and Clarence Lehman. 2005. "Diversity, Productivity and Temporal Stability in the Economies of Humans and Nature." *Journal of Environmental Economics and Management* 49 (3):405–426.

Tscharntke, T., A. Klein, A. Kruess, I. Steffan-Dewenter, and C. Thies. 2005. "Landscape Perspectives on Agricultural Intensification and Biodiversity — Ecosystem Service Management." *Ecology Letters* 8:857–874.

USDA National Agricultural Statistics Service. 2017. "Census of Agriculture." U.S. Department of Agriculture. URL https://www.nass.usda.gov/AgCensus/.

Van Klink, R., D. E. Bowler, K. B. Gongalsky, A. B. Swengel, A. Gentile, and J. M. Chase. 2020. "Meta-Analysis Reveals Declines in Terrestrial but Increases in Freshwater Insect Abundances." *Science* 368 (6489):417–420.

Weitzman, M. 2000. "Economic Profitability Versus Ecological Entropy." *Quarterly Journal of Economics* 115 (1):237–263.

West, Paul C., James S. Gerber, Peder M. Engstrom, Nathaniel D. Mueller, Kate A. Brauman, Kimberly M. Carlson, Emily S. Cassidy, Matt Johnston, Graham K. MacDonald, Deepak K. Ray, and Stefan Siebert. 2014. "Leverage Points for Improving Global Food Security and the Environment." *Science* 345 (6194):325–328.

You, Liangzhi, Stanley Wood, Ulrike Wood-Sichra, and Wenbin Wu. 2014. "Generating Global Crop Distribution Maps: From Census to Grid." *Agricultural Systems* 127:53–60.

# Appendix

This appendix provides proofs of our theoretical predictions (section A), presents a biological foundation for our production model (Section B), and reports additional empirical results, including descriptive statistics, instrument relevance and validity, and robustness (Sections C, D and F).

## A Proofs and Farmland Shares in Equilibrium

### A.1 Proof of Proposition 1

Under perfect competition, farmers are atomistic and take crop prices $p_j^k$, the county-level rental cost of land and labor $c_j$, and the crop allocation $(\boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star})$ as given. For a farmer cultivating crop $k$ in county $j$, expected profit is $\pi_j^k = R_j^k - c_j$, where

$$R_j^k \equiv \max_{\rho} p_j^k \bar{a}_j^k \psi_j^k(\rho; \boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star}) - \tau_j \rho + T_j^k \tag{13}$$

is the maximum expected gross return from allocating one field to crop $k$ in county $j$. Since land can be reallocated freely across crops within county $j$, competitive bidding for land implies that the equilibrium rental cost satisfies $c_j = \max_k R_j^k$. Hence, for any crop grown in equilibrium, i.e. any crop such that $S_j^{k\star} > 0$, we must have $R_j^k = c_j$, i.e. $\pi_j^{k\star} = 0$. By contrast, for any crop not grown in equilibrium, $S_j^{k\star} = 0$, we have $R_j^k \leq c_j$, with strict inequality whenever the crop is strictly dominated, so that its expected profit is weakly negative.

Using that the average tax payment for crop $k$ in area $j$ is $\tau_j \rho_j^{k\star} = T_j^k$ at equilibrium, condition $\pi_j^{k\star} = 0$ leads to

$$p_j^k \psi_j^k(\rho_j^{k\star}; \boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star}) \bar{a}_j^k = c_j \tag{14}$$

for all $k \in \mathcal{K}_j \equiv \{k = 1, \ldots, K | S_j^{k\star} > 0\}$.

Differentiating (13) with respect to $\rho$, the symmetric Nash equilibrium between farmers is such that

$$\psi_j^{k\prime}(\rho_j^{k\star}; \boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star}) = \tau_j / (\bar{a}_j^k p_j^k) \tag{15}$$

for all $k \in \mathcal{K}_j$.

Then, combining (4), (14) and (15), the equilibrium—dominant—strategy of a farmer growing $k$ in area $j$ is

$$\rho_j^{k\star} = \hat{\rho}_j^k - \lambda_j^k \tau_j / c_j.$$

The field resilience to direct biotic factors becomes $\mu_j^k(\rho_j^{k\star}) = 1/t_j^k$, with

$$t_j^k \equiv \exp\left[\lambda_j^k (\tau_j / c_j)^2 / 2\right]$$

an index measuring the extent to which pesticides regulation and pest resistance in area $j$ reduce resilience. *Cross-plot* biotic factors being separable between crops, we also have:

$$B_j^k(\boldsymbol{S_j}, \boldsymbol{\rho_j}) = \frac{\exp\left[-\sum_{\ell\in\mathcal{K}_j}\hat{\kappa}_k(S_j^\ell)\right]}{\prod_{\ell\in\mathcal{K}_j} t_j^\ell}.$$

Eventually, using the developed expressions for all factors, the equilibrium resilience function satisfies

$$\psi_j^k(\rho_j^{k\star}; \boldsymbol{S_j^\star}, \boldsymbol{\rho_j^\star}) = \frac{A_j^k \exp\left[-\sum_{\ell\in\mathcal{K}_j}\hat{\kappa}_k(S_j^{\ell\star})\right]}{t_j^k \prod_{\ell\in\mathcal{K}_j} t_j^\ell}.$$

Writing $a_j^k = \bar{a}_j^k A_j^k$ the *potential* yield absent biotic factors, the *actual* productivity for a given crop $k$ in country $j$ becomes

$$z_j^{k\star} = \frac{a_j^k \exp\left[-\sum_{\ell\in\mathcal{K}_j}\hat{\kappa}_k(S_j^{l\star})\right]}{t_j^k \prod_{\ell\in\mathcal{K}_j} t_j^\ell},$$

which accounts for the direct ($t_j^k$) and indirect ($\prod_{\ell\in\mathcal{K}_j} t_j^\ell$) effects of a reduction in pesticides use. Using $p_j^k z_j^k = c_j$ for all $k \in \mathcal{K}_j$, we also have

$$\frac{p_j^k a_j^k}{p_j^{k'} a_j^{k'}} = \frac{t_j^{k'}}{t_j^k} \exp\left[-\sum_{\ell\in\mathcal{K}_j} \hat{\kappa}_k(S_j^{\ell\star}) - \hat{\kappa}_{k'}(S_j^{\ell\star})\right]$$

for any pair of crops $(k, k') \in \mathcal{K}_j^2$.

## A.2 Farmland Shares in Equilibrium

Given functional forms in Assumption 1, we have

$$\frac{p_j^\ell a_j^\ell}{p_j^k a_j^k} = \frac{t_j^\ell}{t_j^k} \exp\left[\kappa_{0\ell} S_j^{\ell\star} - \kappa_{0k} S_j^{k\star} + (\kappa_k - \kappa_\ell)\bar{H}_j + \kappa_{k'} H_j^{k'\star} - \kappa_k H_j^{k\star}\right].$$

for all pair $(k, \ell) \in \mathcal{K}_j^2$.

Taking logs, dividing by $\kappa_{0l}$, summing over all $\ell \in \mathcal{K}_j$ and using $\sum_{l\in\mathcal{K}_j} S_j^\ell = 1$, it comes

$$\begin{aligned} S_j^{k\star} = \frac{1}{\kappa_{0k}} \Bigg[ & \frac{\bar{\kappa}_0}{K_j} + \left(\ln\left[p_j^\ell a_j^\ell\right] - \frac{1}{K_j}\sum_{\ell\in\mathcal{K}_j} \ln\left[(p_j^\ell a_j^\ell)^{\bar{\kappa}_0/\kappa_{0\ell}}\right]\right) \\ & - \left(\ln t_j^k - \frac{1}{K_j}\sum_{\ell\in\mathcal{K}_j} \ln\left[(t_j^\ell)^{\bar{\kappa}_0/\kappa_{0\ell}}\right]\right) \\ & + \left(\kappa_k\left(\bar{H}_j - H_j^{k\star}\right) - \sum_{\ell\in\mathcal{K}_j} \kappa_\ell \frac{\bar{\kappa}_0}{\kappa_{0\ell}}\left(\bar{H}_j - H_j^{\ell\star}\right)\right)\Bigg], \end{aligned} \tag{16}$$

with $\bar{\kappa}_0 \equiv K_j / \sum_{\ell \in \mathcal{K}_j} (1/\kappa_{0\ell})$ is the harmonic mean of the within-crop parameters.

To illustrate the mechanisms at play, consider the special case where specialization effects are the same for all crops, i.e., $\kappa_{0\ell} = \kappa_0$, and diversity effects are absent, i.e., $\kappa_\ell = 0$ for all $\ell$. As $\bar{\kappa_0} = \kappa_0$, (16) simplifies to

$$S_j^{k\star} = \frac{1}{K_j} + \frac{1}{\kappa_0} \left[ \left( \ln(p_j^k a_j^k) - \frac{1}{K_j} \sum_{\ell \in \mathcal{K}_j} \ln(p_j^\ell a_j^\ell) \right) - \left( \ln t_j^k - \frac{1}{K_j} \sum_{\ell \in \mathcal{K}_j} \ln t_j^\ell \right) \right].$$

In this case, the difference between crop acreages is only due to their differences in potential revenue $p_j^k a_j^k$ and the impact of the environmental tax. The first bracketed term $\ln(p_j^k a_j^k) - 1/K_j \sum_{\ell \in \mathcal{K}_j} \ln(p_j^\ell a_j^\ell)$ corresponds to the difference between the potential revenue of crop $k$ and the average revenue (in the geometric mean sense). Similarly, the second bracketed term measures the relative impact of the pesticides tax on crop $k$ compared to the other crops. The largest areas are then devoted to crops whose difference between relative potential revenue and relative tax impact is the largest. If crop $k$ suffers for higher direct specialization effects than the others, i.e. $\kappa_{0k} > \kappa_{0k'} = \kappa_0$ for $k' \neq k$, then

$$S_j^{k\star} < \frac{1}{K_j} + \frac{1}{\kappa_0} \left( \ln(p_j^k a_j^k / t_j^k) - \frac{1}{K_j} \sum_{\ell \in \mathcal{K}_j} \ln(p_j^\ell a_j^\ell / t_j^\ell)^{\bar{\kappa_0}/\kappa_{0\ell}} \right)$$

and

$$S_j^{k'\star} > \frac{1}{K_j} + \frac{1}{\kappa_0} \left( \ln(p_j^{k'} a_j^{k'} / t_j^{k'}) - \frac{1}{K_j} \sum_{\ell \in \mathcal{K}_j} \ln(p_j^\ell a_j^\ell / t_j^\ell)^{\bar{\kappa_0}/\kappa_{0\ell}} \right).$$

Hence, a reallocation of crops occurs with a relative decrease in the acreage of crop $k$ and a relative increase in the acreage of the other crops.

### A.3 Proof of Proposition 2

For a given county $j$, the social planner chooses the crop allocation $\boldsymbol{S}_j = \{S_j^k\}_{k=1}^K$ to maximize the value of agricultural production per unit of land. The optimization problem is

$$\max_{\boldsymbol{S}_j} \left\{ \sum_{k=1}^K p_j^k z_j^k S_j^k : \sum_{k=1}^K S_j^k = 1 \right\},$$

where effective yields satisfy

$$z_j^k = \frac{a_j^k \exp\left[ -\kappa_0^k S_j^k - \kappa^k \left( \bar{H}_j - H_j^k(\boldsymbol{S}_j) \right) \right]}{t_j^k \prod_{\ell \in \mathcal{K}_j} t_j^\ell},$$

with $\mathcal{K}_j = \Big\{k = 1, \ldots, K : S_j^k > 0\Big\}$,$\bar{H}_j = \ln(|\mathcal{K}_j| - 1)$, and

$$H_j^k(\boldsymbol{S}_j) = -\sum_{\ell \in \mathcal{K}_j \setminus \{k\}} \frac{S_j^\ell}{1 - S_j^k} \ln\left(\frac{S_j^\ell}{1 - S_j^k}\right).$$

The corresponding Lagrangian is

$$\mathcal{L} = \sum_{k=1}^{K} \left(p_j^k z_j^k - \theta_j\right) S_j^k,$$

where $\theta_j$ denotes the shadow value of land. The first-order condition with respect to $S_j^k$ is

$$p_j^k z_j^k \left(1 - \kappa_0^k S_j^k\right) + \sum_{\ell \in \mathcal{K}_j} p_j^\ell z_j^\ell \kappa^\ell S_j^\ell \frac{\partial H_j^\ell}{\partial S_j^k} = \theta_j,$$

for every cultivated crop $k \in \mathcal{K}_j$. Hence,

$$S_j^k = \frac{1}{\kappa_0^k} + \sum_{\ell \in \mathcal{K}_j} \frac{p_j^\ell z_j^\ell \kappa^\ell S_j^\ell}{p_j^k z_j^k \kappa_0^k} \frac{\partial H_j^\ell}{\partial S_j^k} - \frac{\theta_j}{p_j^k z_j^k \kappa_0^k}. \tag{17}$$

Summing over all cultivated crops gives

$$1 = \sum_{k \in \mathcal{K}_j} \frac{1}{\kappa_0^k} + \sum_{k \in \mathcal{K}_j} \sum_{\ell \in \mathcal{K}_j} \frac{p_j^\ell z_j^\ell \kappa^\ell S_j^\ell}{p_j^k z_j^k \kappa_0^k} \frac{\partial H_j^\ell}{\partial S_j^k} - \theta_j \sum_{k \in \mathcal{K}_j} \frac{1}{p_j^k z_j^k \kappa_0^k},$$

hence

$$\theta_j = \frac{\sum_{\ell \in \mathcal{K}_j} \frac{1}{\kappa_0^\ell} - 1 + \sum_{\ell \in \mathcal{K}_j} \sum_{m \in \mathcal{K}_j} \frac{p_j^\ell z_j^\ell \kappa^\ell S_j^\ell}{p_j^m z_j^m \kappa_0^m} \frac{\partial H_j^\ell}{\partial S_j^m}}{\sum_{\ell \in \mathcal{K}_j} \frac{1}{\kappa_0^\ell p_j^\ell z_j^\ell}}.$$

Substituting this expression into (17) yields (10).

To implement the counterfactual simulations, we use the observed crop allocation $\hat{\boldsymbol{S}}_j$ and observed effective yields $\{\hat{z}_j^k\}_{k=1}^K$. Under competitive equilibrium, expected crop revenues are equalized across all cultivated crops,

$$p_j^k \hat{z}_j^k = p_j^\ell \hat{z}_j^\ell = c_j, \qquad \forall\, k, \ell \in \mathcal{K}_j,$$

which implies

$$\frac{p_j^k a_j^k t_j^\ell}{p_j^\ell a_j^\ell t_j^k} = \exp\Big[\kappa_0^k \hat{S}_j^k - \kappa_0^\ell \hat{S}_j^\ell - \Big(\kappa^k H_j^k(\hat{\boldsymbol{S}}_j) - \kappa^\ell H_j^\ell(\hat{\boldsymbol{S}}_j)\Big)\Big] \equiv \hat{e}_j^{k\ell}.$$

Consequently, for any counterfactual crop allocation $\boldsymbol{S}_j$,

$$\frac{p_j^k z_j^k}{p_j^\ell z_j^\ell} = \hat{e}_j^{k\ell} \exp\Big[-\left(\kappa_0^k S_j^k - \kappa_0^\ell S_j^\ell\right) + \kappa^k H_j^k(\boldsymbol{S}_j) - \kappa^\ell H_j^\ell(\boldsymbol{S}_j)\Big].$$

## B From Pest Dynamics to Agricultural Productivity.

Standard models of pest population dynamics in ecology can provide biological foundations to our production model. A large literature models pest populations as following logistic or consumer–resource dynamics, in which population growth is limited by a carrying capacity determined by resource availability (e.g., May, 1976; Murdoch, Briggs, and Nisbet, 2003; Hassell, 1978). In agricultural systems, this carrying capacity depends on the abundance and spatial concentration of host plants, consistent with the resource concentration hypothesis (Root, 1973). At the same time, plant diversity and landscape heterogeneity affect pest regulation through dilution effects and the activity of natural enemies (Andow, 1991; Tscharntke et al., 2005; Letourneau et al., 2011). These mechanisms imply that pest populations increase with the share of the host crop, decrease with crop diversity, and are directly reduced by pesticide use. The simple population model introduced below formalizes these relationships and shows how they can be mapped into the reduced-form productivity equation used in the main text.

Consider the dynamic of the pest population affecting a field growing crop $k \in \mathcal{K}_j$ in county $j$:

$$\dot{P}_j^k = r_k P_j^k \left(1 - \frac{P_j^k}{K_j^k}\right) - \delta_k(\rho_j^k; \boldsymbol{\rho}_j) P_j^k, \tag{18}$$

where $r_k$ is the intrinsic growth rate, $K_j^k$ is the carrying capacity, and $\delta_k(\rho_j^k; \boldsymbol{\rho}_j)$ is the pesticide-induced mortality which depends on the treatment on the field, $\rho_j^k$, and on all average treatments in the county, $\boldsymbol{\rho}_j \equiv \{\rho_j^k, k \in \mathcal{K}_j\}$. Assume that the carrying capacity depends positively on the share of host crop $k$ and negatively on crop diversity in the county:

$$K_j^k = \bar{K}_k S_j^k \exp(-\gamma_k H_j^k), \tag{19}$$

where $H_j^k$ denotes the leave-one-out diversity index. Suppose that during the growing season, the pest population reaches a steady state (i.e., $\dot{P}_j^k = 0$). This equilibrium population is given by

$$P_j^{k*} = K_j^k \left(1 - \frac{\delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{r_k}\right) = \bar{K}_k S_j^k \exp(-\gamma_k H_j^k) \left(1 - \frac{\delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{r_k}\right). \tag{20}$$

Suppose next that expected productivity equals attainable productivity times a survival term decreasing in equilibrium pest population:

$$z_j^k = a_j^k \exp(-\eta_k P_j^{k*}), \tag{21}$$

with $\eta_k > 0$. Substituting for $P_j^{k*}$ gives

$$z_j^k = a_j^k \exp\left[-G_j^k(S, H, \rho_j^k; \boldsymbol{\rho}_j)\right], \tag{22}$$

where

$$G_j^k(S, H, \rho_j^k; \boldsymbol{\rho}_j) \equiv \eta_k \bar{K}_k S \exp(-\gamma_k H) \left(1 - \frac{\delta_k(\rho, \boldsymbol{\rho}_j)}{r_k}\right).$$

Consider a reference point $(\bar{S}_j^k, \bar{H}_j, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)$ where $\hat{\boldsymbol{\rho}}_j \equiv \{\hat{\rho}_j^k, k \in \mathcal{K}_j\}$ corresponds to the vector

of unrestricted pesticides treatment levels, and $\bar{H}_j \equiv \ln(K_j - 1)$ to the minimum value of the leave-one-out diversity index. A first-order Taylor expansion of $G_j^k$ around that point yields

$$G_j^k(S_j^k, H_j^k, \rho_j^k; \hat{\boldsymbol{\rho}}_j) \approx \bar{G}_j^k + \kappa_0^k(S_j^k - \bar{S}_j^k) - \kappa^k(H_j^k - \bar{H}) - \lambda_j^k(\hat{\rho}_j^k - \rho_j^k) - \sum_{\ell \in \mathcal{K}_j} \phi_j^\ell(\hat{\rho}_j^\ell - \rho_j^\ell),$$

where $\bar{G}_j^k \equiv G^k(\bar{S}_j^k, \bar{H}, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)$,

$$\kappa_0^k \equiv \frac{\partial G_j^k(\bar{S}_j^k, \bar{H}, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)}{\partial S} = \eta_k \bar{K}_k \exp(-\gamma_k \bar{H}) \left(1 - \frac{\delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{r_k}\right) > 0,$$

$$\kappa^k \equiv -\frac{\partial G_j^k(\bar{S}_j^k, \bar{H}, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)}{\partial H} = \gamma_k \eta_k \bar{K}_k \bar{S}_j^k \exp(-\gamma_k \bar{H}) \left(1 - \frac{\delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{r_k}\right) > 0,$$

$$\lambda_j^k \equiv -\frac{\partial G^k(\bar{S}_j^k, \bar{H}, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)}{\partial \rho} = -\eta_k \bar{K}_k \bar{S}_j^k \exp(-\gamma_k \bar{H}) \frac{1}{r_k} \frac{\partial \delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{\partial \rho} > 0,$$

and

$$\phi_j^\ell \equiv -\frac{\partial G^k(\bar{S}_j^k, \bar{H}, \hat{\rho}_j^k; \hat{\boldsymbol{\rho}}_j)}{\partial \hat{\rho}_j^\ell} = -\eta_k \bar{K}_k \bar{S}_j^k \exp(-\gamma_k \bar{H}) \frac{1}{r_k} \frac{\partial \delta_k(\rho_j^k; \boldsymbol{\rho}_j)}{\partial \hat{\rho}_j^\ell} > 0.$$

Substituting into (22) and using $\kappa_0^k \bar{S}_j^k = \bar{G}_j^k$ gives

$$z_j^k \approx a_j^k \exp\left[-\kappa_0^k S_j^k - \kappa^k(H_j^k - \bar{H}) - \Lambda_j^k\right],$$

where $\Lambda_j^k = \lambda_j^k(\hat{\rho}_j^k - \rho_j^k) + \sum_{\ell \in \mathcal{K}_j} \phi_j^\ell(\hat{\rho}_j^\ell - \rho_j^\ell)$ collects all pesticide-related effects. In our model, it corresponds at equilibrium to $\Lambda_j^k = \ln t_j^k + \sum_{\ell \in \mathcal{K}_j} \ln t_j^\ell$. Substituting gives (9).

# C Sample and Descriptive Statistics

**Table C1:** Crop Sample Description

| Crop | Mean yield (t/ha) | | Share in harvested area (%) | | Share of counties where crop dominates (%) |
|---|---|---|---|---|---|
| | Actual | Potential | Actual | Potential | |
| Alfalfa | 6.57 | 8.07 | 2.01 | 3.10 | 2.46 |
| Banana | 1.81 | 1.17 | 0.39 | 0.89 | 0.64 |
| Barley | 1.07 | 2.12 | 5.35 | 2.93 | 4.47 |
| Buckwheat | 0.31 | 0.50 | 0.26 | 2.52 | 0.01 |
| Cabbage | 2.89 | 1.80 | 0.27 | 2.85 | 0.03 |
| Carrot | 2.63 | 3.31 | 0.09 | 3.88 | 0.00 |
| Cassava | 1.51 | 2.20 | 1.52 | 2.03 | 2.81 |
| Chickpea | 0.04 | 0.65 | 1.02 | 1.88 | 0.03 |
| Cocoa | 0.10 | 0.28 | 0.66 | 0.81 | 0.92 |
| Coconut | 0.47 | 0.50 | 1.03 | 0.73 | 1.70 |
| Coffee | 0.27 | 0.36 | 1.00 | 0.95 | 3.03 |
| Cotton | 0.16 | 0.22 | 2.90 | 3.33 | 1.37 |
| Cow pea | 0.10 | 0.61 | 0.88 | 2.59 | 0.37 |
| Dry pea | 0.73 | 0.78 | 0.61 | 2.46 | 0.01 |
| Flax | 0.10 | 0.34 | 0.05 | 2.64 | 0.00 |
| Groundnut | 0.41 | 0.76 | 2.20 | 2.82 | 0.44 |
| Maize | 2.14 | 3.37 | 13.38 | 3.78 | 27.76 |
| Millet | 0.33 | 1.22 | 3.33 | 3.41 | 0.85 |
| Oat | 0.78 | 1.14 | 1.31 | 2.83 | 0.34 |
| Oil palm | 0.67 | 0.40 | 0.95 | 0.42 | 0.81 |
| Olive | 0.08 | 0.08 | 0.67 | 0.55 | 1.11 |
| Onion | 3.24 | 3.32 | 0.26 | 3.51 | 0.02 |
| Phaseolus bean | 0.30 | 1.59 | 0.02 | 4.37 | 0.02 |
| Pigeon pea | 0.01 | 0.75 | 0.42 | 2.34 | 0.00 |
| Rapeseed | 0.66 | 1.14 | 2.42 | 2.93 | 0.05 |
| Rice | 1.45 | 1.69 | 14.79 | 2.22 | 15.96 |
| Rubber | 0.26 | 0.17 | 0.76 | 0.59 | 0.34 |
| Rye | 0.57 | 1.34 | 0.94 | 2.91 | 0.48 |
| Sorghum | 0.87 | 2.36 | 3.84 | 5.71 | 2.10 |
| Soybean | 0.78 | 1.75 | 7.45 | 3.71 | 4.48 |
| Sugar beet | 1.18 | 1.61 | 0.60 | 1.71 | 0.05 |
| Sugarcane | 2.32 | 1.87 | 1.92 | 1.15 | 3.62 |
| Sunflower | 0.56 | 1.72 | 2.02 | 3.91 | 0.23 |
| Sweet potato | 1.13 | 2.39 | 0.89 | 2.52 | 0.17 |
| Tea | 0.09 | 0.12 | 0.24 | 0.83 | 0.07 |
| Tobacco | 0.65 | 0.41 | 0.41 | 2.78 | 0.23 |
| Tomato | 6.88 | 2.34 | 0.34 | 3.24 | 0.09 |
| Wheat | 1.33 | 2.47 | 20.55 | 3.56 | 19.18 |
| White potato | 2.21 | 3.45 | 1.90 | 2.73 | 1.83 |
| Yam | 0.98 | 1.82 | 0.36 | 1.87 | 0.08 |

*Notes*: The statistics for each of the 40 crops are computed for our sample of 39,116 counties. The second and third columns indicate the average yield of each crop as estimated by FAO-GAEZ (potential) and Earthstat (actual), respectively; the fourth and fifth columns indicate the average (weighted by county-level total harvested area) and maximum share of each crop in harvested area, respectively; the sixth column indicates the percentage of counties in which the crop has the largest share in harvested area. Computed based on FAO-GAEZ, Earthstat.

**Table C2:** Country Sample Description

| Country | Harvested area | | Effective diversity | |
|---|---|---|---|---|
| | Total (Mha) | Share (%) | Actual | Potential |
| Afghanistan | 2.62 | 0.27 | 2.55 | 12.79 |
| Albania | 0.32 | 0.03 | 5.16 | 26.20 |
| Algeria | 2.31 | 0.23 | 3.26 | 21.47 |
| Angola | 1.85 | 0.19 | 5.06 | 24.51 |
| Argentina | 24.84 | 2.52 | 4.78 | 25.48 |
| Australia | 20.05 | 2.04 | 4.07 | 19.25 |
| Austria | 0.98 | 0.10 | 6.55 | 20.00 |
| Azerbaijan | 0.89 | 0.09 | 3.96 | 23.80 |
| Bangladesh | 12.19 | 1.24 | 2.11 | 28.26 |
| Belarus | 3.14 | 0.32 | 7.18 | 19.88 |
| Belgium | 0.27 | 0.03 | 4.71 | 20.02 |
| Benin | 1.38 | 0.14 | 5.41 | 22.13 |
| Bhutan | 0.08 | 0.01 | 4.29 | 14.94 |
| Bolivia | 1.90 | 0.19 | 6.33 | 16.79 |
| Bosnia and Herzegovina | 0.45 | 0.05 | 5.68 | 23.16 |
| Botswana | 0.11 | 0.01 | 2.70 | 15.57 |
| Brazil | 41.97 | 4.26 | 3.52 | 25.31 |
| Brunei | 0.00 | 0.00 | 2.28 | 17.30 |
| Bulgaria | 2.68 | 0.27 | 5.63 | 23.12 |
| Burkina Faso | 3.36 | 0.34 | 4.48 | 17.27 |
| Burundi | 0.79 | 0.08 | 7.75 | 24.10 |
| Cambodia | 2.20 | 0.22 | 1.88 | 24.66 |
| Cameroon | 2.21 | 0.22 | 7.63 | 23.30 |
| Canada | 28.99 | 2.94 | 4.06 | 17.55 |
| Central African Republic | 0.59 | 0.06 | 6.43 | 24.68 |
| Chad | 1.67 | 0.17 | 3.98 | 15.24 |
| Chile | 0.72 | 0.07 | 6.95 | 14.85 |
| China | 128.30 | 13.03 | 6.73 | 25.21 |
| Colombia | 2.92 | 0.30 | 7.32 | 17.87 |
| Costa Rica | 0.32 | 0.03 | 6.58 | 19.87 |
| Croatia | 0.94 | 0.10 | 6.83 | 23.71 |
| Cuba | 1.72 | 0.17 | 5.03 | 26.49 |
| Czechia | 1.84 | 0.19 | 6.05 | 19.81 |
| Côte d'Ivoire | 5.26 | 0.53 | 7.05 | 24.47 |
| Democratic Republic of the Congo | 5.23 | 0.53 | 5.76 | 24.91 |
| Denmark | 1.70 | 0.17 | 4.27 | 16.48 |
| Dominican Republic | 0.74 | 0.07 | 9.04 | 26.32 |
| Ecuador | 2.05 | 0.21 | 6.85 | 19.41 |
| Egypt | 0.42 | 0.04 | 6.75 | 1.99 |
| El Salvador | 0.62 | 0.06 | 4.41 | 23.56 |
| Equatorial Guinea | 0.10 | 0.01 | 4.41 | 20.89 |
| Eritrea | 0.30 | 0.03 | 3.83 | 13.09 |
| Estonia | 0.34 | 0.03 | 5.12 | 15.66 |
| Ethiopia | 4.46 | 0.45 | 6.59 | 22.06 |
| Finland | 1.25 | 0.13 | 4.23 | 11.70 |
| France | 11.78 | 1.20 | 5.92 | 23.13 |
| French Guiana | 0.01 | 0.00 | 2.52 | 19.98 |
| Gabon | 0.13 | 0.01 | 7.26 | 23.11 |
| Gambia | 0.20 | 0.02 | 3.86 | 17.20 |
| Georgia | 0.49 | 0.05 | 5.99 | 24.21 |
| Germany | 7.46 | 0.76 | 5.51 | 19.80 |

*Continued on next page*

Table C2 – *Continued from previous page*

| Country | Harvested area | | Mean effective diversity | |
|---|---|---|---|---|
| | Total (Mha) | Share (%) | Actual | Potential |
| Ghana | 3.97 | 0.40 | 5.84 | 23.77 |
| Greece | 2.38 | 0.24 | 4.40 | 25.06 |
| Guatemala | 1.26 | 0.13 | 4.84 | 21.34 |
| Guinea | 1.35 | 0.14 | 5.57 | 22.98 |
| Guinea-Bissau | 0.13 | 0.01 | 5.14 | 19.17 |
| Guyana | 0.20 | 0.02 | 3.46 | 21.97 |
| Haiti | 0.83 | 0.08 | 9.28 | 26.52 |
| Honduras | 0.78 | 0.08 | 4.73 | 25.58 |
| Hungary | 3.49 | 0.35 | 6.23 | 22.76 |
| India | 145.19 | 14.74 | 4.85 | 21.96 |
| Indonesia | 27.03 | 2.74 | 6.32 | 21.22 |
| Iran | 9.86 | 1.00 | 4.48 | 16.73 |
| Iraq | 0.32 | 0.03 | 4.48 | 14.80 |
| Ireland | 0.34 | 0.03 | 3.63 | 13.98 |
| Italy | 5.82 | 0.59 | 5.83 | 26.11 |
| Japan | 2.42 | 0.25 | 3.98 | 25.81 |
| Jordan | 0.07 | 0.01 | 4.59 | 14.34 |
| Kazakhstan | 12.18 | 1.24 | 3.96 | 14.96 |
| Kenya | 3.02 | 0.31 | 7.17 | 21.07 |
| Kosovo | 0.31 | 0.03 | 6.11 | 22.33 |
| Kyrgyzstan | 0.82 | 0.08 | 4.72 | 19.00 |
| Laos | 0.83 | 0.08 | 2.28 | 26.10 |
| Latvia | 0.50 | 0.05 | 6.02 | 16.49 |
| Lebanon | 0.16 | 0.02 | 6.23 | 23.10 |
| Liberia | 0.33 | 0.03 | 5.72 | 22.98 |
| Lithuania | 1.17 | 0.12 | 5.86 | 17.31 |
| Luxembourg | 0.02 | 0.00 | 4.57 | 19.88 |
| Madagascar | 1.93 | 0.20 | 4.42 | 26.06 |
| Malawi | 1.32 | 0.13 | 6.42 | 23.00 |
| Malaysia | 5.36 | 0.54 | 3.42 | 19.53 |
| Mali | 2.43 | 0.25 | 3.80 | 14.19 |
| Mauritania | 0.24 | 0.02 | 3.61 | 3.27 |
| Mongolia | 0.24 | 0.02 | 1.41 | 11.17 |
| Morocco | 5.68 | 0.58 | 3.67 | 21.36 |
| Mozambique | 2.47 | 0.25 | 4.63 | 22.69 |
| Myanmar | 8.87 | 0.90 | 5.93 | 25.85 |
| México | 12.61 | 1.28 | 2.60 | 19.76 |
| Namibia | 0.15 | 0.01 | 1.96 | 14.64 |
| Nepal | 3.42 | 0.35 | 4.61 | 27.93 |
| Netherlands | 0.46 | 0.05 | 6.73 | 18.44 |
| New Zealand | 0.21 | 0.02 | 6.37 | 18.93 |
| Nicaragua | 0.64 | 0.06 | 4.91 | 24.13 |
| Niger | 10.18 | 1.03 | 2.91 | 8.92 |
| Nigeria | 32.57 | 3.31 | 5.86 | 20.37 |
| North Korea | 1.92 | 0.19 | 6.96 | 23.02 |
| Norway | 0.20 | 0.02 | 3.67 | 13.71 |
| Oman | 0.00 | 0.00 | 2.66 | 1.60 |
| Pakistan | 16.95 | 1.72 | 5.82 | 17.57 |
| Palestine | 0.07 | 0.01 | 6.84 | 22.07 |
| Panama | 0.25 | 0.03 | 5.85 | 22.96 |
| Papua New Guinea | 0.68 | 0.07 | 7.08 | 20.51 |
| Paraguay | 2.33 | 0.24 | 5.01 | 29.33 |

*Continued on next page*

Table C2 – *Continued from previous page*

| Country | Harvested area | | Mean effective diversity | |
|---|---|---|---|---|
| | Total (Mha) | Share (%) | Actual | Potential |
| Peru | 1.89 | 0.19 | 8.30 | 13.77 |
| Philippines | 10.76 | 1.09 | 3.96 | 21.24 |
| Poland | 8.39 | 0.85 | 6.25 | 19.65 |
| Portugal | 1.02 | 0.10 | 6.19 | 24.38 |
| Republic of the Congo | 0.16 | 0.02 | 6.06 | 24.54 |
| Romania | 7.55 | 0.77 | 5.14 | 21.89 |
| Russia | 47.84 | 4.86 | 5.57 | 18.94 |
| Rwanda | 0.66 | 0.07 | 7.25 | 22.02 |
| Senegal | 1.70 | 0.17 | 4.08 | 14.86 |
| Serbia | 2.32 | 0.24 | 6.09 | 23.57 |
| Sierra Leone | 0.36 | 0.04 | 4.63 | 24.12 |
| Slovakia | 1.13 | 0.12 | 7.31 | 20.03 |
| Slovenia | 0.11 | 0.01 | 5.32 | 21.91 |
| Somalia | 0.41 | 0.04 | 2.84 | 10.62 |
| South Africa | 5.64 | 0.57 | 4.50 | 22.62 |
| South Korea | 1.33 | 0.14 | 2.91 | 25.73 |
| South Sudan | 1.03 | 0.10 | 3.33 | 20.09 |
| Spain | 10.72 | 1.09 | 5.66 | 23.23 |
| Sri Lanka | 1.73 | 0.18 | 3.83 | 23.38 |
| Sudan | 8.07 | 0.82 | 2.67 | 11.62 |
| Suriname | 0.05 | 0.01 | 3.24 | 22.44 |
| Swaziland | 0.15 | 0.02 | 3.68 | 23.33 |
| Sweden | 1.09 | 0.11 | 3.59 | 14.71 |
| Switzerland | 0.22 | 0.02 | 6.28 | 18.60 |
| Syria | 3.70 | 0.38 | 4.33 | 18.78 |
| Taiwan | 0.49 | 0.05 | 4.34 | 28.39 |
| Tajikistan | 0.70 | 0.07 | 3.96 | 20.60 |
| Tanzania | 4.54 | 0.46 | 9.83 | 24.49 |
| Thailand | 15.87 | 1.61 | 4.19 | 24.41 |
| Timor-Leste | 0.15 | 0.02 | 4.33 | 24.67 |
| Togo | 1.16 | 0.12 | 4.88 | 23.14 |
| Tunisia | 1.47 | 0.15 | 3.69 | 23.07 |
| Turkey | 16.96 | 1.72 | 4.55 | 21.95 |
| Turkmenistan | 1.19 | 0.12 | 3.39 | 10.33 |
| Uganda | 3.52 | 0.36 | 10.98 | 25.87 |
| Ukraine | 18.54 | 1.88 | 6.58 | 22.09 |
| United Kingdom | 3.95 | 0.40 | 4.02 | 14.84 |
| United States | 107.56 | 10.92 | 3.40 | 24.79 |
| Uruguay | 0.68 | 0.07 | 5.06 | 31.75 |
| Uzbekistan | 2.63 | 0.27 | 3.63 | 18.23 |
| Venezuela | 1.45 | 0.15 | 7.33 | 23.51 |
| Vietnam | 9.54 | 0.97 | 4.56 | 25.31 |
| Yemen | 0.31 | 0.03 | 7.05 | 10.83 |
| Zambia | 1.04 | 0.11 | 4.43 | 22.50 |
| Zimbabwe | 2.12 | 0.22 | 5.44 | 21.20 |

*Notes*: The statistics for each of the 148 countries are computed over the 40 crops considered. The second and third columns indicate the total harvested area in the country and the share of each country in global harvested area; the fourth and fifth columns indicate average actual and potential effective crop diversity (Hill's numbers), with the Shannon index used as the underlying measure of mean proportional abundance. For instance, a number of 4 indicates that, on average, the country is as diversified as if it was growing 4 crops in equal proportions. Details on the computation of these indexes are available in Subsections 3.2.2 and 3.2.3. Computed based on FAO-GAEZ, Earthstat.

**Table C3:** Number of Pest Species per Crop

| Crop | Scientific name | No. of pest species |
|---|---|---|
| Tomato | *Solanum lycopersicum* | 243 |
| White potato | *Solanum tuberosum* | 225 |
| Maize | *Zea mays* | 155 |
| Phaseolus bean | *Phaseolus vulgaris* | 129 |
| Tobacco | *Nicotiana tabacum* | 118 |
| Rice | *Oryza sativa* | 115 |
| Cotton | *Gossypium hirsutum* | 112 |
| Coffee | *Coffea arabica* | 110 |
| Sugar cane | *Saccharum officinarum* | 101 |
| Sweet potato | *Ipomoea batatas* | 92 |
| Wheat | *Triticum aestivum* | 82 |
| Alfalfa | *Medicago sativa* | 80 |
| Banana | *Musa x paradisiaca* | 80 |
| Soybean | *Glycine max* | 80 |
| Cocoa | *Theobroma cacao* | 78 |
| Coconut | *Cocos nucifera* | 77 |
| Cow pea | *Vigna unguiculata* | 66 |
| Sorghum | *Sorghum bicolor* | 66 |
| Groundnut | *Arachis hypogaea* | 63 |
| Sunflower | *Helianthus annuus* | 63 |
| Cassava | *Manihot esculenta* | 60 |
| Onion | *Allium cepa* | 59 |
| Pigeon pea | *Cajanus cajan* | 56 |
| Cabbage | *Brassica oleracea* | 55 |
| Rapeseed | *Brassica napus* | 54 |
| Tea | *Camellia sinensis* | 53 |
| Olive | *Olea europaea* | 50 |
| Barley | *Hordeum vulgare* | 41 |
| Oil palm | *Elaeis guineensis* | 41 |
| Oat | *Avena sativa* | 39 |
| Rye | *Secale cereale* | 32 |
| Rubber | *Hevea brasiliensis* | 31 |
| Chickpea | *Cicer arietinum* | 28 |
| Flax | *Linum usitatissimum* | 19 |
| Carrot | *Daucus carota* | 18 |

*Continued on next page*

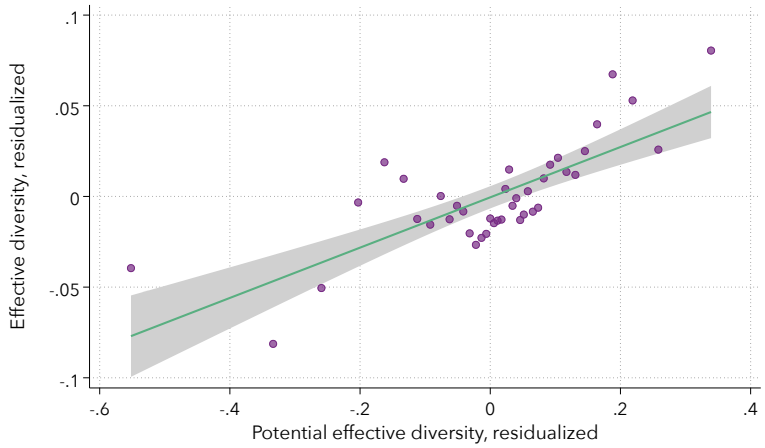
Table C3 – *continued*

| Crop | Scientific name | No. of pest species |
|---|---|---|
| Millet | *Pennisetum glaucum* | 14 |
| Buckwheat | *Fagopyrum esculentum* | 12 |
| Yam | *Dioscorea alata* | 10 |
| Dry pea | *Pisum sativum* | 7 |
| Sugar beet | *Beta vulgaris* | 3 |

*Notes*: Number of pest and pathogen species recorded for each crop in the EPPO Global Database (EPPO, 2026), queried by the crop's host species (scientific name). Crops above the rule form the high-pest group and those below the low-pest group, split at the across-crop median (61.5 species), as used in Table 4.

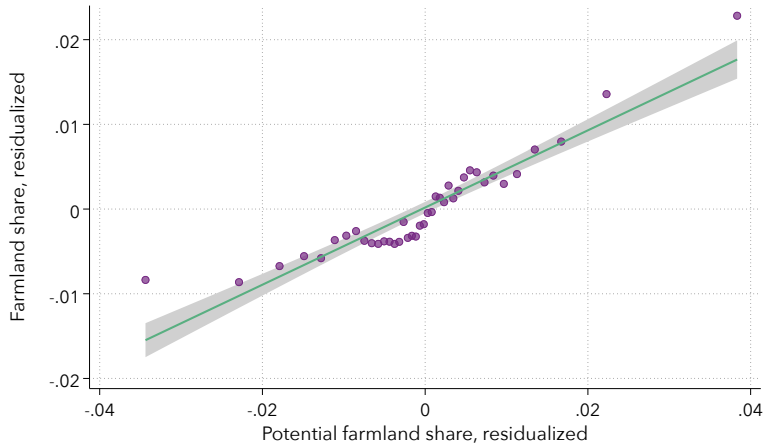
# D Instrument Relevance and Validity

**Figure D1:** First stage: effective diversity and its instrument

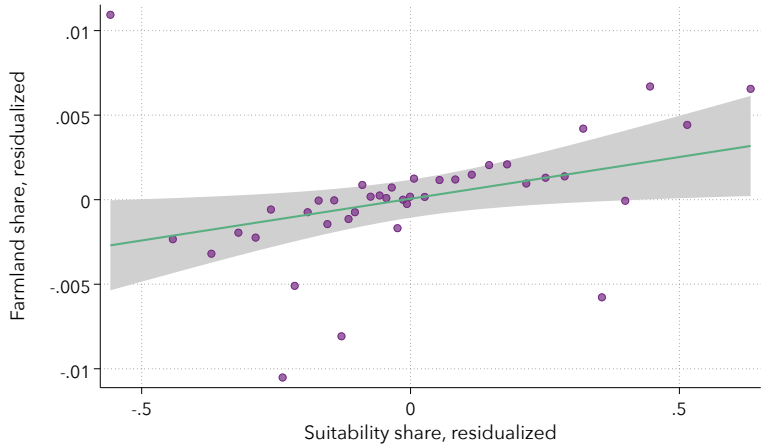



*Notes*: Binned scatterplot with 40 bins of leave-one-out effective diversity against its potential counterpart. Both variables are residualized on country×crop fixed effects, the variation used for identification, but not on controls or other instruments. The figure includes a linear fit with a 95% confidence band. The top and bottom 0.5% of the x-axis are trimmed for display.

**Figure D2:** First stage: farmland share and its two instruments



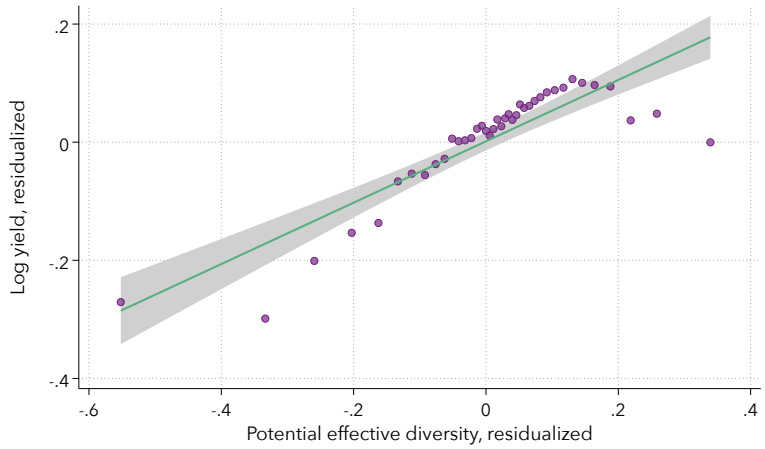

**(a)** Potential farmland share




**(b)** Suitability share

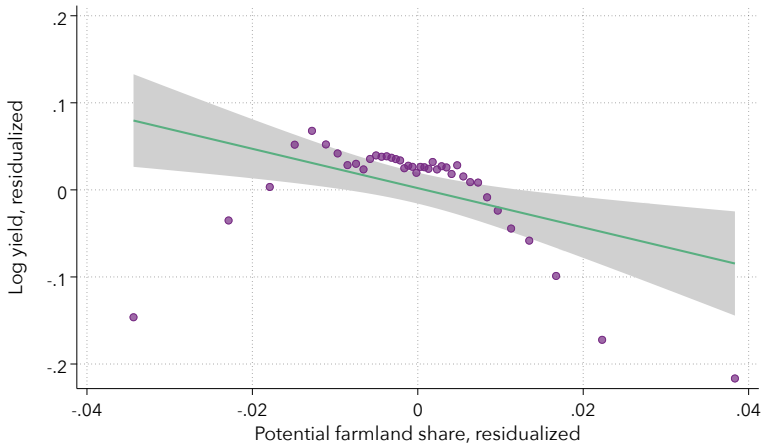
*Notes*: Binned scatterplots with 40 bins of the own-crop farmland share against its potential share (top panel) and suitability share (bottom panel). All variables are residualized on country×crop fixed effects, the variation used for identification, but not on controls or other instruments. Each panel includes a linear fit with a 95% confidence band. The top and bottom 0.5% of the x-axis are trimmed for display.

**Figure D3:** Reduced form: log yield and the diversity instrument

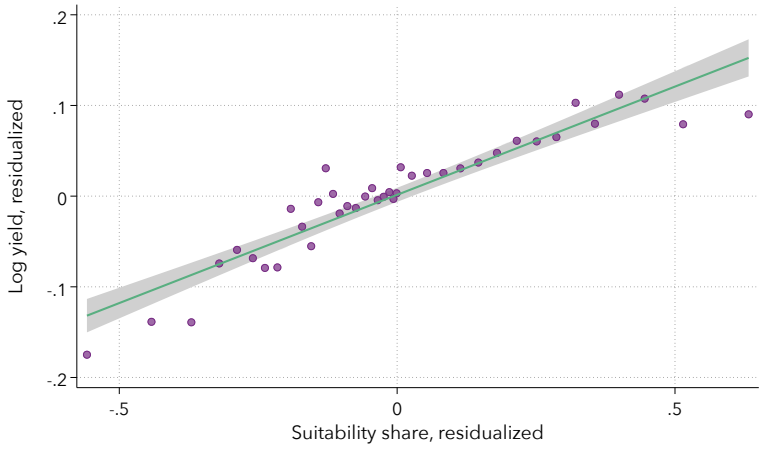



*Notes*: Binned scatterplot with 40 bins of log yield against the potential effective-diversity instrument. Both variables are residualized on country×crop fixed effects, the variation used for identification, but not on controls or other instruments. The figure includes a linear fit with a 95% confidence band. The top and bottom 0.5% of the x-axis are trimmed for display.

**Figure D4:** Reduced form: log yield and the farmland-share instruments




**(a)** Potential farmland share




**(b)** Suitability share

*Notes*: Binned scatterplots with 40 bins of log yield against the potential farmland-share instrument (top panel) and the suitability-share instrument (bottom panel). All variables are residualized on country×crop fixed effects, the variation used for identification, but not on controls or other instruments. Each panel includes a linear fit with a 95% confidence band. The top and bottom 0.5% of the x-axis are trimmed for display.

**Table D1:** Reduced Form: Log Yield on the Excluded Instruments

| | Log yield |
|---|---|
| | (1) |
| Potential effective diversity | 0.336*** |
| | (0.038) |
| Potential crop share | -1.573*** |
| | (0.207) |
| Suitability share | 0.036** |
| | (0.015) |
| Area, agro-climatic + water controls | Yes |
| Country × crop FE | Yes |
| R-squared | 0.770 |
| Observations | 586355 |

*Notes*: Reduced-form regression of log yield on the three excluded instruments (potential effective diversity, potential crop share, and the suitability share), under the baseline specification: country×crop fixed effects, the baseline controls (log attainable yield, log county area, and the agro-climatic and water controls), with standard errors clustered on a 2.5° spatial grid. This is the multivariate counterpart to the reduced-form binned scatters in Figures D3 and D4. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table D2:** Plausibly-exogenous Bounds (Relaxing the Exclusion Restriction)

| | 2SLS | Conley UCI (95%) |
|---|---|---|
| Log effective diversity | 1.655 | [0.400, 3.484] |
| Farmland share | -14.787 | [−28.643, −3.932] |

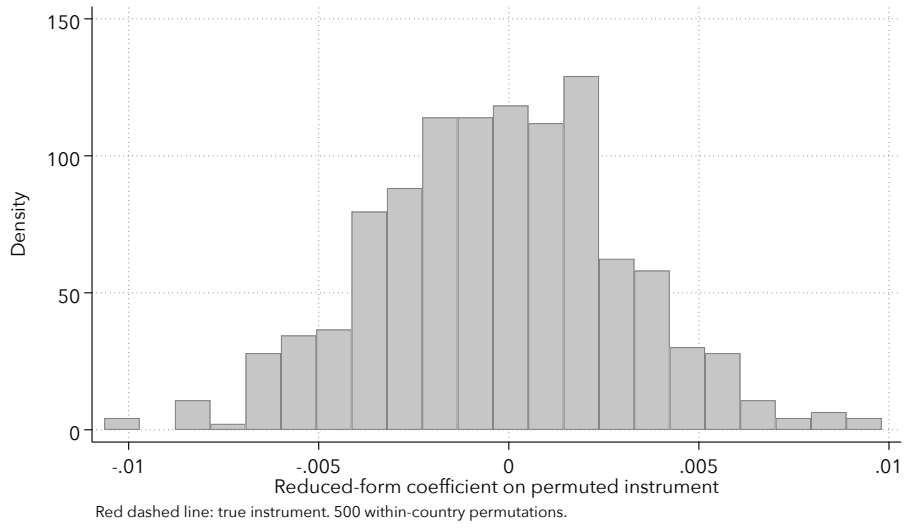
*Notes*: Conley, Hansen, and Rossi (2012) plausibly-exogenous inference. Column "2SLS" is the baseline estimate. Column "Conley UCI" is the 95% union-of-confidence-intervals set obtained by allowing each instrument a direct effect on yield anywhere in $[-\frac{1}{2}, \frac{1}{2}]$ times its reduced-form association with yield and taking the union of the 2SLS confidence intervals over that box. Estimated on the headline sample; standard errors clustered on a 2.5° spatial grid.

**Table D3:** Randomization-inference Placebo on the Diversity Instrument

| | Reduced form coefficient on potential effective diversity |
|---|---|
| Original instrument | 0.336 |
| | ( 0.038) |
| Permuted instruments (mean) | -0.000 |
| 5th–95th percentile | [-0.006, 0.005] |
| Permutation $p$-value | 0.000 |
| Permutations | 500 |
| Observations | 586,355 |

*Notes*: Reduced-form coefficient of log yield on potential effective diversity, partialling out the remaining instruments, controls and country×crop fixed effects, with standard errors clustered on a 2.5° spatial grid. The "original instrument" is the reduced-form coefficient estimated using the actual county-level potential diversity measure. The "permuted instruments" rows summarise the placebo distribution obtained by reassigning each county the potential diversity of a randomly drawn county in the same country (a within country×crop permutation) and re-estimating the reduced form; the permutation $p$-value is the share of permutations whose coefficient is at least as large in absolute value as the original one.

**Figure D5:** Randomization-inference Placebo Distribution




*Notes*: Distribution of the reduced-form coefficient of log yield on potential effective diversity across 500 random reassignments of the instrument across counties within country×crop cells (a placebo permutation). The coefficient obtained with the original instrument has a value of 0.336. The permuted coefficients cluster tightly at zero (5th-95th-percentile band $[-0.006, 0.005]$) and none reaches this value, so the instrument's reduced-form predictive power is not a spurious or mechanical artifact.

# E Heterogeneity

**Table E1:** Heterogeneity of the Baseline Estimates across Regions

| Dependent Variable | Log Actual Yield | | | | | | |
|---|---|---|---|---|---|---|---|
| | EAP | ECA | LAC | MNA | NAC | SAS | SSA |
| Log effective diversity | 0.31 | 0.13 | 4.59** | 1.32 | -1.34** | -0.86 | 2.19** |
| | (0.47) | (0.68) | (2.13) | (1.32) | (0.61) | (0.72) | (1.08) |
| Farmland share | -6.14*** | -46.33*** | -7.17* | 2.91 | -7.23** | -0.96 | -5.55*** |
| | (1.88) | (8.03) | (3.73) | (3.97) | (2.85) | (1.94) | (2.10) |
| Kleibergen-Paap rk Wald $F$ | 3.7 | 8.3 | 2.0 | 0.3 | 7.3 | 1.4 | 3.1 |
| Anderson-Rubin $p$-value | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 |
| Hansen $J$ $p$-value | 0.023 | 0.000 | 0.096 | 0.065 | 0.002 | 0.536 | 0.064 |
| Observations | 105,008 | 171,903 | 162,957 | 20,146 | 42,721 | 20,774 | 58,604 |

*Notes*: Baseline specification (Table 2, column 4) re-estimated separately within each World Bank region (one column per region): EAP (East Asia & Pacific), ECA (Europe & Central Asia), LAC (Latin America & Caribbean), MNA (Middle East & North Africa), NAC (North America), SAS (South Asia), and SSA (Sub-Saharan Africa). Each column reports the 2SLS effective-diversity and farmland-share coefficients (standard errors clustered on a 2.5° spatial grid in parentheses), the Kleibergen-Paap rk Wald $F$, the Anderson-Rubin $p$-value for the joint null of no effect, the Hansen $J$ over-identification $p$-value, and the number of observations. Because each region offers far less identifying variation than the full sample, the region first stages are weaker and inference should rely on the Anderson-Rubin $p$-value rather than the point estimate. The corresponding first-stage estimates for each region are reported in Online Appendix Table O6. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# F Robustness

**Table F1:** Robustness of the Baseline Estimates to the Clustering Scheme

| | | Grid | | | | |
|---|---|---|---|---|---|---|
| | County | 0.75° | 1° | 2° | 2.5° | Region |
| Log effective diversity | 1.655*** | 1.655*** | 1.655*** | 1.655*** | 1.655*** | 1.655*** |
| | (0.133) | (0.229) | (0.266) | (0.335) | (0.353) | (0.415) |
| Farmland share | -14.787*** | -14.787*** | -14.787*** | -14.787*** | -14.787*** | -14.787*** |
| | (0.732) | (1.108) | (1.259) | (1.653) | (1.891) | (2.130) |
| Kleibergen-Paap rk Wald F | 270.0 | 75.2 | 54.1 | 31.3 | 26.6 | 25.1 |
| SW conditional F (diversity) | 246.6 | 66.0 | 49.3 | 27.3 | 27.3 | 23.8 |
| SW conditional F (share) | 560.4 | 165.4 | 117.1 | 60.7 | 44.9 | 39.7 |
| Anderson-Rubin p-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Hansen J p-value | 0.004 | 0.049 | 0.078 | 0.165 | 0.198 | 0.295 |
| Country × crop FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 586,355 | 586,355 | 586,355 | 586,355 | 586,355 | 586,355 |

*Notes*: Baseline specification (Table 2, column 4) re-estimated under progressively coarser clustering scheme: county (admin-2); 0.75°, 1°, 2° and 2.5° spatial grids; and region (admin-1). The 2SLS point estimates are identical across columns; only the standard errors (in parentheses) change with the scheme. For each scheme we report the Kleibergen-Paap rk Wald $F$, the Sanderson-Windmeijer conditional $F$ for each endogenous regressor, the Anderson-Rubin $p$-value, and the Hansen $J$ over-identification $p$-value. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# Online Appendix

This online appendix reports the instrumental-variables estimates of the effective-diversity and farmland-share coefficients, crop by crop, under each of the three candidate instrument sets for the farmland share.

**Table O1:** Crop-specific IV Estimates by Instrument Set

| | Effective diversity | | | Farmland share | | |
|---|---|---|---|---|---|---|
| Crop | Pot. | Suit. | Both | Pot. | Suit. | Both |
| Alfalfa | 2.08* | 2.68* | 2.36** | -3.01 | -0.84 | -0.90 |
| | (1.13) | (1.41) | (1.20) | (2.66) | (1.04) | (0.94) |
| Banana | 0.38 | -0.33 | -0.16 | -0.02 | -0.89 | -0.09 |
| | (2.15) | (1.48) | (1.20) | (2.41) | (4.60) | (2.69) |
| Barley | 1.12 | 0.44 | 0.51 | -11.22* | -13.40** | -13.60** |
| | (1.29) | (0.70) | (0.66) | (5.97) | (5.45) | (5.63) |
| Buckwheat | 0.48 | -0.56 | -0.55 | 67.45 | 77.14*** | 77.67*** |
| | (3.43) | (0.67) | (0.67) | (41.11) | (25.47) | (25.67) |
| Cabbage | -7.90 | 0.96*** | 0.89*** | -229.44 | -38.45** | -36.51** |
| | (16.89) | (0.35) | (0.34) | (363.82) | (15.62) | (15.12) |
| Carrot | 0.87 | 0.85* | 0.86 | -150.54** | -151.13* | -150.51** |
| | (1.30) | (0.52) | (0.54) | (59.90) | (79.99) | (59.30) |
| Cassava | -0.36 | -2.67 | -0.51 | -6.74*** | -6.50 | -6.82*** |
| | (2.93) | (20.13) | (2.94) | (2.35) | (4.30) | (2.41) |
| Chickpea | 1.75 | 1.80 | 1.77 | 60.95 | 69.51 | 67.69 |
| | (1.48) | (1.89) | (1.65) | (107.03) | (79.36) | (69.33) |
| Cocoa | -2.09 | -4.54 | -2.73 | 4.62 | -25.29 | 1.41 |
| | (1.62) | (6.91) | (1.99) | (6.49) | (72.62) | (6.68) |
| Coconut | -2.88 | -0.62 | -0.98 | 3.81 | -3.41 | 0.35 |
| | (2.97) | (0.93) | (1.05) | (7.67) | (4.57) | (3.57) |
| Coffee | 9.01 | -3.96 | 4.71 | -6.58 | 10.13 | -3.70 |
| | (13.95) | (10.52) | (6.83) | (14.41) | (14.78) | (7.42) |
| Cotton | -26.34 | -0.24 | -1.97 | 216.21 | 78.99 | 36.23 |
| | (91.64) | (13.20) | (5.81) | (751.22) | (148.54) | (65.40) |
| Cow pea | 1.66 | 0.17 | 0.77 | 3.13 | 3.42 | 2.32 |
| | (2.31) | (2.13) | (1.26) | (4.97) | (4.36) | (3.87) |
| Dry pea | 3.05*** | 3.36 | 1.96*** | -24.89 | -195.91 | -43.83 |
| | (1.18) | (3.28) | (0.55) | (36.21) | (308.03) | (32.36) |
| Flax | -3.08 | 0.81 | 0.57 | -22.95 | -23.71* | -25.98* |
| | (2.40) | (0.76) | (0.69) | (15.87) | (13.06) | (13.51) |
| Groundnut | -2.70* | -2.63 | -2.85* | 8.29 | 17.08 | 12.71 |
| | (1.50) | (1.60) | (1.61) | (13.36) | (18.93) | (13.28) |
| Maize | -20.19 | 34.29 | 9.05 | -10.77 | -29.86 | -11.12* |
| | (57.98) | (67.18) | (6.96) | (20.13) | (52.21) | (5.71) |
| Millet | -36.36 | -34.37 | -36.38 | 10.30 | 10.39 | 10.38 |
| | (78.95) | (517.81) | (77.91) | (60.64) | (52.08) | (55.38) |
| Oat | 3.44*** | 1.50** | 1.67*** | 5.22 | -29.14* | -2.48 |
| | (1.01) | (0.75) | (0.32) | (7.70) | (16.98) | (5.54) |
| Oil palm | -1.57 | -0.83 | -1.31 | -0.10 | -6.13 | 0.86 |

*Continued on next page*

Table O1 – *continued*

| | Effective diversity | | | Farmland share | | |
|---|---|---|---|---|---|---|
| Crop | Pot. | Suit. | Both | Pot. | Suit. | Both |
| | (1.32) | (1.97) | (0.99) | (4.08) | (23.81) | (2.85) |
| Olive | 1.19 | 1.84 | 1.43 | 3.80 | 3.09 | 3.48 |
| | (3.93) | (7.44) | (3.06) | (4.58) | (8.27) | (3.96) |
| Onion | -0.54 | 0.46 | 0.27 | 12.59 | 7.55 | 4.06 |
| | (0.92) | (0.50) | (0.46) | (16.81) | (12.52) | (11.93) |
| Phaseolus bean | 0.04 | 1.06 | 0.08 | -36.60 | -6.96 | -29.89* |
| | (0.52) | (2.88) | (0.48) | (23.62) | (70.87) | (16.02) |
| Pigeon pea | 0.54 | -5.90 | -4.27 | 38.89 | -70.86 | -76.99 |
| | (7.04) | (9.78) | (5.19) | (188.80) | (162.51) | (136.21) |
| Rapeseed | 25.13 | 0.98 | 0.94 | -1163.70 | 76.14 | 56.39 |
| | (370.07) | (0.98) | (0.73) | (18144.24) | (61.85) | (39.99) |
| Rice | -38.94 | -1.97 | -1.76 | 40.86 | 9.83* | 9.70* |
| | (247.97) | (1.70) | (1.63) | (217.14) | (5.27) | (5.10) |
| Rubber | 18.83 | 2.01 | 0.42 | 89.85 | -23.41 | 3.84 |
| | (122.33) | (1.38) | (0.33) | (590.54) | (28.74) | (3.60) |
| Rye | 13.94 | -36.30 | -7.27 | -23.90 | 428.49 | 99.62 |
| | (29.72) | (186.70) | (8.13) | (190.66) | (2122.43) | (70.34) |
| Sorghum | -0.88 | -1.05 | -0.98* | 9.88** | 9.10*** | 9.22*** |
| | (0.70) | (0.66) | (0.50) | (4.40) | (3.08) | (2.90) |
| Soybean | -5.02 | -5.13 | -3.36 | -10.26 | -2.80 | -2.28 |
| | (7.91) | (5.75) | (3.07) | (19.25) | (3.08) | (2.21) |
| Sugar beet | -134.54 | -280.41 | -5.32 | 132.40 | 1525.72 | 26.28 |
| | (2400.77) | (10274.60) | (4.70) | (2106.78) | (55896.63) | (20.47) |
| Sugar cane | -2.58 | -2.00 | -4.17 | 3.52 | 4.34 | 7.43 |
| | (4.24) | (4.12) | (5.16) | (7.54) | (5.91) | (7.37) |
| Sunflower | 4.42 | -0.04 | 0.33 | -6.30 | 1.93* | 2.06* |
| | (4.09) | (0.50) | (0.48) | (8.82) | (1.08) | (1.11) |
| Sweet potato | -2.67 | -0.42 | -1.43* | 65.34 | 335.50 | 59.98 |
| | (1.77) | (5.16) | (0.85) | (76.92) | (1167.92) | (62.35) |
| Tea | 2.96 | -1.29 | -4.53 | 48.70 | -70.73 | -76.88 |
| | (19.89) | (15.12) | (10.17) | (334.79) | (138.86) | (148.06) |
| Tobacco | 11.59 | -10.33 | -0.69 | 326.09 | -30.81 | 3.46 |
| | (44.50) | (17.80) | (1.11) | (1392.51) | (74.56) | (8.84) |
| Tomato | -2.99 | 4.75 | 5.28 | 14.53 | 139.28 | 156.12 |
| | (13.52) | (11.35) | (12.00) | (270.02) | (349.04) | (373.52) |
| Wheat | 42.26 | 4.85*** | 3.83*** | -71.81 | -0.21 | -0.55 |
| | (218.70) | (1.58) | (1.16) | (404.55) | (1.50) | (1.18) |
| White potato | -10.80 | 1.53*** | 1.26*** | -16.82* | -2.02** | -2.93*** |
| | (8.79) | (0.43) | (0.42) | (10.14) | (0.79) | (0.74) |
| Yam | 20.57 | -0.92 | -0.28 | -15.19 | -97.72 | 1.83 |
| | (145.72) | (2.27) | (0.26) | (126.09) | (133.13) | (4.57) |

*Notes*: Each cell is a separate IV regression for one crop, instrumenting the leave-one-out Shannon index by its potential counterpart throughout and the farmland share by the potential share (Pot.), the suitability share (Suit.), or both. All regressions include country fixed effects; standard errors clustered on a 2.5° spatial grid in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

### Baseline estimates with all coefficients

Table O2 reproduces the baseline estimates of Table 2, reporting every coefficient, including the full set of controls.

**Table O2:** Baseline Estimates with All Coefficients (Table 2)

| Dependent Variable | Log Actual Yield | | | | |
|---|---|---|---|---|---|
| | OLS | | IV | | |
| | (1) | (2) | (3) | (4) | (5) |
| Log effective diversity | 0.105*** | 0.137*** | 1.300*** | 1.655*** | 1.734*** |
| | (0.02) | (0.01) | (0.26) | (0.35) | (0.35) |
| Farmland share | 0.376*** | 0.347*** | -9.529*** | -14.787*** | -13.837*** |
| | (0.04) | (0.04) | (1.54) | (1.89) | (1.69) |
| Log attainable yield | 0.099*** | 0.080*** | 0.112*** | 0.116*** | 0.116*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Log county area | | -0.142*** | | -0.176*** | -0.176*** |
| | | (0.01) | | (0.01) | (0.01) |
| Growing-period length | | 0.000 | | -0.001*** | -0.001*** |
| | | (0.00) | | (0.00) | (0.00) |
| Median altitude | | 0.000*** | | 0.000 | 0.000 |
| | | (0.00) | | (0.00) | (0.00) |
| Median slope | | -0.008* | | -0.024** | -0.025** |
| | | (0.00) | | (0.01) | (0.01) |
| Mean temperature | | -0.011*** | | -0.040*** | -0.041*** |
| | | (0.00) | | (0.01) | (0.01) |
| Potential evapotranspiration | | 0.000*** | | 0.001*** | 0.001*** |
| | | (0.00) | | (0.00) | (0.00) |
| Log precipitation | | -0.043* | | 0.074 | 0.088* |
| | | (0.02) | | (0.05) | (0.05) |
| Soil root conditions | | 0.018* | | -0.033* | -0.032* |
| | | (0.01) | | (0.02) | (0.02) |
| Soil workability | | -0.013 | | 0.021 | 0.021 |
| | | (0.01) | | (0.02) | (0.02) |
| Nutrient retention capacity | | 0.043*** | | 0.046*** | 0.046*** |
| | | (0.01) | | (0.01) | (0.01) |
| Log distance to surface water | | 0.022* | | -0.022 | -0.024 |
| | | (0.01) | | (0.03) | (0.03) |
| Estimator | OLS | OLS | 2SLS | 2SLS | CUE |
| Country × crop FE | Yes | Yes | Yes | Yes | Yes |
| Clustering | 2.5° grid | 2.5° grid | 2.5° grid | 2.5° grid | 2.5° grid |
| Anderson-Rubin p-value | | | 0.000 | 0.000 | 0.000 |
| Hansen J p-value | | | 0.011 | 0.198 | 0.198 |
| Observations | 586,355 | 586,355 | 586,355 | 586,355 | 586,355 |

*Notes*: The five columns of Table 2 – (1) OLS without controls, (2) OLS with controls, (3) 2SLS without controls, (4) 2SLS with controls, (5) CUE – report all coefficients, including the controls suppressed in the main table. Effective diversity and farmland share are instrumented by potential effective diversity, the potential crop share, and the suitability share in columns (3)–(5). Column (5) (CUE) is estimated on variables residualized on the country×crop fixed effects. Standard errors are clustered on a 2.5° spatial grid in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

## First-stage Estimates for the Robustness Specifications

The tables below report the full first-stage estimates underlying the robustness specifications. In each table, Panel A is the effective-diversity equation and Panel B the farmland-share equation, regressing each endogenous regressor on the three excluded instruments (potential effective diversity, potential crop share, and the suitability share). Every column reproduces the exact specification (controls, instruments, fixed effects, and estimation sample) of the corresponding second-stage table.

**Table O3:** First Stage: Robustness Specifications of Table 5

| | Baseline | Census | Additional controls | 40% threshold | Low-input GAEZ |
|---|---|---|---|---|---|
| *Panel A. Effective-diversity equation* | | | | | |
| Potential effective diversity | 0.077*** | 0.250*** | 0.092*** | 0.072*** | 0.097*** |
| | (0.027) | (0.080) | (0.026) | (0.027) | (0.028) |
| Potential crop share | -0.015 | 0.091 | 0.023 | -0.021 | -0.111 |
| | (0.126) | (0.335) | (0.123) | (0.127) | (0.161) |
| Suitability share | 0.077*** | -0.068** | 0.078*** | 0.067*** | 0.060*** |
| | (0.013) | (0.034) | (0.013) | (0.015) | (0.016) |
| KP rk Wald F | 26.6 | 5.1 | 26.5 | 29.7 | 15.2 |
| SW conditional F | 27.3 | 8.0 | 32.1 | 17.8 | 18.0 |
| Observations | 586,355 | 43,802 | 586,082 | 586,355 | 584,939 |
| *Panel B. Farmland-share equation* | | | | | |
| Potential effective diversity | -0.011*** | -0.126*** | -0.010*** | -0.011*** | -0.003 |
| | (0.002) | (0.029) | (0.002) | (0.002) | (0.002) |
| Potential crop share | 0.135*** | 1.380*** | 0.136*** | 0.131*** | 0.272*** |
| | (0.021) | (0.288) | (0.021) | (0.021) | (0.033) |
| Suitability share | 0.005*** | -0.020** | 0.006*** | 0.007*** | -0.000 |
| | (0.002) | (0.010) | (0.002) | (0.002) | (0.002) |
| KP rk Wald F | 26.6 | 5.1 | 26.5 | 29.7 | 15.2 |
| SW conditional F | 44.9 | 11.6 | 43.7 | 48.8 | 26.9 |
| Observations | 586,355 | 43,802 | 586,082 | 586,355 | 584,939 |

*Notes*: First-stage estimates for each column of Table 5 (Baseline, Census, Additional controls, 40% threshold, Low-input GAEZ). Panel A: effective-diversity equation; Panel B: farmland-share equation. Each column includes country×crop fixed effects and that column's controls; the reported Kleibergen-Paap (KP) rk Wald $F$ and Sanderson-Windmeijer (SW) conditional $F$ match the corresponding second stage. Standard errors clustered on a 2.5° spatial grid in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table O4:** First Stage: Heterogeneity by Crop Type (Table 3)

| | Cereals | Legumes | Other annuals | Perennial |
|---|---|---|---|---|
| *Panel A. Effective-diversity equation* | | | | |
| Potential effective diversity | 0.085*** | 0.087** | 0.068** | 0.041 |
| | (0.024) | (0.034) | (0.032) | (0.047) |
| Potential crop share | 0.150 | -0.354 | -0.134 | -0.318 |
| | (0.117) | (0.382) | (0.174) | (0.272) |
| Suitability share | 0.099*** | 0.034 | 0.083*** | 0.043* |
| | (0.014) | (0.022) | (0.016) | (0.025) |
| Kleibergen-Paap rk Wald F | 6.7 | 2.3 | 12.5 | 1.4 |
| Sanderson-Windmeijer conditional F | 25.4 | 3.4 | 17.4 | 2.0 |
| Observations | 166,972 | 84,289 | 259,142 | 75,952 |
| *Panel B. Farmland-share equation* | | | | |
| Potential effective diversity | -0.001 | -0.028*** | -0.017*** | 0.008 |
| | (0.006) | (0.006) | (0.002) | (0.010) |
| Potential crop share | 0.111*** | -0.123** | 0.166*** | 0.083 |
| | (0.034) | (0.059) | (0.031) | (0.070) |
| Suitability share | 0.007* | 0.007 | -0.001 | 0.023* |
| | (0.004) | (0.005) | (0.001) | (0.013) |
| Kleibergen-Paap rk Wald F | 6.7 | 2.3 | 12.5 | 1.4 |
| Sanderson-Windmeijer conditional F | 10.1 | 5.7 | 33.2 | 2.9 |
| Observations | 166,972 | 84,289 | 259,142 | 75,952 |

*Notes*: First-stage estimates for each crop group of Table 3. Panel A: effective-diversity equation; Panel B: farmland-share equation. Each group is estimated on its crops with country×crop fixed effects, the baseline controls, and the same instruments; standard errors clustered on a 2.5° spatial grid. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table O5:** First Stage: Heterogeneity by Crop Pest Pressure (Table 4)

| | High pest | Low pest |
|---|---|---|
| *Panel A. Effective-diversity equation* | | |
| Potential effective diversity | 0.065** | 0.095*** |
| | (0.030) | (0.027) |
| Potential crop share | 0.110 | -0.061 |
| | (0.147) | (0.124) |
| Suitability share | 0.042*** | 0.122*** |
| | (0.013) | (0.016) |
| Kleibergen-Paap rk Wald F | 9.5 | 14.8 |
| SW conditional F | 9.3 | 36.4 |
| Observations | 339,958 | 246,397 |
| *Panel B. Farmland-share equation* | | |
| Potential effective diversity | -0.011*** | -0.013*** |
| | (0.003) | (0.002) |
| Potential crop share | 0.172*** | 0.078*** |
| | (0.033) | (0.024) |
| Suitability share | 0.007*** | 0.001 |
| | (0.003) | (0.001) |
| Kleibergen-Paap rk Wald F | 9.5 | 14.8 |
| Sanderson-Windmeijer conditional F | 25.5 | 23.4 |
| Observations | 339,958 | 246,397 |

*Notes*: First-stage estimates for the high- and low-pest subsamples of Table 4. Panel A: effective-diversity equation; Panel B: farmland-share equation. Specification as in Table O4; standard errors clustered on a 2.5° spatial grid. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table O6:** First Stage: Heterogeneity across Regions (Table E1)

| | EAP | ECA | LAC | MNA | NAC | SAS | SSA |
|---|---|---|---|---|---|---|---|
| *Panel A. Effective-diversity equation* | | | | | | | |
| Potential effective diversity | -0.171** | 0.051 | 0.088* | 0.049 | 0.207** | -0.024 | 0.109** |
| | (0.079) | (0.048) | (0.045) | (0.120) | (0.096) | (0.070) | (0.048) |
| Potential crop share | -0.479 | 0.248 | 0.332 | -0.051 | -0.761* | 0.235 | 0.212 |
| | (0.316) | (0.175) | (0.228) | (0.290) | (0.458) | (0.280) | (0.192) |
| Suitability share | 0.041* | 0.090*** | 0.005 | 0.220*** | -0.027 | -0.099** | 0.043* |
| | (0.022) | (0.019) | (0.026) | (0.080) | (0.040) | (0.042) | (0.023) |
| KP rk Wald F | 3.7 | 8.3 | 2.0 | 0.3 | 7.3 | 1.4 | 3.1 |
| SW conditional F | 3.8 | 14.2 | 3.1 | 0.5 | 5.0 | 1.9 | 5.1 |
| Observations | 105,008 | 171,903 | 162,957 | 20,146 | 42,721 | 20,774 | 58,604 |
| *Panel B. Farmland-share equation* | | | | | | | |
| Potential effective diversity | -0.010*** | -0.004* | -0.019*** | -0.002 | -0.011 | -0.002 | 0.006 |
| | (0.003) | (0.002) | (0.004) | (0.005) | (0.009) | (0.005) | (0.004) |
| Potential crop share | 0.113*** | 0.094*** | 0.099** | -0.074 | 0.351*** | 0.166*** | 0.278*** |
| | (0.032) | (0.026) | (0.041) | (0.070) | (0.102) | (0.059) | (0.062) |
| Suitability share | 0.012*** | 0.000 | 0.018*** | -0.064*** | -0.020*** | 0.024*** | 0.019*** |
| | (0.003) | (0.001) | (0.004) | (0.009) | (0.006) | (0.007) | (0.004) |
| KP rk Wald F | 3.7 | 8.3 | 2.0 | 0.3 | 7.3 | 1.4 | 3.1 |
| SW conditional F | 11.2 | 12.3 | 8.1 | 1.0 | 14.6 | 4.6 | 11.9 |
| Observations | 105,008 | 171,903 | 162,957 | 20,146 | 42,721 | 20,774 | 58,604 |

*Notes*: First-stage estimates for each World Bank region of Table E1. Panel A: effective-diversity equation; Panel B: farmland-share equation. Each region is estimated with country×crop fixed effects, the baseline controls, and the same instruments; standard errors clustered on a 2.5° spatial grid. Splitting the sample sharply reduces within-region identifying variation, so several region first stages are weak. We also report Kleibergen-Paap (KP) rk Wald $F$ and Sanderson-Windmeijer (SW) conditional $F$. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.