

University of Parma Research Repository

Review of oscillating water column converters

This is the peer reviewd version of the followng article:

Original

Review of oscillating water column converters / Delmonte, Nicola; Barater, Davide; Giuliani, Francesco; Cova, Paolo; Buticchi, Giampaolo. - In: IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS. - ISSN 0093-9994. - 52:2(2016), pp. 1698-1710. [10.1109/TIA.2015.2490629]

Availability: This version is available at: 11381/2805957 since: 2021-10-15T18:00:08Z

*Publisher:* Institute of Electrical and Electronics Engineers Inc.

Published DOI:10.1109/TIA.2015.2490629

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

#### 1 The economic, environmental and social performance of European certified food

2 Valentin Bellassen ,\*, Marion Drut , Mohamed Hilal , Antonio Bodini , Michele Donati, Matthieu

3 Duboys de Labarre , Jelena Filipovic Lisa Gauvrit, Jose M. Gil, Viet Hoang, Agata Malak-Rawlikowska,

4 Konstadinos Mattasi, Sylvette Monier-Dilhan j, Paul Muller, Orachos Napasintuwong, Jack Peerlings,

5 Thomas Pomeon, Marina Tomic Maksan, Aron Torok, Mario Veneziani, Gunnar Vittersø, Filippo 6 Arfini c

7 Abstract

8 To identify whether EU certified food – here organic and geographical indications – is more 9 sustainable than a conventional reference, we developed 25 indicators covering the three 10 sustainability pillars. Original data was collected on 52 products at farm, processing and retail levels, 11 allowing the estimation of circa 2,000 indicator values. Most strikingly, we show that, in our sample, 12 certified food outperforms its non-certified reference on most economic and social indicators. On 13 major environmental indicators – carbon and water footprint – their performance is similar. Although 14 certified food is 61% more expensive, the extra-performance per euro is similar to classical policy 15 interventions to improve diet sustainability such as subsidies or taxes. Cumulatively, our findings 16 legitimate the recent initiatives by standards to cover broader sustainability aspects.

## 17 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)

### 27 **1. Introduction**

Consumer-oriented policy is increasingly seen as a key lever (Moran et al., 2018) - if not the 28 29 cornerstone (Springmann et al., 2018) - of a sustainable food system, with diet change and food 30 waste reduction at the forefront. However, environmentally friendly production practices could 31 provide an equally promising way forward, provided that they can be communicated clearly to 32 consumers and thereby inform their choice (Poore and Nemecek, 2018; W. K. Smith et al., 2019). This 33 is precisely the role of certified food. In 2012, the Quality Package (Regulation (EU) No. 1151/2012) 34 was passed in the EU to improve the operation of Geographical Indications (GIs) certification 35 schemes, initially based on product typicality. The Regulation details the rationale for promoting GIs 36 as a means to generate a fair return for farmers and processors and to enable consumers to make 37 better-informed purchasing decisions through effective labelling. Similarly, the organic standard 38 guarantees that neither farmers nor processors used synthetic chemicals. But beyond their initial 39 promise - typicality for GIs and absence of chemicals for organic products - is certified food more 40 sustainable than other food products?

41 Here we focus on organic and GI certifications, the two largest quality food standards in the EU with

42 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.

51 For organic farming, existing impact analyses are more comprehensive, covering both the economic 52 and environmental pillars of sustainability and even some aspects of social sustainability. Several 53 studies have applied Life-Cycle Assessments (LCA) to quantify differences in environmental impacts 54 between organic and conventional agriculture (Nemecek et al., 2011; Thomassen et al., 2008). These 55 assessments are not universally favourable to organic products (Meier et al., 2015a; Seufert and 56 Ramankutty, 2017), in particular when indirect land-use change consequences of lower yield are 57 included in the assessment (Bellora and Bureau, 2016). On the economic side, organic products 58 clearly capture a price premium, which in general allows organic farms to obtain higher net results 59 despite their lower yields (Crowder and Reganold, 2015a; European Commission, 2013; O. M. Smith 60 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., 61 62 2018; Koesling et al., 2008a; Mahé and Lerbourg, 2012). However, the environmental performance 63 assessments of organic farming are usually conducted in isolation from socio-economic assessments 64 (Pimentel and Burgess, 2014), which hinders the assessment of the broad sustainability performance 65 of organic value chains, let alone the synergy or trade-off between different sustainability aspects. 66 Moreover, methodological heterogeneity has been identified as an important pitfall in existing meta-67 analysis of the environmental performance of organic food (Meier et al., 2015b).

68 Here we assess the sustainability performance of EU certified food, questioning whether it 69 outperforms conventional food and at which cost, and identifying synergies and trade-offs between 70 different sustainability indicators. Compared to the existing literature, this assessment thus 71 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.
- 78 ✓ The performance criteria are assessed at the different levels of the value chains, including at
   79 least the farm and the processing levels;
- 80 ✓ It provides the first quantitative evaluation of the environmental and social performance of
   81 GIs.

## 82 2. Material and methods

- 83 **2.1. Data**
- 84

#### 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, 89 while regarding sectors, the unfed seafood/fish sector has much few cases (3) than the vegetal and 90 animal sectors.

91 In order to mitigate the influence of other possible drivers of performance than certification, such as 92 country- or sector-specific features, only the difference between a certified product and its reference 93 product is analysed. This strategy is inspired from the rationale of *controlled trials*. For this reason, 94 detailed guidelines were designed to select only products with a comparable reference. For instance, 95 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 96 97 more driven by the difference between cow and sheep systems than from the difference between 98 certified and non-certified value chains. This product was therefore removed from the analysis as it 99 was impossible to find a reference product meeting our guidelines.

### 100 **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.

- 106 In order to be able to collect data on the 52 products within a reasonable amount of time, the 107 variables requested for each indicator were divided in two categories, based on the literature:
- 108 ✓ Key variables, which are the focus of the primary data collection efforts. Key variables are the
   109 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.
- 112 The values of these variables were collected by in-country scientists according to the following 113 prioritization protocol:
- 114 ✓ Review of existing reports and databases on the value chain providing average values based
   115 on representative samples (secondary data);
- 116 Ad-hoc surveys, in person or online, of a sample of farms and firms (primary data);
- 117 ✓ Expert judgment elicitation, following the IPCC guidance (IPCC, 2006).
- 118 The detailed list of variables, all the spreadsheets including the raw data, their source, and the 119 resulting estimated indicators can be downloaded at xxref.

#### 120 **2.1.3. Quality control procedure**

121 Finally, a quality check procedure was put in place to mitigate the risk of misreporting data. The 122 three main aspects of this procedure were 1) to write down all data, their date and source in a 123 shared spreadsheet, 2) to distinguish the person who collected data from the person who estimated 124 the indicator, and 3) to write down, for each product and indicator, a common interpretation of the 125 estimated differences in performance that made sense for both the person who collected the data 126 and the person who estimated the indicator. These interpretations are recorded in Arfini and 127 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 128 129 a comparison with existing studies for the few indicators and certifications for which they exist 130 (SM7).

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

#### 132 **2.2.** Indicator estimation

#### 133 **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).

140 In order to ensure a common basis of comparison between cases studies despite the heterogeneity 141 in data availability, two orders of priority were established for indicators. The collection of variables 142 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 143 144 economic; three environmental; two social). Altogether, the twenty-five indicators necessitate the 145 collection of 150 variables (see XX for details on indicator design and estimation). We mostly focus 146 here on the "systematic indicators". This subset was selected to be equally distributed over the three 147 pillars, and to use indicators that were most common in their field and for which we had the least 148 missing values. These thirteen indicators cover six of the sustainable development goals of the UN 149 (SM 2, Table S2.1). Results from the entire set are nevertheless provided (SM 3 and SM 4) and used 150 to discuss the key messages where relevant.

151

	Sustainability pillar	Indicator type	Indicator sub-type (code)	Level of analysis along the value chain	
	Economic	Price premium	Price premium (Ec1.1, EUR kg <sup>-1</sup> )	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)		
Systematic	Environmental	Foodmiles	Carbon footprint per unit of product (En1.1, $kgCO_2e\ ton^{-1})$		
		Carbon footprint	Distance travelled per unit of product (En2.1, ton.km ton <sup>-1</sup> )	Single value for the whole value chain	
		Water footprint	Blue water footprint (surface and ground water consumption, En 3.3, m <sup>3</sup> kg <sup>-1</sup> )		
		Water rootprint	Grey water footprint (water pollution by nitrates, En 3.2, m <sup>3</sup> kg <sup>-1</sup> )	round chain	
		Employment	Labour to production ratio (So 1.1, AWU ton <sup>-1</sup> )	One value per level of the value chain	
	Social	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		
		Gender equality (So5.2, no unit)		chain	
Complementar	Economic	Profitability and value added distribution	Gross Value-added (Ec1.2, % of turnover)	One value per	
		Profitability and value added distribution	Net result (Ec1.4, % of turnover)	chain	
			Share of value exported outside Europe (Ec1.6, % of turnover)	Cincle under fam	
		Trade	Share of volume exported within Europe (Ec1.7, % of production)	the whole value chain	
			Share of volume exported outside Europe (Ec1.8, % of production)		
	Environmental	Foodmiles	Carbon footprint per hectare (En1.2, kgCO <sub>2</sub> e ha <sup>-1</sup> )		
1		Carbon footprint	Emissions from transportation per unit of product (En2.2, kgCO <sub>2</sub> e ton <sup>-1</sup> )	One value per level of the value chain	
		Water footprint	Green water footprint (rainwater consumption, En3.1, m <sup>3</sup> kg <sup>-1</sup> )		
	Social	Employment	Turnover to labour ratio (So1.2, EUR AWU <sup>-1</sup> )		
		Governance	Stability of the value chain level (So2.2, no unit)	<sup>D</sup> One value per level of the value	
		Social capital	Wage level (So3.2, EUR AWU <sup>-1</sup> )	chain	
			Gender equality index (So5.3, no unit)		

153	Table 1. List of indicators for sustainability	v assessment
100		y assessment

154

#### 2.2.2. Relative difference and value chain averages

155 Indicators are estimated for different levels of the value chain (farm level, processing level and, 156 where relevant, retail level). To control for the influence of country and product type, we analyse 157 relative differences between the certified product and its reference product rather than absolute 158 values (**Equation 1**).

#### 159 Equation 1.

160

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

where rel\_diff<sub>j</sub> is the relative difference at level j of the value chain, and indic<sub>CERT,j</sub> and indic<sub>REF,j</sub> 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.

166 In a second step, "value chain averages" are computed to evaluate the difference in performance for 167 the entire value chain (Equation 2). For most indicators, "value chain averages" are simple averages 168 of the indicator over the value chain levels for which it was estimated (farm, processing and, where 169 relevant, retail). There are, however, two exceptions.

170 The first exception concerns indicators expressed on a per ton basis, that is the environmental 171 indicators and the labour to production ratio. Because a life cycle assessment underlay these 172 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 173 174 hog at farm level and 1 ton of ham at processing level rather than averaging the footprints of one ton 175 of live hog and one ton of ham. This cumulative aggregation also allocates the footprint to all 176 products (eg. ham versus the rest of the carcass at processing level) based on their relative economic 177 value. For environmental indicators, this economic allocation is embedded in the original indicators. 178 For labour to production ratio, the formula is provided in Equation 2.

179 The indicator on value chain stability is the second exception: its "value chain average" is the 180 minimum of the "bargaining power" indicator across value chain levels (Equation 2).

#### 181 Equation 2.

$$182 \qquad \begin{cases} All indicators except environmental indicators and labour to production: VC_{average} = \frac{\sum_{j=1}^{n} rel\_diff_{j}}{n} \\ Environmental incators: VC_{average} = rel\_diff_{n} \\ VC_{average} = \frac{cum_{indic_{FQS}} - cum_{indic_{REF}}}{cum_{indic_{REF}}} \\ Labour to production ratio: \begin{cases} VC_{average} = \frac{cindic_{x,farm}}{cum_{indic_{REF}}} \\ cum_{indic_{X}} = \frac{indic_{x,farm}}{final_{prod_{ratio}} \times (1 + coproducts_{farm})} + \frac{indic_{x,proc}}{(1 + coproducts_{proc})} \\ Value chain stability: \begin{cases} VC_{average} = \frac{cum_{indic_{REF}}}{cum_{indic_{REF}}} \\ cum_{indic_{X}} = \frac{cum_{indic_{REF}}}{cum_{indic_{REF}}} \\ cum_{indic_{X}} = \frac{cum_{indic_{REF}}}{cum_{indic_{REF}}} \end{cases} \end{cases} \end{cases}$$

183 where VC<sub>average</sub> is the value chain average difference, rel\_diff<sub>i</sub> is the relative difference in performance 184 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\_indicx is the cumulative indicator over 185 186 different value chain levels for product X (either FQS or reference), indic<sub>X,farm</sub> and indic<sub>X,proc</sub> are the 187 indicator value for product X at the farm and processing levels respectively, final\_prod\_ratio is the 188 amount of raw product at farm level (eg. live hog) necessary for one ton of final product (eg. ham), 189 and coproducts\_farm and coproducts\_proc are the value of coproducts (eg. rest of the carcass) 190 expressed as a percentage of the value of the main product (eg. ham) at farm and processing levels 191 respectively.

#### 192 **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<sup>a</sup>. The R code used to compile indicators and conduct statistical tests is provided (SM 9).

### 209 **3. Results and discussion**

#### 210 **3.1.** Sustainability performance of certified food

211 Certified food never performs significantly worse than conventional food on the key 13 indicators 212 (Figure 1). This is still true for the broader set of 25 indicators (Figure S 1): only rainwater use is 213 significantly higher (worse) for certified food, but the relevance of this indicator is debated, especially 214 where rainwater is abundant (Schyns et al., 2019). For two thirds of the key indicators, the 215 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.

230 The environmental performance of certified products is broadly similar to their conventional 231 reference. More precisely, certified products pollute less on a per hectare basis (-27% GHG emissions 232 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 233 234 footprint indicators – which are expressed per ton – this higher performance is diluted by the lower 235 yield of certified farms (-19%, see land\_carbonXX2 for details). This lower productivity is particularly 236 acute for organic products in our sample (-36%), an order of magnitude which is consistent with 237 existing meta-analysis dedicated to the yield of organic farming. Seufert et al. (2012) report an 238 average yield difference of -25%, and -34% when organic and conventional systems are most 239 comparable. Ponisio et al. (2015) report an average different of -19%. In both cases however, the 240 yield gap is found to vary substantially between products and pedo-climatic conditions.

241 Among the key environmental indicators, certified food only performs better on *foodmiles*, thanks to 242 lower exports and local sourcing. While local sourcing is driven by the technical specifications of both 243 geographical indications and organic farming (for animal products), the reason for lower exports 244 differs: many geographical indications are only recognized in their domestic market, leading them to 245 neglect international outlets which do not offer a price premium. For organic farming, there is a 246 divide between high-income countries where domestic supply struggles to match domestic demand 247 (little to no exports) and low-income countries where organic supply chains are almost entirely 248 dedicated to exports.

<sup>&</sup>lt;sup>a</sup> 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. pvalue < 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.

# 262 3.2. Certified food as an efficient way of producing public 263 goods?

The relatively high marginal performance improvement per euro of price premium depict the Quality 264 policy of the European Union – here the certification of organic and GI products – as an efficient way 265 of producing positive economic, environmental and social externalities. Marginal performance 266 267 improvement per euro is obtained by dividing the relative performance difference between certified 268 products and their reference products by their relative price difference. It is an indicator of how 269 much the performance improves for a given cost to the consumer. Where significantly different from 270 zero, we find that the marginal performance improvement per euro, averaged per indicator across 271 the value chain, is between 0.3 and 0.6 (Figure 2), except for local multiplier (0.07) and bargaining 272 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) 273 274 (Guyomard et al., 2018) or the simulated response to a carbon tax (climate impact) (Bonnet et al., 275 2016; Caillavet et al., 2016). For many indicators however, the variability is large with some extreme 276 values. This highlights the absence of direct causality in our experimental design regarding the 277 relationships between price premium and over-performance and moderates the strength of the 278 evidence provided by these pseudo-elasticity estimates. Moreover, these estimates neglect 279 transaction costs (Bellassen et al., 2015) and windfall effects (Chabé-Ferret and Subervie, 2013), both 280 of which have been shown to weigh heavily on the cost-efficiency of any policy, be it certification-281 based or not.

282 Marginal performance improvement per euro of price premium are slightly higher at farm level than 283 at processing level: three dimensions are significantly higher than zero instead of two, and 284 generational change is the only indicator for which the median marginal performance improvement 285 per euro is lower at farm level (0.04) than at processing level (0.06). This argues for focusing policy 286 intervention at farm level rather than processing level, especially as environmental indicators are 287 cumulative, with the bulk of their value already determined at farm level.





292 Figure 2. Marginal performance improvement of certified products per euro of price premium at 293 farm, processing and retail levels. The marginal performance improvement per euro shows how 294 many percentage points are gained for a given indicator per percentage point of price premium of 295 the certified food. It is estimated here in a non-causal way by dividing the relative difference in 296 performance by the relative difference in price for each product and indicator. Indicators pertaining 297 to the entire value chain (export, bargaining power distribution and local multiplier) are displayed in 298 the "retail" frame. Sample size (n) varies between value chain level because not all levels exist in all 299 value chains and because data was not available for some levels in some value chains. The p-value 300 indicates the probability that the median is different from zero (Wilcoxon signed-rank test). 301 Indicators for which the marginal performance improvement per euro is significantly different from 302 zero (i.e. p-value < 0.1) are mentioned in bold. n indicates the number of certified products for which 303 the indicator has been calculated. Three indicators at retail level for which the sample size is smaller 304 than 5 are not displayed. Marginal performance improvement per euro is bounded by [-11;11] for 305 convenience of display. For the three product x level combinations where the price premium of the 306 certified product is - very slightly - lower than zero, the marginal performance improvement is 307 assumed to be infinite (eg.  $+\infty$  when the difference in performance is positive).

### 308 **3.3.** Price threshold for the profitability of certified food

309 Most of the performance improvements rely upon the existence of a price premium, without which 310 firms have no incentive to comply with technical specifications. One way of assessing the scaling-up 311 potential of certified food is therefore the price premium level which perfectly balances the extra 312 costs of production. The median value of this threshold is 25% of the reference price, indicating a 313 good resistance of certified chains to decreasing premiums, especially for organic. The processing 314 level may be the weak link of certified value chains with median extra costs amounting to 52% of the 315 reference price (see SM4 and XX3<sup>b</sup> for details). Scaling-up would of course create other challenges than decreasing price premiums (eg. land requirements associated with reduced yields (Muller et al., 316 317 2017; Searchinger et al., 2018)). On the other hand, all extra-costs need not but covered by the price 318 premium: public subsidies or payments for environmental services can also contribute.

<sup>&</sup>lt;sup>b</sup> Reference anonymized for the purpose of the double-blind peer review process.

### 319 **3.4.** Differences in performance between quality signs

320 PDO and organic value chains both get an edge on attractiveness to young workers, with a 321 significantly better performance of 33% and 26% respectively, compared with a non-significant 18% 322 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. 323 324 price premium of 73%, 58% and 40% for organic, PDO and PGI respectively with a p-value of the 325 Kruskal-Wallis test close to 0.2 for organic and PGI). Other less documented features of certified 326 value chains such as the preservation of cultural heritage could also contribute to attract young 327 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

338 products which would likely qualify as PDOs only seek PGI recognition.

339



341 Figure 3. Sustainability performance of specific subgroups. The median difference in performance 342 for the subgroup (eg. PDO) between certified products and their reference (in %) appears in color 343 above the median difference of the entire sample in brackets. "n" indicates the number of certified 344 products of the subgroup, "p-value" the probability that the median difference in performance 345 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 346 347 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 348 349 than 0.1 while the p-value for the entire population is higher than 0.1.

# 350 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 socialperformance.

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.

364 Finally, foodmiles performance is trivially deteriorating with increasing exports.



365

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).</li>

#### 369 **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.

377 Another crucial improvement would be to increase sample size. 26 paired products, each relying on 378 primary variables averaged over multiple farms/firms, is a large sample compared to existing studies 379 (see SM 7), especially if one excludes meta-analysis which tend to focus on a few indicators and are 380 subject to methodological heterogeneity. However, it remains too modest to draw definitive 381 conclusions on the relative merits of certified food. Moreover, increasing sample size would reduce 382 uncertainty and thus possibly allow to identify other synergies and trade-offs between sustainability 383 indicators. Most importantly, it would pave the way towards a more robust assessment of the 384 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.

401 Finally, while our analysis focuses on the EU, there is no obstacle to using the same method for the 402 assessment of the sustainability of food value chains elsewhere. As a matter of fact, it has been 403 successfully used to assess eight products in South-East Asia. Moreover, our results are consistent 404 with existing literature for the indicators for literature in other parts of the world exists (eg. value 405 added (Arfini et al., 2006; Crowder and Reganold, 2015a; London Economics, 2008a; 406 Vandecandelaere et al., 2018), carbon footprint of organic products (Meier et al., 2015a), ...). This is a 407 good sign for the global relevance of our findings, but extending the geographical coverage our 408 sample would be necessary to ascertain this.

# 409 **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).

## 417 Data availability

418 Both raw data and estimated indicators have been deposited in an open dataset and can be 419 downloaded from XX [link to the Dataverse repository, to be inserted when the manuscript is 420 accepted. The repository contains what is currently provided in SM 1, 4, 8 and the code file 421 mentioned in SM 9. It will therefore replace these supplementary materials when the manuscript is

422 accepted].

### 423 **Code availability**

424 Code to estimate the indicators, perform the statistical tests and display the figures has been 425 deposited in the aforementioned open dataset.

#### 426 **References**

- 427 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.
- 456 Clark, M., Tilman, D., 2017. Comparative analysis of environmental impacts of agricultural production systems, agricultural
   457 input efficiency, and food choice. Environ. Res. Lett. 12, 064016. https://doi.org/10.1088/1748-9326/aa6cd5
- 458 Crowder, D.W., Reganold, J.P., 2015a. Financial competitiveness of organic agriculture on a global scale. PNAS 112, 7611– 459 7616. https://doi.org/10.1073/pnas.1423674112
- 460 Crowder, D.W., Reganold, J.P., 2015b. Financial competitiveness of organic agriculture on a global scale. PNAS 112, 7611–
   461 7616. https://doi.org/10.1073/pnas.1423674112
- 462 European Commission, 2017. Distribution of the added value of the organic food chain.
- 463 European Commission, 2013. Organic versus conventional farming, which performs better financially? (No. 4), Farm 464 Economics Briefs. European Commission, Brussels, Belgium.
- 465 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.
- 468 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
- 473 Guyomard, H., Détang-Dessendre, C., Réquillart, V., Soler, L.G., 2018. La Politique agricole commune doit-elle intégrer des
   474 objectifs de lutte contre le surpoids et l'obésité ? INRA Sciences Sociales 7 p.
- 475 INAO, 2016. L'agro-écologie s'installe dans les SIQO. La lettre de l'INAO 1.

- 476 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.
- 481 Läpple, D., Kelley, H., 2015. Spatial dependence in the adoption of organic drystock farming in Ireland. Eur Rev Agric Econ
   482 42, 315–337. https://doi.org/10.1093/erae/jbu024
- 483 London Economics, 2008a. Evaluation of the CAP policy on protected designations of origin (PDO) and protected 484 geographical indications (PGI).
- 485 London Economics, 2008b. Evaluation of the CAP policy on protected designations of origin (PDO) and protected 486 geographical indications (PGI).
- 487 Mahé, T., Lerbourg, J., 2012. Des agriculteurs bio diplômés, jeunes et tournés vers les circuits courts. Agreste Primeur 4 p.
- 488 Meier, M.S., Stoessel, F., Jungbluth, N., Juraske, R., Schader, C., Stolze, M., 2015a. Environmental impacts of organic and 489 conventional agricultural products – Are the differences captured by life cycle assessment? Journal of 490 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
- Finentel, 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
- 512 Searchinger, T.D., Wirsenius, S., Beringer, T., Dumas, P., 2018. Assessing the efficiency of changes in land use for mitigating 513 climate change. Nature 564, 249. https://doi.org/10.1038/s41586-018-0757-z
- 514 Seufert, V., Ramankutty, N., 2017. Many shades of gray—The context-dependent performance of organic agriculture. 515 Science Advances 3, e1602638. https://doi.org/10.1126/sciadv.1602638
- 516 Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. Nature 485, 517 229–232. 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
- 532 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 533 organic production Netherlands. milk the Agricultural 95-107. and in Systems 96. 534 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.
- 541

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

544 See: SM1\_Method\_indicators.docx

# 545 6. SM 2: correspondence with UN Sustainable Development 546 Goals

UN Sustainable Development Goals (SDGs)		Indicators	Part of the SDGs for which the indicator is relevant
4 QUALITY EDUCATION p k f f	Goal 4. Ensure nclusive and equitable quality education and promote lifelong learning opportunities for all	Educational attainment	<ul> <li>4.3 By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university</li> <li>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</li> </ul>
5 GENDER EQUALITY	Goal 5. Achieve gender equality and empower all women and girls	Gender equality Gender equality index	<ul> <li>5.1 End all forms of discrimination against all women and girls everywhere</li> <li>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</li> <li>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</li> </ul>
6 CLEAN WATER and SANITATION fr	Goal 6. Ensure availability and sustainable management of water and sanitation for all	Grey water footprint Blue water footprint	<ul> <li>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</li> <li>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</li> <li>6.5 By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate</li> </ul>
8 DECENT WORK AND ECONOMIC GROWTH a	Joal 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	<ul> <li>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</li> <li>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</li> </ul>

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	<ul> <li>16.6 Develop effective, accountable and transparent institutions at all levels</li> <li>16.7 Ensure responsive, inclusive, participatory and representative decision-making at all levels</li> </ul>

### 548 7. SM 3: performance difference for all indicators





550 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.

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

561 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
		Hard pressed cooked				Similar uncertified cheese (Emmental)
Comte cheese	France	cheese from cow milk	PDO	Yes	504 191 550	or national average (cow cheese)
Organic flour	France	Wheat flour	Organic	Yes	34 800 000	National average
Saint-Michel bay		Mussels produced on				National average (TSG Bouchot
bouchot mussels	France	"bouchots"	PDO	No	25 450 688	mussels)
Organic rice	France	Rice	Organic	Yes	17 640 000	Non-organic rice (mostly PGI)
Organic pork	Germany	Raw meat Organic yoghurt from	Organic	Yes	69 000 000	National average
Organic yoghurt	Germany	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	another region)
Gyulai sausage	Hungary	Sausage	PGI	Yes	55 000 000	Non-PGI Hungarian sausage
Kalocsai paprika						Imported Chinese pepper milled in
powder	Hungary	Paprika powder	PDO	Yes	10 749 180	Hungary
Parmigiano Reggiano		Hard pressed cooked				Biraghi cheese (similar non-PDO
cheese	Italy	cheese from cow milk	PDO	Yes	1 009 943 894	cheese)
						Conventional processed tomatoes in the
Organic tomatoes	Italy	Organic tomato	Organic	No	68 574 011	same region (Emilia-Romagna)
Opperdoezer Ronde	NT-thenderda	E a alta a a ta ta	BDO	N-	2 771 250	Regular potato in neighbouring
L ofoton stool fish	Nemeriands	Early potato	PDU	INO No	2 //1 230	Disselineerpolders region
Organic salmon	Norway	Salmon	Organic	INU Ves	1 243 333	Conventional salmon
Organic samon	Norway	Samon	organie	105	144 /07 200	Simulated conventional farms with
Organic pasta	Poland	Pasta	Organic	Yes	523 810	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 Unprocessed lamb	PGI	Yes	1 802 520	National average Non-PGI lamb in the same region
Ternasco de Aragon	Spain	meat	PGI	No	16 965 438	(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)
						Non-PGI coffee from the same
Doi Chaang coffee	Thailand	Coffee	PGI	Yes	756 000 000	province
Phu Quoc Fish Sauce	Vietnam	Fish sauce	PDO	Yes	3 426 166	Non-PDO fish sauce from same region
						Non-PGI coffee from Dak Lak province
Buon Ma Thuot coffee	Vietnam	Coffee	PGI	Yes	89 577 676	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<sup>c</sup> 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.

# 574 10. SM 6: classification of products, no partner or country 575 effect

576 SPAD software<sup>d</sup> is used to identify clusters with similar sustainability performances. The basis for the 577 clustering algorithm is the quartile of performance to which a given product belongs for a given 578 indicator (when an indicator could not be estimated for a given product, the product is assigned to a 579 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 580 581 out by chaining a first k-means clustering with around 10 randomly selected centers. The 10 stable 582 classes thus formed are then aggregated by a hierarchical classification followed by consolidation. 583 The objective of consolidation is to reassign elements from one class to another class to which they 584 are closer in order to improve homogeneity within classes.

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

588 Clusters are then characterized with a set of additional variables describing the certified products 589 (partner, country, team in charge of data collection, product type, sector, label type), and 590 geographical context (gross domestic product, population, education level, number of designations, 591 agricultural area, share of organic areas). For continuous variables, a simple t-test is used to identify 592 clusters characterized by a difference for these additional variables. For discrete variables, a test 593 based on a hypergeometric distribution is used instead.

594 This characterization does not identify many additional variables for which a class is significantly 595 different from others. Cluster E is characterized by countries where both the number of quality signs 596 and the share of arable area under organic farming are high. The seafood sector is overrepresented 597 in cluster C and so is one data collection team in cluster B. Overall, the effect of the team which 598 collected data is only one visible for one team in one cluster, which supports the reliability of the 599 data collection methodology and in particular the absence of a significant influence from data 600 collectors.

601

<sup>&</sup>lt;sup>c</sup> Reference anonymized for the purpose of the double-blind peer review process.

<sup>&</sup>lt;sup>d</sup> 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

608 Our findings are consistent with studies relying on a substantial (>5) sample size for the indicator x 609 certification combinations for which literature exists (Table S7.1). Our price premium for organic 610 products is close to a recent European study and higher than a global study, which could be 611 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 612 613 within the range of existing studies, which is therefore likely the case of price premium although the 614 most recent study does not report it (Areté, 2014). Similarly to existing meta-analysis, we find that 615 the carbon footprint of organic products is not significantly different from their conventional 616 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. 617

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% <sup>e</sup> (N = 45, p-value < 0.05) (Finley et al., 2018)
Educational attainment	Organic	+5% (N = 8, p-value = 0.4)	Positive or null <sup>f</sup> (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 (pvalue, 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

627 See SM7\_raw\_data.zip and SM7\_C\_footprint\_calculator.7z

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

630 The R code is provided in SM9\_R\_code.org

<sup>&</sup>lt;sup>e</sup> 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.

<sup>&</sup>lt;sup>f</sup> 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.