The economic, environmental and social performance of European certified food

The economic, environmental and social performance of European certified food

Valentin Bellassen ,*, Marion Drut , Mohamed Hilal , Antonio Bodini , Michele Donati, Matthieu Duboys de Labarre , Jelena Filipovic Lisa Gauvrit, Jose M. Gil, Viet Hoang, Agata Malak-Rawlikowska, Konstadinos Mattasi, Sylvette Monier-Dilhan j, Paul Muller, Orachos Napasintuwong, Jack Peerlings, Thomas Pomeon, Marina Tomic Maksan, Aron Torok, Mario Veneziani, Gunnar Vittersø, Filippo Arfini c

Abstract

To identify whether EU certified food – here organic and geographical indications – is more sustainable than a conventional reference, we developed 25 indicators covering the three sustainability pillars. Original data was collected on 52 products at farm, processing and retail levels, allowing the estimation of circa 2,000 indicator values. Most strikingly, we show that, in our sample, certified food outperforms its non-certified reference on most economic and social indicators. On major environmental indicators – carbon and water footprint – their performance is similar. Although certified food is 61% more expensive, the extra-performance per euro is similar to classical policy interventions to improve diet sustainability such as subsidies or taxes. Cumulatively, our findings legitimate the recent initiatives by standards to cover broader sustainability aspects.

Highlights

- 25 indicators are estimated for 52 food value chains to estimate their economic, environmental and social performance
- The average performance of certified food value chains is significantly higher than their conventional reference value chains on most economic and social indicators
- The average performance of certified food value chains is similar to their conventional reference value chains on the most common environmental indicators, namely carbon and water footprints
- Although certified food is 61% more expensive, its extra-performance per euro is similar to other classical policy interventions than certification in the food sector (eg. taxes, subsidies)

1. Introduction

Consumer-oriented policy is increasingly seen as a key lever (Moran et al., 2018) – if not the cornerstone (Springmann et al., 2018) – of a sustainable food system, with diet change and food waste reduction at the forefront. However, environmentally friendly production practices could provide an equally promising way forward, provided that they can be communicated clearly to consumers and thereby inform their choice (Poore and Nemecek, 2018; W. K. Smith et al., 2019). This is precisely the role of certified food. In 2012, the Quality Package (Regulation (EU) No. 1151/2012) was passed in the EU to improve the operation of Geographical Indications (GIs) certification schemes, initially based on product typicality. The Regulation details the rationale for promoting GIs as a means to generate a fair return for farmers and processors and to enable consumers to make better-informed purchasing decisions through effective labelling. Similarly, the organic standard guarantees that neither farmers nor processors used synthetic chemicals. But beyond their initial promise – typicality for GIs and absence of chemicals for organic products – is certified food more sustainable than other food products?

Here we focus on organic and GI certifications, the two largest quality food standards in the EU with 4% and 5.7% respectively of total retail sales in European countries where data is available (Chever et al., 2012; FiBL, 2017). Two certifications are grouped under GI: Protected Geographical Indication (PGI) which guarantees the location and method of food processing and Protected Designation of

Origin (PDO) which guarantees the location and practices of both farmers and processors. Regarding GIs, previous studies have focused on their economic performance (Arfini et al., 2006; London Economics, 2008a; Vandecandelaere et al., 2018), showing that GI value chains add substantial value to their raw materials and to the labour employed in their production, to the benefit of producers, local and national economies. The existing analysis of the environmental and social performance of GIs is skim and entirely qualitative.

For organic farming, existing impact analyses are more comprehensive, covering both the economic and environmental pillars of sustainability and even some aspects of social sustainability. Several studies have applied Life-Cycle Assessments (LCA) to quantify differences in environmental impacts between organic and conventional agriculture (Nemecek et al., 2011; Thomassen et al., 2008). These assessments are not universally favourable to organic products (Meier et al., 2015a; Seufert and Ramankutty, 2017), in particular when indirect land-use change consequences of lower yield are included in the assessment (Bellora and Bureau, 2016). On the economic side, organic products clearly capture a price premium, which in general allows organic farms to obtain higher net results despite their lower yields (Crowder and Reganold, 2015a; European Commission, 2013; O. M. Smith et al., 2019). Social performance assessments are less common, yet organic value chains have been shown to generate more jobs and to attract younger and better educated workers (Finley et al., 2018; Koesling et al., 2008a; Mahé and Lerbourg, 2012). However, the environmental performance assessments of organic farming are usually conducted in isolation from socio-economic assessments (Pimentel and Burgess, 2014), which hinders the assessment of the broad sustainability performance of organic value chains, let alone the synergy or trade-off between different sustainability aspects. Moreover, methodological heterogeneity has been identified as an important pitfall in existing meta-analysis of the environmental performance of organic food (Meier et al., 2015b).

Here we assess the sustainability performance of EU certified food, questioning whether it outperforms conventional food and at which cost, and identifying synergies and trade-offs between different sustainability indicators. Compared to the existing literature, this assessment thus innovates along three key aspects:

- ✓ The same methodology is applied to a large – 52 – number of cases, providing a uniquely consistent picture of the sustainability of certified and non-certified products. For GI, our sample size is comparable to the largest existing studies which focus on only a few indicators within the same sustainability pillar. Meta-analyses of a few performances of organic farming have larger sample sizes (SM7), but our approach complements them as we trade sample size for consistent methodology and wide array of indicators.
- ✓ The performance criteria are assessed at the different levels of the value chains, including at least the farm and the processing levels;
- ✓ It provides the first quantitative evaluation of the environmental and social performance of GIs.

2. Material and methods

2.1. Data

2.1.1. Choice of products and their references

Twenty-six certified products were selected in thirteen countries (Table S5). Choices aimed at a diversity of sectors – animal, vegetal and unfed seafood/fish – and certifications – organic, PDO, PGI – while taking into account country-specific constraints (some certifications simply do not exist in some countries for some sectors). Ultimately, the cases are evenly distributed across certifications, while regarding sectors, the unfed seafood/fish sector has much few cases (3) than the vegetal and animal sectors.

In order to mitigate the influence of other possible drivers of performance than certification, such as country- or sector-specific features, only the difference between a certified product and its reference product is analysed. This strategy is inspired from the rationale of *controlled trials*. For this reason, detailed guidelines were designed to select only products with a comparable reference. For instance, for a sheep-milk cheese from Serbia initially in the sample, the only possible reference product within Serbia was a cow-milk cheese. For many indicators, the difference in outcome would likely have been more driven by the difference between cow and sheep systems than from the difference between certified and non-certified value chains. This product was therefore removed from the analysis as it was impossible to find a reference product meeting our guidelines.

2.1.2. Data collection

Guidelines were also provided for data collection (eg. relying to the extent possible on secondary data, interviewing key stakeholders in the value chain, …) to improve collection efficiency and data source comparability across case studies (see XXref for details). Note that a reference product can be an actual non-certified product (eg. Kissavos apples) or the average conventional product in the same country.

In order to be able to collect data on the 52 products within a reasonable amount of time, the variables requested for each indicator were divided in two categories, based on the literature:

- ✓ Key variables, which are the focus of the primary data collection efforts. Key variables are the variables which have been shown to be paramount drivers of relevant indicator;
- ✓ And secondary variables, which are collected only if readily available from existing datasets. Otherwise, default values from the literature are used.

The values of these variables were collected by in-country scientists according to the following prioritization protocol:

- ✓ Review of existing reports and databases on the value chain providing average values based on representative samples (secondary data);
- ✓ Ad-hoc surveys, in person or online, of a sample of farms and firms (primary data);
- ✓ Expert judgment elicitation, following the IPCC guidance (IPCC, 2006).

The detailed list of variables, all the spreadsheets including the raw data, their source, and the resulting estimated indicators can be downloaded at xxref.

2.1.3. Quality control procedure

Finally, a quality check procedure was put in place to mitigate the risk of misreporting data. The three main aspects of this procedure were 1) to write down all data, their date and source in a shared spreadsheet, 2) to distinguish the person who collected data from the person who estimated the indicator, and 3) to write down, for each product and indicator, a common interpretation of the estimated differences in performance that made sense for both the person who collected the data and the person who estimated the indicator. These interpretations are recorded in Arfini and Bellassen (2019). The ability of the procedure to result in homogeneous and unbiased data is confirmed by a cluster analysis of the results which shows no country or partner effect (SM 6) and by a comparison with existing studies for the few indicators and certifications for which they exist (SM7).

More details on the data collection procedure are provided in Bellassen et al. (xxin press).

2.2. *Indicator estimation*

2.2.1. Overview of indicators and minimal systematic comparison

Twenty-five indicators were designed to cover the performance of food value chains over the three sustainability pillars: economy, environment and society. The SAFA guidelines (Sustainability Assessment of Food and Agriculture systems) developed by the FAO (2013) formed the basis of indicator choice and design. However, SAFA falls short of detailing a full method to estimate indicators from collected primary data and to interpret them. A subset of SAFA themes were therefore operationalized into 25 actual indicators (Table 1).

In order to ensure a common basis of comparison between cases studies despite the heterogeneity in data availability, two orders of priority were established for indicators. The collection of variables necessary to estimate 13 "systematic indicators" (four economic; four environmental; five social) was thus prioritized over the variables necessary to estimate 11 "complementary indicators" (five economic; three environmental; two social). Altogether, the twenty-five indicators necessitate the collection of 150 variables (see XX for details on indicator design and estimation). We mostly focus here on the "systematic indicators". This subset was selected to be equally distributed over the three pillars, and to use indicators that were most common in their field and for which we had the least missing values. These thirteen indicators cover six of the sustainable development goals of the UN (SM 2, Table S2.1). Results from the entire set are nevertheless provided (SM 3 and SM 4) and used to discuss the key messages where relevant.

	Sustainability pillar	Indicator type	Indicator sub-type (code)	Level of analysis along the value chain
Systematic	Economic	Price premium	Price premium (Ec1.1, EUR kg^{-1})	One value per level of the value chain
		Profitability and value added distribution	Gross Operating Margin (Ec1.3, % of turnover)	
		Trade	Share of value exported within Europe (Ec1.5, % of turnover)	Single value for the whole value chain
		Local multiplier	Local multiplier (Ec2.1, no unit)	
	Environmental	Foodmiles	Carbon footprint per unit of product (En1.1, $kgCO_2e\ ton^{-1}$)	Single value for the whole value chain
		Carbon footprint	Distance travelled per unit of product (En2.1, $ton.km\ ton^{-1}$)	
		Water footprint	Blue water footprint (surface and ground water consumption, En 3.3, $m^3\ kg^{-1}$)	
			Grey water footprint (water pollution by nitrates, En 3.2, $m^3\ kg^{-1}$)	
	Social	Employment	Labour to production ratio (So 1.1, AWU ton^{-1})	One value per level of the value chain
		Governance	Bargaining power distribution (So2.1, no unit)	Single value for the whole value chain
		Social capital	Educational attainment (So3.1, no unit)	One value per

			Generational change (So5.1, no unit)	level of the value chain
			Gender equality (So5.2, no unit)	
Complementary	Economic	Profitability and value added distribution	Gross Value-added (Ec1.2, % of turnover)	One value per level of the value chain
		Profitability and value added distribution	Net result (Ec1.4, % of turnover)	
		Trade	Share of value exported outside Europe (Ec1.6, % of turnover)	Single value for the whole value chain
			Share of volume exported within Europe (Ec1.7, % of production)	
			Share of volume exported outside Europe (Ec1.8, % of production)	
	Environmental	Foodmiles	Carbon footprint per hectare (En1.2, $kgCO_2e\ ha^{-1}$)	One value per level of the value chain
		Carbon footprint	Emissions from transportation per unit of product (En2.2, $kgCO_2e\ ton^{-1}$)	
		Water footprint	Green water footprint (rainwater consumption, En3.1, $m^3\ kg^{-1}$)	
	Social	Employment	Turnover to labour ratio (So1.2, EUR AWU^{-1})	One value per level of the value chain
		Governance	Stability of the value chain level (So2.2, no unit)	
		Social capital	Wage level (So3.2, EUR AWU^{-1})	
			Gender equality index (So5.3, no unit)	

Table 1. List of indicators for sustainability assessment

2.2.2. Relative difference and value chain averages

Indicators are estimated for different levels of the value chain (farm level, processing level and, where relevant, retail level). To control for the influence of country and product type, we analyse relative differences between the certified product and its reference product rather than absolute values (**Equation 1**).

Equation 1.

$$rel\_diff_j = \frac{indic_{CERT,j} - indic_{REF,j}}{indic_{REF,j}}$$

where rel_diff_j is the relative difference at level j of the value chain, and $indic_{CERT,j}$ and $indic_{REF,j}$ are the indicator value at level j of the value chain for the certified and the reference product respectively.

For bargaining power distribution and for environmental indicators, the opposite of the relative difference is used so that a higher performance of certified food (eg. higher gross margin, lower water footprint) consistently corresponds to a positive relative difference.

In a second step, “value chain averages” are computed to evaluate the difference in performance for the entire value chain (Equation 2). For most indicators, “value chain averages” are simple averages of the indicator over the value chain levels for which it was estimated (farm, processing and, where relevant, retail). There are, however, two exceptions.

The first exception concerns indicators expressed on a *per ton* basis, that is the environmental indicators and the labour to production ratio. Because a life cycle assessment underlay these indicators, the value representative for the whole value chain must be calculated cumulatively. If one ton of ham requires 5 tons of live hog, the "value chain average" sums the footprint of 5 tons of live hog at farm level and 1 ton of ham at processing level rather than averaging the footprints of one ton of live hog and one ton of ham. This cumulative aggregation also allocates the footprint to all products (eg. ham versus the rest of the carcass at processing level) based on their relative economic value. For environmental indicators, this economic allocation is embedded in the original indicators. For labour to production ratio, the formula is provided in Equation 2.

The indicator on value chain stability is the second exception: its "value chain average" is the minimum of the "bargaining power" indicator across value chain levels (Equation 2).

Equation 2.

$$\begin{cases} \textbf{\textit{All indicators except environmental indicators and labour to production}}: \boldsymbol{VC_{average}} = \dfrac{\sum_{j=1}^{n} \boldsymbol{rel\_diff_j}}{\boldsymbol{n}} \\ \textbf{\textit{Environmental incators}}: \boldsymbol{VC_{average}} = \boldsymbol{rel\_diff_n} \\ \textbf{\textit{Labour to production ratio}}: \begin{cases} \boldsymbol{VC_{average}} = \dfrac{\boldsymbol{cum_{indic_{FQS}}} - \boldsymbol{cum_{indic_{REF}}}}{\boldsymbol{cum_{indic_{REF}}}} \\ \boldsymbol{cum_{indic_X}} = \dfrac{\boldsymbol{indic_{x,farm}}}{\boldsymbol{final_{prod_{ratio}}} \times (\boldsymbol{1} + \boldsymbol{coproducts_{farm}})} + \dfrac{\boldsymbol{indic_{x,proc}}}{(\boldsymbol{1} + \boldsymbol{coproducts_{proc}})} \end{cases} \\ \textbf{\textit{Value chain stability}}: \begin{cases} \boldsymbol{VC_{average}} = \dfrac{\boldsymbol{cum_{indic_{FQS}}} - \boldsymbol{cum_{indic_{REF}}}}{\boldsymbol{cum_{indic_{REF}}}} \\ \boldsymbol{cum_{indic_X}} = \min(\boldsymbol{indic_{x,farm}}, \boldsymbol{indic_{x,proc}}) \end{cases} \end{cases}$$

where $VC_{average}$ is the value chain average difference, rel_diff_j is the relative difference in performance at level j of the value chain (**Equation 1**), n is the lowest level of the value chain where the indicator could be estimated (most often the processing level), cum_indic_X is the cumulative indicator over different value chain levels for product X (either FQS or reference), $indic_{X,farm}$ and $indic_{X,proc}$ are the indicator value for product X at the farm and processing levels respectively, final_prod_ratio is the amount of raw product at farm level (eg. live hog) necessary for one ton of final product (eg. ham), and coproducts_farm and coproducts_proc are the value of coproducts (eg. rest of the carcass) expressed as a percentage of the value of the main product (eg. ham) at farm and processing levels respectively.

2.3. Statistics

The statistical analysis relies on non-parametric tests based on rank. The Wilcoxon signed-rank test to test whether a median is different from zero and the Kruskal-Wallis test to test whether different certifications (PDO, PGI, organic) belong to the same population. Because of the small sample size – 26 at most – these tests are better suited than classical parametric tests: they are less sensitive to outliers and they don't rely on the normality assumption which is difficult to ascertain in small samples.

These tests do not account for the uncertainty of the indicator estimates themselves, which could come from sampling error of primary variables used to estimate the indicator, modelling uncertainty (eg. carbon and water footprint), … These uncertainties are challenging to quantify exhaustively. The risk of systematic bias is however greatly reduced by the design of the analysis: the use of relative differences of estimates based on the same method is robust to additive modelling errors and even reduces the effect of non-additive biases.

For carbon footprint, a sensitivity analysis has been undertaken on two subjective modelling choices. Although these choices substantially changed the absolute values of the indicator, the relative differences were very similar to those presented here and the overall conclusions were unchanged XX2[a]. The R code used to compile indicators and conduct statistical tests is provided (SM 9).

3. Results and discussion

3.1. Sustainability performance of certified food

Certified food never performs significantly worse than conventional food on the key 13 indicators (Figure 1). This is still true for the broader set of 25 indicators (Figure S 1): only rainwater use is significantly higher (worse) for certified food, but the relevance of this indicator is debated, especially where rainwater is abundant (Schyns et al., 2019). For two thirds of the key indicators, the performance of certified food is significantly higher than conventional food.

Economically, certified products capture a price premium (+61%) and manage to translate it into a higher value added (+14%) and operating margin (+31%). This higher performance trickles down in the territory, although only to a small extent (local multiplier +6%): many feedstocks are locally sourced for both certified food and their reference products so that local sourcing constraints from the technical specifications do not translate into large differences in local multiplier effect. Dairy products are a typical example: cheese or yoghurt factories try to source their milk locally even in conventional value chains due to high transportation costs.

Socially, certified food creates more jobs (+14%) but, thanks to its price premium, its labor productivity – expressed in euros of turnover per unit of labor – is nevertheless higher (+32%) which translates into significantly higher wages (+32%). The higher labor intensity is explained by lower economies of scale – firms involved in certified value chains tend to be smaller than their conventional counterparts – and from technical specifications that often limit the automation of work. However, certified value chains do not attract more female workers and may be attracting more young workers (+18%) but with too much variability for this difference to be significant.

The environmental performance of certified products is broadly similar to their conventional reference. More precisely, certified products pollute less on a per hectare basis (-27% GHG emissions and -23% water pollution, see SM 3 and SM 4), thanks to technical specifications that often limit or – for organic products – forbid the use of synthetic fertilizers for example. But in the more common *footprint* indicators – which are expressed per ton – this higher performance is diluted by the lower yield of certified farms (-19%, see land_carbonXX2 for details). This lower productivity is particularly acute for organic products in our sample (-36%), an order of magnitude which is consistent with existing meta-analysis dedicated to the yield of organic farming. Seufert et al. (2012) report an average yield difference of -25%, and -34% when organic and conventional systems are most comparable. Ponisio et al. (2015) report an average different of -19%. In both cases however, the yield gap is found to vary substantially between products and pedo-climatic conditions.

Among the key environmental indicators, certified food only performs better on *foodmiles*, thanks to lower exports and local sourcing. While local sourcing is driven by the technical specifications of both geographical indications and organic farming (for animal products), the reason for lower exports differs: many geographical indications are only recognized in their domestic market, leading them to neglect international outlets which do not offer a price premium. For organic farming, there is a divide between high-income countries where domestic supply struggles to match domestic demand (little to no exports) and low-income countries where organic supply chains are almost entirely dedicated to exports.

[a] Reference anonymized for the purpose of the double-blind peer review process.

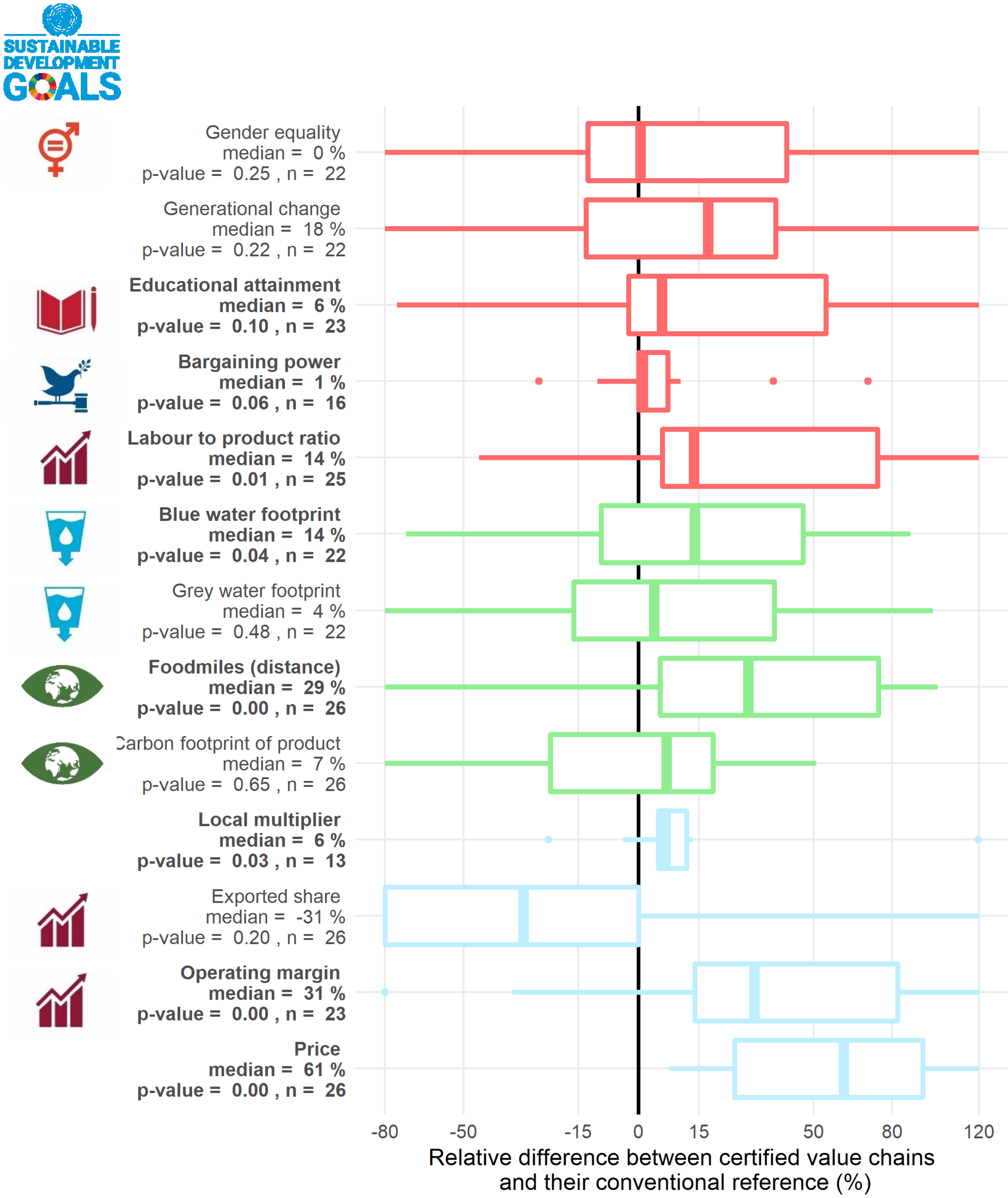


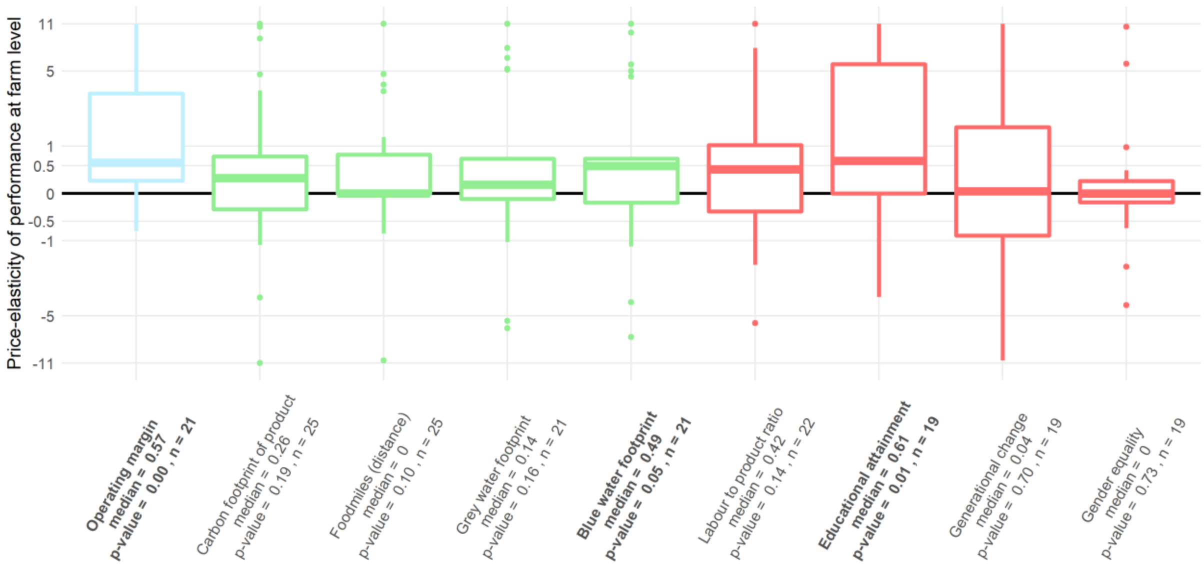
Figure 1. Performance difference between certified products and their conventional reference. The median performance difference between certified products and their conventional reference is expressed in percentage (%) under each key indicator, with positive results suggesting a higher performance and negative results pointing to a lower performance (see SM1 for details and formulas). The p-value indicates the probability that the median is different from zero (Wilcoxon

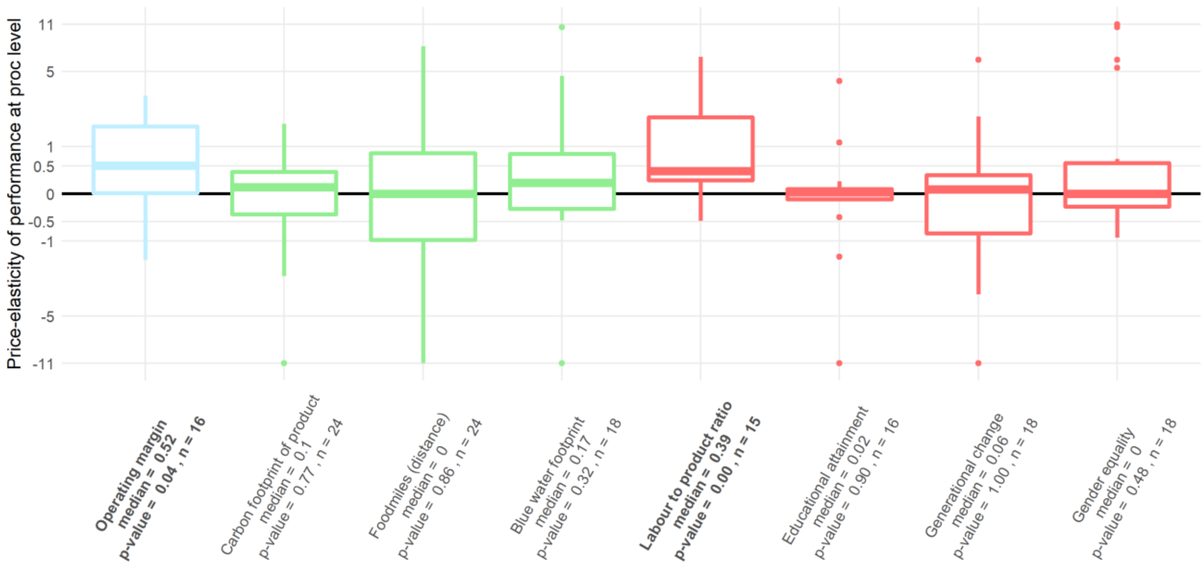
signed-rank test). Indicators for which certified food is significantly different from its reference (i.e. p-value < 0.1) are mentioned in bold. n indicates the number of certified products for which the indicator has been calculated. Boxes indicate the first and third quartiles with the median as a vertical bar within them. Whiskers indicate the largest values which is not further than 1.5 times the interquartile distance from the box. Points are outliers. The logo on the left-hand side indicates for which UN Sustainable Development Goal the indicator is relevant. This relevance is explained in Table S 1.

3.2. Certified food as an efficient way of producing public goods?

The relatively high marginal performance improvement per euro of price premium depict the Quality policy of the European Union – here the certification of organic and GI products – as an efficient way of producing positive economic, environmental and social externalities. Marginal performance improvement per euro is obtained by dividing the relative performance difference between certified products and their reference products by their relative price difference. It is an indicator of how much the performance improves for a given cost to the consumer. Where significantly different from zero, we find that the marginal performance improvement per euro, averaged per indicator across the value chain, is between 0.3 and 0.6 (Figure 2), except for local multiplier (0.07) and bargaining power distribution (0.04). This is only slightly lower than the own-price elasticity of food (Femenia, 2019) and is similar to the more relevant price-elasticity of demand in sugar soda (nutritional impact) (Guyomard et al., 2018) or the simulated response to a carbon tax (climate impact) (Bonnet et al., 2016; Caillavet et al., 2016). For many indicators however, the variability is large with some extreme values. This highlights the absence of direct causality in our experimental design regarding the relationships between price premium and over-performance and moderates the strength of the evidence provided by these pseudo-elasticity estimates. Moreover, these estimates neglect transaction costs (Bellassen et al., 2015) and windfall effects (Chabé-Ferret and Subervie, 2013), both of which have been shown to weigh heavily on the cost-efficiency of any policy, be it certification-based or not.

Marginal performance improvement per euro of price premium are slightly higher at farm level than at processing level: three dimensions are significantly higher than zero instead of two, and generational change is the only indicator for which the median marginal performance improvement per euro is lower at farm level (0.04) than at processing level (0.06). This argues for focusing policy intervention at farm level rather than processing level, especially as environmental indicators are cumulative, with the bulk of their value already determined at farm level.


Price-elasticity of performance at farm level
11
5
1
0.5
0
-0.5
-1
-5
-11
Operating margin median = 0.57 p-value = 0.00 , n = 21
Carbon footprint of product median = 0.26 p-value = 0.19 , n = 25
Foodmiles (distance) median = 0 p-value = 0.10 , n = 25
Grey water footprint median = 0.14 p-value = 0.16 , n = 21
Blue water footprint median = 0.49 p-value = 0.05 , n = 21
Labour to product ratio median = 0.42 p-value = 0.14 , n = 22
Educational attainment median = 0.61 p-value = 0.01 , n = 19
Generational change median = 0.04 p-value = 0.70 , n = 19
Gender equality median = 0 p-value = 0.73 , n = 19


Price-elasticity of performance at proc level
11
5
1
0.5
0
-0.5
-1
-5
-11
Operating margin median = 0.52 p-value = 0.04 , n = 16
Carbon footprint of product median = 0.1 p-value = 0.77 , n = 24
Foodmiles (distance) median = 0 p-value = 0.86 , n = 24
Blue water footprint median = 0.17 p-value = 0.32 , n = 18
Labour to product ratio median = 0.39 p-value = 0.00 , n = 15
Educational attainment median = 0.02 p-value = 0.90 , n = 16
Generational change median = 0.06 p-value = 1.00 , n = 18
Gender equality median = 0 p-value = 0.48 , n = 18

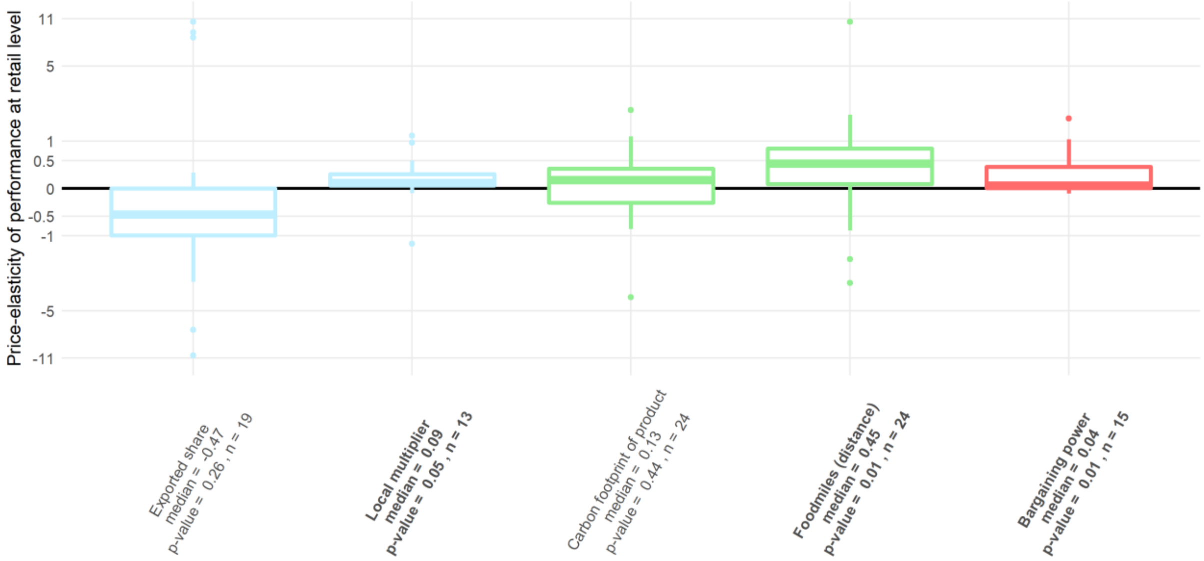


Figure 2. Marginal performance improvement of certified products per euro of price premium at farm, processing and retail levels. The marginal performance improvement per euro shows how many percentage points are gained for a given indicator per percentage point of price premium of the certified food. It is estimated here in a non-causal way by dividing the relative difference in performance by the relative difference in price for each product and indicator. Indicators pertaining to the entire value chain (export, bargaining power distribution and local multiplier) are displayed in the "retail" frame. Sample size (n) varies between value chain level because not all levels exist in all value chains and because data was not available for some levels in some value chains. The p-value indicates the probability that the median is different from zero (Wilcoxon signed-rank test). Indicators for which the marginal performance improvement per euro is significantly different from zero (i.e. p-value < 0.1) are mentioned in bold. n indicates the number of certified products for which the indicator has been calculated. Three indicators at retail level for which the sample size is smaller than 5 are not displayed. Marginal performance improvement per euro is bounded by [-11;11] for convenience of display. For the three product x level combinations where the price premium of the certified product is – very slightly – lower than zero, the marginal performance improvement is assumed to be infinite (eg. +∞ when the difference in performance is positive).

3.3. Price threshold for the profitability of certified food

Most of the performance improvements rely upon the existence of a price premium, without which firms have no incentive to comply with technical specifications. One way of assessing the scaling-up potential of certified food is therefore the price premium level which perfectly balances the extra costs of production. The median value of this threshold is 25% of the reference price, indicating a good resistance of certified chains to decreasing premiums, especially for organic. The processing level may be the weak link of certified value chains with median extra costs amounting to 52% of the reference price (see SM4 and XX3[b] for details). Scaling-up would of course create other challenges than decreasing price premiums (eg. land requirements associated with reduced yields (Muller et al., 2017; Searchinger et al., 2018)). On the other hand, all extra-costs need not but covered by the price premium: public subsidies or payments for environmental services can also contribute.

[b] Reference anonymized for the purpose of the double-blind peer review process.

3.4. Differences in performance between quality signs

PDO and organic value chains both get an edge on attractiveness to young workers, with a significantly better performance of 33% and 26% respectively, compared with a non-significant 18% difference for the entire sample (Figure 3). This confirms and broadens existing evidence for organic farmers and could be related to the better economic performance of these two value chains (eg. price premium of 73%, 58% and 40% for organic, PDO and PGI respectively with a p-value of the Kruskal-Wallis test close to 0.2 for organic and PGI). Other less documented features of certified value chains such as the preservation of cultural heritage could also contribute to attract young workers (Vandecandelaere et al., 2018).

Organic vegetal products perform 16% better than their conventional reference products regarding climate mitigation (SM4 and XXland_carbon_ref). This substantially reinforces existing but weak evidence on the relative merit of organic farming for this sector (Meier et al., 2015a). It is explained by the ban on mineral fertilizers: in the vegetal sector, mineral fertilizers are responsible for 40% of GHG emissions so that even large deficits in yield do not dilute the benefits from their absence.

Finally, for several indicators including bargaining power distribution and generational change, PGI value chains perform far worse than other quality signs, and similarly to conventional products (Figure 3). Indeed, the technical specifications of PGIs are often restricted to processing and do not cover farming practices. However, one must not overemphasize this result: the understanding of the difference between PDO and PGI is not always clear for stakeholders and regulators, so that several products which would likely qualify as PDOs only seek PGI recognition.

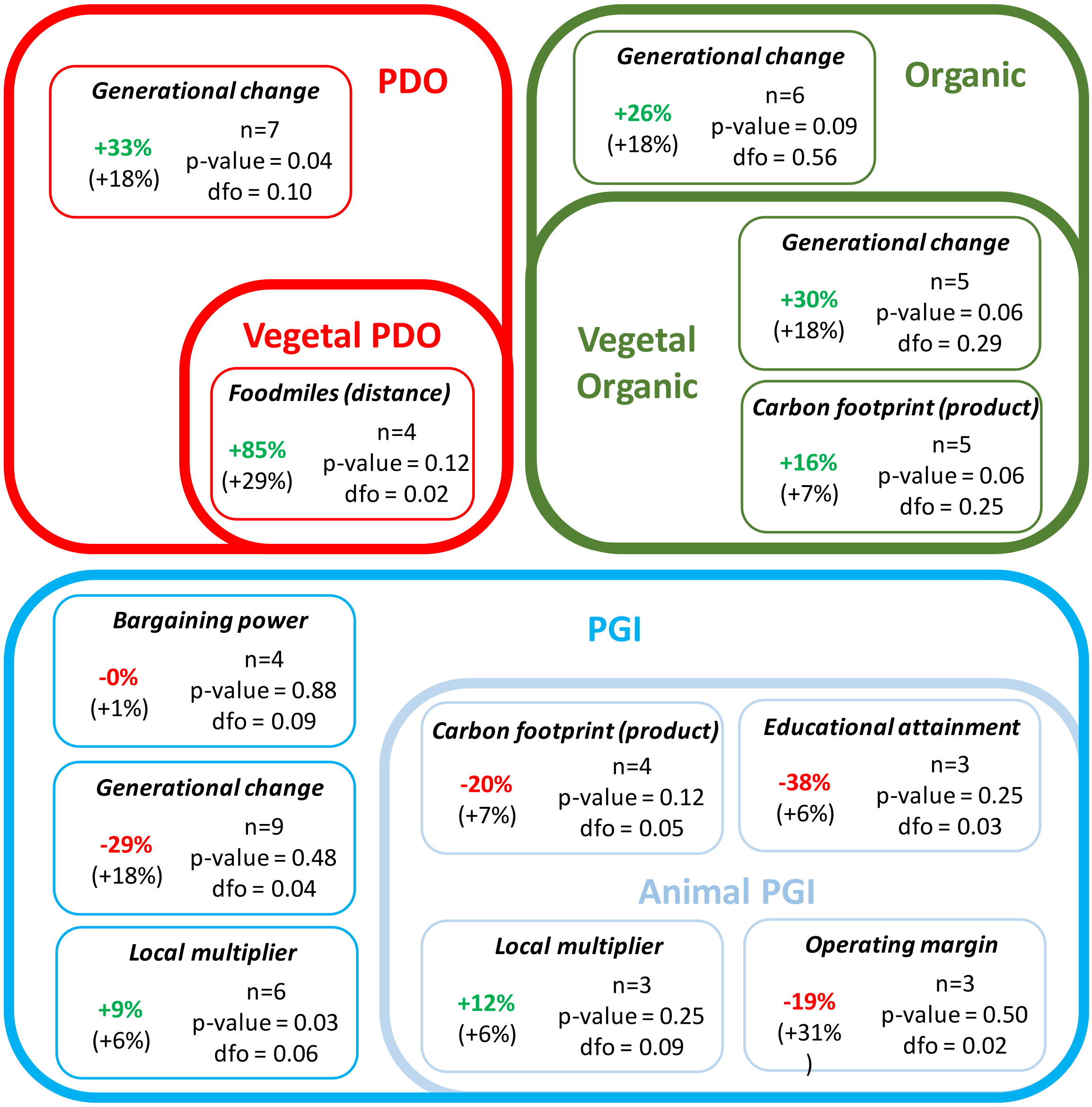


Figure 3. Sustainability performance of specific subgroups. The median difference in performance for the subgroup (eg. PDO) between certified products and their reference (in %) appears in color above the median difference of the entire sample in brackets. "n" indicates the number of certified products of the subgroup, "p-value" the probability that the median difference in performance between the subgroup and their reference products is different from zero (p-value of the Wilcoxon signed rank test) and "dfo" stands for "different from others" and provides the probability that the subgroup is different from the rest of the population (Kruskal-Wallis test). Display is restricted to the subgroup x indicator combinations for which either the latter is lower than 0.1 or the former is lower than 0.1 while the p-value for the entire population is higher than 0.1.

3.5. Synergies and trade-offs between sustainability indicators

High margins are associated with a younger and more educated the workforce (Figure 4): higher margins probably attract these workers, but younger and more qualified workers may also be more efficient and generate higher margins. To the contrary, high margins are negatively associated with a lower local multiplier, which may be explained by a more expensive local supply. Beyond significance level, one can also note that operating margin is positively correlated with almost all other indicators:

when margins are high, it is likely easier for value chains to "invest" in environmental and social performance.

Beside higher margins, younger workers and entrepreneurs are also associated with a more evenly distributed bargaining power and a lower local multiplier. The youth may be more innovative in the contractual relationship with other value chain levels – e.g. through long-term contracts – and more involved in collective undertakings such consortia or unions. This may be mediated by educational attainment, which is also higher for younger workers.

Finally, foodmiles performance is trivially deteriorating with increasing exports.

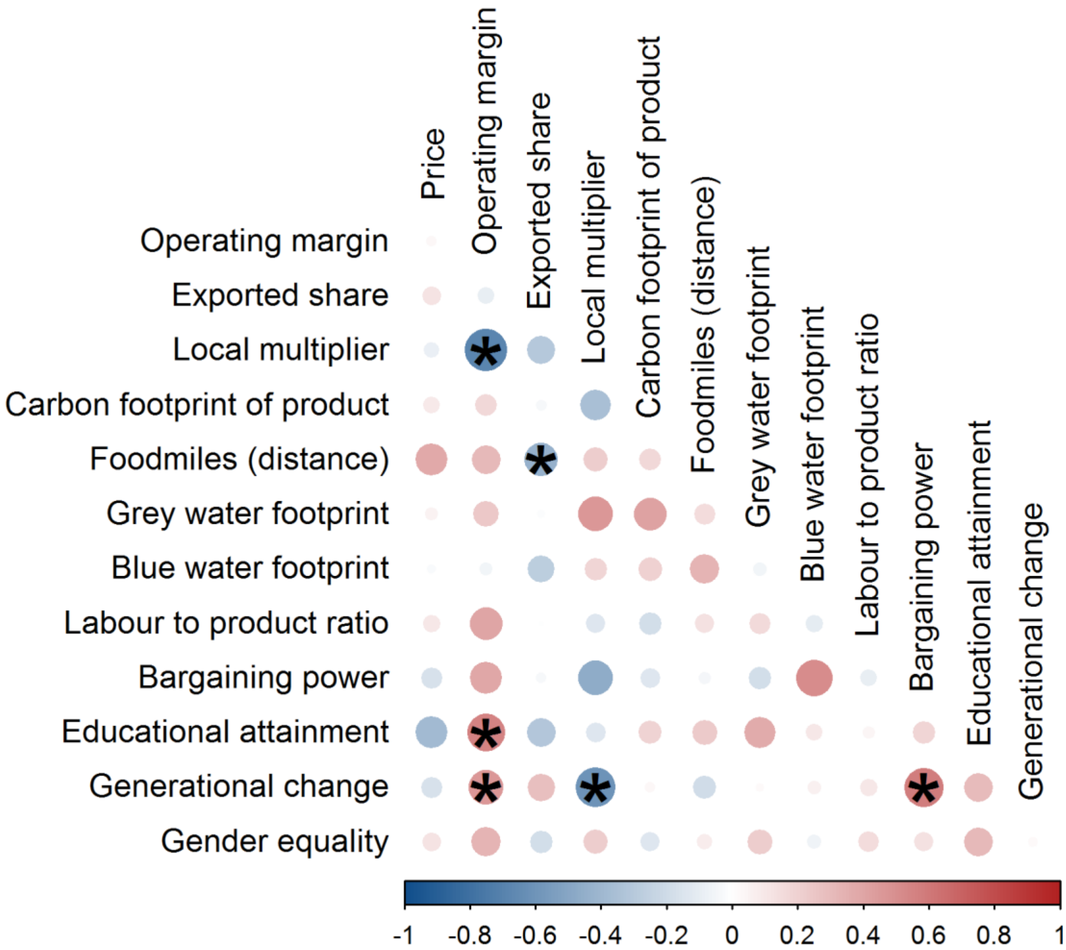


Figure 4. Correlation between indicators. The higher the absolute value of the Spearman correlation coefficient, the larger the corresponding circle. A star indicates that the coefficient is significantly different from zero (p-value < 0.05).

3.6. *Limits and possible improvements*

Several improvements and additions could be undertaken on the set of indicators. Most notably, our indicator for water pollution – grey water footprint – is restricted to nitrates. Pesticides were excluded due to difficulties in data collection in some countries, but proved feasible in several instances. Another interesting addition would be biodiversity. We did not find any biodiversity indicator robust enough to provide relevant information across all sectors: vegetal, terrestrial animal and seafood products. A possibility could be to introduce sector-specific indicators: fish population dynamics, for example, are a key sustainability indicator for wild catch seafood value chains.

Another crucial improvement would be to increase sample size. 26 paired products, each relying on primary variables averaged over multiple farms/firms, is a large sample compared to existing studies (see SM 7), especially if one excludes meta-analysis which tend to focus on a few indicators and are subject to methodological heterogeneity. However, it remains too modest to draw definitive conclusions on the relative merits of certified food. Moreover, increasing sample size would reduce uncertainty and thus possibly allow to identify other synergies and trade-offs between sustainability indicators. Most importantly, it would pave the way towards a more robust assessment of the causality of performance differences.

Two strategies could be explored to increasing sample size. The obvious first strategy would be to enriching our database by repeating the same assessments over new value chains in new countries. The unit cost of assessing one value chain and its reference – around 3 person.months – is accessible for value chain representatives willing to assess their sustainability performance. By opening the method, tools and database, we hope that the database thrive on contributions from future users.

The second strategy would be to estimate our indicators from existing institutional databases (eg. FADN for European farms, Amadeus for European processors). This would however require two key improvements in these databases: the identification of all quality signs with an adequate consideration of representativeness and the addition of a few necessary variables (eg. energy use, age-distribution of the workforce, …) for the assessment of the environmental and social performance.

When this will happen, more causal assessments of the benefits or harms of certification will become possible, using propensity-score matching or quasi-natural experiments. These techniques reduce the subjectivity in the definition of counterfactual/reference firms which is an important limit to all existing studies on the relative performance of GI and organic value chains, including the present one.

Finally, while our analysis focuses on the EU, there is no obstacle to using the same method for the assessment of the sustainability of food value chains elsewhere. As a matter of fact, it has been successfully used to assess eight products in South-East Asia. Moreover, our results are consistent with existing literature for the indicators for literature in other parts of the world exists (eg. value added (Arfini et al., 2006; Crowder and Reganold, 2015a; London Economics, 2008a; Vandecandelaere et al., 2018), carbon footprint of organic products (Meier et al., 2015a), …). This is a good sign for the global relevance of our findings, but extending the geographical coverage our sample would be necessary to ascertain this.

4. Conclusion

In conclusion, EU certified food are shown to perform better than conventional value chains for two thirds of our key indicators, and this performance comes at a reasonable price premium compared with other policy interventions. Cumulatively, our findings justify the policy interventions by the EU to support these standards. Our findings also legitimate recent initiatives by the organic or GI standards to broaden their objectives such as including environmental clauses in the technical specifications of GIs (INAO, 2016) or including social clauses in the technical specifications of private organic standards (The Organic Research Centre, UK and Padel, 2018).

Data availability

Both raw data and estimated indicators have been deposited in an open dataset and can be downloaded from XX [link to the Dataverse repository, to be inserted when the manuscript is accepted. The repository contains what is currently provided in SM 1, 4, 8 and the code file mentioned in SM 9. It will therefore replace these supplementary materials when the manuscript is accepted].

Code availability

Code to estimate the indicators, perform the statistical tests and display the figures has been deposited in the aforementioned open dataset.

References

Areté, 2014. Study on assessing the added value of PDO/PGI products.

Arfini, F., Bellassen, V. (Eds.), 2019. Sustainability of European Food Quality Schemes: Multi-Performance, Structure, and Governance of PDO, PGI, and Organic Agri-Food Systems. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-27508-2

Arfini, F., Boccaletti, S., Caceres Clavero, F., Colinet, P., Desquilbet, M., Garcia, M., Garcia Collado, R., Gay, H., Giacomini, C., Gijsbers, G., Giray, H., Hassan, D., Monier-Dilhan, S., Moro, D., Nahon, D., Orozco, V., Réquillart, V., Roseboom, H., Schneider, A., Sckokai, P., van der Zee, F., 2006. Economics of food quality assurance and certification schemes managed within an integrated supply chain. ETEPS.

Bellassen, V., Stephan, N., Afriat, M., Alberola, E., Barker, A., Chang, J.-P., Chiquet, C., Cochran, I., Deheza, M., Dimopoulos, C., Foucherot, C., Jacquier, G., Morel, R., Robinson, R., Shishlov, I., 2015. Monitoring, reporting and verifying emissions in the climate economy. Nature Climate Change 5, 319–328. https://doi.org/10.1038/nclimate2544

Bellora, C., Bureau, C., 2016. How green is organic ? Indirect effects of making EU agriculture greener (Presented at the 19th Annual Conference on Global Economic Analysis, Washington DC, USA). Global Trade Analysis Project (GTAP), Department of Agricultural Economics, Purdue University, West Lafayette, IN.

Boncinelli, F., Contini, C., Romano, C., Scozzafava, G., Casini, L., 2017. Territory, environment, and healthiness in traditional food choices: insights into consumer heterogeneity. International Food and Agribusiness Management Review 20, 143–157. https://doi.org/10.22434/IFAMR2015.0177

Bonnet, C., Bouamra-Mechemache, Z., Corre, T., 2016. An environmental tax towards more sustainable food consumption: empirical evidence of the French meat and marine food consumption (No. 16–639), TSE Working Paper. Toulouse School of Economics, Toulouse, France.

Caillavet, F., Fadhuile, A., Nichèle, V., 2016. Taxing animal-based foods for sustainability: environmental, nutritional and social perspectives in France. European Review of Agricultural Economics jbv041. https://doi.org/10.1093/erae/jbv041

Chabé-Ferret, S., Subervie, J., 2013. How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. Journal of Environmental Economics and Management 65, 12–27. https://doi.org/10.1016/j.jeem.2012.09.003

Chever, T., Renault, C., Renault, S., Romieu, V., 2012. Value of production of agricultural products and foodstuffs, wines, aromatised wines and spirits protected by a geographical indication (GI). European Commission & AND International, Brussels, Belgium.

Clark, M., Tilman, D., 2017. Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. Environ. Res. Lett. 12, 064016. https://doi.org/10.1088/1748-9326/aa6cd5

Crowder, D.W., Reganold, J.P., 2015a. Financial competitiveness of organic agriculture on a global scale. PNAS 112, 7611–7616. https://doi.org/10.1073/pnas.1423674112

Crowder, D.W., Reganold, J.P., 2015b. Financial competitiveness of organic agriculture on a global scale. PNAS 112, 7611–7616. https://doi.org/10.1073/pnas.1423674112

European Commission, 2017. Distribution of the added value of the organic food chain.

European Commission, 2013. Organic versus conventional farming, which performs better financially? (No. 4), Farm Economics Briefs. European Commission, Brussels, Belgium.

FAO, 2013. SAFA Indicators. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Femenia, F., 2019. A Meta-Analysis of the Price and Income Elasticities of Food Demand. German Journal of Agricultural Economics 68, 77–98.

FiBL, 2017. FiBL Statistics [WWW Document]. URL https://statistics.fibl.org/ (accessed 5.24.19).

Finley, L., Chappell, M.J., Thiers, P., Moore, J.R., 2018. Does organic farming present greater opportunities for employment and community development than conventional farming? A survey-based investigation in California and Washington. Agroecology and Sustainable Food Systems 42, 552–572. https://doi.org/10.1080/21683565.2017.1394416

Guyomard, H., Détang-Dessendre, C., Réquillart, V., Soler, L.G., 2018. La Politique agricole commune doit-elle intégrer des objectifs de lutte contre le surpoids et l'obésité ? INRA Sciences Sociales 7 p.

INAO, 2016. L'agro-écologie s'installe dans les SIQO. La lettre de l'INAO 1.

IPCC, 2006. 2006 IPCC guidelines for national greenhouse gas inventories. IGES, Hayama, Japan.

Koesling, M., Flaten, O., Lien, G., 2008a. Factors influencing the conversion to organic farming in Norway. International Journal of Agricultural Resources, Governance and Ecology 7, 78–95.

Koesling, M., Flaten, O., Lien, G., 2008b. Factors influencing the conversion to organic farming in Norway. International Journal of Agricultural Resources, Governance and Ecology 7, 78–95.

Läpple, D., Kelley, H., 2015. Spatial dependence in the adoption of organic drystock farming in Ireland. Eur Rev Agric Econ 42, 315–337. https://doi.org/10.1093/erae/jbu024

London Economics, 2008a. Evaluation of the CAP policy on protected designations of origin (PDO) and protected geographical indications (PGI).

London Economics, 2008b. Evaluation of the CAP policy on protected designations of origin (PDO) and protected geographical indications (PGI).

Mahé, T., Lerbourg, J., 2012. Des agriculteurs bio diplômés, jeunes et tournés vers les circuits courts. Agreste Primeur 4 p.

Meier, M.S., Stoessel, F., Jungbluth, N., Juraske, R., Schader, C., Stolze, M., 2015a. Environmental impacts of organic and conventional agricultural products – Are the differences captured by life cycle assessment? Journal of Environmental Management 149, 193–208. https://doi.org/10.1016/j.jenvman.2014.10.006

Meier, M.S., Stoessel, F., Jungbluth, N., Juraske, R., Schader, C., Stolze, M., 2015b. Environmental impacts of organic and conventional agricultural products – Are the differences captured by life cycle assessment? Journal of Environmental Management 149, 193–208. https://doi.org/10.1016/j.jenvman.2014.10.006

Moran, D., Wood, R., Hertwich, E., Mattson, K., Rodriguez, J.F.D., Schanes, K., Barrett, J., 2018. Quantifying the potential for consumer-oriented policy to reduce European and foreign carbon emissions. Climate Policy 1–11. https://doi.org/10.1080/14693062.2018.1551186

Muller, A., Schader, C., El-Hage Scialabba, N., Brüggemann, J., Isensee, A., Erb, K.-H., Smith, P., Klocke, P., Leiber, F., Stolze, M., Niggli, U., 2017. Strategies for feeding the world more sustainably with organic agriculture. Nature Communications 8, 1290. https://doi.org/10.1038/s41467-017-01410-w

Nemecek, T., Dubois, D., Huguenin-Elie, O., Gaillard, G., 2011. Life cycle assessment of Swiss farming systems: I. Integrated and organic farming. Agricultural Systems 104, 217–232. https://doi.org/10.1016/j.agsy.2010.10.002

Pimentel, D., Burgess, M., 2014. An Environmental, Energetic and Economic Comparison of Organic and Conventional Farming Systems, in: Pimentel, D., Peshin, R. (Eds.), Integrated Pest Management. Springer Netherlands, Dordrecht, pp. 141–166. https://doi.org/10.1007/978-94-007-7796-5_6

Ponisio, L.C., M'Gonigle, L.K., Mace, K.C., Palomino, J., de Valpine, P., Kremen, C., 2015. Diversification practices reduce organic to conventional yield gap. Proc. R. Soc. B 282, 20141396. https://doi.org/10.1098/rspb.2014.1396

Poore, J., Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. Science 360, 987–992. https://doi.org/10.1126/science.aaq0216

Schyns, J.F., Hoekstra, A.Y., Booij, M.J., Hogeboom, R.J., Mekonnen, M.M., 2019. Limits to the world's green water resources for food, feed, fiber, timber, and bioenergy. Proc Natl Acad Sci USA 116, 4893–4898. https://doi.org/10.1073/pnas.1817380116

Searchinger, T.D., Wirsenius, S., Beringer, T., Dumas, P., 2018. Assessing the efficiency of changes in land use for mitigating climate change. Nature 564, 249. https://doi.org/10.1038/s41586-018-0757-z

Seufert, V., Ramankutty, N., 2017. Many shades of gray—The context-dependent performance of organic agriculture. Science Advances 3, e1602638. https://doi.org/10.1126/sciadv.1602638

Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. Nature 485, 229–232. https://doi.org/10.1038/nature11069

Smith, O.M., Cohen, A.L., Rieser, C.J., Davis, A.G., Taylor, J.M., Adesanya, A.W., Jones, M.S., Meier, A.R., Reganold, J.P., Orpet, R.J., Northfield, T.D., Crowder, D.W., 2019. Organic Farming Provides Reliable Environmental Benefits but Increases Variability in Crop Yields: A Global Meta-Analysis. Front. Sustain. Food Syst. 3, 82. https://doi.org/10.3389/fsufs.2019.00082

Smith, W.K., Nelson, E., Johnson, J.A., Polasky, S., Milder, J.C., Gerber, J.S., West, P.C., Siebert, S., Brauman, K.A., Carlson, K.M., Arbuthnot, M., Rozza, J.P., Pennington, D.N., 2019. Voluntary sustainability standards could significantly reduce detrimental impacts of global agriculture. PNAS 201707812. https://doi.org/10.1073/pnas.1707812116

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B.L., Lassaletta, L., Vries, W. de, Vermeulen, S.J., Herrero, M., Carlson, K.M., Jonell, M., Troell, M., DeClerck, F., Gordon, L.J., Zurayk, R., Scarborough, P., Rayner, M., Loken, B., Fanzo, J., Godfray, H.C.J., Tilman, D., Rockström, J., Willett, W., 2018. Options for keeping the food system within environmental limits. Nature 562, 519. https://doi.org/10.1038/s41586-018-0594-0

The Organic Research Centre, UK, Padel, S., 2018. Setting and reviewing standards for organic farming, in: University of Bonn, Germany, Köpke, U. (Eds.), Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, pp. 345–364. https://doi.org/10.19103/AS.2017.0029.14

Thomassen, M.A., van Calker, K.J., Smits, M.C.J., Iepema, G.L., de Boer, I.J.M., 2008. Life cycle assessment of conventional and organic milk production in the Netherlands. Agricultural Systems 96, 95–107. https://doi.org/10.1016/j.agsy.2007.06.001

Tuomisto, H.L., Hodge, I.D., Riordan, P., Macdonald, D.W., 2012. Does organic farming reduce environmental impacts? – A meta-analysis of European research. Journal of Environmental Management 112, 309–320. https://doi.org/10.1016/j.jenvman.2012.08.018

Vandecandelaere, E., Teyssier, C., Barjolle, D., Jeanneaux, P., Fournier, S., Beucherie, O., FAO, European Bank for Reconstruction and Development, 2018. Strengthening sustainable food systems through geographical indications: an analysis of economic impacts.

5. SM 1: detailed method for data collection and indicator estimation

See: SM1_Method_indicators.docx

6. SM 2: correspondence with UN Sustainable Development Goals

UN Sustainable Development Goals (SDGs)	Indicators	Part of the SDGs for which the indicator is relevant
Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all	Educational attainment	4.3 By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university 4.4 By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship
Goal 5. Achieve gender equality and empower all women and girls	Gender equality Gender equality index	5.1 End all forms of discrimination against all women and girls everywhere 5.5 Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life 4.5 By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations
Goal 6. Ensure availability and sustainable management of water and sanitation for all	Grey water footprint Blue water footprint	6.3 By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally 6.4 By 2030, substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals and supply of freshwater to address water scarcity and substantially reduce the number of people suffering from water scarcity 6.5 By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate
Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	Exported share / Export index Value added Gross operating margin Net result Labour-to-production Turnover-to-labour Wage level	8.2 Achieve higher levels of economic productivity through diversification, technological upgrading and innovation, including through a focus on high-value added and labour-intensive sectors 8.5 By 2030, achieve full and productive employment and decent work for all women and men, including for young people and persons with disabilities, and equal pay for work of equal value

13 CLIMATE ACTION	Goal 13. Take urgent action to combat climate change and its impacts	Carbon footprint Foodmiles	13.2 Integrate climate change measures into national policies, strategies and planning
16 PEACE, JUSTICE AND STRONG INSTITUTIONS	Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	Bargaining power distribution	16.6 Develop effective, accountable and transparent institutions at all levels 16.7 Ensure responsive, inclusive, participatory and representative decision-making at all levels

Table S2.1. Correspondence between our indicators and UN Sustainable Development Goals

7. SM 3: performance difference for all indicators

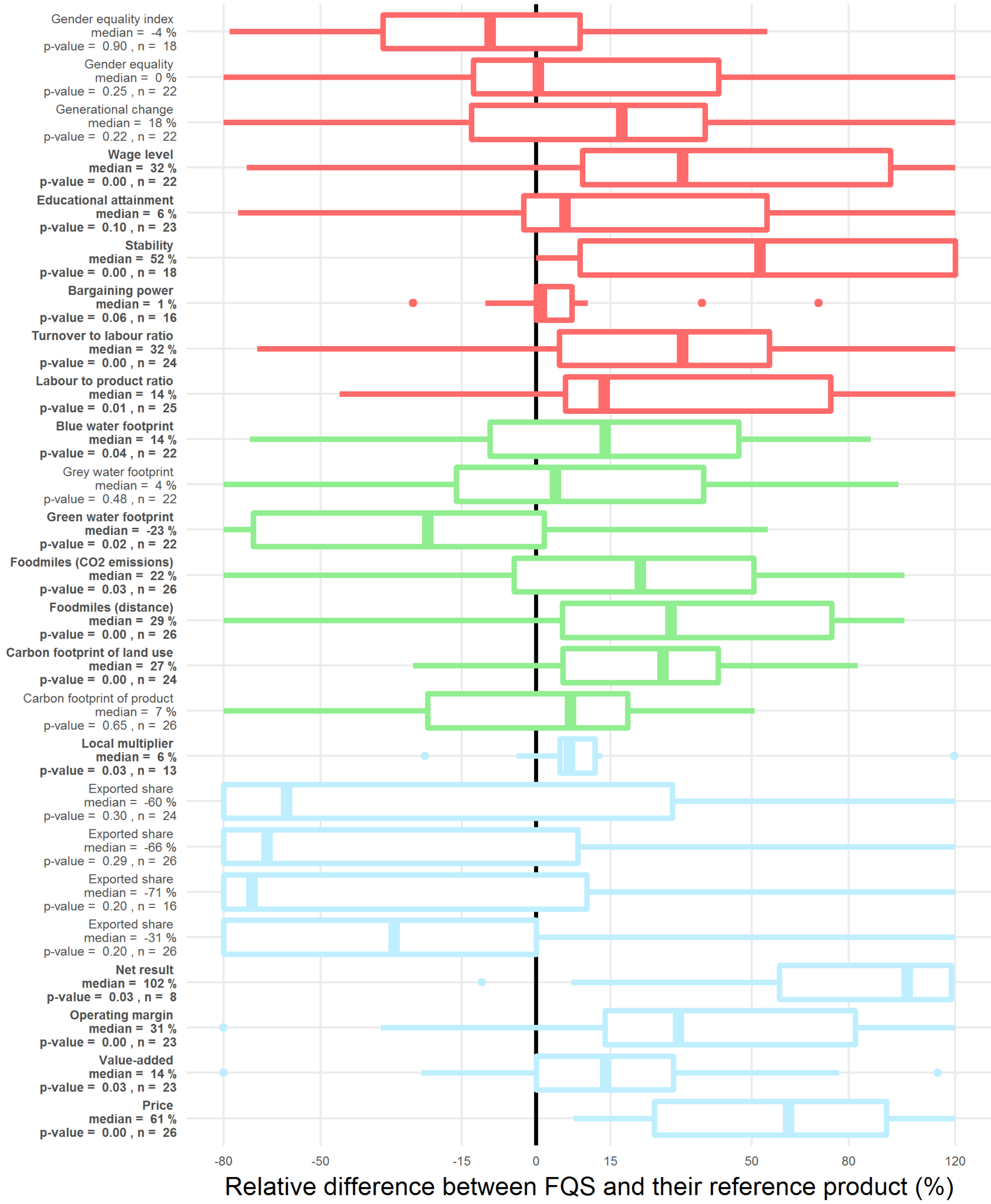


Figure S3. 1. Performance difference between certified products and their conventional reference

The median performance difference between certified products and their conventional reference is expressed in percentage (%) under each key indicator, with positive results suggesting a higher performance and negative results pointing to a lower performance. The p-value indicates the probability that the median is different from zero (Wilcoxon signed-rank test). Indicators for which certified food is significantly different from its reference (i.e. p-value < 0.1) are mentioned in bold. N indicates the number of certified products for which the indicator has been calculated. Boxes indicate the first and third quartiles with the median as a vertical bar within them. Whiskers indicate the largest values which is not further than 1.5 times the interquartile distance from the box. Points are outliers.

8. SM 4: exhaustive list of indicators and statistical tests

See SM4_exhaustive_set_indicators.xlsx

9. SM 5: characteristics of the products in the sample

Product name	Country	Product description	Type of FQS	Processed?	Turnover (EUR yr-1)	Reference product
Dalmatian prosciutto	Croatia	Dry pork ham	PGI	Yes	4 198 792	Local non-PGI firm
PDO olive oil	Croatia	Olive oil	PDO	Yes	254 875	National average
Comte cheese	France	Hard pressed cooked cheese from cow milk	PDO	Yes	504 191 550	Similar uncertified cheese (Emmental) or national average (cow cheese)
Organic flour	France	Wheat flour	Organic	Yes	34 800 000	National average
Saint-Michel bay bouchot mussels	France	Mussels produced on "bouchots"	PDO	No	25 450 688	National average (TSG Bouchot mussels)
Organic rice	France	Rice	Organic	Yes	17 640 000	Non-organic rice (mostly PGI)
Organic pork	Germany	Raw meat	Organic	Yes	69 000 000	National average
Organic yoghurt	Germany	Organic yoghurt from cow milk	Organic	Yes	387 000 000	National average
Zagora apples	Greece	Apple	PDO	No	10 107 900	Kissavos apples (non-GI apples from another region)
Kastoria apples	Greece	Apple	PGI	No	7 500 000	Kissavos apples (non-GI apples from another region)
Gyulai sausage	Hungary	Sausage	PGI	Yes	55 000 000	Non-PGI Hungarian sausage
Kalocsai paprika powder	Hungary	Paprika powder	PDO	Yes	10 749 180	Imported Chinese pepper milled in Hungary
Parmigiano Reggiano cheese	Italy	Hard pressed cooked cheese from cow milk	PDO	Yes	1 009 943 894	Biraghi cheese (similar non-PDO cheese)
Organic tomatoes	Italy	Organic tomato	Organic	No	68 574 011	Conventional processed tomatoes in the same region (Emilia-Romagna)
Opperdoezer Ronde potatoes	Netherlands	Early potato	PDO	No	2 771 250	Regular potato in neighbouring IJsselmeerpolders region
Lofoten stockfish	Norway	Dried fish	PGI	No	71 243 333	Clipfish (cod)
Organic salmon	Norway	Salmon	Organic	Yes	144 707 200	Conventional salmon
Organic pasta	Poland	Pasta	Organic	Yes	523 810	Simulated conventional farms with sample characteristics
Kaszubska strawberries	Poland	Strawberry	PGI	No	637 822	National average
Sjenica cheese	Serbia	Sheep cheese	PGI	Yes	1 213 041	National average (cow cheese)
Organic raspberries	Serbia	Frozen rapsberries	Organic	Yes	4 372 069	National average
Sobrasada Porc Negre	Spain	Raw, cure saussage from pork meat	PGI	Yes	1 802 520	National average
Ternasco de Aragon	Spain	Unprocessed lamb meat	PGI	No	16 965 438	Non-PGI lamb in the same region (Aragon)
Thung Kula Rong-Hai Hom Mali rice	Thailand	Rice	PGI	No	300 739 838	Non certified rice from the same region (90% of GI rice is organic as well)
Doi Chaang coffee	Thailand	Coffee	PGI	Yes	756 000 000	Non-PGI coffee from the same province
Phu Quoc Fish Sauce	Vietnam	Fish sauce	PDO	Yes	3 426 166	Non-PDO fish sauce from same region
Buon Ma Thuot coffee	Vietnam	Coffee	PGI	Yes	89 577 676	Non-PGI coffee from Dak Lak province in Vietnam

Table S5.1. Sample characteristics. Red, green and blue lines highlight the sector (animal, vegetal and seafood respectively). The indicated turnover is either at processing or farm level, whichever is higher. XX1[c] provides a detailed description of each value chain, its structure, its governance and its sustainability performance.

As a result of the quality check procedure described in SM 1, the applicant-PGI Sjenica sheep cheese was removed from the sample: its reference product is a cow cheese, and the difference between cow and sheep was identified as the main drivers for the differences in performance. The procedure also resulted in the exclusion of employment indicators at processing level for PGI Doi Chaang coffee and PGI TKR Hom Mali rice for which differences between certified food and its reference were both high and unexplained.

10. SM 6: classification of products, no partner or country effect

SPAD software[d] is used to identify clusters with similar sustainability performances. The basis for the clustering algorithm is the quartile of performance to which a given product belongs for a given indicator (when an indicator could not be estimated for a given product, the product is assigned to a fifth class for this indicator). The statistical automated processing chain then combines: a coding step followed by a multiple correspondence analysis (MCA) and a mixed classification procedure carried out by chaining a first k-means clustering with around 10 randomly selected centers. The 10 stable classes thus formed are then aggregated by a hierarchical classification followed by consolidation. The objective of consolidation is to reassign elements from one class to another class to which they are closer in order to improve homogeneity within classes.

Clusters A-C are the most performing clusters (9-10 indicators), where C is particularly performing on economic indicators while A is particularly performing on environmental and social indicators (Figure S6.1). Clusters D and E regroup products which combine good economic and social performances.

Clusters are then characterized with a set of additional variables describing the certified products (partner, country, team in charge of data collection, product type, sector, label type), and geographical context (gross domestic product, population, education level, number of designations, agricultural area, share of organic areas). For continuous variables, a simple t-test is used to identify clusters characterized by a difference for these additional variables. For discrete variables, a test based on a hypergeometric distribution is used instead.

This characterization does not identify many additional variables for which a class is significantly different from others. Cluster E is characterized by countries where both the number of quality signs and the share of arable area under organic farming are high. The seafood sector is overrepresented in cluster C and so is one data collection team in cluster B. Overall, the effect of the team which collected data is only one visible for one team in one cluster, which supports the reliability of the data collection methodology and in particular the absence of a significant influence from data collectors.

[c] Reference anonymized for the purpose of the double-blind peer review process.

[d] Coheris Corp., 2017, SPAD Data Mining, v. 9.1 Software Program. Coheris Corp., Suresnes, France.

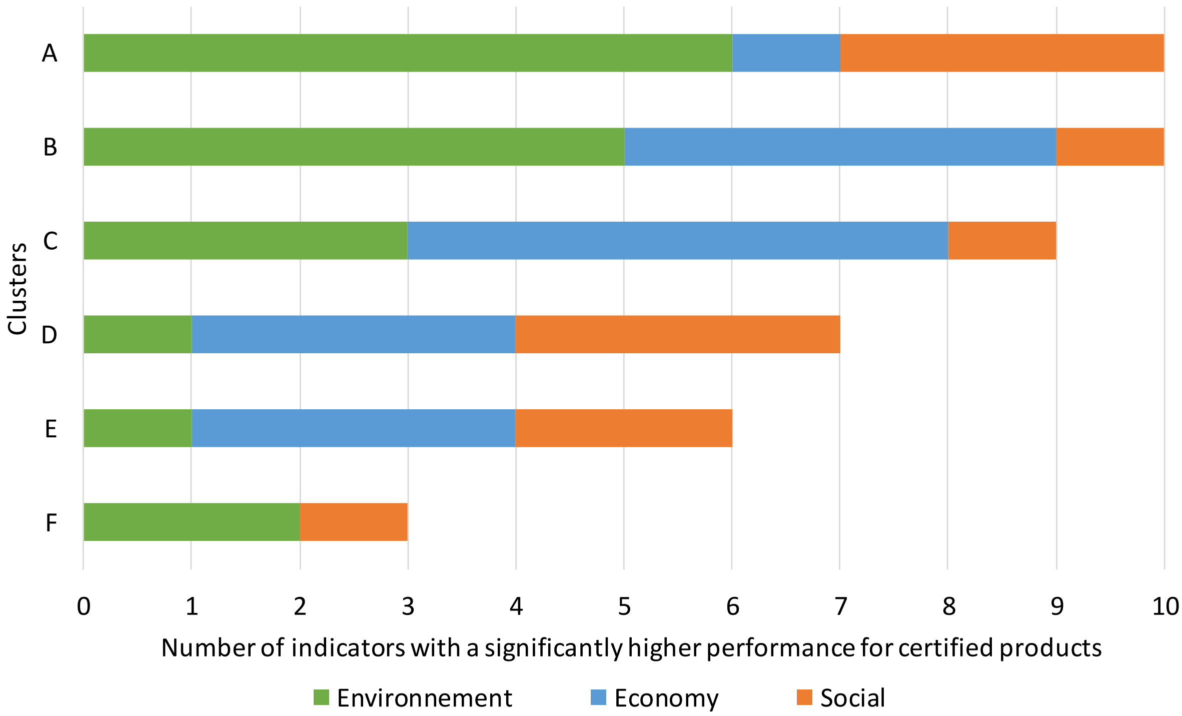


Figure S6.1. Sustainability performance of six homogenous clusters of certified products. Certified products as classified into six homogeneous clusters (A-F) based on their performance for each indicator.

11. SM 7: comparison of our results with existing studies with substantial sample size

Our findings are consistent with studies relying on a substantial (>5) sample size for the indicator x certification combinations for which literature exists (Table S7.1). Our price premium for organic products is close to a recent European study and higher than a global study, which could be explained by a higher willingness to pay of European consumers. This could also explain why we find a higher difference in margin than the global study. For geographical indications, margin difference is within the range of existing studies, which is therefore likely the case of price premium although the most recent study does not report it (Areté, 2014). Similarly to existing meta-analysis, we find that the carbon footprint of organic products is not significantly different from their conventional references. Finally, our finding of a higher labour intensity of organic farming is consistent with (Finley et al., 2018), although the difference is not significantly different from zero in our study.

Indicator	Certification	Our study	Other studies
Price premium	Organic	+73% (N = 8, p-value = 0.01)	+32% (N = 77, p-value = na) (Crowder and Reganold, 2015b) +53% (N = 35, p-value = na) (European Commission, 2017)
Price premium	Geographical indications	+55% (N = 18, p-value = 2e-4)	+13% (N = 15, p-value = na) (London Economics, 2008b)
Gross operating margin	Organic	+59% (N = 8, p-value = 0.01)	+14% (N = 77, p-value = 2e-4) (Crowder and Reganold, 2015b)

Gross operating margin	Geographical indications	+26% (N = 15, p-value = 0.05)	+7% (N=14, p-value = na) (London Economics, 2008b) +32% (N=13, p-value = na) (Areté, 2014)
Carbon footprint	Organic	-14% (N = 8, p-value = 0.11)	+0% (N=33, p-value = na) (Meier et al., 2015b) -4% (N=44, p-value = 0.59) (Clark and Tilman, 2017) +0% (N = 23, p-value > 0.05) (Tuomisto et al., 2012)
Labour to product ratio	Organic	+25% (N = 8, p-value = 0.2)	+6%[e] (N = 45, p-value < 0.05) (Finley et al., 2018)
Educational attainment	Organic	+5% (N = 8, p-value = 0.4)	Positive or null[f] (Boncinelli et al., 2017; Koesling et al., 2008b; Läpple and Kelley, 2015)

Table S7.1. Comparison of our results with existing studies with substantial sample size. These studies are generally focused on a single indicator and a single certification (eg. organic). Most of them are meta-analysis. The values displayed are the median difference between certified products and their reference (in %), the sample size, and the probability that the median difference is null (p-value, where available). Note that what counts as one unit in the sample size is variable (eg. one study for meta-analysis, one product or product x country combination for other studies) but is comparable to our definition of one unit (one product).

12. SM 8: underlying database, raw data and carbon footprint and foodmiles calculators

See SM7_raw_data.zip and SM7_C_footprint_calculator.7z

13. SM 9: R code for figure compilation and statistical tests

The R code is provided in SM9_R_code.org

[e] In [42], the indicator is expressed in AWU/ha. Expressed in AWU/ton as is the case in our study, the difference would likely be higher. The sample size is also not directly comparable as it refers to the number of farm in [42] as opposed to the number of studies or products in all other references in the table.

[f] Our indicator is difficult to compare directly with other studies but out of three identified, two find a positive correlation between farmer education and the organic certification.