ANATOMY OF GREEN SPECIALIZATION: EVIDENCE FROM EU PRODUCTION DATA, 1995-2015

Filippo Bontadini

Francesco Vona

SCIENCES PO OFCE WORKING PAPER n° 21/2020

CONTACT US

OFCE
10 place de Catalogne | 75014 Paris | France
Tél. +33 1 44 18 54 24

www.ofce.fr

WORKING PAPER CITATION

This Working Paper:
Filippo Bontadini and Francesco Vona
Anatomy of Green Specialization: Evidence from EU Production Data, 1995-2015
Sciences Po OFCE Working Paper, n° 21/2020.
Downloaded from URL: www.ofce.sciences-po.fr/pdf/dtravail/WP2020-21.pdf
DOI - ISSN

ABOUT THE AUTHORS

Filippo Bontadini, SciencesPo-OFCE, SPRU-University of Sussex
Email Address: filippo.bontadini@sciencespo.fr
Francesco Vona, Sciences Po-OFCE, Université Côte d'Azur, SKEMA, CMCC Ca' Foscari University of Venice
Email Address: francesco.vona@sciencespo.fr

ABSTRACT

We study green specialization across EU countries and detailed 4-digit industrial sectors over the period of 1995- 2015 by harmonizing product-level data (PRODCOM). We propose a new list of green goods that refines lists proposed by international organizations by excluding goods with double usages. Our exploratory analysis reveals important structural properties of green specialization. First, green production is highly concentrated, with 13 out of 119 4-digit industries accounting for 95% of the total. Second, green and polluting productions do not occur in the same sectors, and countries tend to specialize in either green or brown sectors. This suggests that the distributional effect of European environmental policies can be large. Third, green specialization is highly path dependent, but it is also reinforced by the presence of non-green capabilities within the same sector. This helps explain why economies with better engineering and technical capabilities have built a comparative advantage in green production.

KEYWORDS

Green goods, green specialization, revealed comparative advantage, complementarity, path dependency.

JEL

Q55, L60, O44.

Anatomy of Green Specialization: Evidence from EU Production Data, 1995-2015

Filippo Bontadini
Sciences Po – OFCE
SPRU – University of Sussex
email: filippo.bontadini@sciencespo.fr

Francesco Vona
Sciences Po – OFCE,
Université Côte d'Azur, SKEMA
CMCC Ca' Foscari University of Venice
email: francesco.vona@sciencespo.fr

Abstract

We study green specialization across EU countries and detailed 4-digit industrial sectors over the period of 1995-2015 by harmonizing product-level data (PRODCOM). We propose a new list of green goods that refines lists proposed by international organizations by excluding goods with double usages. Our exploratory analysis reveals important structural properties of green specialization. First, green production is highly concentrated, with 13 out of 119 4-digit industries accounting for 95% of the total. Second, green and polluting productions do not occur in the same sectors, and countries tend to specialize in either green or brown sectors. This suggests that the distributional effect of European environmental policies can be large. Third, green specialization is highly path dependent, but it is also reinforced by the presence of non-green capabilities within the same sector. This helps explain why economies with better engineering and technical capabilities have built a comparative advantage in green production.

Keywords: green goods, green specialization, revealed comparative advantage, complementarity, path dependency.

JEL codes: Q55, L60, O44

Acknowledgement

This work was supported by Horizon 2020 Framework Programme, project INNOPATHS [grant number 730403].

1. Introduction

This paper provides new evidence of the production and specialization of environmentally friendly goods across sectors and European countries over the period of 1995-2015. Understanding comparative advantage in green production is particularly important in light of the growing policy interest around the so-called green economy as a way to reconcile economic growth with environmental preservation and climate change mitigation, which recently culminated in the launch of the European Green Deal by the European Commission. Developing a head start in the green economy was also a strategic goal of the generous fiscal stimulus implemented by President Obama after the great recession, the so-called American Recovery and Reinvestment Act, which sought to build US technological leadership in new high-demand products such as electric cars and solar energy.

Despite its key strategic role in a country's future competitiveness, data constraints have so far limited the scope of empirical research on the green economy. The first contribution of our paper is a consistent measure of green production that varies in the country-year-sector (detailed 4-digit NACE rev. 2 sectors) level for manufactured goods. To this aim, in Section 2, we use a product-level dataset for the manufacturing sector compiled by Eurostat, called PRODCOM, and harmonize it with a methodology proposed by Van Beveren, Bernard and Vandenbussche (2012). To measure green production, we first use different lists of green products that have been proposed during recent international negotiations at the World Trade Organization (WTO). We refine these lists by eliminating green goods with double usages to reach our favourite list of green goods. To the best or our knowledge, we are the first to use PRODCOM to study green production. Previous works have used product-level data to study trade patterns in green production (He *et al.*, 2015; Cantore and Cheng, 2018; Fraccascia, Giannoccaro and Albino, 2018; Tamini and Sorgho, 2018; Mealy and Teytelboym, 2019) and their effects on emission reduction at the country level (Zugravu-Soilita, 2018, 2019). We extend these works by assembling a new dataset that can be used to highlight the fine-grained structure of green production across sectors and countries.

Empirical work on environmental innovation has mostly used patents or self-reported measures of firm innovation to build a proxy of green vs. non-green specialization (Jaffe and Palmer, 1997; Popp, 2002; Veugelers, 2012; Ghisetti and Rennings, 2014; Nesta, Vona and Nicolli, 2014; Calel and Dechezleprêtre, 2016; Horbach, 2016; Conti et al., 2018). This choice is theory-consistent, as most climate (e.g. Nordhaus and Boyer, 2000) or endogenous growth models (e.g. Bovenberg and Smulders, 1995) give prominent importance in reducing the harmful environmental impacts of production to R&D-driven innovation. However, from a policy perspective, the beneficial effects of green specialization in terms of improved environmental quality, job creation and economic growth depend on where production is located (diffusion stage) rather than on where knowledge is created (innovation stage). Additionally, from the perspective of innovation studies, it is well known that not all knowledge is created through R&D investments and then patented (Arundel and Kabla, 1998; Jaffe and Trajtenberg, 2002; Dosi *et al*., 2017;). Other informal channels of knowledge creation, such as learning-by-doing and knowledge spillovers, may be equally, if not more important, in explaining the dramatic improvements in pollution intensity that we have observed in recent decades (Levinson, 2009; Brunel, 2017). While the causality nexus between patented knowledge and production is clearly bidirectional and too complex to be studied in this exploratory study, we show how the use of new production data can provide important insights into key patterns of green specialization.

In accounting for the evolution and structure of green production in Europe, we explore both the industry- and country-level heterogeneity contained in the PRODCOM data. In Section 3, we first focus on industry-level dynamics, which have been ignored by previous work that has treated green production as a unique aggregate sector (Fraccascia, Giannoccaro and Albino, 2018; Mealy and Teytelboym, 2019; Zugravu-Soilita, 2019).

Two findings stand out from our industry-level analysis. First, green production is extremely concentrated in a set of high-tech industries producing capital goods. When it comes to capturing industry concentration, the devil is in the details. At the 2-digit level, 9 out of 26 industries have positive green production. However, of the 119 4-digit industries contained in those green 2-digit industries, only 21 are green, and 13 of those represent 95% of the total green production. We call these 13 industries high-green-potential industries, and these are the focus of our analysis.

Second, we find that polluting and green production occur in two separate sets of industries, which are related only through intra-industry linkages such as the purchase of capital goods. This has an important policy implication for the distributional effect of environmental policies. Not only are the sectors that bear the cost of pollution taxes and standards different from the sectors that can profit from these policies, but green sectors also receive the bulk of the subsidies for the green economy.

In Section 4, we investigate the distribution of green and polluting production across countries to shed light on which countries will be leaders in the green transition. In doing so, we build revealed comparative advantage (RCA) measures of green and brown production. Unsurprisingly, while the average share of green production in Europe increased from 2% to 2.5% from 1995-2015, northern countries, especially Denmark, Austria, Sweden and Germany, have retained a persistent green comparative advantage. In contrast, southern countries specialize in polluting industries. Only Hungary has emerged as a new green player, with production concentrated in parts of turbines and photovoltaic panels.

In Section 5, we investigate the structural properties of green specialization in a multivariate regression framework. We compare the role of three main drivers that have been examined by the literature on green innovation using patents (e.g., Popp, 2002; Nesta et al., 2014; Aghion *et al.*, 2016; Colombelli and Quatraro, 2019; Perruchas, Consoli and Barbieri, 2020): i. path dependency; ii. complementarity with proximate, non-green capabilities; and iii. diversification of the knowledge base. Consistent with the descriptive evidence, green specialization exhibits path dependency: a 10% initial advantage in green specialization is associated with a 6.2% advantage eleven years later. Our regressions also reveal a complementarity between green and non-green specialization within the same narrowly defined 4-digit industry, although the magnitude of the association is much smaller than that for path dependency. Finally, diversification matters in sustaining green specialization, but it is quantitatively less important than having a minimum threshold of productive green competences. In Section 6, we summarize our main findings and the main policy implications.

2. A new measure of green production.

This section is organized as follows. In Section 2.1, we discuss the conceptual issues in measuring green production. In Section 2.2, we present our main source of data, PRODCOM, while in Section 2.3, we discuss how to use PRODCOM to measure green production. Finally, Section 2.4 validates our favourite list of green products—which we refer to as the PRODCOM list henceforth—against other measures.

2.1 Conceptual issues.

The definition of green production presents several conceptual challenges related to the theoretical understanding of what "green" or "environmentally friendly" means and how such definitions can be operationalized to the data.

The first conceptual issue is whether we consider an activity (i.e., a product or a service) green in terms of the effective pollution content of production (*process approach*) or in terms of its potential to minimize the harmful impacts of production on the environment (*output approach*). The first approach is intuitive, as it defines the inverse of product greenness on a continuous scale based on the pollution that is directly or indirectly generated during production. The problem with this approach is that it is very difficult, if not impossible, to reconstruct the pollution content of each product (e.g., Sato, 2014), which depends on the complex nexus of production along global value chains (e.g., Wiedmann *et al.*, 2011). The scholarship has strived to overcome these issues by deploying input-output methodology and constructing several datasets to assess the environmental footprint of production. These datasets, however, include a limited number of countries, years and highly aggregated sectors, yielding rather different estimates of pollution impacts of production (Rodrigues *et al.*, 2018).

The second approach emphasizes the potential functions of certain products that can be beneficial for the environment, and it is the preferred approach for defining most lists of green products or activities. For instance, the Green Goods and Services Survey (GGS) of the Bureau of Labor Statistics in the US was built using this definition.[1] To highlight the difference between these two approaches, one can consider wind turbines: even though they fulfil an unequivocally green function, the process, emission-based, approach would not consider them very green due to the high pollution intensity of the iron that is necessary for their production. A similar approach has been to identify a set of green tasks performed by the workforce in the Occupational Information Network database (Dierdorff *et al.*, 2009; Consoli *et al.*, 2016), but this approach is more suitable for measuring green labour than green production (Vona, Marin and Consoli, 2019).

Additionally, the identification of which functions or tasks are beneficial to the environment remains far from straightforward. Products fulfil functions that have different potential for

[1] See the technical note of the GGS survey at https://www.bls.gov/ggs/ggs_technote_extended.pdf, Deschenes (2013) and Elliott and Lindley (2017) for details. The US Census Bureau also carried out the Survey of Environmental Products and Services in 1992, which also used the output approach to identify green products drawing on input from both government agencies and the private sector; for more detail see Becker and Shadbegian (2009).

reducing pollution based on their underlying technology. Frondel, Horbach and Rennings (2007) distinguish between products relying on end-of-pipe and integrated technologies, also referred to as cleaner production technologies. On the one hand, the former limits the pollution from production processes without changing these processes in essence (e.g., waste-water treatment, air-quality control, catalytic converters or exhaust-gas cleaning equipment). The problem of end-of-pipe technologies is that they may generate cross-media substitution effects, e.g., scrubbers and filters reduce air pollution but can negatively affect water and soil pollution (Bi, 2017; Gibson, 2019). On the other hand, integrated technologies prevent pollution at the source, replacing less clean technologies: wind turbines are a clear example of this kind of product. Such technologies can be considered more beneficial for the environment due to their potential for reducing the environmental impact of production processes across multiple industries.

Other examples include the criteria proposed by Eurostat following an output-based approach: environment protection, i.e., activities that "have as their main purpose the prevention, reduction and elimination of pollution and of any other degradation of the environment" and resource management, i.e., the "preservation, maintenance and enhancement of the stock of natural resources and therefore the safeguarding of those resources against depletion" (Eurostat, 2016, p.15). This rather narrow definition may exclude other products that do not directly fulfil either criterion but that reduce the environmental impact of other sectors.[2]

Finally, the same product can have different usages and thus different environmental impacts. For example, pipes and water tanks may be considered green when used for water and waste management purposes, but they will not be green when used for other activities (Steenblik, 2005), such as textile production that involves intensive water consumption.

These issues make it difficult to find a well-accepted definition of what a green product is. Operationalizing a definition of green products is made even more difficult because standard statistical classifications are not designed to separate environmentally friendly products (Steenblik, 2005; Sauvage, 2014). This increases the likelihood that lists of green products contain false negatives (products that are environmentally friendly but are excluded from the list) and false positives (products that are not environmentally friendly but that are nonetheless included).

[2] For example, LEDs that provide lighting neither have a positive impact on the environment nor directly preserve the depletion of the stock of natural resources. However, they are more energy efficient than traditional light bulbs and therefore correspond to what Eurostat's (2016, p.15) defines as "secondary environmental goods", i.e., products that are "specifically designed to be more environmentally friendly" than existing alternatives.

We propose to overcome, at least partially, the data shortcomings and conceptual ambiguities discussed above using a new dataset, PRODCOM, where product codes and descriptions are available at a highly disaggregated level.

2.2 PRODCOM data.

Eurostat collects very detailed information on manufacturing production in Europe, considering, on average, 4,288 single products per year.[3] The PRODCOM dataset is available for the years between 1995 and 2015 for the core European countries, while detailed data on production in Eastern European countries has been collected from 2001 onwards.[4]

For the purpose of identifying green production across countries and industries, the PRODCOM data present two practical advantages. First, the PRODCOM data are easily linkable to existing lists of product codes that identify green products based on the green potential of their final use. Second, the PRODCOM product classification is nested within the European industrial classification NACE: each PRODCOM code is made of eight digits, the first four of which correspond to NACE industry codes. This feature of the data allows assigning each product to a 4- or 2-digit industry and computing each industry's share of green production, making the data suitable for studying how green production is distributed not only across countries but also across industries.

The use of the PRODCOM data also presents three important challenges. The first is that the product coverage changes over time due to the entry and exit of products[5]. Second, product codes change over time due to constant statistical redefinition, with multiple product codes merging into a single new code or one code splitting into several new or existing codes. Third, in 2008, there was a change in industry classification (from NACE rev. 1 to NACE rev. 2), with some products changing industries at the 4-digit level between the two versions. As a result, by aggregating data

---

[3] As we will discuss in this section, the coverage and number of product codes in the PRODCOM data varies yearly, so we report here the average number of 8-digit product codes contained in the PRODCOM data between 1995 and 2015. It should also be noted that the PRODCOM data only covers manufacturing production, which means that we cannot include environmental services into our analysis.

[4] Countries for which data from 1995 on is available include: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden and the United Kingdom. From 2001 on, our data include: Bulgaria, Croatia, Estonia, Hungary, Latvia, Lithuania, Romania, Slovakia and Slovenia. Poland is included from 2003 onwards.

5 While entries and exits concern few products across all sectors, it should be noted that fuel and coke related products are excluded from the PRODCOM data up until 2005, leading to no information on the production of the whole 2-digit sector "coke and refined petrol".

at the 4-digit industry level, as we do in this study, the combination of changes in product codes and of industry classification may conflate genuine changes in production within an industry with a statistical reallocation of products across industries.

We deal with these issues using the methodology developed by Van Beveren, Bernard and Vandenbussche (2012) (VBBV henceforth) to concord the PRODCOM data over time. In summary, the VBBV methodology identifies chains of product codes that are linked by changes over time and attributes a "synthetic code" to each chain that does not change over time. A key advantage of this methodology is that it solves problematic issues in the conversion from NACE rev. 1 to NACE rev. 2.[6] Indeed, each of these synthetic codes can be easily paired with a NACE rev. 2 industry code at the 4-digit level, since these are the first 4 digits of the PRODCOM codes from 2008 onwards. Because the synthetic codes do not change over time, we can also allocate production values to NACE rev. 2 industries for the years preceding their introduction, covering the whole timespan of the PRODCOM data (1995-2015).

Another key advantage of the VBBV procedure is that it yields a time-consistent measure of green production, taking into account that green products may be split into a green and a non-green product or merged with a non-green product. An important example in the PRODCOM dataset is wind turbines. Until 2007, wind turbines were classified under a residual heading "generating sets n.e.c.", which contained both green and non-green products. Only after 2008 did the code split into a non-green product, "generating sets (excluding wind powered and powered by spark-ignition internal combustion piston engine)", and a green product, "generating sets, wind-powered". As a consequence, we have information on the production of wind-powered generating sets only after the year in which the split occurred (2008), while before then, wind turbines were lumped together with other generating sets. A similar issue applies when a green and non-green product are merged into a unique synthetic code.

To deal with this additional issue and impute the missing data on green production (e.g., wind turbines before 2008), we first compute the average (country-specific) share of the green production of the synthetic code that merged or split over the three years after (before) the merge (split). We then assign production proportionally to this share in the years for which we cannot distinguish between green and non-green production.

[6] Eurostat provides a crosswalk between the two versions of NACE. However, such crosswalk is imperfect as it entails many-to-many correspondences with some NACE rev 1 industries splitting and/or merging into NACE rev 2 industries.

2.3 Measuring green production using the PRODCOM data.

The PRODCOM data are easily linkable to existing lists that identify green goods using the output-based approach described in Section 1. Historically, these lists emerged as part of international negotiations to reduce the tariffs on a set of goods that are crucial for low-carbon transitions and sustainable development in general (WTO, 2001; APEC, 2012). The rationale for this is the idea that reducing tariffs on green products will reduce their cost and thus favour their diffusion (World Bank, 2007; Hufbauer and Kim, 2011), especially in developing countries (Dutz and Sharma, 2012; World Bank, 2012).

The political economy of trade negotiations adds another source of ambiguity in determining what is green to the complex picture described in Section 2.1. Indeed, in compiling green goods lists during trade negotiations, each country bargains to obtain tariff reductions for the goods for which they have a comparative advantage, rather than on the basis of the goods' actual beneficial effect for the environment (Balineau and de Melo 2011; de Melo and Solleder, 2018). The inability to reach an agreement on a final list of green goods was one of the reasons the trade negotiations on environmental goods was interrupted in 2016 (European Commission, 2019).[7]

Despite not reaching a final consensus, the negotiation process has produced several lists of green goods. The most comprehensive is the Combined List of Environmental Goods (CLEG) of the OECD, which is a union of three lists: the Plurilateral Environmental Goods and Services (PEGS) list developed by the OECD itself, the list negotiated within the Asian Pacific Economic Cooperation (APEC) forum and the list agreed upon by the so-called WTO Friends group.[8] These lists are compiled using the Harmonized System (HS), the most widely used product classification system for trade across countries. The HS classifies products with 6-digit codes, and Eurostat provides a crosswalk to the PRODCOM data, which allows the identification of the PRODCOM codes that are considered green under the CLEG.

We use an additional list of green goods that use the Eurostat criteria of environmental protection and resource management described above. Although there is no official list of green products

[7] The case of bicycles has, in particular, been at the centre of controversy within WTO negotiations. China and the European Union have been unable to find a compromise around their inclusion, which contributed to the collapse of trade talks in 2016 (European Commission, 2019).

[8] The group is composed of Canada, the European Union, Japan, Korea, New Zealand, Norway, Switzerland, Chinese Taipei, and the United States.

compiled by Eurostat, the list from the German Statistical Office is used by Eurostat (2009) as an example of a list following its criteria.

We consider the union of the CLEG and German lists to provide a comprehensive list of potential green goods that consists of 902 products. We refine this very broad list to our favourite PRODCOM list of green goods, excluding goods with multiple usages. To do so, we review the product descriptions of the PRODCOM codes and exclude products with both green and non-green usages, such as tanks, industrial ovens, baskets, and mats. Among the goods with double usages, we retain only the machineries that are related to the monitoring and analysis of environmental variables such as thermostats and apparatus equipment for meteorology and the chemical analysis of water. These are included in all three lists that make up the CLEG, signalling a consensus around their high green potential. More generally, instruments that provide information on environmental variables are essential for controlling and minimizing the impact of production processes across most industries.

Our cleaning procedure leaves us with 225 (from 4288 products included in the PRODCOM data and 902 products from the union of the CLEG and German lists) products that we identify as green. In the following section, we compare the PRODCOM list with other existing lists to reinforce our argument for using this list.

2.4 Validation of our measure.

In this section, we compare our favourite list with other alternative lists. These latter lists include the CLEG and German lists already discussed above, as well as other, more restrictive lists. We discuss each of these lists in detail in Appendix A (see Table A.1); broadly speaking, we compare our own PRODCOM list with a set of broader lists (CLEG, Germany, APEC, PEG and WTO2009) and of narrow lists (WTO Core and Core CLEG). Table 1 correlates the dummy variables indicating the presence of a certain product in a given list and reports the number of products included in each list.

[Table 1 about here]

As expected, as we expand our definition of green products, there is an increasing overlap among the lists, but as the definition is narrowed down, lists diverge, identifying different sets of products.

While the correlation across broader lists (PEGS, APEC, WTO2009 and CLEG) tends to be quite high, narrower lists, such as the WTO Core and Core CLEG lists, are weakly correlated with each other. For instance, the WTO Core and Core CLEG lists share only one green product, i.e., spectrometers using optical radiation, confirming once again the importance of environmental monitoring activity.

Our PRODCOM list correlates rather strongly with the WTO2009 list, as well as its narrow version, the WTO Core list (with coefficients of 0.31), and with the PEGS list (0.57). We also find a strong correlation coefficient (0.46) between our PRODCOM list and a core list that we define as a union of the WTO Core and Core CLEG lists. This implies that the PRODCOM list identifies a large set of products that are included in either of the two most restrictive lists, offering additional support to the reliability of our favourite list. To give a few examples, these products include end-of-pipe technologies such as machinery for purifying gases and liquids as well as integrated technologies such as solar cells and monitoring equipment for physical and chemical analysis. Interestingly, the correlation coefficient between the German list (also based on PRODCOM data) and the PRODCOM list is only 0.12, which reflects the fact that the German list follows Eurostat's guidelines on environmental protection and resource management and disregards the issue of multiple usages.

Figure 1 visually shows the overlap between our favourite PRODCOM list, the broadest CLEG list, Germany's list and the narrowest core list. We find that 79 out of 147 products from the German list that are not included in any other list and that the CLEG has several products, 510 out of 605, that are not part of other lists. These products again include multi-usage products such as tanks, industrial ovens and machinery for sorting and grinding material.

[Figure 1 about here]

The narrow core list is fully contained in the CLEG list, but it also shares products with the other two lists. This suggests that there is a consensus around products included in the core list and may make it a credible alternative to our own PRODCOM list. However, we find that important green products are not included in the core list. The core list, in fact, focuses on products whose function is to directly combat pollution through the use of end-of-pipe technologies (i.e., water and waste management equipment) rather than on key integrated technologies (such as wind turbines). It also leaves out secondary environmental products that offer more environmentally sustainable mobility options, such as bicycles, and environmental monitoring equipment.[9]

In conclusion, our own list seems more accurate than other available lists. On the one hand, broad lists, such as the CLEG, German and APEC lists, include products with multiple non-green usages. On the other hand, narrow lists leave out integrated technologies such as wind turbines, electric cars and environmental monitoring equipment. The PRODCOM list we have compiled strikes a balance between these two extremes by focusing on single-usage products and by including both products that directly affect the environment and products that provide greener production processes that reduce pollution across other industries.

3. Green production across industries and high-green-potential industries.

A crucial advantage of the PRODCOM data is that it allows the construction of a measure of green production at fine-grained industry (4-digit NACE) and country levels over two decades (1995-2015). We begin by exploring the industry dimension of the data using the share of green production relative to total production as key statistics. To the best of our knowledge, by observing green production at the product-level within each industry, our paper is the first to compute a continuous measure of industrial greenness that varies along three dimensions (industry-country-year). We show in what follows that using a continuous measure allows us to capture the high degree of heterogeneity in green production across and within industries, especially at the four-digit level.

3.1 Aggregated industries: green vs. brown production.

In Table 2, we first explore the variability of green production, aggregating the data at 2-digit manufacturing sectors because this higher level of aggregation allows the comparison of the output-based and emission-based definitions of green production. We report the mean and standard deviation of green shares for each industry, as well as the average GHG intensity. As mentioned

[9] Gas turbines are included in the WTO list and so are also part of our PRODCOM list. Clearly, their treatment is problematic. On the one hand, they are a transition technology, so they can be considered green. On the other hand, they produce GHG emissions, so they are brown. We choose to keep them in the main analysis, but we confirmed that the results are consistent by excluding them. These results are available upon request by the authors.

in Section 2.2, the number of countries included in the PRODCOM data is unbalanced, so for the sectoral analysis, we focus on 2005, 2010 and 2015, for which we have information for a balanced panel.

We find that green production is highly concentrated in a few industries. While most 2-digit sectors (17 out of 26) have no production of green goods, four industries emerge as the key players in the green transition: i. Computer, electronic and optical equipment, which includes photovoltaic panels; ii. Electrical equipment, which includes equipment for the control and distribution of electricity; iii. Machinery and equipment, which includes wind turbines; and iv. Other transport equipment, which includes railway stocks. Remarkably, these four industries represent 86% of the total green production (column 6).

Within these four industries, we also observe a rather high coefficient of variation (standard deviation), which indicates a high degree of heterogeneity in green production across countries. Over time, the average shares increase in all of the four greenest industries, which contrasts with the stability of the average green share in other green industries. Overall, the diffusion of green production tends to remain highly concentrated in a few industries rather than spreading across industries.

[Table 2 about here]

Importantly, the four industries with a high green potential have a number of other characteristics that make them of strategic interest for industrial policy in general. First, they are all high-tech industries (Eurostat, 2015; Galindo-Rueda and Verger, 2016) that have been found to have large job multipliers (Moretti, 2010; Vona, Marin and Consoli, 2019) and to be conducive to economic growth (Mcmillan, Rodrik and Verduzco-Gallo, 2014; Szirmai and Verspagen, 2015). Second, specialization in these sectors requires a broad set of pre-existing skills (Hidalgo *et al.*, 2007; Mealy and Teytelboym, 2019), particularly engineering and technical skills that have also been found to be prevalent in green jobs (Vona *et al.*, 2018). Third, consistent with the fact that high-tech manufacturing has strong inter-sectoral linkages with the rest of the economy (Hirschman, 1958; Szirmai and Verspagen, 2015), these four green industries are rather upstream and serve intermediate rather than final demand. This is also consistent with evidence pointing to the importance of the suppliers of machinery and electrical equipment for innovation in the renewable energy sector (Jacobsson and Bergek, 2004; Markard and Truffer, 2006).

To help with the comparison of output-based and process-based definitions of green production, the last column of Table 2 reports greenhouse gas (GHG) intensity for the same 2-digit manufacturing industries. We rely on the environmental accounts of the World Input-Output Database (WIOD) that include the energy and GHG content of domestic production of each 2-digit industry for 15 countries between 1995 and 2009.[10] We compute GHG ($C0_2$, N_20 and CH_4, aggregated according to their global warming potential) intensity as the sum of direct and indirect emissions per unit of value added from each industry, country and year. A well-known cluster of brown industries stands out in terms of total (direct and indirect) emissions (Wiebe and Yamano, 2016; de Vries and Ferrarini, 2017): coke and refined petroleum products, other non-metallic mineral products, chemicals and chemical products, basic pharmaceutical products and pharmaceutical preparations, and basic metals and the manufacturing of fabricated metal products, except machinery. In the remainder of this paper, we treat the entire production process of these brown industries as polluting (see Appendix B for details).

Remarkably, comparing columns 3 to 5 with column 8 of Table 2, we observe that there is a strong inverse relation between green production and pollution intensity. This has two main implications for our study. First, from a conceptual point of view, the process and output-based approaches to defining what is green capture completely different aspects of the green economy but are not in contradiction with each other and in fact end up identifying similar "green" sectors. Second, the two approaches are clearly complementary for analyzing policy impacts and understanding the distributional effects of environmental policies. While the competitiveness of brown industries is potentially harmed by an increase in environmental policy stringency (Dechezleprêtre and Sato, 2017), green upstream sectors benefit from the indirect demand for pollution abatement equipment, technical know-how and integrated technologies (Horbach, Rammer and Rennings, 2012; Vona, Marin and Consoli, 2019). Overall, the two well-known channels through which environmental policies affect competitiveness, namely, the cost channel (eventually leading to relocating polluting industries abroad, i.e., the pollution haven hypothesis) and the innovation channel (the so-called Porter hypothesis; Ambec *et al.*, 2013), impact different sets of industries. It follows that ambitious environmental policy will also impact countries differently, depending

[10] We use the 2013 release of the WIOD, for which environmental accounts include information on embodied emissions of GHG for the following countries: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, the Netherlands, Spain, Sweden and the UK. In Appendix B we discuss in detail how we compute GHG intensity using the WIOD, following Marin and Vona (2019).

on their productive structures. The winners of such policies, i.e., countries that already have a comparative advantage in green industries, may be different from the losers, i.e., countries with a comparative advantage in brown industries.

3.2 Disaggregated industries: identifying high-green-potential industries.

We compare green and polluting production at 2 digits of aggregation due to data constraints related to measures of pollution intensity. However, the high level of disaggregation of the PRODCOM data allows us to compute the shares of green production for 4-digit industries. This is important for further understanding which specific industries green production is concentrated in.

[Table 3 about here]

Table 3 reports statistics on 4-digit industries with a green production greater than zero in at least one year. Table 3 confirms that green production is also highly concentrated at the 4-digit level; of the 119 4-digit industries among the 2-digit industries with a green production greater than zero, only 21 are green. Moreover, we find that 11 out of these 21 industries have a maximum green production of 100%, i.e., for at least one country and year, green production was the entirety of the industry's production.

After ranking industries by their average share of green production, we observe a first group of nine extremely green industries, from "bicycle and invalid carriage manufacturing" to "non-domestic cooling and ventilation equipment manufacturing". For these industries, the average green share is above 25% and is not distant from the median, so the outliers do not drive the results. Moreover, there is always at least one industry with a country-year observation with 100% green production, and the absolute long-term changes tend to be positive, with the exception of those in railway production. Finally, these industries represent 83.4% of total green production. We then observe a second group of four industries, including the production of LEDs and PV panels (in "electronic components manufacturing"), that represent another 11.7% of the total green production. The remaining eight industries account for less than 5% of the total green production and always have mean shares of green production below 0.04; thus, they can be considered marginally or indirectly green (like metal industries).[11]

In the remainder of the paper, we study green specialization and comparative advantage, so we focus on the 13 industries included in the first two groups, which we call *high-green-potential industries*. These industries appear the most relevant to understanding how green specialization has evolved in EU countries over the last two decades. Note also that comparing green and non-green production within the same industry when the former represents only a very small share of the industry's total production can yield misleading results when computing country-level averages that are weighted by industry turnover. In fact, marginally green industries may have high total production numbers and be assigned heavy weights that drive the averages at the country level.[12] However, for the sake of completeness, Appendix D replicates the main analyses of Section 4 for the full set of green industries (in particular, see Figures D.1, D.2 and D.3).

4 Concentration and specialization in green production.

In this section, we exploit the cross-country variation in our data to study the spatial concentration and specialization of high-green-potential industries. We analyse green and non-green production within high-green-potential industries so that we can compare green and non-green products that share the same industry-level characteristics.

4.1 Concentration of green production across countries.

We measure the concentration of green production with the Herfindahl-Hirschman index (HHI), which has been widely used in the literature on industrial organization (Daughety, 1990; Matsumoto, Merlone and Szidarovszky, 2012). More recently, Perruchas, Consoli and Barbieri (2020) applied this index to study the geographical concentration of green technologies. We use the HHI index to compare the geographic diffusion of green production with that of non-green production within high-green-potential industries.

---

[11] In Figure C.1 and Figure C.2 in Appendix C we show that high-green-potential products represent a large share of total green production across countries and years. Looking at the product level in Table C.1, we also find that the top three green products are remarkably similar across countries.

[12] We discuss this in more detail in Appendix D when comparing the concentrations of green and non-green production in Figure D.1.

The HHI index is the summation of each country's share in the production of a given sector, and it varies between 0 (uniform distribution) and 1 (concentration of production in only one country):

$$HHI_j = \sum_{i=1}^{N} s_{ij}^2, \tag{1}$$

where s_{ij} is the share of production in country i in industry j. Because our interest is in comparing green and non-green production for each high-green-potential 4-digit industry, we compute a green HHI (HHI_g) using green production only and a non-green HHI (HHI_{ng}) using non-green production only. The concentration of production within an industry depends on industries' characteristics such as economies of scale and capital intensity; thus, we need a relative measure that, like an exact matching procedure, eliminates industry characteristics correlated with the industry's concentration. To this end, we take the ratio between the HHI_g and HHI_{ng}:

$$HHIR_{j,g/ng} = \frac{HHI_{j,g}}{HHI_{j,ng}} \tag{2}$$

The relative HHI ($HHIR_{j,g/ng}$) is above one if the production of green goods has a higher concentration than that of non-green products in the same 4-digit industry. Note that within-industry comparisons are more appropriate than comparisons between industries for understanding whether green production is different from non-green production. Indeed, if we were to compare the concentration of total green and total non-green production at the country level, the results would be driven by the country's industry composition, as green production is prevalent in a few high-tech and capital-intensive industries that are typically more concentrated than non-green industries (see Table 2).

[Figure 2 about here]

In Figure 2, we plot the ratio of the green and non-green HHIR for the weighted average of high-green-potential industries, using industry turnover as a weight to give more importance to larger industries. The results are presented for all countries available in each year (solid line) as well as for balanced panels of countries (dashed and dotted lines). We observe that at the beginning of our sample, green production in high-green-potential industries is, on average, 30% more concentrated relative to non-green production in the same industries. The relative HHI, however, decreases over the two decades covered by our analysis, reaching values near unity in the final years of our sample. The decreasing trend in the HHIR also indicates that green production has diffused across Europe.[13] Our data do not allow us to discern possible explanations for this pattern, although product life cycle theories may help explain the higher concentrations in the production of goods that are at early development stages (Vernon, 1966; Keppler, 1996). In this vein, it appears that green products are at early life cycle stages in the earlier years of our dataset and show higher levels of concentration that decrease as they become more mature over time. By comparing the two lines of the figure, we see that the entry of Eastern European countries (dashed line) does not contribute to explaining this pattern, while the financial crisis of 2008 accelerates the diffusion of green production across countries.

4.2 Specialization of green production across countries.

We compare the share of green production in high-green-potential industries across countries to detect the leaders of the green transition in Europe and the extent to which their advantage is persistent. In Figure 3, we group countries based on size and geographic position to look at large (panel A), small (panel B) and Eastern European (panel C) countries and to plot the evolution of the 3-year moving average of green production shares. As a benchmark, each panel also includes the European (weighted by turnover) average of green shares across all available countries in each year.[14]

Green production shares in high-green-potential industries rarely exceed 4%, with the exception of Denmark, which peaks at 9.5% in 2015. In terms of country rankings, those with the largest shares of green production are Denmark, Germany, the UK, Sweden, Austria and Hungary. Because green production is concentrated in high-tech sectors, this finding resonates with the fact that specialization in such sectors is highly persistent and path dependent. All leaders are, with the exception of Hungary, high-income countries that are at the technological frontier and have strong capabilities in high-tech industries. This suggests that engineering and technical competences, which are core competences for these industries, may be easily reused in green production, a hypothesis that we will explore in the next section**.**

[13] Obviously, this pattern masks significant levels of heterogeneity across sectors that, for sake of space, we discuss in further detail in Figure C.3 and C.4 of Appendix C.

[14] In comparing this with the shares of green production reported in Table 2 and 3, it is important to stress that Figure 3 reports country-level shares, while Table 3 reports the shares of green production within each industry. It is then not surprising that while we find that the green share of production in the high-green-potential industries fluctuates between 8% and 79%, at the country level, green shares are much lower, hovering between 2% and 4%.

Note that we also find high persistence in the green shares of production, which may be due to the fact that we observe a modest increase in the share of green production over the time period considered: for the average green share of our full sample, we observe an increase of 0.5 percentage points from 2% to 2.5%. The lack of widespread diffusion of green production is also related to the emergence of China as a key player in the green economy (Algieri, Aquino and Succurro, 2011; Sawhney and Kahn, 2012; Liu and Goldstein, 2013).

It is also interesting that, compared to patent-based measures of green shares across countries and with the exception of Denmark, the ranking of other countries is rather different, especially for Hungary, which is among the lowest countries in terms of shares of green patents (OECD, 2017). Rather than depending upon the presence of green inventors, as proxied by patents, the location of green production seems to rely on other sources of comparative advantage, such as labour costs and the availability of skilled workers. These issues are clearly beyond the scope of this exploratory study and are left for future work.

[Figure 3 about here]

Green shares of production are informative about the importance that green goods have in industrial production but do not measure green specialization, as they do not entail a comparison with a benchmark. RCA indexes are the most popular approach for defining whether a country is specialized or not in a given production or technology (Balassa, 1965; Cole, Elliott, and Shimamoto, 2005; Hidalgo *et al.*, 2007; Petralia, Balland, and Morrison, 2017). The RCA index is computed as follows:

$$RCA_{ijt} = \frac{y_{ijt} \big/ \sum_j y_{ijt}}{\sum_i y_{ijt} \big/ \sum_j \sum_i y_{ijt}}, \qquad (3)$$

where y_{ijt} is the production of sector *j* in country *i*. The index normalizes the production share of sector *j* in country *i* by dividing it by the production share of sector *j* across all countries. Note that the economically significant threshold in this index is the point of unity, which means that values between 0 and 1 represent non-specialization, while RCA values above 1 show specialization. As a result of this asymmetry, statistical analyses using Balassa's RCA measure may give too much weight to values above one (Dalum, Laursen and Villumsen, 1998; Cole, Elliott and Shimamoto, 2005; Yu, Cai and Leung, 2009). To fix this, Laursen (1998) proposes bounding the index between -1 and 1, making it a symmetric RCA (SRCA) around 0: $SRCA_{ijt} = (RCA_{ijt} - 1)/(RCA_{ijt} + 1)$. In what follows, we use the SRCA, though we refer to it as RCA for the sake of convenience.

We first use this index to assess the correlation between green and brown specializations. Estimating such correlation is important to highlight the winners and losers of ambitious European environmental policies, such as the green new deal plan. The green RCA is computed by treating green production from high-green-potential industries as a unique sector, i.e., y_{ijt} is the total green production from all high-green-potential industries for each country i. Likewise, the brown RCA is computed by treating all of the polluting industries defined in Table 2 as a single sector and by considering all of their production brown.

In Figure 4, we plot green and brown RCA for selected years and divide countries into four quadrants. We choose 2001 as our earliest year because the PRODCOM data are not available for Eastern European countries in previous years. Countries in the top-left quadrant have an RCA in green production but not in polluting production. The top-right quadrant shows countries with an RCA in both types of production, the bottom-right shows countries with an RCA only in polluting production and the bottom-left shows countries with an RCA in neither type of production. We observe that the number of countries with a green RCA (i.e., those above the horizontal dashed line) slowly increases between 2001 and 2010 (with Hungary and then Austria joining Sweden, Germany and Denmark) but remains quite stable overall. Specialization in polluting industries shows a much smaller dispersion, with most countries clustered around 0 (the vertical dashed line), although brown specialization emerges in countries with lower income per capita (such as Romania, Bulgaria, Greece) as well as in some traditional industrial economies (such as Italy and Belgium). Importantly, the green and brown RCAs are negatively correlated, indicating that they often occur in different countries with an estimated slope always beyond -0.45. This evidence, together with the fact that the green leaders are mostly rich countries, indicates that the effect of EU environmental policies may exacerbate the gap between the core and the periphery of Europe in green sectors that will be strategic for future economic development.

In conclusion, green specialization at the country level has, on average, increased over time and is negatively correlated with specialization in polluting industries. However, the green vs. brown comparison is mostly between-sector, so it is silent on the important issue of the within-sector complementarity between green and non-green specialization. To shed more light on this issue, as well as on the issue of path dependency in green specialization, in the next section, we perform a multivariate regression analysis at the sector-by-country level. This analysis allows us to study the relationship between green and non-green RCAs within the same industry, controlling for sector and country level characteristics.

5 Path dependency and complementarity of green comparative advantage.

Our descriptive analysis provides interesting and new insights into the structure and evolution of green specialization. On the one hand, the aggregated increase in the share of green production is associated with both a reduction in the spatial concentration of green production across countries and the consolidation of a few leaders that exhibit a green comparative advantage, suggesting that despite its diffusion across Europe, green production is still characterized by a significant path dependency. On the other hand, the green leaders are countries that already specialize in high-tech sectors producing capital equipment.

Previous research on green innovation seeks to understand the extent to which non-green knowledge can be “recombined” and used to develop new green technologies. Zeppini and van den Bergh (2011) provide several examples: hybrid cars combine, for example, electric propulsion systems with internal combustion engines; similarly, photovoltaic film combines solar cells and thin layer technologies. In the model of Zeppini and van den Bergh (2011), recombining non-green technologies to generate green innovation is the main channel through which companies can escape being locked in to brown technologies and successfully redirect technological change. Theoretically, it is not clear whether what matters for green innovation is the use of non-green knowledge that is similar to green knowledge or the diversity of the knowledge base that increases opportunities for fruitful knowledge recombination (Weitzman, 1998; Olsson and Frey, 2002; Caminati, 2006). In empirical research, the results are still scant and mixed due to the difficulties in measuring separate proxies of knowledge diversity and similarity for the green economy (e.g., Colombelli and Quatraro, 2019; Perruchas et al., 2020).[15]

---

[15] Perruchas, Consoli and Barbieri, (2020) find that countries are more likely to recombine their technological capabilities into green ones that are close to their existing technological specialization. Colombelli and Quatraro, (2019) argue that because of the recombinant and complex nature of green innovation, the diversification of existing

We contribute to this literature using production data rather than patent data, thus focusing directly on green specialization. Our goal is to investigate some structural properties of green specialization in a horse-race regression, comparing the role of three main drivers that have been examined by the literature on green innovation using patents: i. path dependency, ii. complementarity with proximate capabilities, and iii. diversification of the knowledge base. With this aim in mind, we estimate the following equation for high-green-potential sectors over the period 2005-2015:[16]

$$\ln(RCA^{g}_{ij,t}) = \alpha + \sum_t \beta_t \times \ln(\overline{RCA}^{g}_{ij,t_0}) + \gamma \ln(RCA^{ng}_{ij,t-1}) + \delta \ln\left(\#RCA^{g}_{ij,t-1}\right) + \vartheta \ln\left(\#RCA^{ng}_{ij,t-1}\right) + \tau_{it} + \tau_{jt} + \varepsilon_{ijt}. \quad (4)$$

where our dependent variable is (the log of) the RCA indicator for the green production ($RCA^{g}_{ij,t}$) of country *i* in sector *j* at time *t.* The RCA index is now built for each sector and differentiates green and non-green RCAs (henceforth $RCA^{ng}_{ij,t}$).[17] The median share of green production is 0.27 in our estimation sample of high-green-potential industries, implying that there is enough variation in the data to construct the non-green RCA index $RCA^{ng}_{ij,t}$. τ_{it} and τ_{jt} are, respectively, country-year and sector-year dummies that account for unobserved shocks, such as the impact of the great recession on different countries and the diffusion of green technologies in specific sectors. Country-by-year dummies also absorb the effect of environmental policies in a flexible manner, which is beyond the scope of this paper. We take the natural logarithm of all variables of interest

capabilities is also important for the development of green technologies. The authors find that the diversification of the stock of knowledge is conducive to the creation of green start-ups in Italian provinces.

[16] Most Eastern European countries enter in our dataset in 2001, with the exception of Poland, which is included only from 2003 onwards. As a consequence, focusing on the years 2005-2015 allows us to have a balanced panel and to compute pre-sample means for all countries.

[17] We compute an RCA for each high-green-potential industry based on its green and non-green production.

$$RCA^{k}{}_{ijt} = \frac{y^{k}_{ijt} \Big/ \sum_j y_{ijt}}{\sum_i y^{k}_{ijt} \Big/ \sum_j \sum_i y_{ijt}},$$

where $k = g$ (green) or ng (non-green, i.e., $y^{ng}_{ijt} = y_{ijt} - y^{g}_{ijt}$). We refer to these two measures as green and non-green RCA, respectively, and make these indexes symmetric around zero, as above.

because it allows us to deal with the skewedness of certain variables included in equation (4) and, at the same time, interpret the estimated coefficients as elasticities.[18]

For the explanatory variables, the main proxy of path dependency in green specialization is the pre-sample mean of the green RCA ($\overline{RCA}^{g}_{ij,t0}$) computed for the years 2001-2004. We interact the pre-sample mean of green RCA with time dummies to assess how persistent a “head start” over time is, which is more coherent with the notion of path dependency than using the lagged dependent variable, as in standard dynamic models, would be. The pre-sample mean also captures unobserved individual characteristics in a more flexible way than individual fixed effects for variables that are highly persistent (Blundell, Griffith and van Reenen, 1995).

The degree of complementarity between green and non-green capabilities is captured by the level of non-green RCA within the same four-digit sector and lagged one year ($RCA^{ng}_{ij,t-1}$). Taking the level of non-green specialization within the same detailed 4-digit sector represents a natural way to measure capabilities that are similar to green ones. A priori, the effect of having a stronger non-green RCA on green specialization is unclear. It can be positive if the non-green competences can be replicated and successfully used to create a green comparative advantage within the same sector. It can be negative if there is competition between the green and non-green uses of a similar pool of competences. While determining which effect would prevail is an empirical issue that we will explore through equation (4), the unconditional correlation between green $RCA^{g}_{ij,t}$ and non-green $RCA^{ng}_{ij,t}$ is rather high (0.51); thus, we expect stronger non-green capabilities within the same sector to be a driver of green comparative advantage.

To capture diversification in a country’s competences within a particular sector, we include the number of products with comparative advantage, again differentiating between green and non-green RCAs. Green (non-green) diversification is measured with the number of green (non-green) products with a symmetrical RCA>0, i.e., above the threshold to designate a country as having a comparative advantage for that product at time $t-1$ ($\#RCA^{g}_{ij,t-1}$ and $\#RCA^{ng}_{ij,t-1}$ for green and non-green diversification, respectively). We argue, in line with the well-established literature on structural change, that countries that specialize in products based on their productive capabilities (Hidalgo *et al.*, 2007; Hidalgo and Hausmann, 2009), and therefore, the number of green goods

[18] For the RCA, we take the log(2+RCA) so that an RCA equal to -1 is set equal to 0, while for the other variables, we take the log(1+x). While symmetric RCAs are usually transformed in logs, we choose to transform them in order to facilitate the interpretation of the coefficients in terms of elasticities.

produced with an RCA within each country-industry, will capture the breadth of green productive capabilities.

Given the high skewness in the distribution of the number of products with an RCA, we also consider a specification where we replace the number of green (non-green) products with an RCA with a dummy equal to one if the country produces at least one green (non-green) product with an RCA in the previous year. Table E.1 in Appendix E reports summary statistics for the variables included in the regressions, which confirm the high skewness of the distribution of the number of products with an RCA.

Table 4 presents the estimation of equation (4). Note that we do not weigh the estimates by turnover in order to avoid giving excessive importance to larger countries, which have higher turnover in all sectors. The estimated coefficients can thus be interpreted as unweighted average associations. We add the variables of interest sequentially to assess the contribution of each variable of interest and how it interacts with other variables.

[Table 4 about here]

The first finding is a remarkable persistence in the head start in green specialization, although adding proxies for complementarity and diversification significantly reduces the importance of the initial conditions. The first column shows that in eleven years, the elasticity of initial conditions declines from 0.97 to 0.62. This implies that, conditional on country and sector trends, an initial green advantage of 10% continues to explain as much as a 6.2% difference in green RCA after eleven years. In our favourite specification in column 3, although these associations are almost halved, a 10% head start still explains a 3.7% difference in green specialization eleven years later. In Table E.2 in Appendix E, we replicate the same analysis for the non-green RCA in the same high-green-potential industries. Interestingly, non-green specialization is more path dependent than green specialization: after eleven years, the elasticity of initial conditions is 0.45 versus 0.37 in our favourite specification (see Table E.2 in the Appendix). This implies that within a group of technologically advanced sectors such as the high-green-potential sectors, the comparative advantage of green products is slightly more fluid than that of non-green products.

Our second finding is that green and non-green specialization reinforce each other. As expected, after we control for path dependency and country and sector trends, the degree of complementarity is substantially lower than the unconditional correlation of 0.51. An increase of 10% in the non-green RCA explains a difference of 0.95% in the green RCA (column 3), which is, however, significantly larger than the degree of complementarity between green and non-green RCA obtained in Table E.2 (i.e., 0.7%). In the appendix, we show that the complementarity effect holds when the dependent variable is not log transformed (Table E.3) and without transforming the RCA in a symmetric index (Table E.4). In contrast, the effect is weakened when we weight the regression based on the size of the sector (Table E.5) or include marginally green sectors in our analysis (Table E.6). Recall, however, that marginally green sectors represent less than 5% of total green production.

The third result is that only green-related diversification matters. The role of green diversification is sizeable, as testified by the size of the estimated elasticity, e.g., 0.26 (column 3). Importantly, this correlation is conditional on the initial level of green specialization; thus, only countries that are successful in diversifying green productions are able to keep their initial head start. When we look at whether a country sector has at least one product with an RCA, be it green or non-green, instead of at its level of diversification, we find that having at least one green product with an RCA is positively associated with subsequent green specialization (column 4). This suggests that it is important to have a minimum threshold of green capabilities to sustain and reinforce a green specialization path. Indeed, this threshold effect is quantitatively more important than green diversification per se when we add both to the regression (column 5).

Overall, leading in at least one green product and diversifying green production is important for sustaining green specialization, but the former is significantly more important than the latter. While these regressions provide interesting insights into the structure and evolution of green comparative advantage, we do not examine in detail the causes of such advantages or their interactions with the structural characteristics highlighted here. Such work is left to future research.

6 Conclusions.

This paper presents new stylized data on the structure and evolution of specialization in green productions by assembling a new dataset based on the PRODCOM dataset of Eurostat, which allows us to examine variation in green production across detailed sectors (4-digit NACE), countries (in the EU) and years (1995-2015). We construct a favourite list of green products by comparing and synthetizing several existing lists of green goods proposed during recent international negotiations at the WTO. Our main criterion is excluding green goods with double usages from our final list, as this is the most challenging issue in the debate on the definition of what is green.

By exploiting the richness of our data, we are the first to study the distribution of green production across both sectors and countries. Our first finding is that there is virtually no overlap between green production and the (direct and indirect) pollution intensity across two-digit NACE industries. This result has two important implications. First, in the debate on the definition of what is green, the process and output-based approaches capture different aspects of the green economy but are not in contradiction with each other and end up identifying similar "green" sectors. Second, in the design of environmental policies, the winners and losers will be different, raising the issue of the distributional effects of such policies because the sectors receiving green subsidies are different from those paying environmental taxes. The analysis of the revealed green and brown comparative advantage indicates that, indeed, European countries tend to specialize either in green or brown sectors.

The second result is that green production is and remains highly concentrated in a few sectors despite an average increase of 25% (from 2% to 2.5%) over the time period considered: out of 119 4-digit manufacturing sectors, 13 of them represent 95% of European green production and are those where green production has been most diffused. This result qualifies the policy implication above, as these high-green-potential sectors are high-tech, produce capital goods, are large job multipliers and have strong inter-sectoral linkages with the rest of the economy. They are also relatively upstream sectors that can enhance the environmental sustainability of other industries by making production processes less harmful to the environment.

Third, since green production is likely to require competences and skills similar to those used in non-green production in high-green-potential industries, our RCA measure indicates that green leaders are countries such as Germany, Denmark, Sweden and Austria where high-tech sectors were already strong. However, we also observe that the existence of persistent green leaders coexists with the general fact that, on average, green production in the 13 high-green-potential industries has become less concentrated over time. In this respect, the fact that our data cover only European countries represents a limitation because the concentration of production and catching-up are affected by other major players in green industries such as China, Japan and the US that we do not observe and cannot take into account in our analysis.

Finally, we consolidate these facts in a multivariate regression framework where we seek to compare the structural determinants of green specialization. Not surprisingly, we find a remarkable path dependency in green specialization that is, however, lower than the path dependency in non-green specialization in the same high-green-potential industries, indicating that the lock-in in green specialization is less rigid than that in non-green specialization. Second, within similar 4-digit industries, green and non-green specializations complement and reinforce each other. The role of such complementarities is clearly smaller than that of path dependency but corroborates the descriptive analysis pointing to the pre-existing advantage in certain high-tech sectors. Third, acting as a leader in at least one green product and diversifying green production are important for sustaining green specialization, but the former is significantly more important than the latter. As a result, it is better for a country to first specialize in one green good and then move on to creating new comparative advantages.

A possible shortcoming of our analysis is that the country coverage is limited to European countries. Because the index of comparative advantage is relative in nature and depends on the number of countries available in the data, there is limited cross-country variation in our data. While an extension of our analysis to trade data is planned, the advantage of studying green specialization with limited cross-country variation is that we can be the first to study green production and not only green trade in detail. Another limitation of the PRODCOM data is that it only covers the production of manufactured goods and thus excludes the service sector. Leaving services out of our analysis also means ignoring the largest part of European economies, some of which may have a significant positive role in the green economy. Finally, our analysis identifies green products based on their potential to benefit the environment, and comparison with pollution intensity production is possible only at 2 digits of aggregation. Future research will greatly benefit from more disaggregated information on the pollution content of production.

Acknowledgement

This work was supported by Horizon 2020 Framework Programme, project INNOPATHS [grant number 730403].

References

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R. and Van Reenen, J., 2016. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), pp.1-51.

Algieri, B., Aquino, A. and Succurro, M., 2011. Going "green": trade specialisation dynamics in the solar photovoltaic sector. *Energy Policy*, *39*(11), pp.7275-7283.

Ambec, S., Cohen, M.A., Elgie, S. and Lanoie, P., 2013. The Porter hypothesis at 20: can environmental regulation enhance innovation and competitiveness? *Review of environmental economics and policy*, *7*(1), pp.2-22.

APEC, 2012. APEC Leaders' Declaration ANNEX C APEC List of Environmental Goods, Vladivostok, 8–9 September.

Arundel, A. and Kabla, I., 1998. What percentage of innovations are patented? Empirical estimates for European firms. *Research policy*, *27*(2), pp.127-141.

Balineau, G. and de Melo, J., 2011. Stalemate at the negotiations on environmental goods and services at the Doha Round. *Fondation Pour Les Études Et Recherches sur le Développement International Working Paper* No. 28.

Becker, R.A. and Shadbegian, R.J., 2009. Environmental products manufacturing: A look Inside the Green Industry. *The BE Journal of Economic Analysis & Policy*, *9*(1).

Bi, X., 2017. "Cleansing the air at the expense of waterways?" Empirical evidence from the toxic releases of coal-fired power plants in the United States. *Journal of Regulatory Economics*, *51*(1), pp.18-40.

Blundell, R., Griffith, R. and Reenen, J.V., 1995. Dynamic count data models of technological innovation. *The Economic Journal*, *105*(429), pp.333-344.

Bovenberg, A.L. and Smulders, S., 1995. Environmental quality and pollution-augmenting technological change in a two-sector endogenous growth model. *Journal of public Economics*, *57*(3), pp.369-391.

Brunel, C., 2017. Pollution offshoring and emission reductions in EU and US manufacturing. *Environmental and Resource Economics*, *68*(3), pp.621-641.

Calel, R. and Dechezlepretre, A., 2016. Environmental policy and directed technological change: evidence from the European carbon market. *Review of economics and statistics*, *98*(1), pp.173-

191.

Caminati, M., 2006. Knowledge growth, complexity and the returns to R&D. *Journal of Evolutionary Economics*, *16*(3), p.207.

Cantore, N. and Cheng, C.F.C., 2018. International trade of environmental goods in gravity models. *Journal of environmental management*, *223*, pp.1047-1060.

Caro, D., Pulselli, F.M., Borghesi, S. and Bastianoni, S., 2017. Mapping the international flows of GHG emissions within a more feasible consumption-based framework. *Journal of cleaner production*, *147*, pp.142-151.

Cole, M.A., Elliott, R.J. and Shimamoto, K., 2005. Why the grass is not always greener: the competing effects of environmental regulations and factor intensities on US specialization. *Ecological Economics*, *54*(1), pp.95-109.

Colombelli, A. and Quatraro, F., 2019. Green start-ups and local knowledge spillovers from clean and dirty technologies. *Small Business Economics*, *52*(4), pp.773-792.

Consoli, D., Marin, G., Marzucchi, A. and Vona, F., 2016. Do green jobs differ from non-green jobs in terms of skills and human capital?. *Research Policy*, *45*(5), pp.1046-1060.

Conti, C., Mancusi, M.L., Sanna-Randaccio, F., Sestini, R. and Verdolini, E., 2018. Transition towards a green economy in Europe: Innovation and knowledge integration in the renewable energy sector. *Research Policy*, 47(10), pp.1996-2009.

Dalum, B., Laursen, K. and Villumsen, G., 1998. Structural change in OECD export specialisation patterns: de-specialisation and 'stickiness'. *International Review of Applied Economics*, *12*(3), pp.423-443.

Daughety, A.F., 1990. Beneficial concentration. *American Economic Review*, *80*(5), pp.1231-1237.

Dechezleprêtre, A. and Sato, M., 2017. The impacts of environmental regulations on competitiveness. *Review of Environmental Economics and Policy*, *11*(2), pp.183-206.

Deschenes, O., 2013. *Green jobs* (No. 62). IZA Policy Paper.

Dierdorff, E., Norton, J., Drewes, D., Kroustalis, C., Rivkin, D., and Lewis, P., 2009. Greening of the World of Work: Implications for O*NET-SOC and New and Emerging Occupations. National Center for O*NET Development.

Dosi, G., Grazzi, M. and Moschella, D., 2017. What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification. *Small Business Economics*, *48*(2), pp.413-429.

Dutz, M.A. and Sharma, S., 2012. *Green growth, technology and innovation*. The World Bank.

Elliott, R.J. and Lindley, J.K., 2017. Environmental jobs and growth in the United States. *Ecological Economics*, *132*, pp.232-244.

European Commission, 2019. *A6 A Balanced and Progressive Trade Policy to Harness Globalisation*..

Eurostat, 2015. *Glossary : High-tech classification of manufacturing industries*, *Eurostat Statistics Explained*.

Eurostat, 2016. *Environmental goods and services sector accounts manual : 2016 edition. Eurostat, Luxemburg*

Fankhauser, S., Bowen, A., Calel, R., Dechezleprêtre, A., Grover, D., Rydge, J. and Sato, M., 2013. Who will win the green race? In search of environmental competitiveness and innovation. *Global Environmental Change*, *23*(5), pp.902-913.

Fleming, L. and Sorenson, O., 2001. Technology as a complex adaptive system: evidence from patent data. *Research policy*, *30*(7), pp.1019-1039.

Fraccascia, L., Giannoccaro, I. and Albino, V., 2018. Green product development: What does the country product space imply? *Journal of cleaner production*, *170*, pp.1076-1088.

Frondel, M., Horbach, J. and Rennings, K., 2007. End-of-pipe or cleaner production? An empirical comparison of environmental innovation decisions across OECD countries. *Business strategy and the environment*, *16*(8), pp.571-584.

Galindo-Rueda, F. and Verger, F., 2016. OECD taxonomy of economic activities based on R&D intensity. OECD Publishing, Paris.

Ghisetti, C. and Rennings, K., 2014. Environmental innovations and profitability: How does it pay to be green? An empirical analysis on the German innovation survey. *Journal of cleaner production*, *75*, pp.106-117.

Gibson, M., 2019. Regulation-induced pollution substitution. *Review of Economics and Statistics*, *101*(5), pp.827-840.

He, Q., Fang, H., Wang, M. and Peng, B., 2015. Trade liberalization and trade performance of environmental goods: evidence from Asia-Pacific economic cooperation members. *Applied Economics*, *47*(29), pp.3021-3039.

Hidalgo, C.A., Klinger, B., Barabási, A.L. and Hausmann, R., 2007. The product space conditions the development of nations. *Science*, *317*(5837), pp.482-487.

Hidalgo, C.A. and Hausmann, R., 2009. The building blocks of economic complexity. *Proceedings of the national academy of sciences*, *106*(26), pp.10570-10575.

Hirschman, A. O., 1958. *Strategy of economic development*. New Haven, Connecticut and London: Yale University Press.

Horbach, J., 2016. Empirical determinants of eco-innovation in European countries using the community innovation survey. *Environmental Innovation and Societal Transitions*, *19*, pp.1-14.

Horbach, J., Rammer, C. and Rennings, K., 2012. Determinants of eco-innovations by type of environmental impact—The role of regulatory push/pull, technology push and market pull. *Ecological economics*, *78*, pp.112-122.

Hufbauer, G.C. and Kim, J., 2010. Reaching a global agreement on climate change: what are the obstacles?. *Asian Economic Policy Review*, *5*(1), pp.39-58.

Jacobsson, S. and Bergek, A., 2004. Transforming the energy sector: the evolution of technological systems in renewable energy technology. *Industrial and corporate change*, *13*(5), pp.815-849.

Jaffe, A.B. and Palmer, K., 1997. Environmental regulation and innovation: a panel data study. *Review of economics and statistics*, *79*(4), pp.610-619.

Jaffe, A.B. and Trajtenberg, M., 2002. *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

Klepper, S., 1996. Entry, exit, growth, and innovation over the product life cycle. *American economic review*, pp.562-583.

Laursen, K., 1998. Revealed comparative advantage and the alternatives as measures of international specialisation (No. 98-30). *DRUID, Copenhagen Business School, Department of Industrial Economics and Strategy - Aalborg University, Department of Business Studies*.

Levinson, A., 2009. Technology, international trade, and pollution from US manufacturing. *American Economic Review*, *99*(5), pp.2177-92.

Liu, J. and Goldstein, D., 2013. Understanding China's renewable energy technology exports. *Energy Policy*, *52*, pp.417-428.

Marin, G. and Vona, F., 2019. Climate policies and skill-biased employment dynamics: Evidence from EU countries. *Journal of Environmental Economics and Management*, *98*, p.102253.

Markard, J. and Truffer, B., 2006. The promotional impacts of green power products on renewable energy sources: direct and indirect eco-effects. *Energy Policy*, *34*(3), pp.306-321.

Matsumoto, A., Merlone, U. and Szidarovszky, F., 2012. Some notes on applying the Herfindahl–Hirschman Index. *Applied Economics Letters*, *19*(2), pp.181-184.

Mcmillan, M., Rodrik, D. and Verduzco-Gallo, I. (2014) 'Globalization, Structural Change and Productivty Growth, with an Update on Africa', *World Development*, 63, pp. 11–32.

Mealy, P. and Teytelboym, A., 2017. 'Economic complexity and the green economy' INET Oxford Working Paper No. 2018-03, Oxford University.

de Melo, J. and Solleder, J., 2018. The EGA Negotiations: why they are important, why they are stalled and challenges ahead. *Fondation Pour Les Études Et Recherches sur le Développement International Working Paper* No. 236.

Moretti, E., 2010. Local multipliers. *American Economic Review*, *100*(2), pp.373-77.

Nesta, L., Vona, F. and Nicolli, F., 2014. Environmental policies, competition and innovation in renewable energy. *Journal of Environmental Economics and Management*, *67*(3), pp.396-411.

Nordhaus, W.D. and Boyer, J., 2000. *Warming the world: economic models of global warming*. MIT press.

OECD, 2017. *Green Growth Indicators 2017,* OECD Publishing, Paris

Olsson, O. and Frey, B.S., 2002. Entrepreneurship as recombinant growth. Small Business Economics, 19(2), pp.69-80.

Perruchas, F., Consoli, D. and Barbieri, N., 2020. Specialisation, diversification and the ladder of green technology development. *Research Policy*, *49*(3), p.103922.

Petralia, S., Balland, P.A. and Morrison, A., 2017. Climbing the ladder of technological development. *Research Policy*, *46*(5), pp.956-969.

Popp, D., 2002. Induced innovation and energy prices. *American economic review*, *92*(1), pp.160-180.

Quitzow, R., 2015. Dynamics of a policy-driven market: The co-evolution of technological innovation systems for solar photovoltaics in China and Germany. *Environmental Innovation and Societal Transitions*, *17*, pp.126-148.

Rodrigues, J.F., Moran, D., Wood, R. and Behrens, P., 2018. Uncertainty of consumption-based carbon accounts. *Environmental science & technology*, *52*(13), pp.7577-7586.

Sawhney, A. and Kahn, M.E., 2012. Understanding cross-national trends in high-tech renewable power equipment exports to the United States. *Energy Policy*, *46*, pp.308-318.

Sato, M., 2014. Product level embodied carbon flows in bilateral trade. *Ecological* *economics*, *105*, pp.106-117.

Sauvage, J., 2014, The stringency of environmental regulations and trade in environmental goods, OECD Trade and Environment Working Papers, OECD Publishing, Paris.

Steenblik, R., 2005. *Environmental goods: A comparison of the APEC and OECD lists* (No. 2005/4). OECD Publishing, Paris.

Szirmai, A. and Verspagen, B., 2015. Manufacturing and economic growth in developing countries, 1950–2005. *Structural Change and Economic Dynamics*, *34*, pp.46-59.

Tamini, L.D. and Sorgho, Z., 2018. Trade in environmental goods: evidences from an analysis using elasticities of trade costs. *Environmental and Resource Economics*, *70*(1), pp.53-75.

Tran, T. M., 2019. Green Havens and Trade in Environmental Goods. EAERE Conference Paper.

Van Beveren, I., Bernard, A.B. and Vandenbussche, H., 2012. *Concording EU trade and production data over time* (No. 18604). National Bureau of Economic Research.

Vernon, R., 1966. International Investment and International Trade in the Product Life Cycle, The *Quarterly Journal of Economics*, 80(2) , pp. 190–207.

Veugelers, R., 2012. Which policy instruments to induce clean innovating?. *Research policy*, *41*(10), pp.1770-1778.

Vona, F., Marin, G. and Consoli, D., 2019. Measures, drivers and effects of green employment: evidence from US local labor markets, 2006–2014. *Journal of Economic Geography*, *19*(5), pp.1021-1048.

Vona, F., Marin, G., Consoli, D. and Popp, D., 2018. Environmental regulation and green skills: an empirical exploration. *Journal of the Association of Environmental and Resource Economists*, *5*(4), pp.713-753.

de Vries, G.J. and Ferrarini, B., 2017. What accounts for the growth of carbon dioxide emissions in advanced and emerging economies? The role of consumption, technology and global supply chain participation. *Ecological economics*, *132,* pp.213-223.

Weitzman, M.L., 1998. Recombinant growth. *The Quarterly Journal of Economics*, *113*(2), pp.331-360.

Wiebe, K.S. and Yamano, N., 2016. Estimating CO2 emissions embodied in final demand and trade using the OECD ICIO 2015, OECD Publishing, Paris.

Wiedmann, T., Wilting, H.C., Lenzen, M., Lutter, S. and Palm, V., 2011. Quo Vadis MRIO? Methodological, data and institutional requirements for multi-region input–output analysis. *Ecological Economics*, *70*(11), pp.1937-1945.

World Bank, 2007. *International Trade and Climate Change*, The World Bank, Washington

World Bank, 2012. *Inclusive Green Growth The Pathway to Sustainable Development,* The World Bank, Washington

WTO, 2001. *Doha Ministerial Declaration*, *Ministerial Declaration Adopted on 14 November 2001*.

Yu, R., Cai, J. and Leung, P., 2009. The normalized revealed comparative advantage index. *The Annals of Regional Science*, *43*(1), pp.267-282.

Zeppini, P. and van den Bergh, J.C., 2011. Competing recombinant technologies for environmental innovation: extending Arthur's model of lock-in. *Industry and Innovation*, *18*(03), pp.317-334.

Zugravu-Soilita, N., 2018. The impact of trade in environmental goods on pollution: what are we learning from the transition economies' experience? *Environmental Economics and Policy Studies*, *20*(4), pp.785-827.

Zugravu-Soilita, N., 2019. Trade in Environmental Goods and Air Pollution: A Mediation Analysis to Estimate Total, Direct and Indirect Effects. *Environmental and Resource Economics*, *74*(3), pp.1125-1162.

Tables and Figures

Table 1: Correlation table among green product lists.

	CLEG	WTO 2009	PEGS	PRODCOM	APEC	Germany	Core (WTO + CLEG)	WTO Core	CLEG Core
CLEG	1								
WTO 2009	0.84**	1							
PEGS	0.73**	0.47**	1						
PRODCOM	**0.5****	**0.31****	**0.57****	**1**					
APEC	0.46**	0.49**	0.41**	**0.32****	1				
Germany	0.16**	0.15**	0.14**	**0.12****	0.17**	1			
Core (WTO + CLEG)	0.37**	0.37**	0.35**	**0.46****	0.44**	0.13**	1		
WTO Core	0.29**	0.25**	0.27**	**0.31****	0.16**	0.04**	0.77**	1	
CLEG Core	0.23**	0.28**	0.24**	**0.35****	0.51**	0.16**	0.65**	0.03**	1
Number of goods	820	605	471	**225**	206	147	123	78	47

Notes: authors' own calculation on PRODCOM data. The table reports correlation coefficients of dummy variables indicating the presence of a certain product in a given list across different lists. The last row reports the number of PRODCOM product codes within each green product list. For further details about the lists of green goods, see Appendix A. *p<0.05 ** p<0.01.

Table 2: Green and polluting production by 2-digit industries.

NACE	Label	Mean green share 2005	Mean green share 2010	Mean green share 2015	Share of total green production	Absolute Change 2005-2015	Average GHG intensity
28	Machinery and equipment n.e.c.	0.088 (0.086)	0.101 (0.102)	0.115 (0.115)	0.31	0.027	0.54
26	Computer, electronic and optical products	0.069 (0.06)	0.121 (0.131)	0.103 (0.076)	0.21	0.034	0.3
27	Electrical equipment	0.108 (0.166)	0.103 (0.078)	0.162 (0.217)	0.21	0.054	0.3
30	Other transport equipment	0.281 (0.292)	0.346 (0.318)	0.38 (0.334)	0.13	0.098	0.61
33	Repair, installation of machinery	0.022 (0.031)	0.033 (0.024)	0.028 (0.026)	0.04	0.006	0.74
29	Motor vehicles, trailers and semi-trailers	0.002 (0.01)	0.007 (0.031)	0.003 (0.011)	0.01	0.001	0.61
31	Furniture	0 (0)	0 (0)	0 (0)	0	0	0.74
32	Other manufacturing	0 (0)	0 (0)	0 (0)	0	0	0.74
16	Products of wood, cork, straw, plaiting	0 (0)	0 (0)	0 (0)	0	0	0.88
22	Rubber and plastic products	0 (0)	0 (0)	0 (0)	0	0	0.94
13	Textiles	0 (0)	0 (0)	0 (0)	0	0	0.97
14	Wearing apparel	0 (0)	0 (0)	0 (0)	0	0	0.97
15	Leather and related products	0 (0)	0 (0)	0 (0)	0	0	0.97
17	Paper and paper products	0 (0)	0 (0)	0 (0)	0	0	1.18
18	Printing and reproduction of recorded media	0 (0)	0 (0)	0 (0)	0	0	1.18
10	Food products	0 (0)	0 (0)	0 (0)	0	0	1.45
11	Beverages	0 (0)	0 (0)	0 (0)	0	0	1.45
12	Tobacco products	0 (0)	0 (0)	0 (0)	0	0	1.45
	Polluting industries						
19	Coke and refined petroleum products	.	0 (0)	0 (0)	0	.	44.99
23	Other non-metallic mineral products	0.029 (0.029)	0.033 (0.022)	0.033 (0.026)	0.05	0.003	7.78
20	Chemicals and chemical products	0 (0)	0 (0)	0 (0)	0	0	5.11
21	Basic pharma. products, preparations	0 (0)	0 (0)	0 (0)	0	0	5.11
25	Fabricated metal products, exc. machinery	0.018 (0.018)	0.019 (0.016)	0.017 (0.014)	0.04	-0.001	4.23
24	Basic metals	0.006 (0.021)	0.007 (0.023)	0.008 (0.03)	0.01	0.002	4.23

Notes: Authors' elaboration on PRODCOM data. The definition of green products used here is explained in section 2 and it is the one called PRODCOM in Figure 1. Columns 3 to 5 report the mean green share of production with the standard deviation in brackets of each industry for the years 2005,2010 and 2015, respectively. Coke and refined petroleum products is not included in PRODCOM until 2005, as PRODCOM coverage is not stable over time and doesn't include fuel related products. Column 6 reports the share that green production of each industry represents in total green production. Absolute changes 2005-2015 refer to industries' average green shares of production. Polluting industries are identified as the 5 industries with the highest average GHG intensity computed with WIOD, for further detail see Appendix B.

Table 3: Distribution of green production shares across green industries at 4 digits NACE

NACE	Label	Mean	Median	Max	Standard deviation	Change 1995-2015	Change 2001-2015	Share of green production
	High green potential industries							
3092	Manufacture of bicycles and invalid carriages	0.79	0.82	1	0.24	0.03	0.02	2.98
3020	Manufacture of railway locomotives and rolling stock	0.7	0.8	1	0.28	-0.02	-0.1	9.53
2530	Manufacture of steam generators, except central heating hot water boilers	0.55	0.54	1	0.35	0.11	0.2	1.82
2312	Shaping and processing of flat glass	0.39	0.34	1	0.3	0.02	0.05	4.61
2712	Manufacture of electricity distribution and control apparatus	0.39	0.34	1	0.23	0.01	0.04	16.09
2651	Manufacture of instruments and appliances for measuring, testing and navigation	0.37	0.37	1	0.19	0.01	-0.01	17.46
2811	Manufacture of engines and turbines, except aircraft, vehicle and cycle engines	0.29	0.19	1	0.31	0.19	0.1	12.8
2829	Manufacture of other general-purpose machinery n.e.c.	0.29	0.24	1	0.22	0.04	0.03	7.4
2825	Manufacture of non-domestic cooling and ventilation equipment	0.28	0.28	1	0.18	0.01	0.01	10.67
2611	Manufacture of electronic components	0.14	0.01	1	0.27	0.09	0.09	3.67
2740	Manufacture of electric lighting equipment	0.13	0.12	0.63	0.1	0.05	0.04	3.32
2752	Manufacture of non-electric domestic appliances	0.11	0.03	0.5	0.14	-0.12	-0.1	1.01
3320	Installation of industrial machinery and equipment	0.08	0.06	0.67	0.08	0.07	0.08	3.8
	Marginally green industries							
2410	Manufacture of basic iron and steel and of ferro-alloys	0.04	0	1	0.18	0.02	0.02	0.62
2751	Manufacture of electric domestic appliances	0.04	0	0.91	0.11	0.01	0.01	0.46
2511	Manufacture of metal structures and parts of structures	0.03	0.03	0.19	0.03	0	0	1.99
2599	Manufacture of other fabricated metal products n.e.c.	0.02	0.01	0.29	0.04	0.01	0.01	0.59
2351	Manufacture of cement	0.01	0	0.34	0.05	0.04	0.02	0.23
2910	Manufacture of motor vehicles	0.01	0	0.51	0.04	0.01	0.01	0.76
2711	Manufacture of electric motors, generators and transformers	0.00047	0	0.04	0	0	0	0.03
2899	Manufacture of other special-purpose machinery n.e.c.	0.00239	0	0.1	0.01	0.01	0.01	0.19

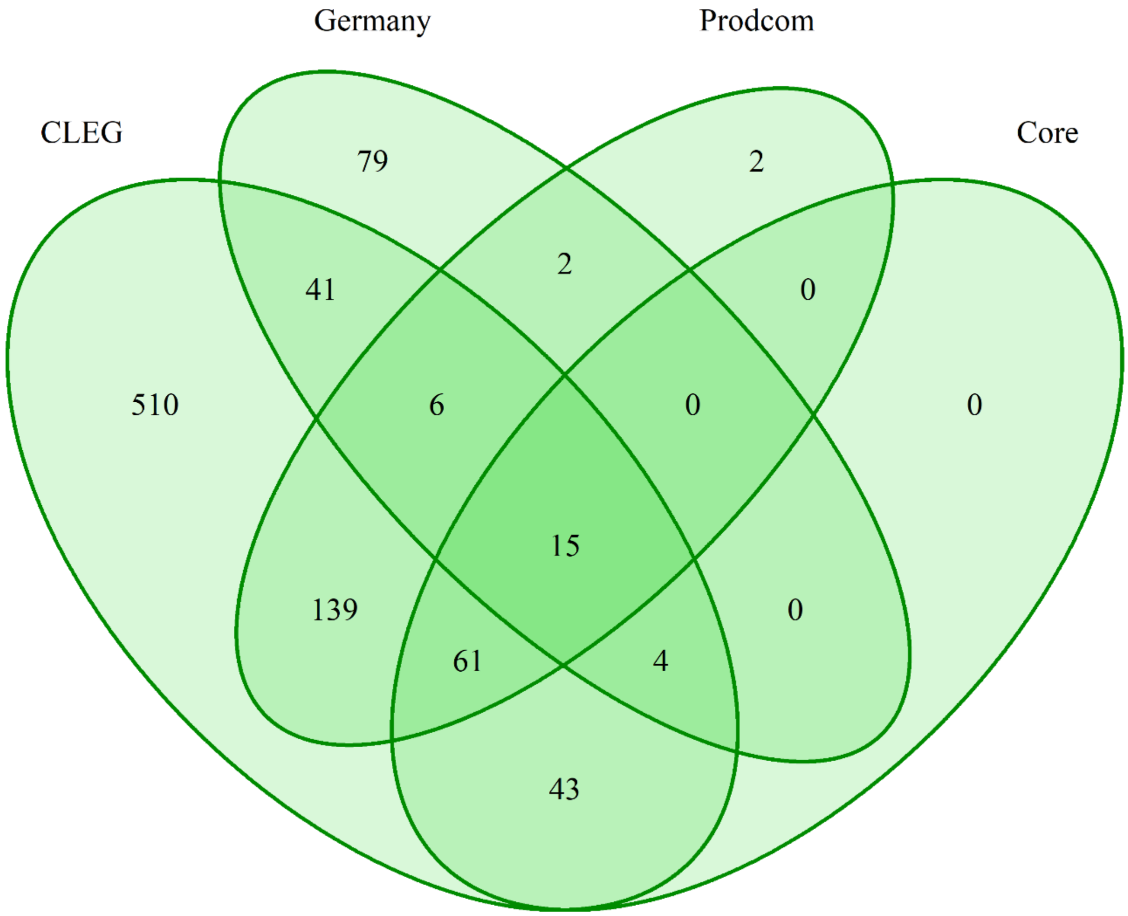
Notes: Authors' elaboration on PRODCOM data. The definition of green products used here is explained in section 2 and it is the one called PRODCOM in Figure 1. Average, median, maximum and standard deviation are computed over all available countries and years (1995-2015), columns 7 and 8 report changes in the average green share for 1995-2015 and 2001-2015 respectively. The last column reports for each industry the share it represents in total green production across all industries, countries and years.

Table 4: Path-dependency of green RCA and complementarity with non-green RCA.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA (log) * 2005	0.974***	0.923***	0.622***	0.541***	0.526***
	(0.0291)	(0.0335)	(0.0432)	(0.0413)	(0.0418)
Pre-sample green RCA (log) * 2006	0.916***	0.864***	0.591***	0.522***	0.509***
	(0.0415)	(0.0450)	(0.0507)	(0.0494)	(0.0495)
Pre-sample green RCA (log) * 2007	0.865***	0.811***	0.568***	0.502***	0.492***
	(0.0496)	(0.0528)	(0.0566)	(0.0570)	(0.0565)
Pre-sample green RCA (log) * 2008	0.830***	0.773***	0.546***	0.485***	0.475***
	(0.0545)	(0.0573)	(0.0601)	(0.0592)	(0.0590)
Pre-sample green RCA (log) * 2009	0.810***	0.754***	0.499***	0.443***	0.428***
	(0.0578)	(0.0606)	(0.0625)	(0.0583)	(0.0584)
Pre-sample green RCA (log) * 2010	0.776***	0.723***	0.479***	0.416***	0.405***
	(0.0558)	(0.0576)	(0.0605)	(0.0580)	(0.0579)
Pre-sample green RCA (log) * 2011	0.738***	0.687***	0.444***	0.388***	0.375***
	(0.0587)	(0.0599)	(0.0622)	(0.0578)	(0.0580)
Pre-sample green RCA (log) * 2012	0.707***	0.656***	0.457***	0.425***	0.412***
	(0.0611)	(0.0620)	(0.0580)	(0.0505)	(0.0510)
Pre-sample green RCA (log) * 2013	0.679***	0.627***	0.420***	0.376***	0.365***
	(0.0642)	(0.0650)	(0.0628)	(0.0555)	(0.0559)
Pre-sample green RCA (log) * 2014	0.649***	0.599***	0.405***	0.365***	0.355***
	(0.0670)	(0.0674)	(0.0635)	(0.0572)	(0.0576)
Pre-sample green RCA (log) * 2015	0.625***	0.573***	0.367***	0.332***	0.319***
	(0.0725)	(0.0725)	(0.0677)	(0.0609)	(0.0614)
Non-green RCA_{t-1} (log)		0.155***	0.0947**	0.107***	0.0743*
		(0.0435)	(0.0449)	(0.0394)	(0.0426)
Number of green products with RCA_{t-1} (log)			0.260***		0.0794***
			(0.0237)		(0.0221)
Number of non-green products with RCA_{t-1} (log)			-0.00533		0.0130
			(0.0154)		(0.0153)
Dummy for at least one green product with RCA_{t-1}				0.300***	0.237***
				(0.0211)	(0.0254)
Dummy for at least one non green product with RCA_{t-1}				0.0104	0.00978
				(0.0186)	(0.0208)
Constant	0.117***	0.0678***	0.0805***	0.0634***	0.0646***
	(0.0236)	(0.0260)	(0.0216)	(0.0200)	(0.0198)
Observations	3,289	3,289	3,289	3,289	3,289
R-squared	0.676	0.687	0.767	0.789	0.793

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only for 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. RCA are symmetrical around 0 and the logarithm is taken of RCA+2. Country-year and sector-year fixed effects are included in all estimates. Standard errors are clustered at the country level and reported in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Overlap of PRODCOM product codes among selected lists of green goods.

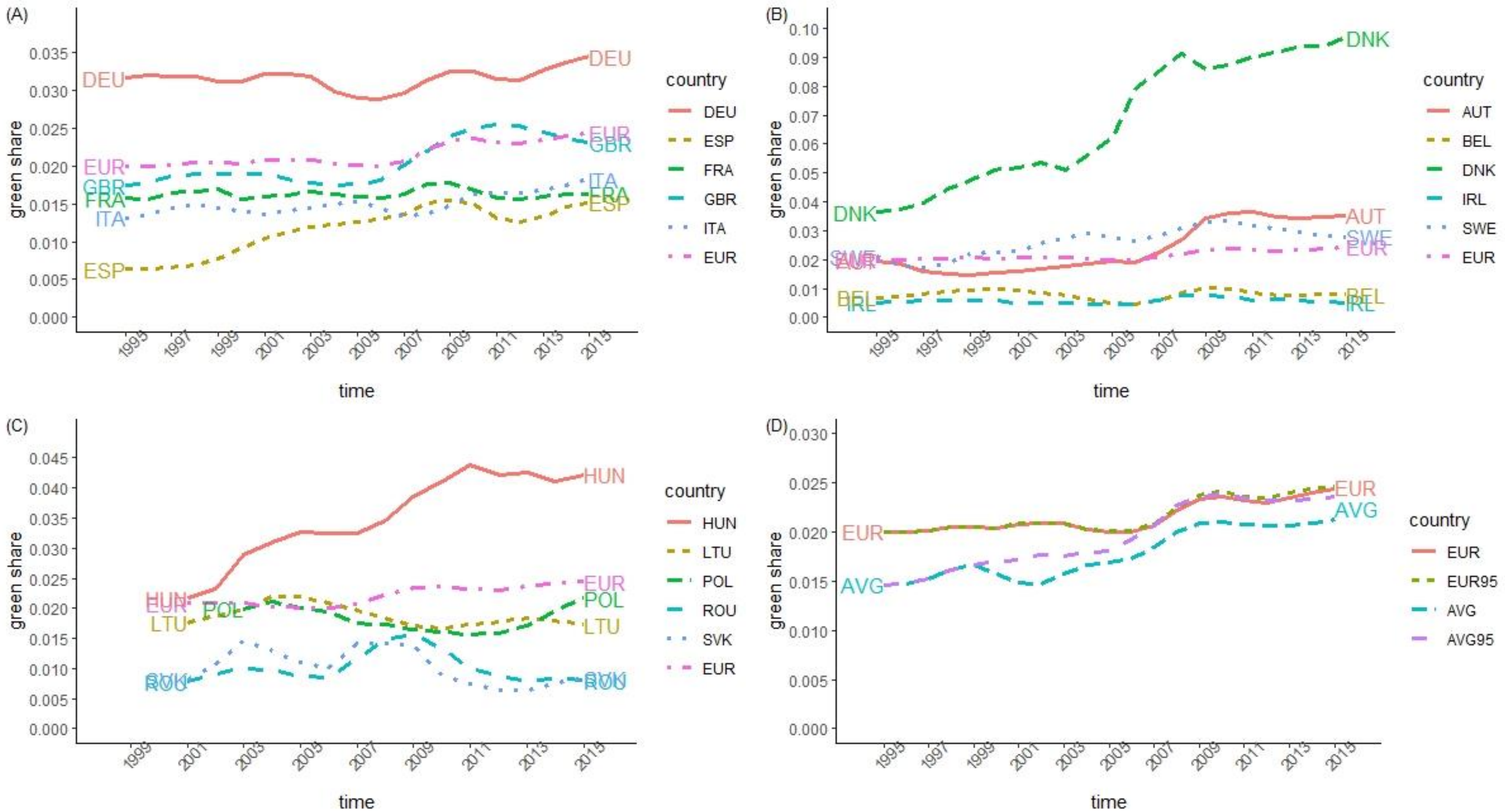


Notes: Authors' elaboration on PRODCOM data. The figure depicts the overlap among four existing lists of green goods, the numbers represent the number of PRODCOM product code that fall within each category. For further details about the lists of green goods see Appendix A.

Figure 2: Average ratio of green and non-green HHI in high-green potential industries, weighted on industries' turnover.

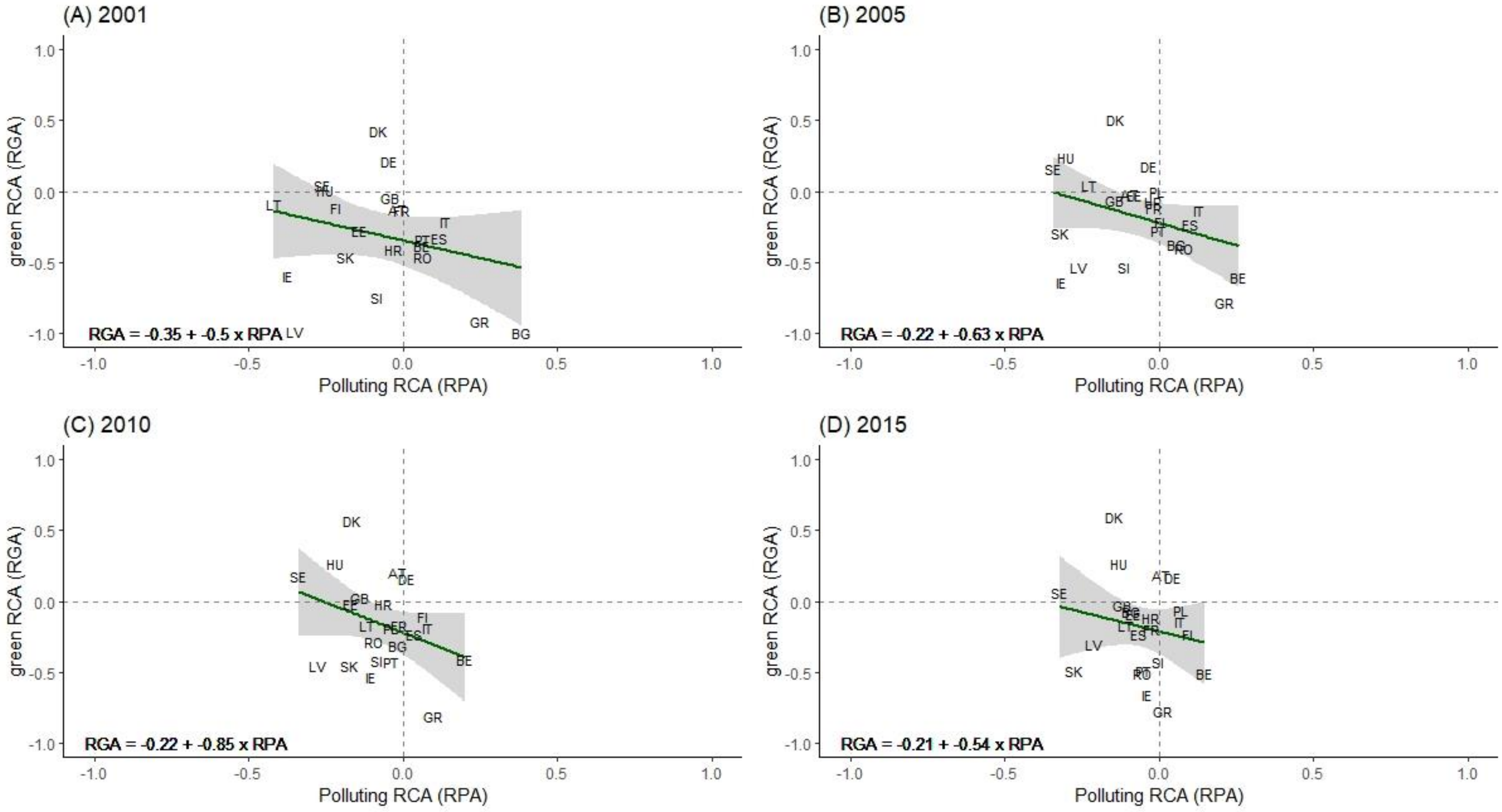
Notes: Authors' elaboration on PRODCOM data. The Figure reports the average of HHIR in high-green potential industries using industries' turnover as weights. We compute HHIR across all available countries in each year in our sample, as well as only for countries for which we have a balanced panel from 1995: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, United Kingdom. The horizontal dashed line indicates the unity.

Figure 3: Evolution of green production shares for selected European countries



Notes: Panel A, B and C report green production shares over time for large, small and Eastern European countries, respectively. These have been smoothened by taking 3-years moving averages. We only use green production from high-green potential industries as identified in Table 3. EUR is the European average of green shares across all available countries in each year, weighted on countries' output. In panel D we compare it with the unweighted average (AVG) which is not affected by countries' size, especially Germany. Because data on Eastern countries is available only from 2001 onwards, and 2003 onwards for Poland, we report both these measures computed for each year for all available countries as well as only for countries for which we have a balanced panel since 1995, i.e.: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, United Kingdom, (EUR95 and AVG95).

Figure 4: Green and polluting RCA across countries and over time.



Notes: Authors' elaboration on PRODCOM data. We plot countries' green and polluting RCA. Green RCA are based solely on green production from high-green potential industries, as identified in Table 3. Polluting production is total production from polluting industries identified in Table 2. The RCAs are computed following formula 3 are made symmetrical around 0 and bounded between -1 and 1, the value of 0 indicates therefore whether a country has successfully specialized in green production. We also report the coefficient of a regression of green RCA on polluting RCA for each year.

Appendix (for online publication)

A. Lists of green products.

In this Appendix, we provide additional information on the lists used to identify our favourite PRODCOM list and for the validation analysis of Section 2.4. As we detail in section 2 our universe of potential lists is the union of the CLEG list and German list. CLEG is the result of the union of three broader lists of the Asia and Pacific Economic Cooperation (APEC) forum, WTO Friends' list and Plurilateral Environmental Goods and Services (PEGS).

In 2012, the APEC members have committed to reduce tariffs on green goods to 5% at the most, in the Vladivostok declaration (APEC, 2012).[19] The APEC list is one of the most commonly used list in investigating the role of trade in green products on pollution (Zugravu-Soilita, 2018; Mealy and Teytelboym, 2019). Negotiations within the WTO have led to the creation of several lists, of which the WTO Friends' list from 2009 and its more narrow subset WTO core have also received considerable attention (Sauvage, 2014; Mealy and Teytelboym, 2019). Finally, the PEGS list has been developed by the OECD in preparation for the Toronto G20 summit in 2010 and among the three lists included in CLEG is the only one that is not the outcome of international trade negotiations, which as we have discussed in section 2 can impact what products are included in the final list.

As we have discussed in the main text, a key challenge with these product lists is that the HS classification is not designed to isolate green products and therefore there is the risk that green and non-green products may be lumped together under the same product code. In other words, it is possible that a given product code may cover both green and non-green products. In order to deal with this the OECD has relied on experts' advice and has examined all products codes included in the CLEG list to identify those that are less likely to be affected by this issue. OECD experts have provided an estimate of the proportion of trade flows taking place under each product code that corresponds to trade of green goods. They have used two thresholds, 2/3 and 1/3 to put forward two narrow lists: CLEG Core and CLEG Core Plus, respectively.

[19] APEC members are: Australia, Brunei, Canada, Indonesia, Japan, South Korea, Malaysia, New Zealand, Philippines, Singapore, Thailand, the United States, Taiwan, Hong Kong, China, Mexico, Papua New Guinea, Chile, Peru, Russia and Vietnam.

To give an example of how these two lists treat products differently, we can think of vacuum pumps that include both pumps that can be used for environmentally friendly functions, such as water management, as well as in other production processes that have no positive impact on the environment. In this specific case the OECD experts have estimated that more than 33% but less than 66% of all traded vacuum pumps are actually used to fulfil environmental activities. Therefore the OECD has included this product in the CLEG Core plus list but not in the CLEG Core. In light of this ambiguity, vacuum pumps are not included in our own list, as they do not respect the criterium of no multiple usage.

To recap, we have a set of broad lists (CLEG, WTO2009, APEC, PEGS and German list) and a set of narrow lists (CLEG Core, CLEG Core Plus and WTO Core). This multitude of lists reflects the lack of agreement on a definition of green products. We present all the lists we have discussed here in Table A.1.

Table A.1 – Green lists

List	Year	N. of Products	Description	Negotiated	Organization
CLEG	2014	820	The list has been compiled by Sauvage (2014) merging WTO Friends, PEGS and APEC.	No	OEDC
WTO Friends	2009	605	This list has been negotiated by a smaller group of high-income economies within the WTO	Yes	WTO
PEGS	2010	471	The list has been compiled by OECD with a focus on renewable energies	No	OECD
APEC	2012	206	Countries member of APEC have negotiated this list agreeing to reduce tariffs on the products included down to at least 5%	Yes	APEC
WTO Core	2011	78	This is more restrictive list that has been negotiated within WTO during negotiations towards a comprehensive free trade agreement on environmental goods.	Yes	WTO
CLEG Core	2014	163	This is a more restrictive version of CLEG compiled by OECD experts with the aim of dealing with the issue of multiple usage. It only includes product codes for which at least 1/3 of the associated trade flows consists of green products.	No	OECD
CLEG Core	2014	47	This is an even more restrictive version of CLEG compiled by OECD experts with the aim of dealing with the issue of multiple usage. It only includes product codes for which at least 60% of the associated trade flows consists of green products.	No	OECD
German list	2009	147	The list has been compiled by Germany's statistical office in accordance with Eurostat's criteria of environmental protection and resource management.	No	German National Statistical Office

Notes: Authors' elaboration on PRODCOM data. For each list we report its name, the year in which it was compiled, the number of PRODCOM codes it contains, a brief description of the list, whether it is the outcome of trade negotiations and which organization has compiled it. All lists in the table are based on the HS product classification, except for Germany's list that is compiled with PRODCOM product codes. To obtain the number of products for each list we have relied on crosswalks between HS and Eurostat's Combined Nomenclature (CN) and between PRODCOM and CN, provided by Eurostat.

B. More details on polluting industries

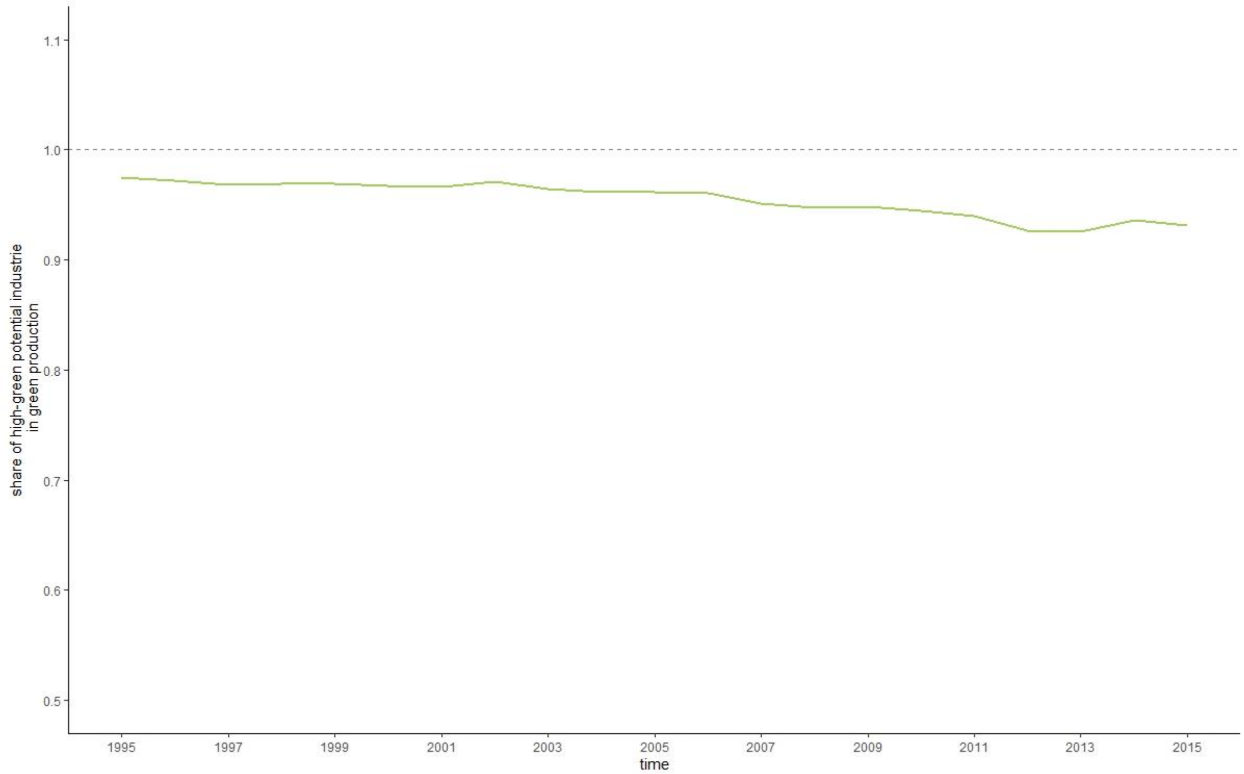
We compute our measure of polluting industry using the 2013 WIOD release, which includes information on countries' and industries' greenhouse gas (GHG) emissions as well as energy intensity. We follow Marin and Vona (2019) and compute GHG (C02, N20 and CH4, aggregated according to their global warming potential) intensity as the sum of direct and indirect GHG emissions per unit of value added of each industry, country and year. Direct emissions are those associated to the production of each sector, indirect emissions are those embodied in the purchases of electricity from the power sector of each industry (which we compute using input-output technical coefficients).

The WIOD classifies industries using the ISIC rev 3.1, for which an official crosswalk only exists with NACE rev. 1, given the high level of aggregation (less than two digits NACE rev.2), it is also straightforward to match WIOD data with NACE rev. 2 industries, which is based on ISIC rev. 4. Because of the high level of aggregation of WIOD we consider that the entire production of brown industries is polluting. However, we exploit our fine-grained data at the 4-digit level data to slightly refine this coarse classification of brown industries by excluding the processing of nuclear fuel from basic metal manufacturing and the production of pharmaceutical products and preparation from the chemical sector. The pharmaceutical and chemical sector have the same pollution intensity in the WIOD data, because the two sectors are lumped together in the ISIC rev. 3 industry classification. However, chemical industries are well-known to be significantly more polluting than pharmaceutical ones. The processing of nuclear fuel is contained within basic metals manufacturing at 2-digits of the NACE rev. 2 classification, so we identify the corresponding 4-digit code (2446) in PRODCOM and we exclude it from our computation of polluting production.

C. More results on green production

Figure C.1 plots the evolution over time of the share of green production from high-green potential industries in total green production. Despite a mildly decreasing trend, high-green potential industries account for the majority of green production in our observed time period. On average, 96% of green production is concentrated in 13 out of 235 4-digit NACE industries.

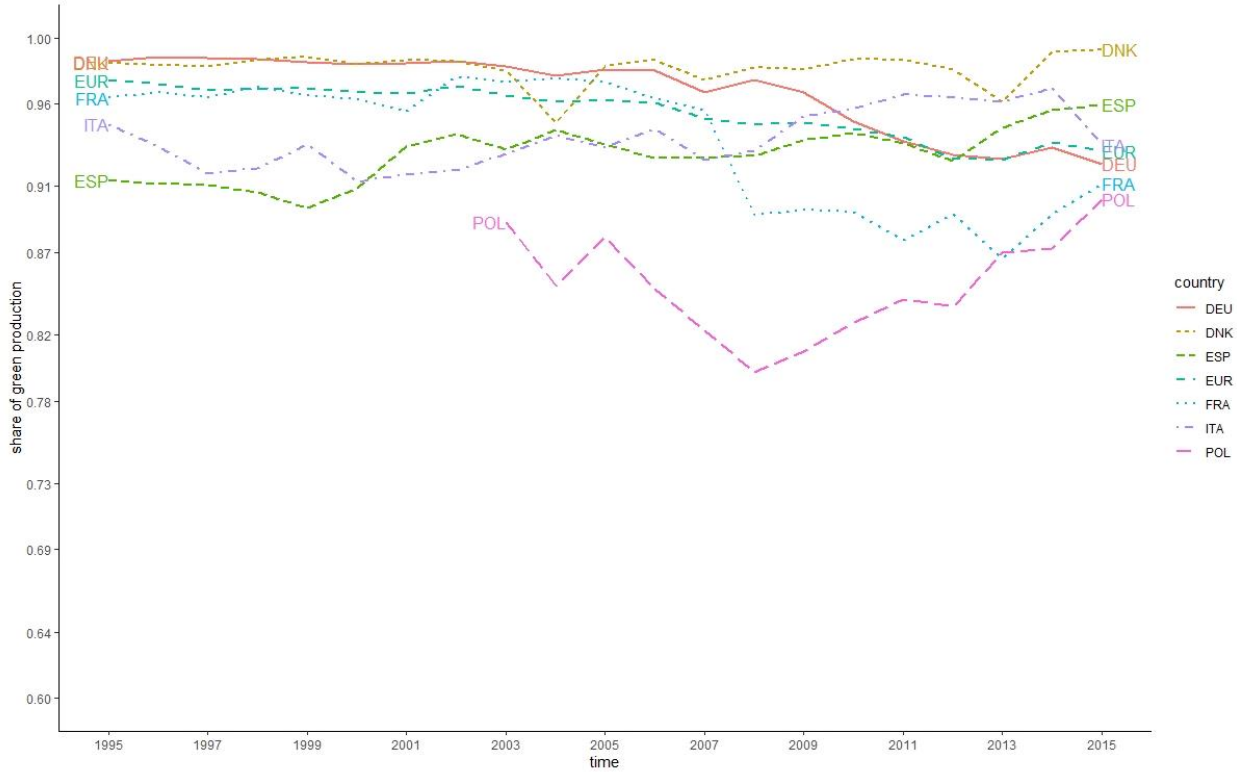
Figure C.1: High-green potential industries' share of total green production over time.



Notes: Authors' elaboration on PRODCOM data. We report the evolution over time of the share of green production from high-green potential industries, as identified in Table 3 as a share of total green production based on the list of green products PRODCOM discussed in section 2.

In Figure C.2 we plot this same measure for selected countries, finding some heterogeneity. We find in particular that high-green potential industries in Denmark and Poland represent an increasing share of green production, while there is a decreasing trend for France. There are also countries like Germany and Italy that are closer to the European share we observe in Figure C.1. Overall green production from high-green potential industries never represents less than 78% of the country's total green production in any of the countries considered here, confirming that high-green potential industries account for the bulk of green production.

Figure C.2: Share of green production from high-green potential industries in total green production for selected countries and Europe.



Notes: Authors' elaboration on PRODCOM data. We report for selected countries the evolution over time of the share that green production from high-green potential industries, as identified in Table 3 represents of total green production based on the list of green products PRODCOM discussed in section 2.

To illustrate which are the most important green products for each country, Table C.1 presents the top three green products and their share in total green production for each country. Remarkably, we find that top products are rather similar across countries. They mostly concern integrated technologies for renewable energy, appliances to increase energy efficiency, as well as insulating material.

Table C.1: Top three green products across countries and shares of green production.

Country	First product	Second product	Third product	Share of total green production
Austria	Programmable memory controllers for a voltage <= 1 kV	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Railway material (of steel)	0.00748
Belgium	Multiple-walled insulating units of glass.	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Bicycles and other cycles (including delivery tricycles), non-motorised	0.00438
Bulgaria	Non-motorized bicycles and other cycles with ball bearings (including delivery tricycles)	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Multiple-walled insulating units of glass	0.00086
Croatia	Parts for steam turbines and other vapor turbines	Photosensitive semiconductor devices; solar cells, photodiodes, phototransistors, etc	Other bases for electric control, distribution of electricity, voltage <= 1 kV	0.00066
Denmark	Programmable memory controllers for a voltage <= 1 kV	Generating sets, wind-powered	Parts of vapor generating boilers and super-heater water boilers	0.00585
Estonia	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Multiple-walled insulating units of glass	Parts and accessories for automatic regulating or controlling instruments and apparatus	7.00e-04
Finland	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Heat exchange units	Machinery and apparatus for solid-liquid separation/ purification excluding for water and beverages, centrifuges and centrifugal dryers, oil/petrol filters for internal combustion engines	0.00681
France	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Heat exchange units	Multiple-walled insulating units of glass	0.02843
Germany	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Programmable memory controllers for a voltage <=1 kV	Photosensitive semiconductor devices; solar cells, photodiodes, phototransistors, etc	0.10737
Greece	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Bicycles and other cycles (including delivery tricycles), non-motorised	Vapour generating boilers (including hybrid boilers) (excluding central heating hot water boilers capable of producing low pressure steam, watertube boilers)	0.00024
Hungary	Parts of gas turbines (excluding turbojets and turbo-propellers)	Photosensitive semiconductor devices; solar cells, photodiodes, phototransistors, etc.	Other bases for electric control, distribution of electricity, voltage <= 1 kV	0.01091
Ireland	Machinery and apparatus for filtering or purifying water	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Parts for filtering and purifying machinery and apparatus, for liquids or gases (excluding for centrifuges and centrifugal dryers)	0.00239

Italy	Heat exchange units	Gas turbines (excluding turbojets and turboprops)	Machinery and apparatus for filtering or purifying air (excluding intake filters for internal combustion engines)	0.03125
Latvia	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Multiple-walled insulating units of glass	Machinery and apparatus for filtering or purifying water	0.00023
Lithuania	Non-motorized bicycles and other cycles with ball bearings (including delivery tricycles)	Multiple-walled insulating units of glass	Other bases for electric control, distribution of electricity, voltage > 1 kV	0.00076
Poland	Multiple-walled insulating units of glass	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Parts for steam turbines and other vapor turbines	0.00662
Portugal	Non-motorized bicycles and other cycles with ball bearings (including delivery tricycles)	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Multiple-walled insulating units of glass	0.00265
Romania	Multiple-walled insulating units of glass	Railway or tramway goods vans and wagons, not self-propelled	Other bases for electric control, distribution of electricity, voltage > 1 kV	0.00133
Slovakia	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Heat exchange units	Multiple-walled insulating units of glass	0.00085
Slovenia	Machinery and apparatus for filtering or purifying air (excluding intake filters for internal combustion engines)	Multiple-walled insulating units of glass	Heat exchange units	0.00026
Spain	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Generating sets (excluding wind-powered and powered by spark-ignition internal combustion piston engine)	Photosensitive semiconductor devices; solar cells, photodiodes, phototransistors, etc.	0.00981
Sweden	Heat exchange units	Instruments and apparatus using optical radiations, n.e.c.	Other bases for electric control, distribution of electricity, voltage <= 1 kV	0.00935
United Kingdom	Parts of gas turbines (excluding turbojets and turbo-propellers)	Other bases for electric control, distribution of electricity, voltage <= 1 kV	Multiple-walled insulating units of glass	0.02336

Notes: Authors' elaboration on PRODCOM data. The table reports for each country the three green products with largest green production and the total share of green production that these three products combined represent in countries' total green production. Products are identified here with the synthetic, time-invariant product codes derived from VBBV methodology.

Figure 2 in the main text shows average relative concentration across all high-green potential industries, it is however interesting to explore whether there is significant heterogeneity across industries. We plot in Figure C.3 the HHI ratio for four selected industries, finding rather heterogeneous results.

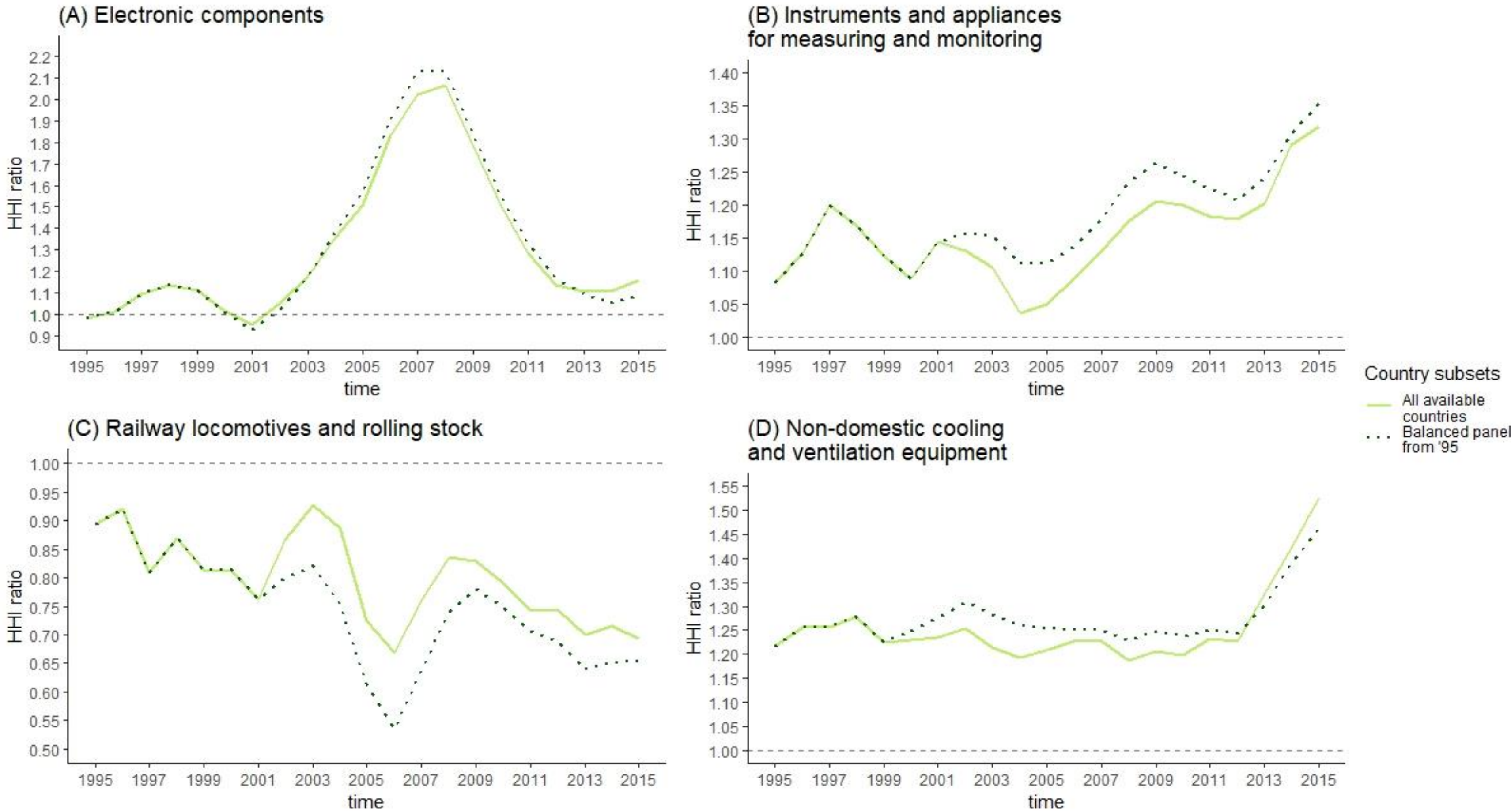
In panel A we see that green production in the electronic components manufacturing, which includes PV and LEDs, experiences a sustained increase in its relative concentration until 2008 followed by an equally steep decline. These stark changes in the relative concentration of green production reflect the rise and fall of Germany in the production of photovoltaic panels as well as the emergence of non-European players such as China (Algieri, Aquino and Succurro, 2011; Sawhney and Kahn, 2012; Liu and Goldstein, 2013). Because our data only covers European countries, the shift of production of photovoltaic panels from Germany to China and other non-EU countries results in a reduction in Germany's total production of photovoltaic panels and an apparent reduction of production concentration within Europe. In order to shed further light on this we report in Figure C.4 the evolution of the two HHIs for green and non-green production: the trend of the relative HHI we observe in panel A Figure C.3 is essentially driven by a steep increase and then decline in the concentration of green production, while the concentration of non-green production remains rather stable.

The instruments and appliances manufacturing for measuring and monitoring in panel B follows a rather different pattern. There are two peaks in 1997 and 2001 that are similar to the average HHIR in Figure 2, but after a decline in 2005 the concentration of green production increases significantly, relative to non-green production. Importantly, the index remains constantly above the unity.

In panel C, locomotives and rolling stock follows yet another pattern, the relative HHI is in fact on a clearly downwards path and always well below the unity suggesting that green production in this industry is less concentrated than its non-green counterpart. We should note that this is the second sector for average green share of production at 70%, with a median green share of 80% (see Table 3). Because non-green production within this industry is so small it is more likely to be concentrated in fewer countries than the green production, which in contrast represents the bulk of the sector's production and is thus likely to be distributed across more countries. These differences in absolute values of production may therefore explain the fact that non-green production is more concentrated than green production within this specific industry.

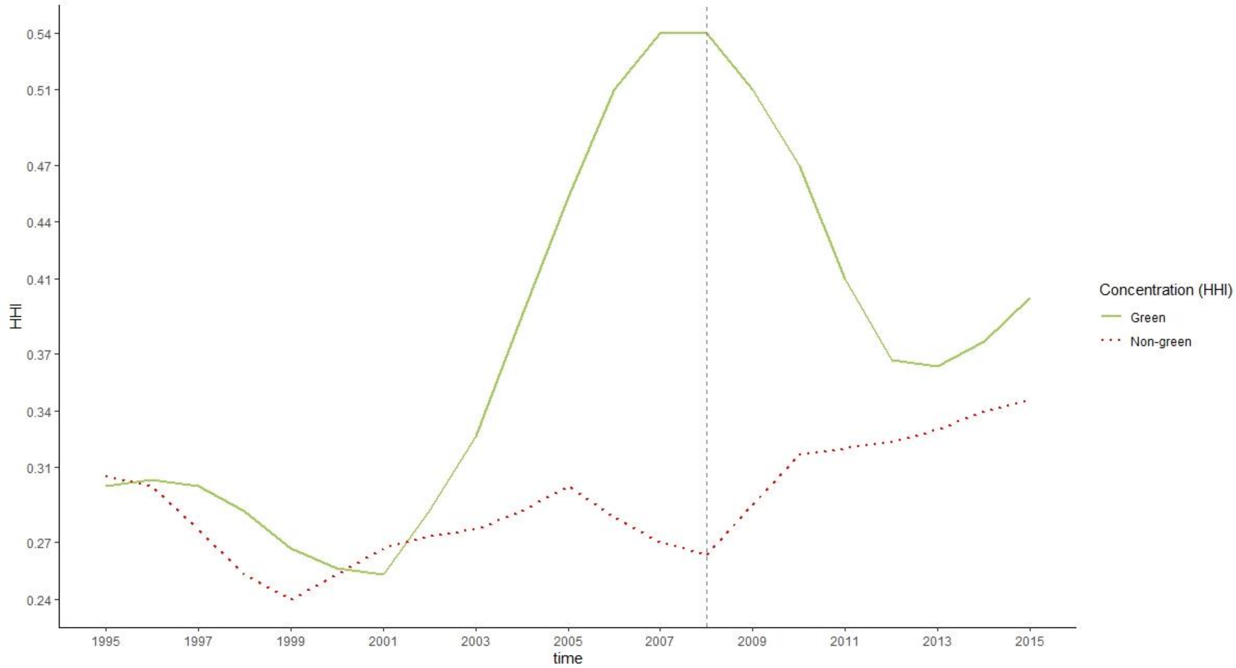
Finally, the relative HHI in Panel D for non-domestic cooling and ventilation equipment, follows a similar pattern to that of the instruments and appliances manufacturing for measuring and monitoring. The relative HHI is always above the unity and experiences an increase after 2012, suggesting that the green production within this sector, which includes both heat pumps and filtering equipment, is conglomerating in few leading countries such as Germany, Italy, the United Kingdom and Italy.

Figure C.3: Ratio of green and non-green HHI over time in selected high-green potential industries.



Notes: authors' elaboration on PRODCOM data. Panel A reports the average HHIR for electronic components manufacturing, Panel B reports the same measure for instruments and appliances for measuring and monitoring, Panel C for railway locomotives and Panel D for non-domestic cooling and ventilation equipment. We compute HHI for all countries in our sample as well as only for countries for which we have a balanced panel from 1995, i.e.: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, United Kingdom.

Figure C.4: Green and non-green HHI over time for electronic components manufacturing.



Notes: authors' elaboration on PRODCOM data. HHI indexes have been computed using all countries available in all years for the electronic components manufacturing. The vertical dotted line indicates 2008.

D. Robustness checks using all green industries.

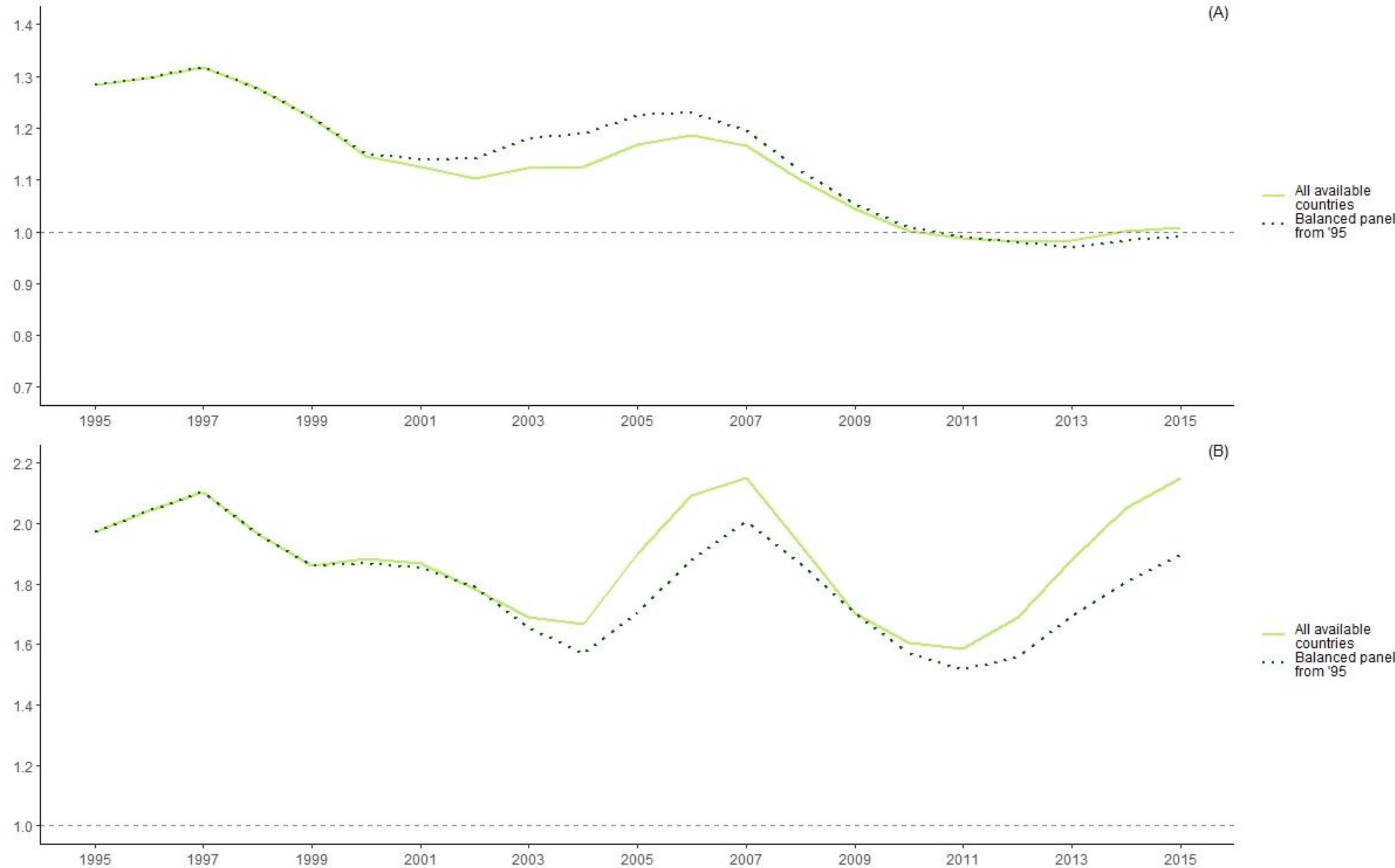
In this section of the Appendix we report robustness checks on our analysis in section 4, using all of green industries identified in Table 2 in the main text. Overall, we find very similar results, so we only comment on the differences we find providing insights on their possible origin.

In Figure D.1 we compare the average HHIR from Figure 2 in the main text, which is the weighted average across high-green potential industries of the concentration of green production, relative to non-green production, with the same measure computed using all green industries from Table 2. We find a very similar pattern over time, with a significant difference in levels, however. The average HHIR is much higher when we include all green industries, being consistently above the unity. This is due to the fact that, by taking the weighted average of all green industries, we are also considering industries that have very little green production, which is therefore more likely to be highly concentrated and drives the average HHIR upwards.

This difference is a good example of how the inclusion of green industries with very low shares of green production can bias our results, when looking at the concentration of production because small production values are, almost by construction, more concentrated and can erroneously lead to think that green production has on average much higher concentration relative to non-green production than it is the case for high-green production industries.

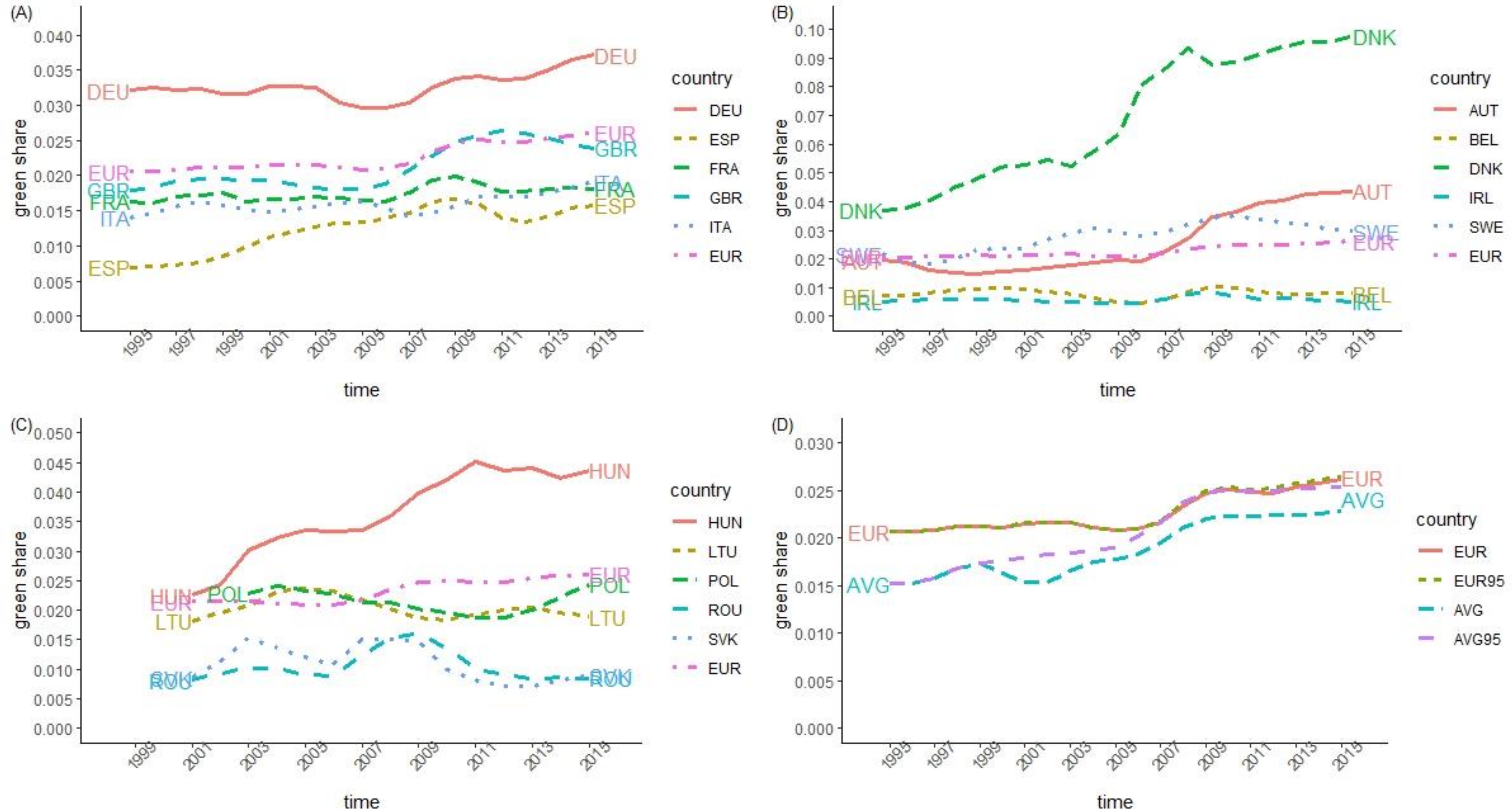
We now turn to how the inclusion of all green industries, rather than only high-green potential, changes our results concerning green specialisation across countries and over time. In Figure D.2 and D.3 we replicate Figure 3 and 4 from the main text, respectively, finding strikingly similar results. This doesn’t come as a surprise as high-green potential account for 96% of total green production and because neither of these figures are production-weighted averages and therefore the inclusion of marginally green industries with little green production volumes leads to negligible changes in our results.

Figure D.1: Ratio of green and non-green HHI over time in high-green potential and all green industries, average weighted on production.



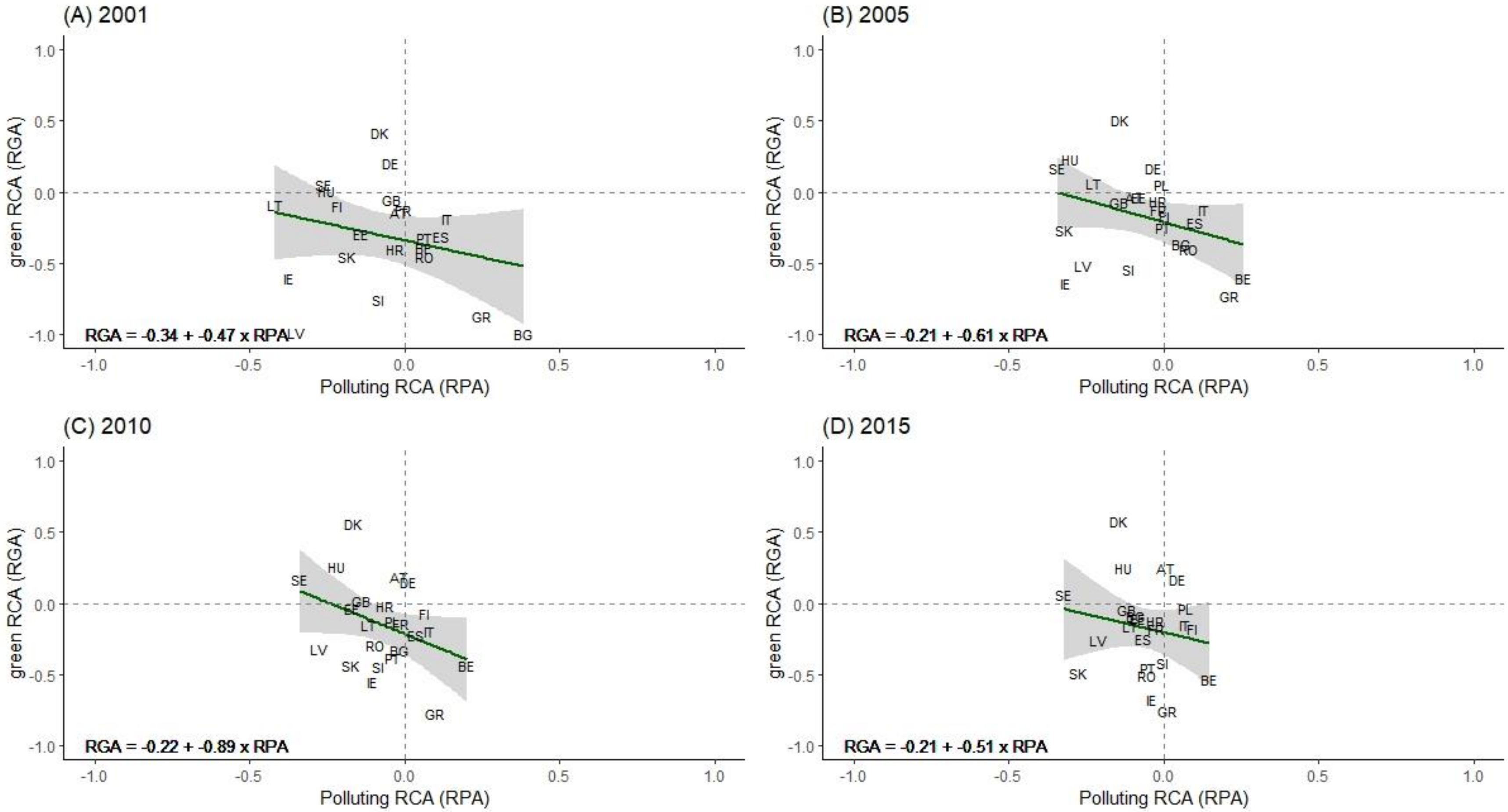
Notes: Authors' elaboration on PRODCOM data. Panel A reports the average HHIR across all high-green industries, weighted on industry output, Panel B reports the same measure but including all green industries. We compute HHI for all countries in our sample as well as only for countries for which we have a balanced panel from 1995, i.e.: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, United Kingdom

Figure D.2: Evolution of green production shares for selected European countries for all green industries.



Notes: Authors' elaboration on PRODCOM data. Panel A, B and C report green production shares over time for large, small and Eastern countries respectively, these have been smoothened by taking 3-years moving averages. We only use green production from all green industries from Table 3 in the main text. EUR is the European average of green shares across all available countries in each year, weighted on production shares, in panel D we compare it with the unweighted average (AVG) which is not affected by countries' size, especially Germany. Because data on Eastern countries is available only from 2001 onwards, and 2003 onwards for Poland, we report both these measures computed for each year for all available countries as well as only for countries for which we have a balanced panel since 1995, i.e.: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, United Kingdom, (EUR95 and AVG95).

Figure D.3: Green and polluting RCA across countries and over time, using green production from all green industries.



Notes: Authors' elaboration on PRODCOM data. We plot countries' green and polluting RCA. Green RCA is based on green production of all green industries. Polluting production is total production from polluting industries identified in Table 2. The RCAs are computed following equation 3 and are made symmetrical around 0 and bounded between -1 and 1, the value of 0 indicates therefore whether a country has, on average, successfully specialized in green production. We also report the coefficient of a regression of green RCA on polluting RCA for each year

E. Robustness checks on regressions of Section 5

We report in Table E.1 descriptive statistics of the variables we use in our econometric analysis of section 5. We report both the level and the log of the number of green and non-green products with RCA to show the skewedness in the distribution of these variables.

Table E.1: Descriptive statistics

Variables	Min	1st Qu.	Median	Mean	3rd Qu.	Max	St. Dev.
Green RCA (log)	0.00	0.11	0.52	0.45	0.73	1.07	0.32
Green RCA, pre-sample mean (log)	0.00	0.10	0.46	0.43	0.70	1.05	0.31
Non-green RCA (log)	0.00	0.19	0.52	0.46	0.71	1.07	0.30
Number of green products with RCA (log)	0.00	0.00	0.00	0.47	0.69	3.09	0.57
Number of non-green products with RCA (log)	0.00	0.00	1.10	0.97	1.61	3.53	0.82
Number of green products with RCA	0.00	0.00	0.00	0.98	1.00	21.00	1.94
Number of non-green products with RCA	0.00	0.00	2.00	2.78	4.00	33.00	4.00

Notes: Authors' elaboration on PRODCOM data. The table reports the distribution of the key variables from equation 4. In addition, the last two rows report the number of green and non-green products with RCA not in logs, to show the difference in distribution of the variable compared to when we take the logarithm.

We also report in this Appendix a battery of robustness checks that are described in detail in section 5. As a reminder, Table E.2 reports the results when looking at path-dependency of non-green RCA and its complementarity with green RCA. We also present the results without log transforming the dependent variable (Table E.3), without making the RCA index symmetrical around 0 (Table E.4). Finally, Tables E.5 and E.6 present the results weighting the regression based on the size of the sector and including both high-green potential and marginally green sectors, respectively.

Table E.2: Path-dependency of non-green RCA and complementarity with green RCA.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA (log) * 2005	0.976***	0.928***	0.686***	0.638***	0.584***
	(0.0307)	(0.0371)	(0.0414)	(0.0427)	(0.0427)
Pre-sample green RCA (log) * 2006	0.936***	0.887***	0.654***	0.634***	0.577***
	(0.0399)	(0.0442)	(0.0457)	(0.0436)	(0.0441)
Pre-sample green RCA (log) * 2007	0.868***	0.823***	0.594***	0.578***	0.521***
	(0.0469)	(0.0509)	(0.0547)	(0.0520)	(0.0531)
Pre-sample green RCA (log) * 2008	0.829***	0.785***	0.567***	0.549***	0.494***
	(0.0530)	(0.0563)	(0.0622)	(0.0577)	(0.0598)
Pre-sample green RCA (log) * 2009	0.792***	0.751***	0.552***	0.536***	0.486***
	(0.0581)	(0.0616)	(0.0635)	(0.0543)	(0.0550)
Pre-sample green RCA (log) * 2010	0.754***	0.712***	0.508***	0.494***	0.442***
	(0.0612)	(0.0643)	(0.0666)	(0.0567)	(0.0573)
Pre-sample green RCA (log) * 2011	0.700***	0.657***	0.461***	0.458***	0.405***
	(0.0674)	(0.0709)	(0.0691)	(0.0602)	(0.0601)
Pre-sample green RCA (log) * 2012	0.677***	0.636***	0.439***	0.424***	0.372***
	(0.0686)	(0.0714)	(0.0692)	(0.0578)	(0.0580)
Pre-sample green RCA (log) * 2013	0.673***	0.634***	0.450***	0.430***	0.384***
	(0.0699)	(0.0724)	(0.0673)	(0.0579)	(0.0570)
Pre-sample green RCA (log) * 2014	0.675***	0.637***	0.451***	0.447***	0.400***
	(0.0691)	(0.0707)	(0.0651)	(0.0581)	(0.0563)
Pre-sample green RCA (log) * 2015	0.665***	0.628***	0.448***	0.430***	0.388***
	(0.0701)	(0.0716)	(0.0647)	(0.0607)	(0.0578)
Green RCA_{t-1} (log)		0.0986***	0.0692*	0.0841**	0.0581
		(0.0370)	(0.0366)	(0.0370)	(0.0357)
Number of green products with RCA_{t-1} (log)			-0.00937		0.0488**
			(0.0192)		(0.0229)
Number of non-green products with RCA_{t-1} (log)			0.166***		0.0758***
			(0.0146)		(0.0135)
Dummy for at least one green product with RCA_{t-1}				-0.0168	-0.0493**
				(0.0173)	(0.0229)
Dummy for at least one non green product with RCA_{t-1}				0.279***	0.218***
				(0.0211)	(0.0235)
Constant	0.115***	0.0892***	0.0389*	0.0109	0.00778
	(0.0234)	(0.0238)	(0.0208)	(0.0181)	(0.0182)
Observations	3,289	3,289	3,289	3,289	3,289
R-squared	0.723	0.729	0.788	0.806	0.818

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. RCA are symmetrical around 0 and the logarithm is taken of RCA+2. Country-year and sector-year fixed effects are included in all estimates. This table replicates the same model as in Table 4, but the outcome variable is now non-green RCA and our main variable of interest is the lag of the green RCA. Standard errors are clustered at the country level and reported in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table E.3: Path-dependency of green RCA and complementarity with non-green RCA, not in log.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA * 2005	0.992***	0.939***	0.622***	0.553***	0.534***
	(0.0288)	(0.0333)	(0.0440)	(0.0417)	(0.0424)
Pre-sample green RCA * 2006	0.940***	0.884***	0.598***	0.538***	0.521***
	(0.0390)	(0.0429)	(0.0494)	(0.0485)	(0.0485)
Pre-sample green RCA * 2007	0.891***	0.832***	0.578***	0.517***	0.504***
	(0.0482)	(0.0517)	(0.0550)	(0.0560)	(0.0552)
Pre-sample green RCA * 2008	0.852***	0.790***	0.550***	0.495***	0.482***
	(0.0537)	(0.0567)	(0.0591)	(0.0588)	(0.0582)
Pre-sample green RCA * 2009	0.821***	0.760***	0.492***	0.444***	0.425***
	(0.0582)	(0.0608)	(0.0620)	(0.0584)	(0.0582)
Pre-sample green RCA * 2010	0.782***	0.725***	0.473***	0.418***	0.404***
	(0.0567)	(0.0582)	(0.0602)	(0.0576)	(0.0574)
Pre-sample green RCA * 2011	0.745***	0.690***	0.436***	0.388***	0.372***
	(0.0592)	(0.0599)	(0.0616)	(0.0572)	(0.0573)
Pre-sample green RCA * 2012	0.716***	0.661***	0.450***	0.424***	0.406***
	(0.0611)	(0.0615)	(0.0580)	(0.0508)	(0.0515)
Pre-sample green RCA * 2013	0.689***	0.634***	0.418***	0.380***	0.366***
	(0.0646)	(0.0650)	(0.0628)	(0.0556)	(0.0562)
Pre-sample green RCA * 2014	0.666***	0.612***	0.408***	0.372***	0.359***
	(0.0671)	(0.0671)	(0.0633)	(0.0570)	(0.0574)
Pre-sample green RCA * 2015	0.648***	0.592***	0.376***	0.347***	0.331***
	(0.0714)	(0.0714)	(0.0666)	(0.0603)	(0.0608)
Non-green RCA_{t-1}		0.161***	0.105**	0.131***	0.0962**
		(0.0436)	(0.0444)	(0.0392)	(0.0428)
Number of green products with RCA_{t-1} (log)			0.441***		0.158***
			(0.0400)		(0.0414)
Number of non-green products with RCA_{t-1} (log)			-0.0142		0.0170
			(0.0256)		(0.0267)
Dummy for at least one green product with RCA_{t-1}				0.493***	0.370***
				(0.0348)	(0.0435)
Dummy for at least one non green product with RCA_{t-1}				-0.00260	3.91e-05
				(0.0301)	(0.0339)
Constant	-0.0391	-0.00584	-0.313***	-0.365***	-0.416***
	(0.0252)	(0.0257)	(0.0492)	(0.0406)	(0.0468)
Observations	3,289	3,289	3,289	3,289	3,289
R-squared	0.659	0.671	0.760	0.779	0.784

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. Only the number of green and non-green products with RCA are in logs in this specification. RCA are symmetrical around 0 and the logarithm is taken of RCA+2. Country-year and sector-year fixed effects are included in all estimates. All green industries are included. Standard errors are clustered at the country level and reported in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table E.4: Path-dependency of green RCA and complementarity with non-green RCA, using asymmetric RCAs.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA (log) * 2005	1.010***	0.971***	0.746***	0.702***	0.688***
	(0.0356)	(0.0396)	(0.0575)	(0.0565)	(0.0592)
Pre-sample green RCA (log) * 2006	0.978***	0.936***	0.731***	0.690***	0.678***
	(0.0418)	(0.0470)	(0.0641)	(0.0639)	(0.0659)
Pre-sample green RCA (log) * 2007	0.921***	0.876***	0.694***	0.649***	0.641***
	(0.0551)	(0.0603)	(0.0732)	(0.0734)	(0.0742)
Pre-sample green RCA (log) * 2008	0.861***	0.814***	0.643***	0.606***	0.596***
	(0.0657)	(0.0701)	(0.0815)	(0.0805)	(0.0817)
Pre-sample green RCA (log) * 2009	0.800***	0.754***	0.562***	0.530***	0.516***
	(0.0766)	(0.0800)	(0.0880)	(0.0849)	(0.0863)
Pre-sample green RCA (log) * 2010	0.762***	0.719***	0.544***	0.506***	0.497***
	(0.0777)	(0.0797)	(0.0890)	(0.0861)	(0.0877)
Pre-sample green RCA (log) * 2011	0.750***	0.708***	0.528***	0.493***	0.482***
	(0.0805)	(0.0821)	(0.0908)	(0.0865)	(0.0883)
Pre-sample green RCA (log) * 2012	0.738***	0.695***	0.543***	0.523***	0.510***
	(0.0791)	(0.0810)	(0.0877)	(0.0828)	(0.0847)
Pre-sample green RCA (log) * 2013	0.731***	0.688***	0.536***	0.508***	0.498***
	(0.0813)	(0.0838)	(0.0903)	(0.0852)	(0.0870)
Pre-sample green RCA (log) * 2014	0.728***	0.685***	0.537***	0.507***	0.499***
	(0.0817)	(0.0841)	(0.0896)	(0.0853)	(0.0868)
Pre-sample green RCA (log) * 2015	0.729***	0.685***	0.526***	0.507***	0.495***
	(0.0848)	(0.0872)	(0.0897)	(0.0865)	(0.0877)
Non-green RCA_{t-1} (log)		0.136***	0.0855*	0.107**	0.0836*
		(0.0483)	(0.0476)	(0.0471)	(0.0483)
Number of green products with RCA_{t-1} (log)			0.361***		0.147**
			(0.0466)		(0.0569)
Number of non-green products with RCA_{t-1} (log)			-0.00943		0.0133
			(0.0259)		(0.0305)
Dummy for at least one green product with RCA_{t-1}				0.391***	0.273***
				(0.0392)	(0.0491)
Dummy for at least one non green product with RCA_{t-1}				0.00104	-0.00239
				(0.0330)	(0.0370)
Constant	0.115***	0.0697**	0.0167	-0.00768	-0.0139
	(0.0288)	(0.0302)	(0.0283)	(0.0293)	(0.0297)
Observations	3,289	3,289	3,289	3,289	3,289
R-squared	0.649	0.659	0.727	0.737	0.742

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. Country-year and sector-year fixed effects are included in all estimates. Standard errors are clustered at the country level and reported in parentheses*** p<0.01, ** p<0.05, * p<0.1.

Table E.5 reports our results, weighting for industries' size, measured as industry total production in 2004. We find that the results on path-dependency are robust, while in contrast the complementarity effect vanishes when we control for the number of green and non-green products with RCA (column 3). Similarly, our results lose significance when we consider all green industries in our analysis in Table E.6. Our preferred specification (column 3) shows complementarity between non-green RCA and green RCA only at 10% significance, and when we simultaneously control for our diversification measures and the threshold of having at least one (green or non-green) product with RCA (column 6), we find no significant complementarity effect. However, one should bear in mind that marginally green industries only represent a very small share of total green production, which means that our main specification discussed in the main text concerns the vast majority of green production.

Table E.5: Path-dependency of green RCA and complementarity with non-green RCA, weighting for the size of the sector.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA (log) * 2005	0.939***	0.886***	0.622***	0.545***	0.522***
	(0.0404)	(0.0466)	(0.0533)	(0.0461)	(0.0480)
Pre-sample green RCA (log) * 2006	0.890***	0.834***	0.587***	0.523***	0.502***
	(0.0440)	(0.0490)	(0.0561)	(0.0491)	(0.0509)
Pre-sample green RCA (log) * 2007	0.836***	0.777***	0.557***	0.492***	0.476***
	(0.0474)	(0.0519)	(0.0579)	(0.0548)	(0.0548)
Pre-sample green RCA (log) * 2008	0.773***	0.712***	0.504***	0.445***	0.427***
	(0.0545)	(0.0577)	(0.0648)	(0.0605)	(0.0617)
Pre-sample green RCA (log) * 2009	0.734***	0.674***	0.469***	0.418***	0.399***
	(0.0607)	(0.0633)	(0.0651)	(0.0585)	(0.0587)
Pre-sample green RCA (log) * 2010	0.686***	0.628***	0.424***	0.375***	0.359***
	(0.0627)	(0.0636)	(0.0689)	(0.0639)	(0.0641)
Pre-sample green RCA (log) * 2011	0.655***	0.600***	0.403***	0.363***	0.347***
	(0.0627)	(0.0630)	(0.0686)	(0.0619)	(0.0626)
Pre-sample green RCA (log) * 2012	0.603***	0.550***	0.388***	0.379***	0.363***
	(0.0679)	(0.0679)	(0.0658)	(0.0526)	(0.0543)
Pre-sample green RCA (log) * 2013	0.579***	0.526***	0.350***	0.328***	0.313***
	(0.0721)	(0.0720)	(0.0735)	(0.0632)	(0.0637)
Pre-sample green RCA (log) * 2014	0.541***	0.492***	0.330***	0.316***	0.302***
	(0.0758)	(0.0752)	(0.0752)	(0.0647)	(0.0665)
Pre-sample green RCA (log) * 2015	0.528***	0.474***	0.307***	0.298***	0.284***
	(0.0791)	(0.0789)	(0.0774)	(0.0686)	(0.0697)
Non-green RCA_{t-1} (log)		0.174***	0.0724	0.123**	0.0602
		(0.0486)	(0.0531)	(0.0481)	(0.0519)
Number of green products with RCA_{t-1} (log)			0.217***		0.0622***
			(0.0241)		(0.0233)
Number of non green products with RCA_{t-1} (log)			0.0122		0.0291*
			(0.0169)		(0.0168)
Dummy for at least one green product with RCA_{t-1}				0.274***	0.227***
				(0.0194)	(0.0246)
Dummy for at least one non green product with RCA_{t-1}				0.0170	0.00794
				(0.0215)	(0.0226)
Constant	0.141***	0.0833***	0.0922***	0.0664***	0.0678***
	(0.0243)	(0.0256)	(0.0246)	(0.0238)	(0.0235)
Observations	3,289	3,289	3,289	3,289	3,289
R-squared	0.693	0.704	0.772	0.799	0.804

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. RCA are symmetrical around 0 and the logarithm is taken of RCA+2. Estimates are weighted on industry size, using industry total production in 2004. Country-year and sector-year fixed effects are included in all estimates. All green industries are included. Standard errors are clustered at the country level and reported in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table E.6: Path-dependency of green RCA and complementarity with non-green RCA, for all green industries.

	(1)	(2)	(3)	(4)	(5)
Pre-sample green RCA (log) * 2005	0.964***	0.943***	0.672***	0.588***	0.583***
	(0.0274)	(0.0286)	(0.0406)	(0.0412)	(0.0417)
Pre-sample green RCA (log) * 2006	0.923***	0.901***	0.648***	0.567***	0.562***
	(0.0336)	(0.0346)	(0.0435)	(0.0437)	(0.0440)
Pre-sample green RCA (log) * 2007	0.813***	0.791***	0.567***	0.494***	0.491***
	(0.0503)	(0.0507)	(0.0555)	(0.0556)	(0.0557)
Pre-sample green RCA (log) * 2008	0.771***	0.747***	0.538***	0.470***	0.466***
	(0.0530)	(0.0534)	(0.0567)	(0.0556)	(0.0558)
Pre-sample green RCA (log) * 2009	0.726***	0.704***	0.485***	0.428***	0.422***
	(0.0532)	(0.0531)	(0.0564)	(0.0535)	(0.0542)
Pre-sample green RCA (log) * 2010	0.695***	0.674***	0.465***	0.404***	0.399***
	(0.0529)	(0.0524)	(0.0551)	(0.0531)	(0.0535)
Pre-sample green RCA (log) * 2011	0.648***	0.628***	0.431***	0.384***	0.379***
	(0.0542)	(0.0536)	(0.0545)	(0.0517)	(0.0522)
Pre-sample green RCA (log) * 2012	0.620***	0.600***	0.420***	0.381***	0.375***
	(0.0553)	(0.0547)	(0.0539)	(0.0506)	(0.0512)
Pre-sample green RCA (log) * 2013	0.587***	0.567***	0.384***	0.345***	0.339***
	(0.0562)	(0.0557)	(0.0550)	(0.0508)	(0.0514)
Pre-sample green RCA (log) * 2014	0.563***	0.540***	0.367***	0.327***	0.322***
	(0.0583)	(0.0578)	(0.0539)	(0.0487)	(0.0490)
Pre-sample green RCA (log) * 2015	0.553***	0.529***	0.350***	0.310***	0.304***
	(0.0607)	(0.0601)	(0.0547)	(0.0490)	(0.0494)
Non-green RCA_{t-1} (log)		0.125***	0.0628*	0.0434	0.0390
		(0.0330)	(0.0349)	(0.0322)	(0.0340)
Number of green products with RCA_{t-1} (log)			0.294***		0.0590**
			(0.0252)		(0.0241)
Number of non green products with RCA_{t-1} (log)			-0.0151		-0.00798
			(0.0143)		(0.0150)
Dummy for at least one green product with RCA_{t-1}				0.329***	0.283***
				(0.0216)	(0.0272)
Dummy for at least one non green product with RCA_{t-1}				0.0222	0.0308
				(0.0170)	(0.0204)
Constant	0.119***	0.0662***	0.0883***	0.0649***	0.0682***
	(0.0176)	(0.0212)	(0.0183)	(0.0159)	(0.0161)
Observations	5,001	4,999	4,999	4,999	4,999
R-squared	0.667	0.674	0.751	0.778	0.779

Notes: Pre-sample mean computed for the years 2001-2004, for Poland only 2003-2004 due to data constraints. All explanatory variables, except the pre-sample mean, are lagged by one year. Estimation time span is 2005-2015. RCA are symmetrical around 0 and the logarithm is taken of RCA+2. We include in this estimation all green industries. Country-year and sector-year fixed effects are included in all estimates. All green industries are included. Standard errors are clustered at the country level and reported in parentheses*** p<0.01, ** p<0.05, * p<0.1.

ABOUT OFCE

The Paris-based Observatoire français des conjonctures économiques (OFCE), or French Economic Observatory is an independent and publicly-funded centre whose activities focus on economic research, forecasting and the evaluation of public policy.

Its 1981 founding charter established it as part of the French Fondation nationale des sciences politiques (Sciences Po), and gave it the mission is to "ensure that the fruits of scientific rigour and academic independence serve the public debate about the economy". The OFCE fulfils this mission by conducting theoretical and empirical studies, taking part in international scientific networks, and assuring a regular presence in the media through close cooperation with the French and European public authorities. The work of the OFCE covers most fields of economic analysis, from macroeconomics, growth, social welfare programmes, taxation and employment policy to sustainable development, competition, innovation and regulatory affairs.

ABOUT SCIENCES PO

Sciences Po is an institution of higher education and research in the humanities and social sciences. Its work in law, economics, history, political science and sociology is pursued through ten research units and several crosscutting programmes.
Its research community includes over two hundred twenty members and three hundred fifty PhD candidates. Recognized internationally, their work covers a wide range of topics including education, democracies, urban development, globalization and public health.
One of Sciences Po's key objectives is to make a significant contribution to methodological, epistemological and theoretical advances in the humanities and social sciences. Sciences Po's mission is also to share the results of its research with the international research community, students, and more broadly, society as a whole.

PARTNERSHIP

SciencesPo