Journal of Cleaner Production 415 (2023) 137847

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

The productivity-environment nexus in space. Granularity bias, aggregation issues and spatial dependence within Italian farm-level data

Edoardo Baldoni [a], Silvia Coderoni [b,*], Roberto Esposti [c]

[a] European Commission, Joint Research Centre, Seville, Spain
[b] Department of Bioscience and Agro-Food and Environmental Technology, University of Teramo, Italy
[c] Department of Economics and Social Sciences, Università Politecnica delle Marche, Ancona, Italy

ARTICLE INFO

Handling Editor: Zhifu Mi

JEL classification:
O13
Q12
D24

Keywords:
Farm-level data
Granularity and aggregation
Joint production and total factor productivity
GHG emissions
Crop diversity
Dynamic spatial panel models

ABSTRACT

This paper looks for empirical support to the existence of a positive nexus between economic and environmental performance in farming as implied by the Sustainable Intensification hypothesis. As the ecological scale at which this nexus actually occurs is unobservable, the paper juxtaposes its estimation at three spatial scales, the farm level and two regional levels. Starting with a common theoretical background, the paper estimates a dynamic spatial panel model on these alternative scales. Identification issues (granularity, aggregation bias and spatial dependence) may generate significantly discrepant estimates eventually questioning the reliability of these results. The empirical study investigates the relationship between total factor productivity, greenhouse gas emissions and crop diversity using a 2008–2018 panel of Italian farms. Results show that the productivity-environment nexus changes and may even revert its sign when passing from farm-level to aggregate data. The implications of these results for evidence-based policy making are discussed.

1. Introduction

Ensuring sustainable food production is the heart of the "Farm to Fork Strategy" of the European Union (EU). This strategy recognizes the importance of achieving greater sustainability standards while ensuring higher returns to farmers, by creating added value and reducing costs (European Commission, 2020a). Eventually, this strategy can be intended as the EU variant of the so-called Sustainable Intensification (SI) of agriculture, that is, the production of more food with fewer resources and lower emissions (Godfray et al., 2010). At the farm level, SI can be considered a win-win farm management strategy that assists the balance between environmental sustainability and resource productivity (Firbank et al., 2013; Gadanakis et al., 2015; Yu and Wu, 2018). The SI hypothesis thus implies a positive nexus between economic (i.e, productive) and environmental performance in agriculture (Omer et al., 2010; Le Mouël et al., 2018).

The present paper deals with the empirical support for this hypothesis and focuses on the suitable spatial scale of analysis in this respect. Empirical literature on the existence and the form of this nexus is relatively abundant (Di Falco and Perrings, 2003; Di Falco and Chavas, 2008; Omer et al., 2007, 2010), but it provides contrasting evidence (Koiry and Huang, 2023; Sidhoum et al., 2023) and, above all, it substantially disregards the issue of the appropriate scale of analysis informing the policy decision making. This can be surprising as a significant part of the wide and multidisciplinary literature on the environmental implications of farming, and on the associated policy measures, stresses its site specificity and, therefore, the relevance of the spatial scale of analysis (Leip et al., 2008; Gocht and Röder, 2014; Clough et al., 2020). In fact, data concerning the production decisions that eventually induce the productivity-environment nexus are typically collected at the farm scale and can only be aggregated further to investigate the nexus at coarser regional scales.

The main contribution of the present paper is twofold. First, it aims to develop a conceptual framework deriving the productivity-environment nexus within a coherent farm level production decision making but also explicating the possible sources of discrepancy across the "ideal" scale of analysis and the available levels of observation. These sources originate from the spatial dimension of data: space

* Corresponding author. Department of Bioscience and Agro-Food and Environmental Technology, University of Teramo, Via Renato Balzarini, 1, 64100, Teramo, Italy.
E-mail address: scoderoni@unite.it (S. Coderoni).

https://doi.org/10.1016/j.jclepro.2023.137847
Received 3 January 2023; Received in revised form 30 May 2023; Accepted 17 June 2023
Available online 19 June 2023

matters and, therefore, it must be properly considered in the empirical assessment of the productivity-environment nexus. In particular, three aspects will be highlighted: the granularity bias, the aggregation issues and the spatial dependence.

Second, the paper derives a spatially explicit empirical specification of this theoretical framework and applies it to real data in order to assess if and how largely this discrepancy occurs. The lack of robustness on the existence of the productivity-environment nexus, in turn, may seriously question viability and consistency of an evidence-based policy making in this field.

The empirical application concerns the productivity-environment nexus within a 2008–2018 panel of Italian farms. Italian agriculture is usually considered an interesting case study for the wide heterogeneity of its farming conditions also with specific reference to their environmental implications (Coderoni and Esposti, 2018). At this farm scale identification issues due to granularity may arise. They can be either mitigated or exacerbated by aggregating these data a higher-level scale of analysis, that of 90 NUTS3 or 21 NUTS2 Italian regions over the same time period.

The productivity indicator here used is the Total Factor Productivity (TFP). Productivity performance is universally measured as TFP growth (OECD, 2001a). However, traditional "Unadjusted" TFP (UTFP) measures do not account for non-marketable outputs, like environmental "goods" (positive externalities) and "bads" (negative externalities) produced by agricultural activities. Nonetheless, some of these "goods" or "bads" can be accurately measured by appropriate environmental indicators to accompany the UTFP thus allowing the investigation of their reciprocal interdependence.

Among these good and bad outputs associated with agricultural activity, here we focus on the provision of ecosystem services and the emission of greenhouse gases (GHG) whose reduction is a primary EU climate objective (European Commission, 2020b). A suitable indicator for the provision of ecosystem services by agricultural activities can be given by crop diversification at the farm level, which has shown to have positive effects on soil fertility, nutrient cycling, carbon sequestration, water regulation, pest control, and small positive effects on biodiversity (Tamburini et al., 2020). These two non-marketable goods express the two possibly competing dimensions of the environmental implications of farming and, therefore, of the productivity-environment nexus (Le Mouël et al., 2018).

The rest of the paper is structured as follows. Section 2 overviews the recent literature on the productivity-environment nexus with specific reference to agriculture and the role of the spatial scale of analysis. The contribution the present paper aims to provide to this literature is highlighted. Section 3 develops the theoretical background and the consequent empirical model. Section 4 details how space, i.e. the scale of analysis, may interfere with this conceptual modelling, discusses the three aforementioned potential sources of bias and derives the dynamic spatial panel model that is eventually estimated. Section 5 presents the adopted datasets, the productivity measure, and the environmental indicators. Econometric implications are also discussed. Section 6 illustrates the estimation results and, in particular, discusses the discrepancy emerging between the three scales of analysis. On the basis of this, Section 7 concludes by drawing some implications for policy making.

2. Contribution to the literature in the field

The co-existence of higher productivity and sustainability in agriculture represents a global policy objective (Bureau and Antón, 2022; Stetter et al., 2022). Pursing this objective requires an appropriate empirical support, that is, a reliable and regular assessment of both productivity and environmental progresses in farming and of their independence. Over the last decades, a growing literature has tried to deal with this empirical challenge by incorporating non-marketable environmental outputs into traditional production theory to obtain an estimation of the so-called Environmentally-Adjusted TFP (EATFP) (for a review see Scheel, 2001; Zhou et al., 2008; OECD, 2022). These adjusted performance measures are empirically obtained through modifications of standard parametric and non-parametric productivity and efficiency analysis techniques (Coelli et al., 2007; Sidhoum et al., 2023).

This empirical literature can be roughly classified into three groups depending on how environmental performances are taken into account:[1] *i*) entering environmental bads/goods as additional inputs outputs and then adopting conventional methods of productivity measurement such as index number (or growth accounting) approaches, Data Envelopment Analysis or stochastic functions estimation (Färe et al., 1989; Reinhard et al., 2000); *ii*) the frontier eco-efficiency models, that derive eco-efficiency measures within the production frontier framework as the ratio between the economic output value and an indicator of the environmental pressures generated by the production processes (Picazo-Tadeo et al., 2014); *iii*) the nutrients balance-based models, that rely on the so-called materials balance principle establishing that the total amount of materials must equalize in either desirable or undesirable inputs and outputs (Kuosmanen and Kuosmanen, 2013).

The present analysis can be brought back to the first group of studies, but it does not aim to estimate the EATFP. Though following the same theoretical framework, it rather aims to derive the elasticities of non-marketable good and bad outputs with respect to the conventional TFP calculation. Eventually, the present work focuses on that productivity-environment nexus that is required by a proper EATFP calculation. In this respect, the main methodological contribution of this study with respect to the EATFP literature consists in dealing with the identification issues (i.e., estimation biases) often overlooked in conventional EATFP studies. They all have to do with the role of space and the consequent scale of analysis.

Whether and how productivity and environmental performances affect each other is largely an empirical issue, mostly because this nexus is highly place dependent, thus it is heterogeneous across farms and farm typologies. It follows that the main methodological challenge in carrying out such an investigation concerns how to take this scale and place dependency properly into account. In particular, several studies have shown that the observed relationship between productivity and environmental performance is largely dependent on the aggregation level of the data used (Baldoni et al., 2017, 2018). Moving from micro to macro data, sign and statistical significance of this relationship can change. This should not surprise as the different aggregation levels actually provide different information (Fuglie *et al.,* 2016; Baldoni and Esposti, 2021). One reason for this discrepancy is the so-called aggregation bias that can occur because spatial aggregation (i.e., aggregating farm-level data at some geographical scale) usually affects the measurements (Jansen and Stoorvogel, 1998; Wade et al., 2019) and thus can conceal different micro behaviours in both TFP and environmental performance calculation.

Aggregation seems particularly problematic for environmental performance indicators as they are strongly scale dependent, thus they can interact with productivity differently whether they are measured at the farm or at some aggregation level. In the case of ecosystem services, an example can be water quality. It is best assessed at the catchment level since it is the result of many farm practices. Thus, higher aggregation levels (i.e., NUTS 2 or NUTS3 regions) can mask substantial local variations thus diverting the attention from areas or localities where unsustainable and unresilient agriculture is practiced (Fuglie et al., 2016). Another interesting example is provided by agricultural GHG emissions. At the micro level, farms can be both net emitters (i.e., when emissions are higher than carbon sequestration in soil and biomasses) or net sinks, and net emitters and net sinks can locally co-exist. However, when some regional level is considered, the whole region can become a net emitter

[1] There have also been attempts to smartly combine some of the features of these different approaches. See, for instance, Pethig (2006), Murty et al. (2012), Murty and Russell (2022) and Sidhoum et al. (2023).

or sink, thus missing the co-existence of different micro-level performances.

To avoid the aggregation bias, recent literature has strongly suggested to focus on farm-level analysis (Kimura and Sauer, 2015; Serra et al., 2014; Gadanakis et al., 2015; Sheng et al., 2015; Koiry and Huang, 2023; Sidhoum et al., 2023) as microdata allow identifying the heterogeneous productivity-environment nexus across farms and, consequently, across space (Cui et al., 2016). However, though with less emphasis, recent literature in the field evokes that also the use of spatially explicit micro-data may raise a major issue due to the presence of place effects. The latter may induce a bias in the estimation of the productivity-environment nexus whenever data are surveyed, or, in general, not randomly distributed across space (Dingel and Tintelnot, 2021; Schoefer and Ziv, 2021). Moreover, when these micro data are used the presence of spatial dependence cannot be ruled and, in any case, it occurs differently compared to the macro-level data (Baldoni and Esposti, 2021).

Although all these empirical issues (see Section 4 for further details) attracted much attention in recent literature, the problem of the scale of analysis in assessing the productivity-environment nexus has never been explicitly put in the forefront of the investigation. This paper aims to fill this gap by estimating the same dynamic specification of the productivity-environment nexus on both micro (i.e., farm-level) and macro data (i.e., some geographical aggregation) thus making explicit the actual relevance of these biases, and the possible trade-off between them at different scales.

3. The theoretical framework

A nexus between agricultural productivity and environmental performance occurs because, in farming, any production decision directly or indirectly induces an environmental consequence. The scale at which this consequence takes place is here called the Ecological Scale (ES). Ideally, at this scale the “real” productivity-environment nexus should be investigated (Chakir, 2009; Chaudhary et al., 2016; Gerber et al., 2016). In fact, the underlying production decisions are taken, and possibly observed, at a different scale, what we call here the Behavioural Scale (BS). With a very good approximation, the farm level evidently corresponds with this latter scale. ES and BS do not correspond mainly because the ES may vary depending on the specific environmental aspect under consideration and on the specific local context. In most cases the ES is finer than the BS as it mostly coincides with the field, the crop or the herd level. In other cases, however, it can be coarser since it emerges as a combination of multiple farm choices (sometime, this is called the *landscape scale*) (Herzog et al., 2006).

Due to this case-by-case varying scale of occurrence, the ES is not normally observed while at the BS observations are usually and systematically available. Nonetheless, discrepancy between these two levels (or *scale discrepancy*) may create problems in the empirical identification and estimation of this nexus. One main reason is that within and among farms, this nexus can show significant diversity and heterogeneity that the analyst can hardly control for. This heterogeneity may mitigate, or even vanish, whenever farms are aggregated at a higher level (here called the Regional Scale, RS), that has also the advantage to take the different farm size into account, thus giving more importance to units that are actually more relevant for the overall productivity-environment nexus.

Since the origin of the nexus is behavioural, it remains true that the conceptual framework on which to ground its investigation must concern the farmers' decision making. The objective here is to connect this theoretical modelling with the empirical investigation at the core of the present study. Fig. 1 diagrammatically illustrates the conceptual linkages occurring among the three abovementioned different scales and the identification issues (or biases) emerging when passing from one level to another. For a full understanding of Fig. 1, the reader can refer to Section 4 where the discrepancies and pros and cons of the different scales of analysis are scrutinized.

In order to derive the decision-making theoretical framework at the BS (farm level), consider a panel of N production units observed over T periods. Represent the unit-specific production technology with the production set Y_i, $\{Y_i = y_i \in \mathbb{R}^L : F_i(y) \leq 0\}$, where: $y_i = (y_{i1}, y_{i2}, \ldots, y_{il}, \ldots, y_{iL}) \in \mathbb{R}^L$ indicates a generic production plan with L goods representing either inputs or outputs. $F_i(y)$ is the *transformation function*. Any y_i such that $F_i(y) = 0$ lies on the boundary of the transformation function, and it is usually designed as *transformation frontier*.

As most of the empirical literature in the field (OECD, 2001b; Brandt et al., 2013, 2014; Cárdenas Rodríguez et al., 2018a,b), we consider the transformation function an appropriate conceptual tool for the present analysis. Since to F_i the usual assumptions and restrictions on non-joint production technologies (including free disposability) do not apply, it seems suitable to accommodate all the possible different forms of production jointness that can be encountered in agriculture, especially when dealing with a large sample of heterogenous farms. The not exhaustive overview of possible forms of jointness reported by OECD (2001b, pp. 124–131) clearly demonstrates that any more specific definition of the production set would exclude some of the possible forms of jointness possibly occurring in agriculture. Such more restrictive (usually parametric) specifications would be needed when the objective of the analysis is the identification and estimation (either parametric or non-parametric) of the production technology and of all the underlying relations across inputs and outputs (Murty, 2015). But the focus here is not on technology estimation, but only on deriving a theoretically consistent and empirically tractable productivity-environmental relationship. Therefore, the implicit nature of $F_i(y)$ minimizes the restriction to be imposed on the relationship between inputs and outputs and makes it general and generic enough to admit a wide range of production jointness.

The fact that we do not impose restrictions on the technology does not mean that for some of the non-market (i.e., environmental) outputs some technological relationships can't be expected ex ante. This is evidently the case of pollution phenomena that usually depend on the use of some specific inputs to which they are linked by a well-defined materials balance, like the GHG emissions here considered. A relevant part of GHG emissions (the part concerning CO_2 emissions) can be evidently linked to the use of fossil fuels and energy as production inputs. However, this representation wouldn't be helpful for that other part of farm-level GHG emissions (e.g., the part concerning CH_4 emissions). In this latter case, the relevant relationship does not only concern inputs but rather livestock activities and the respective market outputs. More in general, as will be clarified in Section 5 and Appendix 3 (see Table A1), agricultural GHG emissions can hardly be modelled within a pollution-generating technology representation, that is, driven by a single set of pollution-generating inputs, but may rather originate from many different sources, in different forms, under different production choices and types of farming. Moreover, in principle, it would be possible to explicitly consider those farm-level mitigating or abatement technologies eventually affecting this materials-balance linkage between inputs, market and non-market outputs (Pethig, 2006; Murty et al., 2012).[2] But the multiplicity of linkages would make the modelling

[2] We wish to thank an anonymous referee for helpful comments on this issue and for some suggestions on possible directions of future research in this respect.

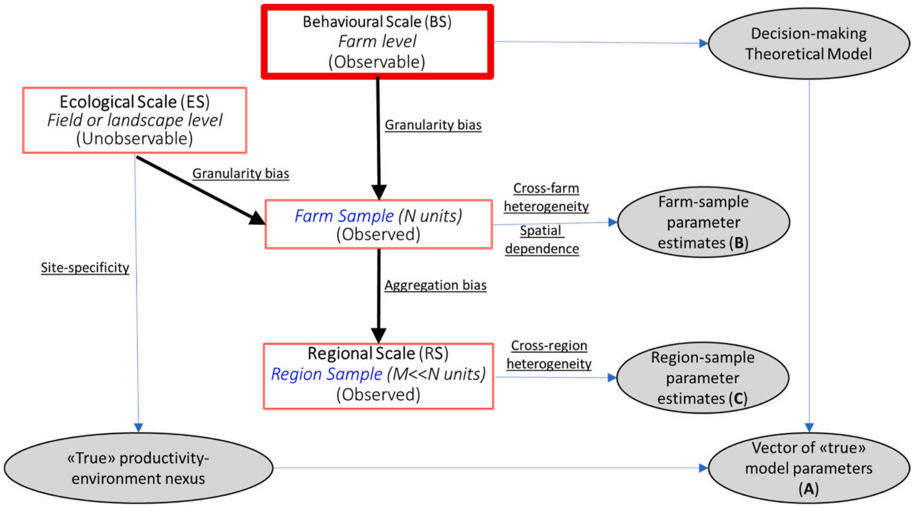

Fig. 1. Diagrammatic conceptualization of the methodological and empirical issues under analysis
Boxes represent the different possible scales of investigation and thick arrows the flow of information or data among them. The BS box is thicker as it represents the scale the theoretical model refers to.
Grey circles express the structural linkages among variables as emerging from either theoretical modelling or empirical observation or estimation. The thin arrows indicate their expected correspondence or concordance. Terms along the arrows indicate the possible identification issues and sources of estimation bias.

also of these abatement technologies quite complex.[3] In addition, within a largely heterogeneous sample, these farm-level linkages may substantially weaken and may even vanish or become unobservable when the analysis is performed at some aggregation level.[4]

In order to preserve this generality, consider a unit-specific transformation function $F_i(Y,G,B,K,L,M,t)$, where Y indicates the aggregate of firm market outputs, G a desirable non-market output (that is, a positive externality like ecosystem services), B an undesirable non-market output (that is, a negative externality like GHG emissions), K the capital input, L the labour input, M all the remaining production factors and t is the usual time trend proxying the unobserved level of the technology. By defining A_t and A_0 the technological level at time t and in the initial period, respectively, λ_i the conventional unit-specific exogenous Hicks-neutral technological change rate, and assuming that all production units are technically efficient, we can write the transformation function as (Cárdenas Rodríguez et al., 2018a, 2018b):

$$A_{it}F_i(Y,G,B,K,L,M) = 1, \text{with } A_{it} = A_0 e^{\lambda_i t} \quad (1)$$

This transformation function can be used to specify the productivity-environment nexus expressed by the elasticity of the market output Y with respect to the two non-marketed outputs G and B. By totally log-differentiating (1) over time (Hulten, 1992; Cárdenas Rodríguez et al., 2018a, 2018b) and rearranging terms[5], we eventually obtain the following relation:[6]

$$y_{it} - [\varepsilon_{Kit}(k_{it}-m_{it}) + \varepsilon_{Lit}(l_{it}-m_{it}) + m_{it}] = \lambda^*_{it} + \varepsilon_{Git}g_{it} + \varepsilon_{Bit}b_{it} \quad (2)$$

where: y_{it}, k_{it}, m_{it}, l_{it}, g_{it}, b_{it} are the growth rates of Y, K, M, L, G and B, respectively; $\varepsilon_X = (\partial lnY_i/\partial lnX_i) = (\partial Y_i/\partial X_i)(X_{it}/Y_{it})$ are the elasticities of Y with respect to the various inputs; ε_{Git} and ε_{Bit} are the elasticities of Y with respect to the two non-marketed outputs; $\lambda^*_{it} = -\lambda_i/\theta_{Yit}$ is the EATFP, that is, the productivity growth that also takes into account the growth of non-marketed outputs, where $\theta_{Yit} = (\partial lnF_i/\partial lnY_i) = (\partial F_i/\partial Y)(Y_{it}/F_{it})$ denotes the elasticity of the transformation function with respect to Y.

As the two non-market outputs are costless or priceless, they are unintended, that is, not considered in the i-th unit maximization choices.[7] Therefore, the difference between UTFP and EATFP in (2) does not rely on the different optimizing behavior of farms but on the joint production of the environmental outputs as implied by the underlying farm-specific technology F_i. According to (2), the observed UTFP growth can be decomposed into the two unit-specific terms on the right-hand side: the idiosyncratic Hicks-neutral exogenous technological change corrected by term θ_{Yit}, λ^*_{it}; the relationship between market-outputs, Y, and the two non-market outputs (G and B), as expressed by term $(+\varepsilon_{Git}g_{it} + \varepsilon_{Bit}b_{it})$. This latter has not to be intended as a causal linkage, that is, as the growth of G and B having an impact on Y. This term rather

[3] Non-farming activities may allow a more explicit modelling of the production jointness underlying environmental performance. The typical case is that of a pollution-generating technology where one (or more) pollution-generating input contributes to both the intended output and to the unintended pollution. In such case, an abatement technology can also be admitted as a sort of an additional output. Murty et al. (2012) present an interesting approach where pollution takes the form of a by-production rather than a joint-production technology and it is more properly modelled as the intersection between two productions sets (or transformation functions), that is, the conventional intended production technology and an unintended-output sub--technology. However, though extendible to other forms of by-production (Murty and Russell, 2022), this interesting approach is distant from several cases encountered in farming, where not only good and bad unintended outputs coexist, but also a unidirectional linkage between a single input and one of these unintended outputs is hardly detectable. Even when the bad output is some form of pollution, in the agricultural case this pollution may originate from quite diverse activities and production choices. Therefore, this by-production approach seems to be suitable in farming only when a very specific and well confined production activity and pollutant are considered. This is the case, for instance, of milk production and nitrogen pollution investigated by Sidhoum et al. (2023).

[4] This incongruity across the different scales can occur in both directions. For instance, a farm highly specialized in permanent crops can provide a net sink of carbon (with Land Use, Land-Use Change and Forestry, LULUCF, sequestration) offsetting its GHG emissions, but if these latter are summed with those of other farms at the NUTS3 or NUTS 2 level, this mitigation effect disappears. Similarly, some mitigation solutions can be captured at the landscape level but not at the farm level. For instance, if livestock farms are organized in networks for biogas recovery, the gas emission balance of a manure delivering farm takes this manure into account but not the respective biogas recovery.

[5] For details on the derivation, refer to Appendix 1 of the Supplementary Material.

[6] As clarified in Appendix 1 of the Supplementary Material, ε_{Git} and ε_{Bit} should enter (2) with opposite signs. However, as the objective here is not to provide a correction of the TFP measure (the EATFP) but only to assess the nexus between conventional productivity and these unintended environmental performances, here ε_{Git} and ε_{Bit} are reported with a positive sign. Eventually, they become conventional regression coefficients with either positive or negative sign to be established by the econometric estimation.

[7] If not the case, Cárdenas Rodríguez et al. (2018b, pp. 159–160) show that the UTFP would require a further adjustment in order to properly adjust the elasticities of market output (Y) with respect to production inputs (K, L and M).

expresses the production jointness between the market outputs Y (on which the profit maximizing behavior of the *i-th* unit applies) and G and B, that is, the generation of the positive and negative externalities (Cárdenas Rodríguez et al., 2018a,b; OECD, 2022).

The abovementioned multiplicity and complexity of the relationships between inputs, market outputs and non-market outputs, combined with the wide cross-farm heterogeneity and the aggregation issues, severely questions the predictability of the sign of elasticities ε_{Git} and ε_{Bit} of interest here, even in the case of a conventional pollution process like GHG emissions. In any case, even though their sign is not easily predictable ex ante, elasticities ε_{Git} and ε_{Bit} still express the production jointness between Y and G and B, respectively. A positive (negative) elasticity indicates that a greater Y level brings about a higher (lower) level of G or B. Therefore, a positive productivity-environment nexus would imply a positive elasticity for G and a negative elasticity for B.

It is worth noticing that elasticities ε_{Git} and ε_{Bit} can be also interpreted as the shadow values of G and B in terms of Y: they correspond to the units of Y associated to a unit increase of these environmental goods. From the derivation above, it follows that $\varepsilon_{Git} = -MRT_{GYit}(G_{it}/Y_{it})$ and $\varepsilon_{Bit} = -MRT_{BYit}(B_{it}/Y_{it})$, where MRT indicates the Marginal Rate of Transformation between market and non-market goods. More explicitly, if multiplied by the market price of Y (p_{Yit}), these elasticities express the additional cost (G) and the additional value (B) associated to the production of Y, i.e. their shadow prices $p_{Git} = -p_{Yit}MRT_{GYit}$ and $p_{Bit} = -p_{Yit}MRT_{BYit}$ (Cárdenas Rodríguez et al., 2018a, p. 70). These shadow prices reflect the cost (benefit) incurred (received) by the farmers because of the externalities associated to the production of Y and, thus, they represent unobserved private costs or private values to be distinguished from social costs and social values.

The left-hand side of (2) can be observed as it corresponds to the conventional UTFP, while the unknown terms on the right-hand side, λ^*_{it}, ε_{Git} and ε_{Bit}, can be econometrically estimated given the observed terms UTFP, g_{it} and b_{it}. Therefore, the empirical strategy here adopted, does not concentrate on computing the EATFP given some estimated values of ε_{Git} and ε_{Bit}, but on estimating ε_{Git} and ε_{Bit} given the observed UTFP. Nonetheless, the econometric identification and estimation of the right-hand side terms of (2) requires appropriate specifications and data availability. First, it requires the measurement of its left-hand term, i.e., the UTFP. As here a farm panel dataset is used, the index numbers approach allows computing the UTFP without the need of recovering the underlying technology under a large and unknown cross-farm heterogeneity (Baldoni et al., 2021). The adopted methodological approach thus consists of two stages. In the first stage, the left-hand side (2), the *UTFP* index, is measured through index number techniques[8]. In the second stage, this *UTFP* is regressed on the right-hand side terms of (2) (Chambers, 1988: 232; O'Donnell, 2016: 328–329).

In this second stage of the analysis all the possible determinants of the observed UTFP, the stochastic terms included, are made explicit. Time and space (i.e., cross-farm) heterogeneity within the panel, as well as the stochastic nature of the i-th unit productivity in the right-hand side of (2), can be concentrated in the idiosyncratic term as $\lambda^*_{it} = \mu_0 + \mu_t + \mu_i + \mu_{it}$, where μ_0 represents the mean level across units and over time; μ_t the *t*-th time specific level common to all units; μ_i the *i*-th unit time-invariant specific level; μ_{it} represents the *i*-th unit time-variant (possibly stochastic) specific level. To allow parameter identification, in this second stage it will also be assumed that $\varepsilon_{Git} = \varepsilon_{Gjs} = \varepsilon_G$ and $\varepsilon_{Bit} = \varepsilon_{Bjs} = \varepsilon_B, \forall i,j \in N, \forall t,s \in T$.[9] It follows that (2) is rewritten as:

$$\ln UTFP_{it} = \mu_0 + \mu_t + \mu_i + \mu_{it} + \varepsilon_G g_{it} + \varepsilon_B b_{it} \tag{3}$$

where $\ln UTFP_{it}$ is the i-th unit t-th time Theil-Tornquist discrete-time approximation of the respective Divisia productivity index.

(3) can be interpreted as a regression equation where μ_{it} is the usual disturbance term (see below), μ_0, μ_t and μ_i constant terms, ε_G and ε_B two regression coefficients whose sign can be either positive or negative. According to what discussed above, however, for a positive productivity-environment nexus to occur we expect ε_G to be positive and ε_B negative. However, interpreting (3) as a regression equation brings about a major econometric implication. It concerns the required assumption of g_{it} and b_{it} as exogenous regressors. In fact, since G and B quantities are still the consequence (though not an object) of an optimizing decision making, g_{it} and b_{it} cannot be considered exogenous regressors as they depend on the left-hand side of the equation. This endogeneity issue must be properly considered in the estimation stage.

Not all unit- and time-specific determinants in (3) are unobserved. To take this into account, Eberhardt and Helmers (2010) propose the following general specification to explicitly distinguish between observed and unobserved productivity determinants:

$$\ln UTFP_{it} = \mu_0 + \mu_t + \mu_i + \varepsilon_G g_{it} + \varepsilon_B b_{it} + \mathbf{Z}_i\Gamma + \mathbf{X}_{it}\Pi + \mu_{it} \tag{4}$$

where $\mathbf{Z}_i$ and $\mathbf{X}_{it}$ are (1xk) and (1xh) vectors of time invariant and time-variant observable productivity determinants, respectively, and Γ and Π are the correspondent (kx1) and (hx1) vectors of unknown parameters to be estimated.

As Griliches and Mairesse (1995) point out, term μ_{it} is unobserved by the analyst but is known by the decision maker. While from the econometrician's perspective it simply represents a typical stochastic error term, for the farmer it is an available information (expressing some time-variant farm characteristic or condition) that affects input use decisions. Therefore, inputs are partially determined by the unobserved time-variable characteristics contained in μ_{it}, so the usual exogeneity assumptions are unlikely to hold. Following Blundell and Bond (2000), Baldoni and Esposti (2021) and Baldoni et al. (2021), a theoretically and econometrically viable solution to this endogeneity problem consists in specifying the dynamic stochastic process generating μ_{it} as follows:

$$\mu_{it} = \rho\mu_{it-1} + u_{it} \tag{5}$$

where u_{it} is an i.i.d. $\sim N(0,\sigma^2)$ error term expressing deviations from the idiosyncratic productivity mean due to measurement errors, unexpected delays or other external circumstances (Van Beveren, 2010). The introduction of the AR(1) term aims to capture the impact of past productivity shocks on current input decisions, so it internalizes that producers might react with delay to changes in productivity (Bond and Söderbom, 2005).

As from (4) it follows that $\mu_{it-1} = lnUTFP_{it-1} - \mu_0 - \mu_i - \mu_{t-1} - \varepsilon_G g_{it-1} - \varepsilon_B b_{it-1} - \mathbf{Z}_i\mathbf{\Gamma} - \mathbf{X}_{it-1}\Pi$, by replacing in (5) and rearranging we obtain:

$$\begin{aligned}\ln UTFP_{it} = {} & \overline{\mu}_0 + \overline{\mu}_t + \overline{\mu}_i + \rho lnUTFP_{it-1} + \varepsilon_G g_{it} + \varepsilon_B b_{it} + \overline{\varepsilon}_G g_{it-1} \\ & + \overline{\varepsilon}_B b_{it-1} + \mathbf{Z}_i\overline{\Gamma} + \mathbf{X}_{it}\Pi + \mathbf{X}_{it-1}\overline{\Pi} + u_{it}\end{aligned} \tag{6}$$

where $\overline{\mu}_0 = (1-\rho)\mu_0$, $\overline{\mu}_t = \mu_t - \rho\mu_{t-1}$, $\overline{\mu}_i = (1-\rho)\mu_i$, $\overline{\varepsilon}_G = -\rho\varepsilon_G$, $\overline{\varepsilon}_B = -\rho\varepsilon_B$, $\overline{\Gamma} = (1-\rho)\Gamma$, $\overline{\Pi} = -\rho\Pi$.[10]

4. The spatial issues

Starting from the sample of N farms, a relationship like (6) can be also estimated by aggregating farms at some geographical level (region) thus obtaining a panel of M regions (with $M \ll N$) observed over T periods. In any case, in both the micro (N x T) and the macro (M x T) spatial

[8] See Appendix 2 and 3 of Supplementary Material.

[9] This assumption also responds to the need of providing an univocal, thus relevant from a policy perspective, estimation of the parameters of interest, ε_G and ε_B.

[10] These restrictions on parameters, especially those associated to terms g_{it} and b_{it} and to terms $\mathbf{X}_{it}$ and $\mathbf{X}_{it-1}$ cannot be imposed ex ante and have to be statistically tested in post estimation.

panels the presence of a place effect is a major but often disregarded issue. This place effect has two major implications in the identification and estimation of the productivity-environment nexus. On the one hand, it may generate biases whose nature and relevance substantially differ between the micro and the macro panel. On the other hand, it may bring about spatial dependence across contiguous units that also may essentially differ between micro and macro-level data.

4.1. The granularity bias

The place effect consists in the fact that, within the adopted modelling approach, the *i-th* unit productivity performance and the productivity-environment nexus itself are idiosyncratic in the sense that they depend on the location of the unit itself. This effect of the location is partially stochastic and is expressed, in (6), by the term $(\overline{\mu}_i + u_{it})$. The idiosyncratic time-invariant deterministic term (fixed effect) $\overline{\mu}_i$ expresses to what extent the place brings about a specific productivity performance. It captures a variety of specific environmental and socio-economic mechanisms and determinants that manifest themselves in the idiosyncratic productivity of any unit within that location. At the same time, stochastic terms u_{it} are simply deviations around the true place effect. u_{it} captures all the place-specific exogenous shocks in the productivity performance, thus deviations in the productivity-environment nexus, as well as unmeasured exogenous shocks (or measurement errors) of terms g_{it} and b_{it}.

Estimating this place effect with the observed spatial units, and not at the unobservable ES at which it actually occurs, may generate a bias that we designate here as *granularity bias* (Schoefer and Ziv, 2021).[11] In fact, the granularity bias is some combination of two distinct possible biases. To appreciate this, it is worth formulating a statistical definition of the true place effects for finite counts of idiosyncratically heterogeneous units (Schoefer and Ziv, 2021). Assume that agricultural production is heterogenous across a set L of locations indexed by $l \in L$. Each location has a count of potential P^l units, $i^l \in P^l$. P^l denotes the location-l infinite superpopulation of units, from which the finite observed population (census data, as in the real economy, or a possibly representative sample) is drawn. Assume that any unit in location l (i^l) is characterised by an idiosyncratic productivity $(\overline{\mu}^l_{it} + u^l_{it})$. Also assume that any u^l_{it} is a time-t draw from a potentially location-specific distribution $u^l_{it} \sim \Xi_l(u)$ with zero expected value, $E[u^l_{it}|l] = 0, \forall t \in T$. This latent data generating process $\Xi_l(u)$ does not describe the given finite population of observed units in location l but the respective unobserved superpopulation. Given this stochastic data generation process across space we can define $E[\overline{\mu}^l_{it} + u^l_{it}|l] = \overline{\mu}^l$ as the *true place effect*, that is, the expected value of idiosyncratic productivity in location l within the respective superpopulation. This *true place effect* is unobservable as it refers to the ES superpopulation which is itself unobservable but can be estimated from the observed units $i \in N^l \in P^l$ here representing the farm level (i.e., the BS).

Given the description above, if the observed units are randomly sampled from P^l, by averaging across units N^l we may obtain an unbiased and consistent estimation of the true place effect $\overline{\mu}^l$. However, for phenomena showing high site-specificity (as, arguably, agricultural production) the place effect is itself farm-specific, but it does not correspond to the farm effect as any farm can contain several place effects. In this case, a random sample of farms (as well as any realized population) cannot be fully representative over the spatial dimension. As a given point in the space can be obviously occupied only by one unit, if this unit is not drawn the respective place cannot be represented. In other words, any observed spatial sample or population inevitably implies spatial gaps and, therefore, brings about a bias in the estimation of the true place effect.[12]

But even if these imperfections were excluded, real data still bring about a second possible source of estimation bias. This latter bias concerns the dispersion (i.e., the variance) of the true place effect. In fact, the true place effect variance can be decomposed as $\mathrm{Var}(\overline{\mu}^l_{it} + u^l_{it}|l) = \mathrm{Var}\left(\overline{\mu}^l + 1/N^l \sum_{i \in N^l} u^l_{it}\right) = \mathrm{Var}(\overline{\mu}^l) + (1/N^l)^2 \sum_{i \in N^l} \sigma^l(u^l_{it})^2$, where $\sigma^l(u^l_{it})^2$ is the variance of the unit-specific deviations u^l_{it} from the true place effect in location l.[13] Therefore, with finite populations or samples within a location, the variance of the place averages is an *upward biased* estimator of the variance of true place effect, thus of the true location-specific *UTFP*. This bias arises under finite observed units N^l combined with large idiosyncratic variance $\sigma^l(u^l_{it})^2$.[14] The major consequence for the present analysis is that also the variance of the estimated parameters in (6) can be itself upward biased and this may substantially jeopardize their statistical significance (Goldberger, 1991: 165–267).

4.2. The aggregation bias

One possible way to escape the granularity bias simply consists in eliminating granularity itself. Granularity disappears when the superpopulation and the observed units, therefore the true place effects and the estimated place effects, correspond. This is what happens when a model like (6) is estimated not on the farm sample but on a regional sample of M units (regions), with $M \ll N$. In such case, units become regions and place effects become region effects. In practice, unlike micro data, in aggregate units it is $i \equiv l, \forall i \in M, \forall l \in L$; thus $M = L$. Moreover, in macro data the observed variance of the disturbance term corresponds to the variance of the true place effects and the idiosyncratic dispersion (σ^l) does not occur (namely, it is $\sigma^l = 0$).

Even though aggregating units to some regional level gets rid of granularity, aggregate spatial analyses are not exempt from statistical problems as they may suffer from a *spatial aggregation bias*. A detailed analysis of this bias is beyond the scope of the present paper (see Anselin, 2002, for an in-depth discussion). Nonetheless, two main aspects of this possible bias are of major interest here. First, an aggregation bias may occur simply because aggregating micro units to some geographical level always re-proportions the variables themselves whenever micro units show heterogenous size. In such a circumstance (which evidently occurs in most real-case analyses), any aggregation of units across space implicitly weights respective units on the base of their size.[15] As anticipated, this kind of implicit weighting may be desirable for a size-dependent environmental performance (i.e., most pollution phenomena), while it may generate a bias for a size-independent environmental performance (as can be the case of biodiversity protection).

The second aspect is that the determinants of the productivity-environment nexus may substantially differ passing from the farm level to a given aggregate geographical scale, even when not particularly large (for instance, the landscape scale; García Cidad et al., 2003: 14; Herzog et al., 2006; Stetter et al., 2022). Therefore, if the micro scale

[11] In general terms, it corresponds to what is also designated, within this literature, as *spatial sampling bias* (Baldoni and Esposti, 2021).

[12] These gaps may reflect underlying economic forces like selection on entry and exit (within the population), as well as practical issues like the sampling design (within the sample) or errors or uncertainties in units' location (Anselin, 2001, 2002; Arbia et al., 2015).

[13] Where, by construction, $\mathrm{Cov}\left(\overline{\mu}^l, \frac{1}{N^l} \sum_{i \in N^l} u^l_{it}\right) = 0$.

[14] The size of the idiosyncratic variance seems particularly critical here as the dispersion of productivity across farms is typically quite high and this is true for the Farm Accountancy Data Network (FADN) sample under investigation (Baldoni and Esposti, 2021).

[15] For instance, if some large or dominant units (i.e., farms) are present, the aggregate analysis tend to mostly capture the relationship among variables occurring in these units.

presents evidence that is closer to the "true" effect to be identified (that of the ES), working at the macro level may produce spurious evidence as some farm-level determinants vanishes over spatial aggregation while other determinants emerge. As emphasized in the recent empirical literature (Baldoni and Esposti, 2021), only using a dense-enough aggregation level the micro-level spatial properties can be preserved and the aggregation bias prevented or minimized.

The combination of granularity bias and spatial aggregation bias may eventually generate a discrepancy in the productivity-environment nexus found across alternative scales of analysis and this raises the question on the most appropriate level of investigation for policy making. Given this possible discrepancy between the farm and the regional levels, it seems worth to perform the empirical analysis at these different scales also to carry out a comparison among the respective results and to assess the robustness of empirical evidence across spatial scales. But there is a third major empirical complication, that of spatial dependence, that suggests repeating the investigation at different spatial scales.

4.3. Spatial dependence

A further implication of place effects concerns the assumption that, in (6), u_{it} is an i.i.d. $\sim N(0,\sigma^2)$ disturbance term. If this term is place-specific, we cannot exclude that these terms are also space-dependent, that is, $E(u_{it}, u_{jt}) \neq 0$ for some contiguous units $i, j \in (1, \ldots, N)$ and $i \neq j$. In order to admit both productivity dynamics and space dependence, we can augment (5) as follows:

$$\mu_{it} = \rho\mu_{it-1} + \delta \mathbf{W}\, \mu_{it} + e_{it} \tag{7}$$

where $\mathbf{W}$ is the NxN (or MxM) spatial weight matrix $\mathbf{W}$, expressing the degree of contiguity of any i-th unit with the surrounding space and δ is the spatial correlation parameter. Spatial lag $\delta\mathbf{W}$ is aimed to capture the spatial dependence of productivity generated by productivity spillovers. Thus, e_{it} returns to be the usual spherical disturbance i.i.d. $\sim N(0,\sigma)$ as the original spatial dependence in u_{it} is now made explicit in term $\delta\mathbf{W}$.

It follows that (6) can be rewritten in compact vector notation as:

$$\begin{aligned}\mathbf{lnUTFP}_t = &\overline{\mu}_0 \boldsymbol{I}_N + \overline{\mu}_t \boldsymbol{I}_N + \rho\mathbf{lnUTFP}_{t-1} + \delta\mathbf{WlnUTFP}_t + \varepsilon_{Git}\mathbf{g}_{it} + \varepsilon_{Bit}\mathbf{b}_{it} \\ &+ \overline{\varepsilon}_{Git}\mathbf{g}_{it-1} + \overline{\varepsilon}_{Bit}\mathbf{b}_{it-1} + \varepsilon_{Git}\mathbf{Wg}_{it} + \varepsilon_{Bit}\mathbf{Wb}_{it} + \mathbf{Z}\alpha + \mathbf{WZ}\overline{\alpha} + \mathbf{X}_t\beta + \mathbf{WX}_t\overline{\beta} \\ &+ \mathbf{X}_{t-1}\beta + \overline{\boldsymbol{\mu}} - \delta\mathbf{W}\boldsymbol{\mu} + \mathbf{e}_t\end{aligned} \tag{8}$$

where: $\boldsymbol{I}_N$ ($\boldsymbol{I}_M$) is the Nx1 (Mx1) identity vector; $\boldsymbol{\mu}$ is the Nx1 (Mx1) vector of the time-invariant unit-specific productivity μ_i; $\mathbf{lnUTFP}_t$ is the Nx1 (Mx1) vector of time t productivity levels; $\mathbf{g}_t$ and $\mathbf{b}_t$ are the Nx1 (Mx1) vectors of time t environmental good and bad output levels; $\mathbf{Z}$ and $\mathbf{X}_t$ are the Nxk (Mxk) and Nxh (Mxh) matrices of time-invariant and time-variant observable productivity determinants, respectively. Coefficients in (8) are defined as follows: $\overline{\mu_0} = (1-\rho-\delta)\mu_0, \overline{\mu_t} = (1-\delta)\mu_t - \rho\mu_{t-1}, \overline{\rho} = (1-\rho), \varepsilon_G = -\delta\varepsilon_G, \varepsilon_B = -\delta\varepsilon_B, \overline{\alpha} = -\delta\alpha, \overline{\beta} = -\delta\beta, \beta = -\rho\beta, \overline{\boldsymbol{\mu}} = \overline{\rho}\boldsymbol{\mu}$. ρ and δ are the two unknown autoregressive coefficients, α and $\overline{\alpha}$ are the two (kx1) vectors of unknown coefficients associated with the exogenous time-invariant variables $\mathbf{Z}$ and $\beta, \overline{\beta}, \beta$ are (hx1) vectors of unknown coefficients associated with the exogenous time-variant variables contained in $\mathbf{X}$. $\mathbf{e}_t$ is the Nx1 (Mx1) vector of disturbances i.i.d. $\sim N(\mathbf{0}, \sigma^2\mathbf{I})$. It is worth noticing that in (8), since temporary shocks in $\mathbf{g}_t$, $\mathbf{b}_t$ and $\mathbf{X}_t$ may be location specific, they may also be directly transmitted across space (via terms $\varepsilon_G\mathbf{Wg}_t, \varepsilon_B\mathbf{Wb}_t$ and $\mathbf{WX}_t\overline{\beta}$). The consequence is that the productivity-environment nexus itself can diffuse over space.

In (8) parameters cannot be identified and estimated unless some further restrictions are imposed (Baldoni and Esposti, 2021). In fact, the representation of the productivity-environment nexus transmission across space can be achieved through simpler specifications (LeSage and Pace, 2009; Elhorst, 2010; Baldoni and Esposti, 2021). In particular, the so-called *Dynamic Spatial Lag Model* (DSLM) with fixed effects (Debarsy et al., 2012), which admits only the endogenous spatial interaction, represents an informative and simpler alternative:

$$\begin{aligned}\mathbf{lnUTFP}_t = &\overline{\mu_0}\boldsymbol{I}_N + \overline{\mu_t}\boldsymbol{I}_N + \rho\mathbf{lnUTFP}_{t-1} + \delta\mathbf{WlnUTFP}_t + \varepsilon_G\mathbf{g}_t + \varepsilon_B\mathbf{b}_t \\ &+ \overline{\varepsilon}_G\mathbf{g}_{t-1} + \overline{\varepsilon}_B\mathbf{b}_{t-1} + \mathbf{Z}\alpha + \mathbf{X}_t\beta + \mathbf{X}_{t-1}\beta + \overline{\boldsymbol{\mu}} + \mathbf{e}_t\end{aligned} \tag{9}$$

The assumption is that there is no direct spatial diffusion of the productivity-environment nexus, as well as of the other characteristics of the cross-sectional units represented by $\mathbf{X}_t$. In other words, in (9) the productivity-environment nexus can diffuse over space only indirectly via the spatial dependence of TFP. This is the empirical specification that will be considered henceforth.

In principle, it could be argued that this spatial dependence is more likely to occur at the ES, thus also at the BS, rather than at the RS. In fact, even in the simplified specification (9) the meaning and implications of the space-dependent productivity-environment nexus is different between micro level and macro level data as it captures different phenomena (Anselin, 2002). On the one hand, the complex linkages among fields and farms are typically local and highly affected by structural and production similarity. At an aggregate level, such complexity vanishes. On the other hand, aggregate units may generate a "gravity" impact on neighbours for which there is no correspondence at the micro level. Therefore, there can be a substantial difference in empirically investigating spatial dependence in the productivity-environment nexus at these alternative scales (García Cidad et al., 2003). Making this difference emerge is a further motivation for applying and estimating model specification (9) in both the farm-level and the two regional-level datasets, the only difference being the definition of matrix $\mathbf{W}$ (see below).

Fig. 1 recaps all the issues that emerge, moving through the different spatial scales, in passing from the theoretical model to the econometric estimation. It elucidates why the actually estimated vector of model parameters (thus, the productivity-environment nexus) may differ between the farm and the regional scales (vectors **B** and **C** in Fig. 1, respectively) and both may differ from the "true" unknown vector (**A** in Fig. 1; thus: $\mathbf{B} \neq \mathbf{A} \neq \mathbf{C}$). Some of these sources of under-identification or estimation bias may be controlled for by the analyst. Within the DSLM specification (9), the sources of heterogeneity to be controlled for are both observables ($\mathbf{Z}$, $\mathbf{X}_t$) and unobservables (fixed-effects $\overline{\boldsymbol{\mu}}$ and disturbance terms $\mathbf{e}_t$). Other sources, however, cannot be controlled for, that is, the granularity and aggregation bias. Consequently, we can only observe their impact *ex post* by comparing **B** and **C** estimates. This difference reveals whether there is concordance or contradiction between the estimated nexus at the three scales of analysis. An inconsistency would severely question the viability of an evidence-based policy making in this field.

5. The empirical study

5.1. The datasets

The 2008–2018 Italian Farm Accountancy Data Network (FADN) dataset is used. It consists of an unbalanced panel of farms, ranging from 11,389 farms (in 2008) to 10,386 farms (in 2018), with a total amount of 119,229 observations over the whole period. A balanced panel can be also extracted. It consists of 1,658 farms observed for the entire period (11 years). This balanced panel is used in the present study for the farm-level estimation of model (9), while the unbalanced panel is used to construct the regional datasets (see below). The FADN sample is only a small fraction of the population of commercial farms in Italy (Baldoni et al., 2021). Therefore, whenever the investigation concerns aspects for which space matters (i.e., a place-effect very likely occur), the use of this farm-level dataset is likely to incur the abovementioned granularity bias.

Following Baldoni and Esposti (2021), a possible empirical strategy to overcome this problem consists in aggregating these farm-level data at some geographical scale (typically, administrative levels). The resulting regional data has a lattice structure; thus, it does not suffer from granularity. Due to the aggregation bias, however, at the regional scale the productivity-environment nexus might emerge as an artefact. In the present case, both Italian NUTS3 and NUTS2 regions represent suitable aggregation levels. These two macro datasets consist of a balanced panel of 90 and 21 regions, respectively. To form these macro panels, the full FADN unbalanced sample is aggregated, and model (9) then estimated at these geographical levels.

5.2. The performance indicators

Relative levels of UTFP for units and years in the panel are derived as ratios of output quantity indexes on input quantity indexes at each corresponding aggregation level (either farms or NUTS2/NUTS3 regions). The output index contains crop and livestock products while inputs include labour, fertilizers, pesticides, external services, water, energy, seeds, feeding stuff, capital services, farm reuses and other costs. Aggregation of outputs and inputs is obtained using Fisher indexes and transitivity of the indicator is achieved through the minimum spanning tree approach (Hill, 1999, 2004).

For the calculation of the GHG emission indicator, we follow the IPCC methodology (IPCC, 2006), as in Coderoni and Vanino (2022), where a linear relationship is assumed between emission factors and activity data. For any *i-th* unit an any time *t*, the Carbon Footprint Index (CFI_{it}) is calculated by firstly deriving the GHG emissions at the farm level and then eventually aggregating them at the relevant higher level (NUTS3 and NUTS2 regions). Secondly, as emissions are size-dependent while UTFP is not, an index is computed by fixing as basis the initial year emission level. This CFI_{it} expresses the environmental bad output (B) and its growth rate (cfi_{it}) the consequent b_{it} variable in (9).

As regards ecosystem services, a crop diversity index is used as a reliable proxy.[16] By reviewing 98 meta-analyses and performing a second-order meta-analysis based on 41,946 comparisons between diversified and simplified practices, Tamburini et al. (2020) have recently shown that crop diversity is strongly associated with the provision of different ecosystem services. A Crop Diversity Index ()CDI_{it} is here computed as a Shannon diversity index. This index is separately computed on the three datasets (farm-level, NUTS3 and NUTS2 regions) using the respective crop shares on total Utilized Agricultural Area (UAA). This CDI_{it} variable expresses the environmental good output (G), and its growth rate (cdi_{it}) the consequent g_{it} variable in (9). More details on the methodologies adopted to compute these performance indicators and the respective descriptive statistics are provided in Appendices 3 and 4 (Supplementary Material).

5.3. Model specification and estimation

Model in (9) is a dynamic, AR(1), spatial panel model whose estimation raises several issues and requires an appropriate strategy. Baldoni and Esposti (2021) compare the pros and cons of alternative estimation approaches suggesting that the two-step GMM-SYS estimator represents a suitable and robust solution on both micro and macro datasets. This solution is thus adopted to estimate model (9) parameters.

This estimation approach still encounters the possible endogeneity of some regressor, beside term $\ln UTFP_{it-1}$. In particular, the well-known endogeneity issue implied by production function estimation (Griliches and Mairesse, 1995) remains a concern. Here it applies to the non-market outputs G and B. As already mentioned, terms g_{it} and b_{it} in (9) are possibly endogenous as they depend on the input use decisions and on the consequent level of market output Y that also affect the left-hand side of (9), *i.e.*, the *UTFP* (Hulten, 1992: 968). Consequently, in the above-mentioned estimation g_{it} and b_{it} are instrumented by the respective suitable lagged values.

The list of regressors included in (9) is completed by the following additional variables. $\mathbf{Z}_i = Year_{i,s}$ where $Year_s$ are time dummies (s indicates the year); $\overline{\mathbf{X}}_{it} = (FOR_{it}, PS_{it})$ where: FOR_{it} expresses the farm/region forest area on total land; PS_{it} is the farm's physical size expressed by the UAA. PS_{it}, may itself express the farmers' choices depending on current and past productivity levels, as well as the possible returns to scale at the farm level. Consequently, endogeneity cannot be excluded also for PS_{it} and in model estimation this variable is instrumented by the respective suitable lagged values.

Finally, the NxN (MxM) spatial matrix $\mathbf{W}$ is specified to express the proximity among units. Proximity is here intended as both contiguity and distance. Therefore, neighbourhoods are identified by a combination of radial distance and a queen contiguity matrix: the i-th row/j-th column element is fixed at 1 if the j-th and i-th units are contiguous (corner or edge neighbour), or the former falls within the predetermined radial distance from the latter, and at 0 otherwise. For any given dataset, alternative matrices $\mathbf{W}$ can be thus specified by varying the radial distance that defines the contiguity. Alternative specifications are thus adopted to assess the robustness of results.

6. Results and discussion

Model (9) is estimated on the three datasets and over alternative definitions of the spatial matrix $\mathbf{W}$. Respective estimation results are reported, in sequence, in Tables 1–3.[17] The usual tests performed on these GMM-SYS estimations (bottom of Tables 1–3) are concordant across the three levels of analysis. AR(1) and AR(2) tests confirm that the dynamic model (9) is properly specified, while the adopted instruments result to be valid as indicated by the Hansen test (Arellano, 2003).

Comparing Tables 1 and 2 it emerges that some results are robust passing from the farm-level to the denser geographical aggregation (NUTS3 regions) while others are not. The former seems quite robust also across alternative definitions of matrix $\mathbf{W}$. They concern the time and spatial correlation of the UTFP.

These correlations emerge as the main drivers of the observed productivity performance. Time correlation is positive and indicates some persistence of the exogenous productivity shocks, thus confirming previous evidence on agricultural productivity also in the Italian case (Esposti, 2000). In magnitude, this effect remains quite stable over the different radial distances ($\mathbf{W}$ matrices) and is similar in the farm-level and in the NUTS3-level datasets though a little higher in the latter case. Spatial dependence is also significantly positive and even more relevant, in magnitude. As it could be expected, productivity transmits to neighboring units. Therefore, if we combine this effect with time dependence, also this spatial transmission tends to have some persistence over time. The size of this spatial dependence is similar in the micro and the NUTS3 datasets and in both cases, as expected, it increases by augmenting the radial distance defining proximity.

After controlling for time and spatial dependence, however, there are no other clear linkages with the productivity performance. As expected, size matters only at the micro-level evidently as possible effect of returns to scale on the productivity performance. At the NUTS3 regional level, the physical size of farms, expressed as the farm average size within the

[16] Appendix 3 of the Supplementary Material discusses more in detail why a crop diversity index instead of a crop rotation index has been considered in the analysis.

[17] Due to space limitations, the estimated coefficients of the time (year) dummies are not reported. Most time dummies are statistically significant further supporting the dependence of the productivity performance on short-term shocks.

Table 1
Model (9) estimates on the farm-level dataset under alternative **W** (radial distance in Km) - Estimated standard errors in parenthesis.

Km	25	50	100	150	200
Model Variable					
$lnUTFP_{t-1}$	0.127***	0.127***	0.128***	0.134***	0.134***
	(0.016)	(0.015)	(0.015)	(0.016)	(0.016)
$WlnUTFP_t$	0.470***	0.664***	0.779***	0.787***	0.793***
	(0.109)	(0.111)	(0.108)	(0.113)	(0.113)
cfi_{it}	0.001**	0.001**	0.001**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cfi_{it-1}	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cdi_{it}	0.029	0.006	0.007	−0.015	−0.013
	(0.076)	(0.074)	(0.080)	(0.073)	(0.074)
cdi_{it-1}	−0.003	−0.005**	−0.005**	−0.006**	−0.006**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
FOR_{it}	−0.225	−0.108	−0.185	−0.201	−0.211
	(0.356)	(0.353)	(0.356)	(0.346)	(0.343)
FOR_{it-1}	−0.110	−0.266	−0.240	−0.216	−0.198
	(0.349)	(0.346)	(0.350)	(0.339)	(0.336)
PS_{it}	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
PS_{it-1}	−0.002*	−0.002*	−0.002*	−0.002*	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N. of units (N)	1,658	1,658	1,658	1,658	1,658
N. of observations (NxT)	14,922	14,922	14,922	14,922	14,922
AR(1) test (p-value)	0.000***	0.000***	0.000***	0.000***	0.000***
AR(2) test (p-value)	0.194	0.280	0.258	0.204	0.185
Hansen test (p-value)	0.307	0.417	0.510	0.255	0.195

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 2
– Model (9) estimates on the NUTS3 regions dataset under alternative **W** (radial distance in Km) - Estimated standard errors in parenthesis.

Km	50	100	150	200	250	300
Model Variable						
$lnUTFP_{t-1}$	0.151***	0.176***	0.171***	0.169***	0.142**	0.138***
	(0.051)	(0.054)	(0.057)	(0.057)	(0.056)	(0.053)
$WlnUTFP_t$	0.456***	0.662***	0.780***	0.825***	0.749***	0.936***
	(0.154)	(0.222)	(0.261)	(0.222)	(0.245)	(0.253)
cfi_{it}	0.037	0.040	0.029	0.050*	0.052	0.050
	(0.033)	(0.031)	(0.040)	(0.029)	(0.034)	(0.033)
cfi_{it-1}	−0.027	−0.027*	−0.033*	−0.027*	−0.025	−0.027
	(0.018)	(0.016)	(0.020)	(0.016)	(0.019)	(0.020)
cdi_{it}	0.663***	0.529**	0.458	0.484*	0.582*	0.575*
	(0.256)	(0.252)	(0.315)	(0.276)	(0.318)	(0.322)
cdi_{it-1}	0.238	−0.013	0.003	−0.030	0.138	0.125
	(0.335)	(0.249)	(0.351)	(0.290)	(0.343)	(0.336)
FOR_{it}	−0.359	−0.554	−0.450	−0.361	−0.353	−0.475
	(0.432)	(0.574)	(0.543)	(0.553)	(0.514)	(0.516)
FOR_{it-1}	0.174	0.446	0.397	0.419	0.063	0.153
	(0.465)	(0.549)	(0.559)	(0.515)	(0.566)	(0.535)
PS_{it}	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PS_{it-1}	0.000*	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N. of units (N)	90	90	90	90	90	90
N. of observations (NxT)	810	810	810	810	810	810
AR(1) test (p-value)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
AR(2) test (p-value)	0.117	0.122	0.092*	0.117	0.121	0.098*
Hansen test (p-value)	0.450	0.861	0.367	0.932	0.471	0.560

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3
Model (9) estimates on the NUTS2 regions dataset under alternative **W** (radial distance in Km) - Estimated standard errors in parenthesis[a].

Km	50	100	150	200	250	300
Model Variable						
$lnUTFP_{t-1}$	0.086	0.086	0.130	0.149*	0.074	0.079
	(0.140)	(0.140)	(0.102)	(0.090)	(0.120)	(0.096)
$WlnUTFP_t$	6.736	6.736	3.635	5.570	7.430	10.218
	(9.307)	(9.307)	(8.911)	(8.050)	(8.615)	(10.796)
cfi_{it}	0.195	0.195	0.206	0.193	0.198	0.191
	(0.313)	(0.313)	(0.318)	(0.294)	(0.271)	(0.232)
cfi_{it-1}	−0.022	−0.022	−0.000	0.015	−0.031	0.198
	(0.251)	(0.251)	(0.267)	(0.271)	(0.257)	(0.350)
cdi_{it}	−1.211	−1.211	−0.785	−0.354	−1.165	1.160
	(3.130)	(3.130)	(2.982)	(2.772)	(1.703)	(3.860)
cdi_{it-1}	−0.100	−0.100	0.041	0.182	−0.221	0.516
	(1.670)	(1.670)	(1.754)	(1.680)	(1.267)	(1.821)
FOR_{it}	−1.091	−1.091	−0.600	0.017	−0.987	4.359
	(3.688)	(3.688)	(3.463)	(3.254)	(1.061)	(7.113)
FOR_{it-1}	1.711	1.711	1.146	0.463	1.570	−3.635
	(3.867)	(3.867)	(3.605)	(3.412)	(1.273)	(6.960)
PS_{it}	−0.000	−0.000	−0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PS_{it-1}	0.000	0.000	0.000	0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N. of units (N)	21	21	21	21	21	21
N. of observations (NxT)	189	189	189	189	189	189
AR(1) test (p-value)	0.010**	0.010**	0.001***	0.000***	0.00288*	0.0741*
AR(2) test (p-value)	0.119	0.119	0.149	0.296	0.0776	0.321
Hansen test (p-value)	0.737	0.737	0.681	0.810	0.856	0.972

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
[a] The estimates obtained with 50 and 100 km of radial distance are identical as they imply, for the NUTS2 regions, the same spatial matrix **W**.

region, does not bring about any linkage with productivity. Also, the presence of forest area does not statistically affect the productivity performance regardless the level of the investigation.[18]

What really matters here, however, is the nexus between productivity and the two environmental indicators. The evidence emerging from estimates is mixed between the farm level and the NUTS3 dataset. At the farm level, the nexus of GHG emissions (*cfi)* with productivity is positive and statistically significant. Nonetheless, this nexus is weak as the estimated elasticity is actually about 0.1%. Moreover, this elasticity declines to 0 and loses statistical significance by expanding matrix **W** through a higher radial distance. The same pattern is observed for the lagged values of the *cfi*.

These results are not completely unexpected as they confirm what obtained in previous studies in the field (Stetter et al., 2022, p. 733). Therefore, rather than being interpreted as evidence of some underlying model misspecification, this statistically weak and somehow counterintuitive evidence has to be intended as the consequence of the quite complex and heterogenous relationship occurring at the farm-level between GHG emission, conventional input use and market-output supply discussed in previous sections. These results evidently do not exclude the existence of farms whose higher technological level is compatible with higher TFP and the adoption of advanced technical solutions allowing a more efficient saving input use (fuel and energy, in particular) and/or mitigating or abatement technologies. For these farms a negative nexus between *cfi* and productivity should be observed. Evidently, however, this relationship is more than compensated by those farms whose more intense input use and the lacking adoption of mitigating solutions justifies the co-existence of higher TFP and higher GHG emissions.

[18] According to the model theoretical derivation, the parameters associated to the lagged values of *FOR* and *PS* should correspond to the parameters associated to the respective current values multiplied by -ρ, that is, the parameter of the time correlation dependence (see section 3). These parameter restrictions can be tested ex-post. Limiting the attention only to the statistically significant parameters, these tests accept the validity of these theoretical restrictions. Tests results are available upon request.ρ

In the case of crop diversity (*cdi*) an opposite effect is observed. The elasticity associated to the current *cdi* value is not statistically different from 0 but it gains significance with the lagged *cdi.* The respective elasticity is still quite low (about 0.5%) and quite constant across alternative **W** matrices. The sign of the productivity-*cdi* nexus goes in the opposite direction compared to *cfi*: higher productivity is associated with lower *cdi* levels. In fact, if we combine the current and lagged *cdi* values the overall linkage with the unadjusted productivity is questionable as the current *cdi* shows an elasticity that, though not statistically significant from zero, is larger in magnitude but with the opposite sign than the lagged *cdi.* Therefore, by limiting the attention to the farm-level data, one could conclude that no productivity-environment nexus is found while a trade-off seems to rather emerge. However, the overall statistical significance of these relations seems quite limited, therefore caution is needed in drawing conclusions on these estimates.

If we move to the NUTS3 level, we observe that only part of the farm-level evidence about the productivity-environment nexus remains and survives aggregation. The current value of the *cfi* still shows a positive relationship with productivity but the respective elasticity loses statistical significance even though its magnitude its much larger than what observed at the farm level. Moreover, this positive effect is almost entirely offset by the lagged value of *cfi*. Also, for the *cdi* we observe substantial differences with respect to the farm level. The elasticity associated to the current value is positive and statistically significant while the lagged value is not. Moreover, the magnitude of this elasticity is much higher indicating a very strong relationship between productivity and *cdi.* This nexus observed at the NUTS3 regional level seems to partially support the existence of a positive productivity-environment nexus: a statistically significant positive relation between productivity and crop diversity emerges in most of the estimated models, while the relation between productivity and emissions is less clear and seems to point at a negative lagged relation.

None of these results is confirmed at the NUTS2 regional level (Table 3). This level of aggregation implies a very poor statistical quality of model estimates. No parameter is statistically significant but those associated to time dummies. Moreover, several parameter estimates assume unreliable size and seem highly unstable. Eventually, we can conclude that pushing aggregation at this level (which is, by the way, the administrative level at which most EU policies are planned and implemented) destroys any evidence on the linkages between agricultural productivity and environmental performance.

This conclusion should not surprise since, as noticed, these linkages are already weak at the farm-level due to the large farm heterogeneity and the co-existence of contrasting relationships. When moving at an aggregate level some of these farm-level relationships, like the adoption of mitigating or abatement solutions, may disappear and even become unobservable while other effects may surface.[19] Moreover, in the case of GHG emissions, at the aggregate level the change in output composition over time, particularly the relative decline of livestock activities within the regional agriculture, may assume high relevance while it is irrelevant at the farm scale. Eventually, at this coarser aggregation level, not only no nexus with the environmental indicators is detectable, but also the role of space becomes barely discernible. At this aggregation level agricultural productivity can be apparently explained only as the result of short-term exogenous shocks.

The comparison of estimation results across these different scales of analysis eventually confirms that space matters but, at the same time, it is not conclusive with respect to be best choice to be made in this respect, that is, the most reliable scale. On the one hand, if we assume that the real productivity-environment nexus is that occurring at the farm level, the present study suggests that the granularity of the adopted farm sample may prevent from a clear identification of this nexus. If any, however, the relationship emerging at this level of the analysis would rather suggest a trade-off between productivity and environment in farming. On the other hand, aggregating micro data at some geographical level might preserve micro properties while strengthening their statistical robustness, but it might also alter the farm-level relationship and surface different evidence. When aggregation is pushed to a higher level most of the information about the underlying relationships is lost and the consequent empirical evidence seems statistically poor and unreliable.

7. Concluding remarks

According to many analysts, the future of agriculture consists in the adoption of techniques and solutions able to reconcile more food production with the preservation of environmental resources it interacts and interferes with. Whether this kind of farming (also designated as Sustainable Intensification) already exists and prevails, or it has still to come, is debatable. Consequently, also whether current policies already provide enough incentives in this direction and should be just maintained and reinforced is widely discussed. In fact, if and where a nexus between higher productivity and better environmental indicators actually occurs requires an appropriate empirical assessment. This empirical investigation is challenging mostly because it has to do with the role of space.

Three possible spatial scales of analysis can be considered: the ecological scale (usually unobservable), the behavioral scale (i.e., the farm level), the regional scale. The present paper develops a conceptual framework and a consequent empirical specification making explicit which identification issues and sources of bias can arise moving across these different scales. These biases may eventually motivate why inconsistent or even contrasting evidence is found at different levels.

By using the Italian FADN sample data over the 2008–2018 period, the present study investigates the productivity-environment nexus making the role of space explicit and thus showing how the spatial scale of analysis may affect the empirical evidence. This is done by adopting the same dynamic spatial panel model specification both on the farm sample and on the region sample obtained by aggregating the farm data up to the NUTS3 and NUTS2 Italian regions. For their relevance and diversity, two jointed environmental indicators (GHG emissions and crop diversity) are considered and their relationship with the conventional TFP measure explored.

Results obtained confirm that space and geographical scale matter, namely, that the empirical evidence about the productivity-environment nexus is space-dependent. At the farm level, the productivity-environment nexus shows statistically poor evidence and, in any case, it rather emerges as a trade-off. Regional data, on the contrary, can provide more robust results when aggregation is maintained at a dense-enough scale (NUTS3 regions). In this case, the presence of a positive productivity-environment nexus finds some support though it could surface as an artefact of aggregation itself.

The implication of such scale-dependent evidence for policy making should not be understated. At a first glance, it could be interpreted as the need for a careful evaluation on the most appropriate scale of policy design and implementation. In fact, the problem seems more serious than this: whatever is the scale of policy intervention, the lack of a robust evidence on the productivity-environment nexus puts at risk the existence itself of an evidence-based policy making in the field.

Results here obtained are evidently only indicative and deserve further confirmation and modelling developments, including a more specific and sophisticated representation of the underlying technology, refinements of the econometric approach and extension of the analysis to other and larger datasets. Though the Italian agriculture shows a remarkable diversity in the farming conditions, the repetition of the present approach to other possibly more distinctive agricultural contexts may be helpful in assessing the generalizability of the present results.

CRediT authorship contribution statement

Edoardo Baldoni: Data curation, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Silvia Coderoni:** Conceptualization, Data curation, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Roberto Esposti:** Supervision, Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2023.137847.

References

Anselin, L., 2001. Spatial effects in econometric practice in environmental and resource economics. Am. J. Agric. Econ. 83 (3), 705–710.
Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. Agric. Econ. 27, 247–267.
Arbia, G., Espa, G., Giuliani, D., 2015. Dirty spatial econometrics. Ann. Reg. Sci. 56, 177–189.
Arellano, M., 2003. Panel Data Econometrics. Oxford University Press, Oxford.
Baldoni, E., Coderoni, S., Esposti, R., 2017. The productivity and environment nexus through farm-level data. The case of Carbon Footprint applied to Lombardy FADN farms. Bio base Appl. Econ. 6 (2), 119–137. https://doi.org/10.13128/BAE-19112.
Baldoni, E., Coderoni, S., Esposti, R., 2018. The complex farm-level relationship between environmental performance and productivity. The case of carbon footprint of Lombardy FADN farms. Environ. Sci. Pol. 89, 73–82. https://doi.org/10.1016/j.envsci.2018.07.010.

[19] See footnote 4 for some examples in this respect.

Baldoni, E., Coderoni, S., Esposti, R., 2021. Immigrant workforce and agriculture productivity. Evidence from Italian farm-level data European Review of Agricultural Economics 48 (4), 805–834, 101093/erae/jbaa033.

Baldoni, E., Esposti, R., 2021. Agricultural productivity in space: an econometric assessment based. on Farm-Level Data American Journal of Agricultural Economics 103 (4), 1525–1544, 101111/ajae12155.

Blundell, R., Bond, S., 2000. GMM estimation with persistent panel data: an application to production functions. Econom. Rev. 19, 32 1–340.

Bond, S., Söderbom, M., 2005. Adjustment Costs and the Identification of Cobb Douglas Production Functions *IFS Working Papers* W05/04. Institute for Fiscal Studies, London.

Brandt, N., Schreyer, P., Zipperer, V., 2013. Productivity Measurement with Natural Capital, OECD Economics Department Working Papers. OECD Publishing. No 1092.

Brandt, N., Schreyer, P., Zipperer, V., 2014. Productivity Measurement with Natural Capital and Bad Outputs OECD Economics Department Working Papers. OECD Publishing. No 1154.

Bureau, J., Antón, J., 2022. Agricultural Total Factor Productivity and the Environment: A Guide to Emerging Best Practices in Measurement, OECD Food, Agriculture and Fisheries Papers, No. 177. OECD Publishing, Paris. https://doi.org/10.1787/6fe2f9e0-en.

Cárdenas Rodríguez, M., Haščič, I., Souchier, M., 2018a. Environmentally Adjusted Multifactor Productivity: Methodology and Empirical Results for OECD and G20 Countries OECD Green Growth Papers. OECD Publishing, Paris. No 2018/02.

Cárdenas Rodríguez, M., Haščič, I., Souchier, M., 2018b. Environmentally adjusted multifactor productivity: methodology and empirical results for OECD and G20 countries. Ecol. Econ. 153, 147–160.

Chakir, R., 2009. Spatial downscaling of agricultural land-use data: an econometric approach using cross entropy. Land Econ. 85 (2), 238–251.

Chambers, R.G., 1988. Applied Production Analysis. Cambridge University Press, Cambridge.

Chaudhary, A., Pfister, S., Hellweg, S., 2016. Spatially explicit analysis of biodiversity loss due to global agriculture, pasture and forest land use from a producer and consumer perspective. Environ. Sci. Technol. 50 (7), 3928–3936.

Clough, Y., Kirchweger, S., Kantelhardt, J., 2020. Field sizes and the future of farmland biodiversity in European landscapes. Conservation letters 13 (6), e12752.

Coderoni, S., Vanino, S., 2022. The farm-by-farm relationship among carbon productivity and economic performance of agriculture. Sci. Total Environ. 819, 153103 https://doi.org/10.1016/j.scitotenv.2022.153103, 2022.

Coderoni, S., Esposti, R., 2018. CAP payments and agricultural GHG emissions in Italy A farm-level assessment. Sci. Total Environ. 627, 427–437. https://doi.org/10.1016/j.scitotenv.2018.01.197.

Coelli, T., Lauwers, L., Van Huylenbroeck, G., 2007. Environmental efficiency measurement and the materials balance condition. J. Prod. Anal. 28 (1), 3–12.

Cui, J., Lapan, H., Moschini, G., 2016. Productivity export and environmental performance: air pollutants in the United States. Am. J. Agric. Econ. 98 (2), 447–467.

Debarsy, N., Ertur, C., LeSage, J.P., 2012. Interpreting dynamic space-time panel data models. Stat. Methodol. 9, 158–171.

Di Falco, S., Chavas, J., 2008. Rainfall shocks, resilience and the effects of crop biodiversity on agroecosystem productivity. Land Econ. 84 (1), 83–96.

Di Falco, S., Perrings, C., 2003. Crop genetic diversity, productivity and stability of agroecosystems. A theoretical and empirical investigation. Scot. J. Polit. Econ. 50, 207–216.

Dingel, J.I., Tintelnot, F., 2021. Spatial Economics for Granular Settings NBER Working Paper 27287. National Bureau of Economic Research, Cambridge Mass.

Eberhardt, M., Helmers, C., 2010. Discussion paper series n. 513. In: Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimator. University of, Oxford.

Elhorst, J.P., 2010. Applied spatial econometrics: raising the bar. Spatial Econ. Anal. 5 (1), 9–28.

Esposti, R., 2000. Stochastic technical change and procyclical TFP: the Italian agriculture case. J. Prod. Anal. 14 (2), 117–139.

European Commission, 2020a. A Farm to Fork Strategy for a Fair, Healthy and Environmentally-Friendly Food System, Brussels, 2052020. COM, 2020) 381 final.

European Commission, 2020b. The European Green Deal, Brussels, 11122019. COM (2019) 640 final.

Färe, R., Grosskopf, S., Lovell, C., Pasurka, C., 1989. Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev. Econ. Stat. 71.

Firbank, L.G., Elliott, J., Drake, B., Cao, Y., Gooday, R., 2013. Evidence of sustainable intensification among British farms. Agric. Ecosyst. Environ. 173, 58–65.

Fuglie, K., Benton, T., Sheng, Y., et al., 2016. Metrics of Sustainable Agricultural Productivity G20 MACS White. Paper 25th April 2016.

Gadanakis, Y., Bennett, R., Park, J., Areal, F.J., 2015. Evaluating the sustainable intensification of arable farms. J. Environ. Manag. 150, 288–298, 2015. http://dxdoiorg/101016/jjenvman201410005.

García Cidad, V., De Blust, G., Maljean, J.-F., Peeters, A., 2003. Overview of Biodiversity Indicators Related to Agriculture in Belgium in: OECD, Agriculture and Biodiversity: Developing Indicators for Policy Analysis OECD Paris, pp. 133–151.

Gerber, J.S., Carlson, K.M., Makowski, D., Mueller, N.D., Garcia de Cortazar-Atauri, I., Havlík, P., Herrero, M., Launay, M., O'Connell, C.S., Smith, P., West, P.C., 2016. Spatially explicit estimates of N2O emissions from croplands suggest climate mitigation opportunities from improved fertilizer management. Global Change Biol. 22 (10), 3383–3394.

Gocht, A., Röder, N., 2014. Using a Bayesian estimator to combine information from a cluster analysis and remote sensing data to estimate high-resolution data for agricultural production in Germany. Int. J. Geogr. Inf. Sci. 28 (9), 1744–1764.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Science 327, 812–818.

Goldberger, A.S., 1991. A Course in Econometrics. Harvard University Press, Cambridge Mass.

Griliches, Z., Mairesse, J., 1995. Production Functions: Search for Identification *NBER Working Paper 5076*. National Bureau of Economic Research, Cambridge.

Herzog, F., Steiner, B., Bailey, D., Baudry, J., Billeter, R., Bukácek, R., Blust, G., Dirksen, J., Dormann, C.F., de Filippi, R., Frossard, E., Liira, J., Schmidt, T., Stöckli, R., Thenail, C., van Wingerden, W., Bugter, R., 2006. Assessing the intensity of temperate European agriculture at the landscape scale. Eur. J. Agron. 24 (2), 165–181.

Hill, R.J., 1999. International comparisons using spanning trees. In: Heston, A., Lipsey, R. E. (Eds.), International and Interarea Comparisons of Income, Output and Prices. University of Chicago Press, Chicago, pp. 109–120.

Hill, J., 2004. Constructing price indexes across space and time: the case of the European Union. Am. Econ. Rev. 94, 1379–1410.

Hulten, C.R., 1992. Growth accounting when technical change is embodied in capital. Am. Econ. Rev. 82 (4), 964–980.

IPCC, 2006. In: Eggleston, H.S., Buendia, L., Miwa, K., et al. (Eds.), 2006 Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories Prepared by the National Greenhouse Gas Inventories Programme (Japan).

Jansen, H.G.P., Stoorvogel, J.J., 1998. Quantification of aggregation bias in regional agricultural land use models: application to guacimo county, Costa Rica. Agric. Syst. 58 (3), 417–439.

Kimura, S., Sauer, J., 2015. Dynamics of Dairy Farm Productivity Growth: Cross-Country Comparison OECD Food Agriculture and Fisheries Papers No 87. OECD Publishing Paris.

Koiry, S., Huang, W., 2023. Do ecological protection approaches affect total factor productivity change of cropland production in Sweden? Ecol. Econ. 209 https://doi.org/10.1016/j.ecolecon.2023.107829 (in press).

Kuosmanen, N., Kuosmanen, T., 2013. Modeling cumulative effects of nutrient surpluses in agriculture: a dynamic approach to material balance accounting. Ecol. Econ. 90, 159–167. https://doiorg/101016/jecolecon201303016.

Le Mouël, C., de Lattre-Gasquet, M., Mora, O. (Eds.), 2018. *Land Use and Food Security in 2050: a Narrow Road. Agrimonde-Terra* Versailles Cedex, France: Éditions Quæ.

Leip, A., Marchi, G., Koeble, R., Kempen, M., Britz, W., Li, C., 2008. Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. Biogeosciences 5 (1), 73–94.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. Taylor & Francis, Boca Raton, FL.

Murty, S., 2015. On the properties of an emission-generating technology and its parametric representation. Econ. Theor. 60, 243–282.

Murty, S., Russell, R.R., Levkoff, S.B., 2012. On modelling pollution-generating technologies. J. Environ. Econ. Manag. 64, 117–135.

Murty, S., Russell, R.R., 2022. Bad outputs. In: Ray, S.C., Chambers, R.G., Kumbhakar, S. C. (Eds.), Handbook of Production Economics. Springer, Singapore.

O'Donnell, C.J., 2016. Using information about technologies, markets and firm behaviour to decompose a proper productivity index. J. Econom. 190, 328–340.

OECD, 2001a. *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth* Paris. OECD Publishing.

OECD, 2001b. Multifunctionality: towards an Analytical Framework. OECD Publishing, Paris.

OECD, 2022. Insights into the measurement of agricultural total factor productivity and the environment. https://www.oecd.org/agriculture/topics/network-agricultural-productivity-and-environment/.

Omer, A., Pascual, U., Russell, N., 2010. A theoretical model of agrobiodiversity as a supporting service for sustainable agricultural intensification. Ecol. Econ. 69, 1926.–1933.

Omer, A., Pascual, U., Russell, N.P., 2007. Biodiversity conservation and productivity in intensive agricultural systems. J. Agric. Econ. 58, 308–329.

Pethig, R., 2006. Non-linear production, abatement, pollution and materials balance reconsidered. J. Environ. Econ. Manag. 51 (2), 185–204.

Picazo-Tadeo, A., Castillo, J., Beltrán-Esteve, M., 2014. An intertemporal approach to measuring environmental performance with directional distance functions: greenhouse gas emissions in the European Union. Ecol. Econ. 100, 173–182. https://doiorg/101016/jecolecon201402004.

Reinhard, S., Knox Lovell, C.A., Thijssen, G.J., 2000. Environmental efficiency with multiple environmentally detrimental variables: estimated with SFA and DEA. Eur. J. Oper. Res. 121 (2).

Scheel, H., 2001. Undesirable outputs in efficiency valuations. Eur. J. Oper. Res. 132.

Schoefer, B., Ziv, O., 2021. Productivity, Place, and Plants: Revisiting the Measurement CEPR Discussion Paper Series, N 15676. Centre for Economic Policy Research, London.

Serra, T., Chambers, R.G., Oude Lansink, A., 2014. Measuring technical and environmental efficiency in a state-contingent technology. Eur. J. Oper. Res. 236, 706–717.

Sheng, Y., Jackson, T., Gooday, P., 2015. Resource reallocation and its contribution to productivity growth in Australian broadacre agriculture. Aust. J. Agric. Resour. Econ. 59, 1–20.

Sidhoum, A.A., Mennig, P., Sauer, J., 2023. Do agri-environment measures help improve environmental and economic efficiency? Evidence from Bavarian dairy farmers. Eur. Rev. Agric. Econ. https://doi.org/10.1093/erae/jbad007 (in press).

Stetter, C., Mennig, P., Sauer, J., 2022. Using machine learning to identify heterogeneous impacts of agri-environment schemes in the EU: a case study. Eur. Rev. Agric. Econ. 49 (4), 723–739.
Tamburini, G., Bommarco, R., Cherico Wanger, T., Kremen, C., van der Heijden, M.G.A., Liebman, M., Hallin, S., 2020. Agricultural diversification promotes multiple ecosystem services without compromising yield. Sci. Adv. https://doi.org/10.1126/sciadv.aba1715, 2020; 6 (45): eaba1715.
Van Beveren, I., 2010. Total factor productivity estimation: a practical review. J. Econ. Surv. 10, 1–38.
Wade, C.M., Baker, J.S., Latta, G., Ohrel, S.B., 2019. Evaluating potential sources of aggregation bias with a structural optimization model of the U.S. Forest sector. J. For. Econ. 34 (3–4), 337–366. https://doi.org/10.1561/112.00000503.
Yu, J., Wu, J., 2018. The sustainability of agricultural development in China: the agriculture–environment nexus. Sustainability 10 (6), 1776.
Zhou, P., Ang, B.W., Poh, K.L., 2008. A survey of data envelopment analysis in energy and environmental studies. Eur. J. Oper. Res. 189.