Firm Heterogeneity in Carbon Productivity: Evidence from Representative Cross-Country Micro Data

Josh De Lyon[1,2] and Antoine Dechezleprêtre[1,2]

[1] OECD Directorate for Science, Technology and Innovation

[2] Centre for Economic Performance, London School of Economics and Political Science

Abstract

This paper explores the heterogeneity in firm environmental performance and its relationship with economic performance based on representative data from Croatia, France, Indonesia, Lithuania, and Sweden. It documents extensive firm heterogeneity in carbon productivity within narrowly defined industries, which significantly exceeds the extent of heterogeneity in labour productivity. On average, the 90th percentile firm is 24 times more carbon productive than the 10th percentile firm in the same industry, compared with five times for labour productivity. Even comparing firms producing the same 8 digit product in France, there is huge dispersion in carbon productivity. This heterogeneity has important implications for aggregate emissions: raising the carbon productivity of the least productive firms to the carbon productivity of the 25th percentile firm in their industry would reduce carbon emissions by half for the same level of output. Furthermore, a growing carbon productivity dispersion is associated with lower carbon productivity growth. Industries that are more dispersed in carbon productivity are also more dispersed in labour productivity and firms that are more carbon productive are also more labour productive, controlling for other factors. These correlations suggest that structural characteristics of both firms and industries may jointly explain both economic and environmental outcomes. Firm-level regressions show that an in energy prices causes a fall in CO2 and an improvement in CO2 productivity, without detrimental economic effects. This evidence suggests that there is a significant untapped potential of improved environmental performance and reduction in industrial emissions – by boosting the performance of unproductive firms and improving allocative efficiency - and that these improvements could be achieved without harming economic performance.

Acknowledgements

This project has benefited from very helpful comments from Chiara Criscuolo, Guy Lalanne and Francesco Manaresi, as well as Delegates of CIIE and WPIA, especially discussions by Paul Cusson, Danny Leung, Mika Maliranta, and Michael Polder. The project benefited greatly from work of previous OECD micro-distributed data projects, and especially from the help of Rudy Verlhac and Martin Reinhard. Gabriela Garcia Sandoval and Xavier Koch provided excellent research assistance relating to this project and Isabelle Desnoyers-James provided invaluable help with data collection and meta data. The authors are very grateful to Ulrika Stavlöt for running the programmes on Swedish data, and to Statistics Lithuania and the Croatian Bureau of Statistics for helpful discussions relating to data. Access to some confidential data on France used in this work has been made possible within a secure environment offered by CASD – Centre d'accès sécurisé aux données (Ref. 10.34724/CASD). This project is partly financed by the European Commission under Work Programme Horizon Europe – 2021-2022 Horizontal support expenditure for Horizon Europe and Euratom Programmes under grant no. LC-01965484. The authors declare to have no conflict of interest to disclose regarding the research on this paper.

1 Introduction

The industrial sector accounts for 23% of all emissions in OECD Member countries (Aulie et al., 2023[1]). In recent years, manufacturing firms in many OECD Member countries have become more carbon efficient; in 2018, the ratio of real output to CO_2 emissions was 56% higher than it was in 2000 on average across the manufacturing sector of OECD Member countries. This pattern holds even after accounting for carbon embodied in intermediate inputs and sectoral structural change.[1]

Policy has played a role in this clean-up of manufacturing. For example, there is robust evidence suggesting that carbon pricing, such as the European Union (EU) Emissions Trading Scheme (ETS) reduced the emissions and increased carbon productivity of treated firms.[2] Yet many policies, including the ETS, only apply to a subset of firms – typically those with high pollution in certain sectors. Today, 24% of global emissions are subject to a carbon tax or an emissions trading system. Despite the recent improvements in the carbon efficiency of manufacturing in OECD Member countries, there is a long way to go to achieve the necessary and ambitious net zero commitments adopted by around 150 countries around the world, representing 90% of global greenhouse gas emissions. Policies will need to become more stringent as well as increase coverage across firms, sectors, and countries.

In a world aiming for net zero, carbon efficiency is set to become a key driver of firms' competitiveness. A recent literature has documented huge differences in efficiency across firms, especially in terms of their labour productivity (see for example Andrews, Criscuolo and Gal (2019[2])). These may be caused by a lack of technology transfer between firms, excessive barriers to entry, or a lack of potential for productive startups to scale-up and can result in a drag on aggregate productivity growth. Yet, while there is extensive evidence on the extent and determinants of the differences between firms' economic performance, much less is known on differences in their environmental performance. A handful of papers suggest there is also significant dispersion across firms in their environmental performance (Wagner et al. (2020[3]); Lyubich, Shapiro, and Walker (2018[4])), but they do not investigate the drivers of this dispersion. It could, for instance, be related to similar factors as dispersion in labour productivity, as well as under-pricing of carbon and energy comprising a small share of total costs. Understanding this firm heterogeneity in environmental performance is key to help design policies to improve the efficiency of firms across the board, from the most efficient firms to the least efficient.

This paper uses cross-country firm-level data to document differences across firms in their environmental and economic performance. The project develops new methodologies to collect, harmonise, and analyse micro data for multiple OECD and non-OECD Member countries. Building on the OECD MultiProd project (Berlingieri et al., 2017[5]), it develops a set of programmes which run flexibly on confidential micro datasets on the environmental and economic performance of firms. Carbon emissions from production are constructed from data on energy use which is used to measure carbon productivity – defined as dollars of value added per tonne of CO_2 – alongside measures of economic performance such as labour productivity. Thorough procedures of data cleaning, harmonisation, and checks are developed to ensure that data are

[1] Authors' calculations using OECD data (Yamano and Guilhoto, 2020[96]). Also see Shapiro and Walker (2018[11]) and Levinson (2015[12]) for a discussion of the manufacturing cleanup in the United States.

[2] See for example Dechezleprêtre, Nachtigal, and Venmans (2023[31]), Colmer et al. (2020[70]), or Petrick and Wagner (2014[33]). A thorough discussion is presented in the literature review of this paper.

comparable across countries, including a new algorithm to estimate carbon emissions when only total energy expenditure is available, new procedures to identify data entry errors (which appear to be more common in environmental surveys than in balance sheet data), and harmonised use of sampling weights. The project currently uses data for Croatia, France, Indonesia, Lithuania, and Sweden and is securing access to data in various other countries.

The data reveal that there is huge dispersion in carbon productivity between firms within narrowly defined industries. On average, the 90th percentile firm is 24 times more carbon productive than the 10th percentile firm. There is far more dispersion in carbon productivity than in labour productivity in all countries included in the sample. On average, the 90th percentile firm is five times more labour productive than the 10th percentile firm.

Addressing this dispersion in carbon productivity could have important aggregate implications. A counterfactual exercise is conducted, whereby the carbon productivity of all firms is improved to at least that of the median firm in their industry. By improving the carbon productivity of the least productive firms to the median level, CO_2 emissions could be reduced by 71% to achieve the same total level of output, aggregating across the five countries. Bringing the unproductive firms up to the 25th percentile firm would already halve aggregate emissions. This result is amplified by the empirical observation that less productive firms tend to be larger. Furthermore, regression analysis shows that industries with more dispersion in carbon productivity have worse aggregate carbon productivity, and that as industries become more dispersed, they also become more carbon intensive.

The dispersion in carbon productivity is not due to firms in the same industry producing different products. Using data on firms' sales by 8 digit products for France, there is still large dispersion in carbon productivity when comparing firms producing the same products, and this dispersion is still magnitudes greater than dispersion in labour productivity. Comparing firms producing the same 8 digit product, the 90th percentile firm is 14 times more carbon productive than the 10th percentile firm. In comparison, in terms of labour productivity, the 90th percentile firm is three times more productive than the 10th percentile firm.

Average dispersion has generally been increasing in Indonesia, France, and Sweden over the past decade. There are differences in dispersion patterns over time across industries: in some industries, dispersion has increased while in others it has fallen, and this may be driven by improvements at either the top or the bottom of the distribution.

The characteristics of industries that are more dispersed in carbon productivity are examined. There is a strong relationship between dispersion in carbon productivity and dispersion in labour productivity: industries with greater dispersion in labour productivity are also more dispersed in carbon productivity, and this also holds in changes over time. This suggests that the factors that have been proposed to explain dispersion in labour productivity, such as a lack of competitive pressure, limited technology diffusion, or barriers to entry and scale-up, may also explain dispersion in carbon productivity. Energy comprises a lower share of firms' total costs than labour – energy accounts for 7% of total costs on average compared with 31% for labour costs in the manufacturing sample considered – meaning that there is less competitive pressure on firms to be energy productive compared with labour productive. In addition, carbon is notoriously underpriced, especially for less carbon productive firms which are, on average, larger emitters and typically benefit from lower fossil fuel prices and energy cost compensation schemes. Changes in carbon productivity dispersion are also related to changes in capital intensity and capital productivity dispersion, as well as dispersion in the cost shares of other factors of production (labour and intermediate inputs).

Cross-sectional firm-level regressions reveal that firms with higher carbon productivity also have higher labour productivity and that, as firms improve their carbon productivity over time, they also improve their labour productivity, even controlling for other firm-level characteristics. This suggests that there may be synergies between improving aggregate labour productivity – crucial for achieving growth and wage

increases – while also reducing carbon emissions from the industrial sector. Firms with higher carbon productivity also tend to be smaller, younger, and less capital intensive.

Finally, the impact of energy price changes, which roughly proxy for carbon pricing policies, on firms' economic and environmental performance is examined. An established instrumental variables strategy is adopted from the existing literature, which exploits variation in the types of fuels that firms use as differences in exposure to macro energy price changes. The results show that increases in energy prices cause firms to significantly reduce their total carbon emissions and improve their carbon productivity. In Indonesia, higher energy prices cause an increase in value added, employment, and labour productivity, while in France there is no evidence of any causal negative effects on economic performance.

The heterogeneity of effects across firms of different initial carbon productivity, age, and size (in terms of employment) are shown. While there are some differences in the estimated elasticities by firm group, firms of all types respond to higher energy prices by reducing carbon emissions and improving carbon productivity. This suggests that rolling out carbon pricing to have broader coverage of firms could be beneficial in achieving the transition to net zero without significant economic costs.

This evidence suggests that there is a significant untapped potential of improved environmental performance and reduction in industrial emissions, and that this is unlikely to come at the expense of economic performance. First, while policies should keep incentivising innovation in green technologies to push the frontier of environmental efficiency, there is huge potential to reduce emissions by improving the carbon productivity of unproductive firms, as well as by improving allocative efficiency in the long run. This suggests that policies that have been proposed to increase technology diffusion are critically important for improving carbon productivity. Second, structural policies or policies to improve labour productivity in general may have co-benefits for carbon efficiency and aggregate emissions. Third, expansions of carbon pricing are likely to come with improved carbon efficiency with potentially positive impacts on firms' labour productivity on average.

The paper proceeds with an in-depth review of the literature and the contribution of this paper. It then provides detail on the datasets used and the procedures for harmonising micro data across countries in Section 2. In Section 3, the extent of dispersion in carbon productivity is documented, and the factors correlated with this dispersion are explored. The characteristics of firms with higher carbon productivity are documented in Section 4. Section 5 presents results of regressions of changes in energy prices on firms' environmental and economic performance, including heterogeneity across different types of firms. Finally, Section 6 concludes and provides a discussion of policy implications.

Literature review

This paper contributes to four branches of the literature. First, there is a small literature documenting the extent of heterogeneity between firms in their environmental performance. Most closely related to the present paper is Wagner et al. (2020[3]) who document the distributions of CO_2 per worker for seven countries using representative data for manufacturing sector plants. They show that there is high dispersion in CO_2 per job within 2-digit industries to a similar degree across countries, but that this dispersion has declined over time. Also in a cross-country setting, Capelle et al. (2023[6]) document extensive within-industry dispersion in environmental performance using self-reported data from 4 000 large, listed firms. They explain these differences across firms with a model whereby firms choose different vintages of capital and how much to spend on R&D.

Looking at individual countries, Lyubich, Shapiro, and Walker (2018[4]) document the extent of heterogeneity in carbon productivity across firms in the United States, showing that it is significantly greater than heterogeneity in labour, capital, and multi-factor productivities. Similar evidence is documented for Estonia (Criscuolo et al., 2024[7]) and Germany (Von Graevenitz and Rottner (2020[8]) and Petrick,

Rehdanz, and Wagner (2010[9])). Klenow, Pastén, and Ruane (2024[10]) document that there are differences in fossil fuel intensity across firms in Chile, and that this is negatively correlated with their total-factor productivity.

This paper contributes to this literature by providing new detailed evidence on the extent of heterogeneity across firms within industries in multiple countries. It documents extensive dispersion even within very narrowly defined 4-digit industries, measuring carbon productivity as value added per unit of CO_2 with accurate estimates of CO_2 computed from energy consumption or energy expenditures. It shows the evolution of dispersion over time, and detailed measures of dispersion by industry. The characteristics of industries with more dispersion are also investigated, providing indicative evidence on the potential frictions related to dispersion, as well as the correlation between dispersion and aggregate outcomes.

The second literature to which this paper contributes explains the recent "clean-up" of manufacturing production in many OECD Member countries, whereby less CO_2 is emitted to produce the same or increasing outputs. In the United States, aggregate improvements in carbon efficiency can be attributed to within-industry or within-product improvements in carbon productivity, rather than a change in composition of production (Shapiro and Walker (2018[11]) and Levinson (2015[12])). Petrick (2013[13]) uses firm-level data for Germany to show that improvements in carbon productivity are mainly driven by innovative new entrants rather than improvements within firms that already operate. Looking at the role of product mix in Germany, Rottner and Von Graevenitz (2022[14]) show that, since 2011, a shift to a cleaner product composition explained the entire clean-up of manufacturing. In contrast to the findings of Petrick (2013[13]), Kuosmanen and Maczulskij (2024[15]) use data from Finland to show that existing firms were the biggest contributor to improvements in carbon efficiency, while the net contribution of entry and exit was negative during the Great Financial Crisis. Barrows and Ollivier (2018[16]) show that reallocation across firms in India played an important role in the improvement of aggregate carbon productivity, with the product mix playing a counteracting role.

This paper contributes to this literature by showing how firms respond to energy price shocks, which may proxy for carbon pricing policies, and therefore explains one mechanism through which the clean-up of manufacturing may have occurred. Higher carbon prices may incentivise firms to improve the carbon efficiency of their production processes, leading to an aggregate clean-up of manufacturing.

Third, a broader literature documents the extent of firm heterogeneity in other metrics of performance, mainly productivity. Berlingieri, Blanchenay, Criscuolo (2017[17]) use representative cross-country data to document large and increasing dispersion in labour and multi-factor productivity. Andrews, Criscuolo, and Gal (2019[2]) show that the increasing divergence is related to weakening catch-up and market selection below the frontier, while Berlingieri et al. (2020[18]) document the characteristics of the less productive laggard firms. Canonical papers on this topic include Syverson (2011[19]) on the drivers of productivity and Hsieh and Klenow (2009[20]) for the implications of extensive misallocation of resources between productive and unproductive firms.

This paper contributes to the literature on heterogeneity between firms by providing further evidence on the huge differences between firms even within narrowly defined industries, namely their differences in carbon productivity. Furthermore, it shows that dispersion in labour productivity is correlated with dispersion in carbon productivity, both in levels and changes, and controlling for other industry characteristics. This suggests that factors such as a lack of technological diffusion or credit constraints related to dispersion in both types of productivity. This suggests that addressing these factors may help to achieve the net zero transition alongside strong economic performance.

Finally, this paper contributes to an extensive literature estimating the impacts of energy prices, carbon pricing, and broader environmental policies on firms' economic and environmental performance. Numerous studies estimate the impacts of energy prices, carbon prices, and environmental policy on firm outcomes. Papers looking at the effects of firm prices typically use a shift-share instrumental variables methodology, and sometimes country-sector-year average energy price data, outlined in Sato et al. (2019[21]). The

methodology consists of using the shares of each energy type used by the firm or industry as exposure to changes in prices of each energy type. These papers typically find that higher energy prices – which may proxy for carbon pricing – tend to reduce CO_2 emissions from production and, hence, increase carbon productivity (see Boyd and Lee (2019[22]) for the United States, Brucal and Dechezleprêtre (2021[23]) for Indonesia, Dussaux (2020[24]) and Marin and Vona (2021[25]) for France, and Steinbuks and Neuhoff (2014[26]) for a cross-country study).

Some studies have found that increases in energy prices cause an improvement in economic productivity. André et al. (2023[27]) use cross-country data and show that productivity initially declines as firms reduce their capacity utilisation but then improves in the medium term in response to energy price shocks. Amann et al. (2021[28]) use data from Oman to show that increases in fossil fuel energy prices, following subsidy reforms, lead to significant improvements in productivity.[3] On employment, the evidence suggests that there is a reallocation effect leading to close to no net impact on employment, whereby larger or less energy efficient firms reduce their employment slightly but this is offset by increases in employment in other firms. Brucal and Dechezleprêtre (2021[23]) show this for Indonesia, while Dussaux (2020[24]) and Marin and Vona (2021[25]) show this for France. Fetzer, Palmou, and Schneebacher (2024[29]) use real-time energy price data to show that increasing energy prices are not associated with employment layoffs. On the other hand, Cox et al. (2014[30]) estimate that increasing energy prices slightly reduces employment when output is held constant (isolating the substitution effect only) but reduces employment more significantly as a result of firms downscaling through reduced overall output.[4,5]

There are also many papers estimating the effects of various carbon pricing policies. The most common policy studied is the European Union Emissions Trading Scheme (EU ETS). Looking at the effects of the EU ETS on multiple European countries, Dechezleprêtre, Nachtigal, and Venmans (2023[31]) find that the carbon price was effective in reducing carbon emissions while causing no significant change in employment and profits, with regulated firms' revenue and fixed assets increasing. Colmer et al. (2024[32]) find similar results for French manufacturing firms, while in Germany the EU ETS caused a fall in emissions and no negative effects on employment, turnover, or exports (Petrick and Wagner, 2014[33]), as well as on productivity (Lutz, 2016[34]). Slight improvements in carbon intensity are found in Italy (Maria D'arcangelo, Pavan and Calligaris, 2022[35]) and Lithuania (Jaraite and Di Maria, 2016[36]). In Norway, Klemetsen, Rosendahl and Jakobsen (2020[37]) find that the ETS caused weak emissions reductions and no effect on emissions intensity. The effects of the ETS on energy efficiency also disseminate to other firms through the production network (Criscuolo et al., 2024[7]).[6]

[3] Grover and Karplus (2020[58]) use cross-country data from World Bank Enterprise Surveys to show that firms' energy intensity is related to the quality of their management practices, suggesting a possible mechanism through which energy price changes might improve economic performance.

[4] There is contrasting evidence on the passthrough of energy cost shocks to consumer prices, with Alpino et al. (2023[59]) using Italian survey data to show that the short term passthrough is close to zero due to fixed contracts, while Arquié and Thie (2023[60]) find more than 100% passthrough for France and Ganapati, Shapiro, and Walker (2020[61]) find 70% passthrough in the United States.

[5] Curuk and Rozendaal (2022[62]) study energy price effects in 15 European countries, finding a negative association with the labour share of income, but no association with industry concentration, markups, or the capital share. Mertens, Müller, and Neuschaffer (2022[63]) find that higher energy prices are associated with a decrease in wages, especially in small firms. Finally, Dussaux and Monjon (2023[64]) find that energy price increases are associated with a decline in exporting in France.

[6] There is evidence that increasing the price of carbon can promote innovation, which may explain improvements in carbon and economic productivity. Calel and Dechezleprêtre (2016[56]) show that the ETS increased low-carbon innovation across multiple European countries, while Borghesi et al. (2012[68]) find the same result for energy efficiency

Several studies have examined the impacts of country-specific carbon trading schemes. In Canada, the carbon tax introduced by British Columbia is found to have reduced emissions and increased production (Ahmadi, Yamazaki and Kabore (2022[38]), Kumbhakar, Badunenko and Willox (2022[39]) and a review in Murray and Rivers (2015[40])), with a positive effect on productivity (Yamazaki, 2022[41]). Yamazaki (2017[42]) shows that the revenue-neutral carbon tax caused employment to rise in clean services industries and fall in dirtier industries, with the aggregate average effect on employment being small and positive.[7] Looking at the effects of a carbon tax policy in the UK, Martin, de Preux and Wagner (2014[43]) find a strong negative effect on energy intensity and electricity use and no statistically significant impacts on employment, revenue, or plant exit. Carbon pricing is also found to have a negative impact on emissions in Sweden (Martinsson et al., 2022[44]) and the People's Republic of China (hereafter, 'China') (Zhang and Wu, 2022[45]). D'Arcangelo et al. (2022[46]) use the OECD Effective Carbon Rates database to show that, in a cross-country setting, carbon pricing is associated with a reduction in emissions.[8]

Various papers study the impacts of environmental regulations beyond climate policy, showing that they effectively reduce emissions across various countries and sectors. Najjar and Cherniwchan (2021[47]) found that air quality standards in Canada caused a significant reduction in $PM_{2.5}$ emissions, largely through reduced production, while also decreasing the number of pollution-intensive plants. Greenstone, List, and Syverson (2012[48]) show that US air quality regulations, especially ozone regulations, caused a 2.6% drop in productivity in heavily regulated sectors, while Albrizio, Kozluk, and Zipperer (2017[49]) note that although environmental regulations across OECD Member countries have increased, their overall effect on productivity has been minimal, with short term disruptions where advanced firms gain slightly while less productive firms see declines. For literature reviews on the relationship between environmental regulations and business performance, see Dechezleprêtre and Kruse (2017[50]) and Koźluk and Zipperer (2013[51]).[9]

---

innovations in Italy and (Calel, 2020[69]) shows that UK carbon patenting and R&D increased in response to the ETS. Känzig (2023[67]) uses high-frequency data from the European carbon market, focusing on carbon futures price movements, regulatory event studies, and household-level data to show that the ETS led to higher energy prices, lower emissions and more green innovation, but at the expense of lower economic activity which disproportionately affected poorer households. Tian (2020[74]) shows that carbon pricing in China increased innovation.

[7] Also for Canada, Gu, Hussain and Willox (2019[73]) estimate environmentally adjusted multi-factor productivity which had higher growth than standard measures of MFP because of declines in emissions intensity.

[8] Also see Bernard and Kichian (2021[72]) and Hintermann et al. (2024[71]) for passthrough estimates to prices of the effects of carbon pricing schemes. Alpay, Buccola, and Kerkvliet (2002[80]) and Lanoie et al. (2011[79]) show that technological innovation in pollution-heavy sectors can spur productivity growth. Similarly, Hamamoto (2006[81]) finds in Japan that environmental regulation boosts R&D and productivity. Gray and Shadbegian (2003[82]) observe that integrated pulp and paper mills face significant productivity declines due to abatement costs, while non-integrated mills show little effect, while Koźluk and Zipperer (2013[51]) and Lu and Pless (2022[83]) emphasize the importance of policy flexibility in maintaining productivity growth.

[9] Chen and Xu (2021[75]) found that stricter air pollution controls in China reduced exports from pollution-intensive industries through firm exits, aligning with Broner, Bustos, and Carvalho (2019[76]), who show that laxer environmental policies give countries a comparative advantage in polluting industries. Brunel (2017[77]) and Harrison and Antweiler (2003[78]) reinforce that while emissions reductions occur, pollution-intensive firms bear competitiveness costs. Employment impacts of environmental regulations are modest, with shifts toward greener jobs. Vona et al. (2018[84]) show that labour demand shifts without significant overall employment changes. Grolleau, Mzoughi and Pekovic (2012[85]) and Horbach (2010[86]) link job growth to cleaner technologies, particularly in R&D-intensive firms. Horbach and Rennings (2013[87]) confirm that cleaner process innovations foster employment, while Curtis (2018[88]) and Walker (2013[89]) find younger workers face slower hiring under regulations.

This paper contributes to this literature by estimating the impact of energy prices on firm outcomes in a cross-country setting, and by highlighting the differences and similarities across firms in their responses to energy price shocks. It shows that energy prices cause an increase in carbon productivity in three diverse countries – France, Indonesia, and Sweden. There is evidence that energy prices may be complementary with capital and substitutable with labour. Some of the effects differ across countries: in Indonesia, higher energy prices are associated with a rise in value added, while in France there is no statistically significant effect on value added.

2 Data

The project relies on firm-level data for multiple countries. As this data is typically confidential and therefore only accessible under strict conditions, a key contribution of the project is to develop a flexible set of codes that can be run on the firm-level micro data in a variety of settings. These codes clean and harmonise the micro data, conduct analysis at the firm-level, and produce industry-year level datasets that contain within-industry moments of the distribution as well as aggregates on which further analysis can be conducted.

Data on five countries are used in the analysis: Croatia, France, Indonesia, Lithuania, and Sweden.[10] The data are obtained from a combination of production surveys and surveys on energy use and cover the period 1999-2021, depending on the country. All the sources used are either the population of firms or a representative sample of industry aggregates. The baseline sample for the analysis focuses on the manufacturing sector and on firms with at least 20 employees to harmonise the samples across countries.

The two key variables used throughout the analysis are labour productivity and carbon productivity. Labour productivity is defined as value added per worker. In line with this, carbon productivity is defined as value added per tonne of energy-related carbon dioxide (CO_2) emitted through the use of electricity and fuels in the production process. Methane and nitrous oxide emissions are also accounted for and are converted into CO_2 equivalents – throughout the paper, "carbon emissions" will be used to refer to total emissions of carbon dioxide, nitrous oxide, and methane. Process emissions are not available. Carbon productivity also does not include the amount of emissions embodied in intermediate inputs, just as the labour embodied in intermediate inputs is also not included in the measure of labour productivity.[11]

This section is divided into three parts. First, the sources of raw data for each of the five countries are discussed. Second, details on the data collection process are provided, including the process used to estimate CO_2 from different types of data on energy use, and the procedures to harmonise the data across countries. Third, summary statistics are provided for the firm-level datasets and for the industry-level distributive micro data.

Raw data sources

Croatia

Data for Croatia are from the Register of Annual financial reports of enterprises obtained from the Financial Agency (Financijska agencija, Fina). This is a mandatory administrative register: all legal and natural persons subject to corporate income tax must submit their balance sheets and profit and loss accounts, as well as some additional data such as on employment (GFI-POD forms).

[10] More countries will be added to the project when possible.

[11] A potentially important margin of adjustment, whereby firms outsource emissions to improve their own carbon performance, is therefore omitted. Where applicable, the value of intermediate inputs or intermediate input intensity is controlled for. A full analysis of the role of outsourcing emissions is left for future work.

The Fina data cover the whole non-financial business sector, and the statistical unit is the enterprise. An addendum to the administrative register requests voluntary information on total expenditure on energy used. Most firms provide this information, and those that do not respond are mainly very small (mostly one employee). The data cover 2002 to 2019.

France

Energy use data for France come from the Annual Survey of Energy Consumption in the Industry (EACEI) carried out by INSEE, the French statistical office.[12] The data cover the manufacturing sector and are based on a stratified sample for establishments between 20 and 250 employees; survey weights are provided. All establishments with more than 250 employees are included. From 2013 onwards, only establishments with more than 20 employees are covered (threshold of 10 employees up to 2012). The data used in the analysis cover the period 2005 to 2020.

It collects detailed information on the quantity of different energy types used, as well as the total expenditure and average price paid by type. A wide range of energy types is covered, of which electricity, natural gas, refinery gas, biofuels, heat, coal and coke.

The energy data are collected at the establishment level. They are therefore supplemented with the business register of establishments, the Répertoire des entreprises et des établissements (REE), which contains the employment of all registered establishments and the corresponding identifier of the firm.[13]

Detailed balance sheet data are obtained for the universe of firms from Fichier complet unifé de SUSE (FARE) and Fichier approché des résultats Ésane (FICUS).

Indonesia

Statistics Indonesia conducts a survey of all manufacturing establishments with more than 20 employees, although the response rate varies year-on-year. The data contain detailed information on the quantities of different energy types used, as well as financial information of the establishment. Unfortunately, firm identifiers are not available, so the data are used at the establishment level (for all variables). Additional cleaning on the raw data is carried out for Indonesia to drop a small number of establishments with extremely volatile employment. In the analysis, data between 1999 and 2015 are currently used, with post-2015 data set to be added in the next version.

Lithuania

Data for Lithuania come from the sampled part of the Structural Business Statistics. The data source is the questionnaire F-01 (annual enterprise activity report) administered by Statistics Lithuania. The survey population is comprised of all active public and private companies, state and municipal enterprises, foreign affiliates, all types of partnerships (with non-agricultural activity), individual enterprises, natural persons carrying out economic (commercial) activity (included from 2005) and public bodies covering the major share of their expenses with income from selling goods and providing services. The economic activities of the enterprises surveyed cover sections B-J, L-N and P-R and divisions A02, A03, S95 and S96 of EVRK Rev. 2, and data are collected at the firm-level. The variable used is total energy expenditure, and data between 2004 and 2020 are used.

[12] https://www.casd.eu/en/source/survey-on-energy-consumption-in-industries/

[13] https://www.insee.fr/en/metadonnees/source/serie/s1346

As survey weights are not available in the raw data for all years, estimated survey weights are constructed based on the stratification system of the data collection method. These weights are refined by trying to match population aggregates on variables that are observed for all firms (such as employment).

Sweden

Data for on energy use by type for Sweden is obtained from Industrins energianvändning (ISEN), collected by the Swedish Energy Agency. The data cover all manufacturing firms with at least 10 employees, over the period 2005-2021. The data is collected at the plant level and converted to the firm-level by the data owners. The data can be matched with Structural Business Statistics to obtain balance sheet information.

Data collection procedure

The data collection procedure is flexibly designed to utilise confidential data on firms' environmental and production activities from varying types of sources across countries. This project builds on similar procedures that have been developed at OECD to collect data on productivity (Berlingieri et al., 2017[5]), business dynamics (Calvino, Criscuolo and Verlhac, 2020[52]), AI adoption (Calvino and Fontanelli, 2023[53]), R&D (Appelt et al., 2022[54]), and employment reallocation (Calligaris et al., 2023[55]). Data is provided at the firm-year or establishment-year level, before a set of harmonised cleaning, estimation, and aggregation procedures are conducted.

Multiple data sources may be used within each country to obtain different variables or to help with harmonisation. Three main types of data can be included in the module: an energy use survey; a production survey; and a business register. It must be possible to match the data sources at the firm-level. Note that for each country it is possible to match every firm with energy data to its production data, either because they are obtained from the same source (Croatia, Lithuania, Indonesia) or because the production data is available for the census of firms (France and Sweden). Importantly, this means that there is no need to recompute sampling weights to maintain representativeness. The possible uses of each source are described in this section.

The data construction begins by conducting basic checks and cleaning on each data source provided. These include, for example, confirming the monetary units and basic properties of the data, checking for negative values in variables that should be non-negative, and checking for missing values. It is up to the discretion of the data owner to assign whether non-negative and missing values of each variable should be left as they are (which, for missing values, means they are effectively dropped), or replaced as 0. Industry codes are converted to ISIC Revision 4 and any changes in the classification system throughout the period are converted using an internally-constructed crosswalk that exploits the panel data structure of the data.

It processes the environmental data in numerous ways. First, if any information on energy use is available, it computes estimated emissions (see below). Second, if the environmental data is at the facility or establishment level, it converts it to the firm-level, scaling appropriately, such that all final datasets are at the firm-level. This process also includes a procedure to convert survey weights at the plant level to the firm-level while maintaining aggregate representativeness.

All data sources are then combined, choosing the preferred variables from each source based on expected data quality in the cases where variables (e.g. employment or industry codes) are available from multiple sources. Sample cleaning is conducted, including deflation. If a business register is available, it is used to construct strata and associated weights, though these would not be chosen as the baseline if sample weights are already available. Variables such as value added per CO_2 emissions are constructed. Suspicious outliers – most likely to be data entry errors - are detected using a new methodology that combines comparisons of within-firm variation over time with contemporaneous differences in key variables

within the firm. Particular attention is made to avoid over cleaning of genuine frontier and laggard firms. Remaining severe outliers are detected using cross-sectional variation, which is especially useful if panel data is not available. This cross-sectional detection has two main components. First, if the level of a key variable (value added, output, energy spending, carbon) is more than 15 times the difference between the within-industry 75th and 25th percentile values (Tukey method) *and* the ratio of the variable to employment is also 15 times bigger, the observation is dropped. By requiring both these conditions, the aim is to avoid dropping firms that are simply excessively large. Second, on the key ratio variables – value adder per worker and value adder per carbon – observations with extreme outlier (following the same criteria) are dropped. All monetary variables are deflated and converted to US Dollars using deflators from OECD STAN at the 2-digit industry-level. Analysis can then be conducted on the cleaned firm-level data, such as density plots and regressions. See Annex A for more details.

Finally, the data is aggregated to the industry-year level, taking moments of the firm distribution within each cell. Confidentiality checks are applied to the aggregated data before returning the data to the OECD. The final outputs can be produced at different levels of industry aggregation and may be further disaggregated according to firm characteristics such as age, employment, energy efficiency, or productivity.

Throughout the process, diagnostic statistics, notes, and detailed meta data are collected to ensure that the procedure is running correctly and there are no major issues with the process. Figure A A.5 shows the data collection procedure diagrammatically.

The remainder of the sub-section provides more detail on the estimation of CO_2 emissions from energy use data. In Annex A, details are provided on the method to convert industry codes, external data sources used to compute emissions, conversion of establishment to firm-level data (used for France only), detection and treatment of extreme outliers (used to try to remove data entry errors), and confidentiality requirements.

Estimating CO_2 emissions

Emissions from production can be computed from different variables, depending on data availability. The most accurate data from which production emissions can be computed is the quantity of energy used by energy type (electricity, gas, coal, etc). In this case, emissions factors for each energy type, collected from the IEA, are used to obtain the volume of CO_2 emitted, as well as other greenhouse gases which are converted into CO_2 equivalents and referred to in their totality as carbon throughout for simplicity. Quantity data by energy type is available for France, Indonesia, and Sweden.

In contrast, if only the total spending on all energy is available, publicly available data on energy type usage by industry and energy prices, collected from IEA and Eurostat, must be used alongside the emissions factors to compute estimates of production emissions at the firm-level. These data are available for Croatia and Lithuania.

The following sub-sections describe the procedure to estimate production emissions under each type of data availability, and discuss the assumptions required in each case. A comparison of the emissions estimated from total energy spending and from energy use by type is provided using data for France, where both are available. Annex A provides details on the external datasets used to compute emissions as well as the cleaning and imputation procedures used to prepare them. It also provides details on how a continuous relationship for prices as a function of quantity consumed is estimated from available data, including evidence on the accuracy of this procedure.

Estimating CO_2 emissions with energy use by fuel type data

When data on quantity of energy by type is available, it is straightforward to compute production emissions. Quantities are converted to tera joules (TJ), then emissions factors from TJ to tonnes of CO_2 equivalent

are used. This is done for each energy type, then total emission is simply the sum of emissions from each energy type.

Estimating CO_2 emissions when only total energy spending is available

If only total energy spending is available, emissions can be estimated under certain assumptions, described below and in more detail in Annex A, using external datasets. The procedure takes three main steps: 1) use the industry of the firm to match with energy use by type of fuel, available at the industry-year level (either at the 2, 3, or 4 digit level depending on the level of detail in the energy use data); 2) match energy types with their prices (which can vary by quantity consumed) and emissions factors in each country-year; 3) combine firm-level spending with the external data sources to compute estimates of emissions.

Step 3 can be conducted using either prices that do not vary with the quantity of energy consumed (i.e. they are fixed at the fuel-country-year level), or prices that vary according to an estimate of the quantity of the fuel that is consumed. When using prices that do not vary with quantity consumed, the process is straightforward: total energy spending is divided by an industry-country-year level energy price index. However, in practice, prices paid for energy vary with quantity of energy purchased, and some data is available on this. These data are used to measure emissions more accurately by developing an algorithm that numerically solves for firms' quantity of each type of energy purchased, using information on their total size and estimated parameters on the price of each energy type as a function of quantity consumed. The procedure is described in detail in Annex A and is used as the baseline estimate of CO_2 when only total energy spending is available.

Robustness of estimating emissions from total energy spending

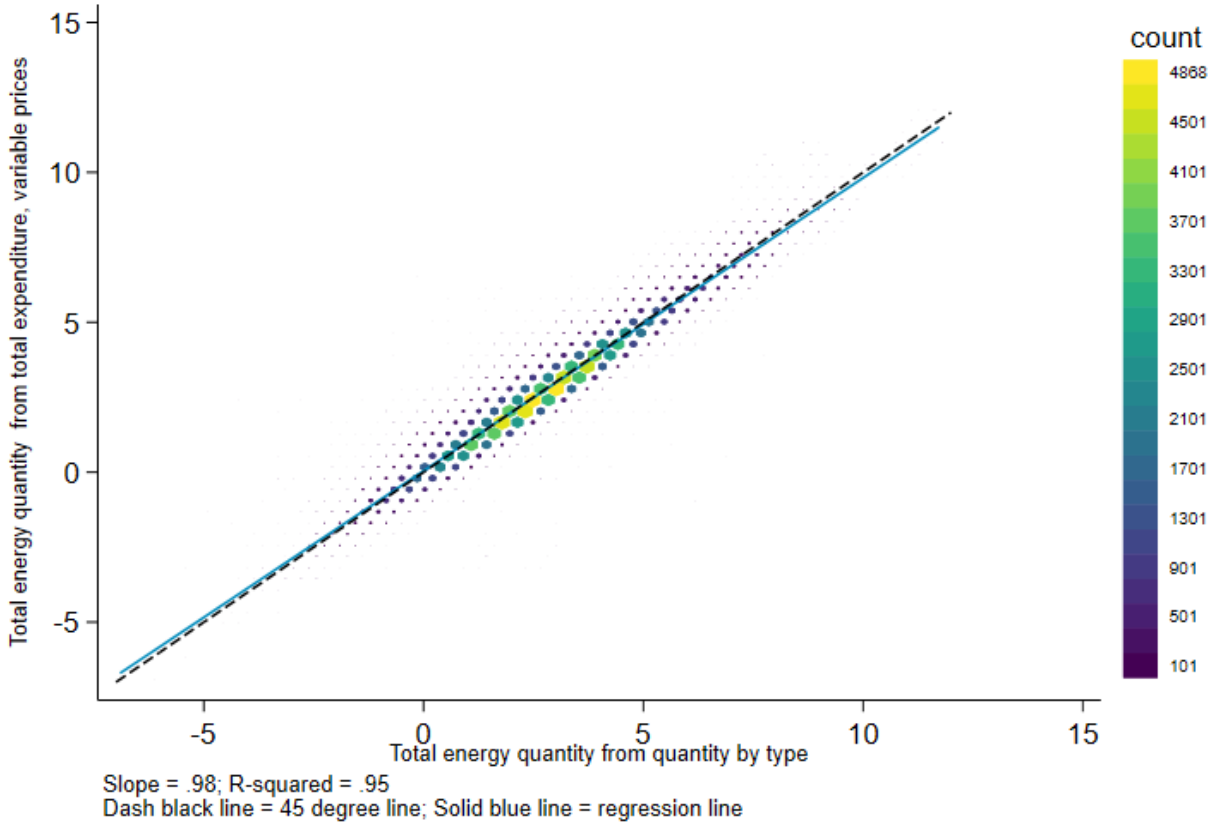
Using French data, which contains both total energy spending and quantity used of different fuels, it is possible to test the accuracy of the procedures to estimate CO_2 emissions from only data on total energy spending. Figure 2.1 shows that there is a close relationship between the log of estimated quantity from total energy spending (y axis) and the log of actual spending (x axis) in the French data, where data on energy quantity and energy spending by type are available. The slope is 0.98, close to the 45-degree line, and the data points are close to the estimated line, with an R-squared of 0.95. It is important to account for variable energy prices: when only using average prices, the slope of the relationship is 0.89, reflecting that larger consumers pay lower prices for energy.

The relationship for emissions is close but less accurate than for quantity, as shown in Figure 2.2. This is driven by mismeasurement of quantity shares across energy types because there is a good fit for both energy quantity and energy emissions when emissions are computed using spending by type (not discussed here for brevity). The slope is 0.89, compared with 0.81 when only using prices that do not vary with quantity consumed. This suggests that firms that consume a large quantity of energy tend to be more likely to use more polluting energy types than the average firm, and vice versa for small firms. It also suggests that the price per joule is similar across energy types, while emissions vary more by energy types.

Note that the slope lower than one for emissions means that emissions are slightly underestimated for large consumers of energy and slightly overestimated for small energy consumers in countries where only total energy spending is available (Croatia and Lithuania). This would lead to an underestimate of dispersion in emissions and, because emissions are negatively correlated with carbon productivity, to an underestimate of dispersion in carbon productivity. This is shown in Figure A A.2.

Figure 2.1. Energy quantity is similar to true quantity when estimated using total energy spending

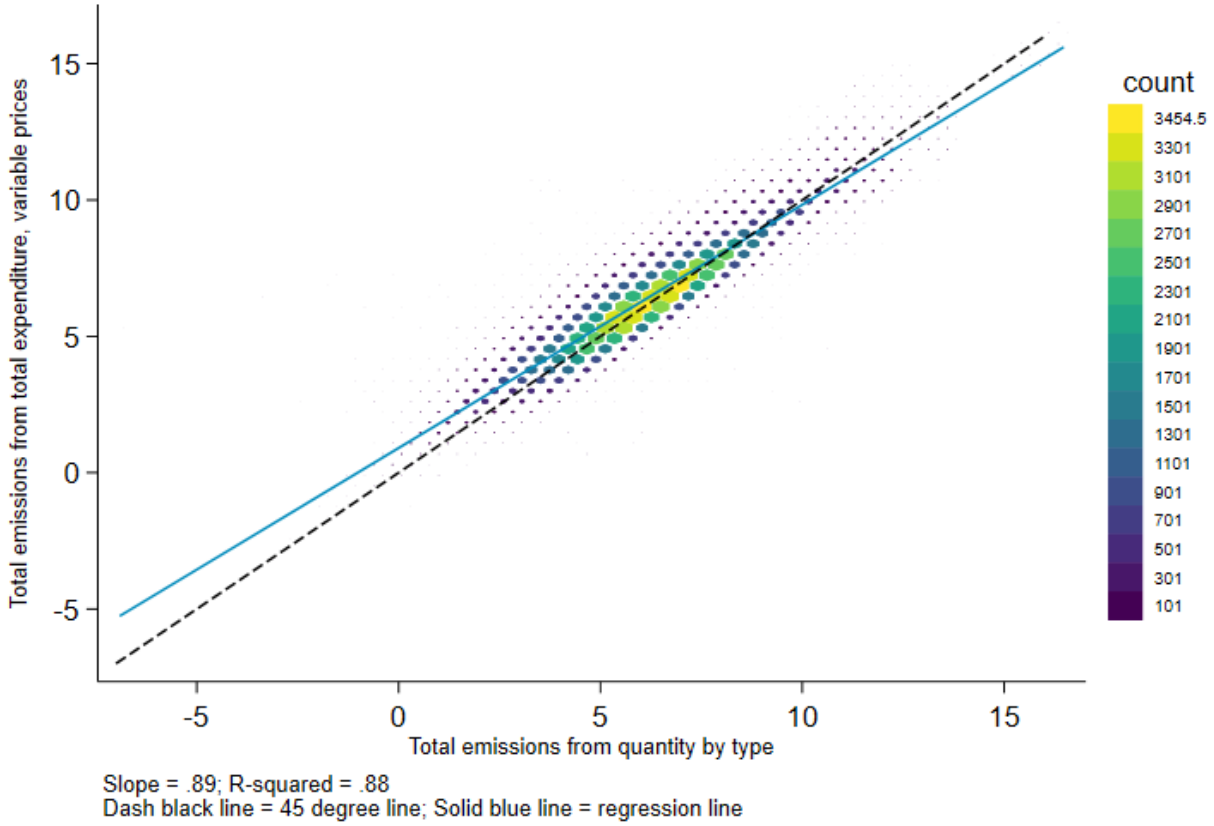
Log quantity of energy from total energy spending against log quantity of energy from data on energy by type, establishment level data for France



Note: The chart plots the log of total energy quantity estimated using only total expenditure on energy (y axis) against the total energy quantity estimated from detailed information on quantities used of different fuel types and electricity (x axis). The dashed black line is the 45-degree line while the blue line is the fitted OLS regression estimate from y on x. The heat map shows the underlying data points where the size and colour of the areas represent the number of underlying data points. The data is establishment level information from France pooled over 2002 to 2020.
Source: authors' calculations.

Figure 2.2. Energy emissions is similar to true emissions when estimated using total energy spending

Log emissions from total energy spending against log emissions from data on energy by type, establishment level data for France



Note: The chart plots the log of total emissions estimated using only total expenditure on energy (y axis) against the total emissions estimated from detailed information on quantities used of different fuel types and electricity (x axis). The dashed black line is the 45-degree line while the blue line is the fitted OLS regression estimate from y on x. The heat map shows the underlying data points where the size and colour of the areas represent the number of underlying data points. The data is establishment level information from France pooled over 2002 to 2020.
Source: authors' calculations.

Summary statistics

Industry summary statistics

Table 2.1 presents summary statistics on the distributed micro data at the country-industry-year (cell) level. The main sample used throughout the analysis is restricted to firms with at least 20 employees in 3-digit manufacturing industries. The within-cell percentiles and dispersion measures are only computed in cells with at least 10 firms; hence the smaller cell counts in these cells.[14] Some cells must also be blanked for confidentiality reasons, which can be variable-specific, so cell counts vary slightly across variables. The mean, standard deviation, minimum, maximum, percentiles, and count of cells are provided for each

[14] Cell counts for France are not available – only the sample weighted cell counts can be released, so these are used instead. This explains why the minimum dispersion is 0 – it is computed on a cell where the 10th and 90th percentile firm is the same.

variable. Note that the columns provide different statistics *across* industry-country-years (e.g. the median industry-country-year has value added of 161 million), while some variables are specific moments of the distribution *within*-industry-country-years (e.g. the mean of the 10th percentile firm's log value added per CO_2 within an industry-country-year is 5.46).[15]

Industries vary in terms of their overall size, as measured by the count of firms, total employment, and total gross output. The distributed micro data approach provides cross-country data on within-industry moments of the firm distribution, providing information on dispersion of firms conducting similar production activities. The variables beginning with [p10], [p50] and [p90] present the within-industry moments for both log value added per CO_2 (carbon productivity) and log value added per worker (labour productivity).

These variables can also be used to measure within-industry dispersion, measured as the log of the 90th percentile minus the log of the 10th percentile of a variable within an industry. In the mean country-industry-year, the 90th percentile firm has 24 times (3.18 log points) higher carbon productivity than the 10th percentile firm. The average dispersion in labour productivity is smaller, at five times higher (1.71 log points). There is a lot of variation across industries in the extent of dispersion in both carbon productivity and labour productivity. These patterns are explored in detail in the next section.

[15] Note that there are a small number of cases where value added is negative. These will be excluded from any analysis that takes logs.

Table 2.1. Country-industry-year level summary statistics

	mean	sd	min	p10	p25	p50	p75	p90	max	count
Aggregate statistics computed for all cells										
Firm Count	123	295	1	2	6	22	99	330	3296	5525
[sum] Employment	19650	45600	20	236	966	4314	16265	50613	526714	5075
[sum] Value Added (millions)	866	1700	-141	4	27	161	809	2713	18040	4729
Dispersion statistics computed on cells with at least 10 firms										
[p10] Log Value added / CO_2	7.66	1.68	2.57	5.46	6.31	7.64	8.95	9.81	14.96	3505
[p50] Log Value added / CO_2	9.21	1.57	5.53	7.24	7.95	9.08	10.39	11.23	14.96	3505
[p90] Log Value added / CO_2	10.85	1.63	6.51	8.87	9.59	10.64	11.99	13.06	16.26	3505
[p10] Log Value added / labour	9.37	1.51	4.89	7.06	8.18	9.66	10.68	10.95	12.93	3544
[p50] Log Value added / labour	10.25	1.15	6.23	8.47	9.48	10.49	11.16	11.43	13.59	3544
[p90] Log Value added / labour	11.09	0.94	7.90	9.67	10.52	11.31	11.73	12.12	13.65	3544
log(VA/CO_2) p90-p10	3.18	0.96	0.00	1.97	2.54	3.19	3.80	4.37	7.30	3505
log(VA/L) p90-p10	1.71	0.96	0.00	0.78	1.00	1.37	2.29	3.23	5.70	3544

Note: The table presents summary statistics on the industry-year level data for 3-digit ISIC industries in the manufacturing sector. Only firms with at least 20 employees are included in the sample. The statistics are unweighted across industries (but weighted by sample weights to the industry-level in France and Lithuania), and cover Croatia, France, Indonesia, Lithuania, and Sweden. The variables beginning [p10], [p50], and [p90] represent within-industry percentiles of those variables while the variables p90-p10 measure within-industry dispersion. These within-industry percentiles and dispersion measures are only computed for cells with at least 10 firms.
Source: authors' calculations.

Firm summary statistics

It is also important to examine summary statistics on the underlying firm-level data for each country used in the project. These data underpin the industry-level micro-distributed datasets and are used in the firm-level analysis such as the regressions in Section 5. In particular, given that the project uses some micro data that cannot be directly accessed by the researchers, the summary statistics provide another important check on the consistency of the data within and across countries.

The following tables show the firm-level summary statistics for France (Table 2.2), Croatia (Table 2.3), Lithuania (Table 2.4), Indonesia (Table 2.5), Sweden (Table 2.6). Reassuringly, these key variables exhibit broadly similar distributions across countries, with some differences to be expected. For example, the average share of energy costs in total costs (across all sectors) is similar across the European countries and higher in Indonesia, while average labour and carbon productivity are highest in Sweden – the country with the highest GDP per capita of the countries included in the sample.

The firm-level summary statistics are also indicative of the cross-country differences in the extent of dispersion in carbon productivity across firms. The 90th percentile firm is significantly more productive than the 10th percentile firm: this difference ranges from a factor of 22 (2.93 log points) in Croatia to a factor of 77 (4.34 log points) in Indonesia.[16]

Table 2.2. Firm-year-level summary statistics for France

	mean	sd	p10	p50	p90	count
Employment	141	1 171	23	45	237	122 290
Value Added (millions)	11.1	70.6	1.0	2.8	17.0	122 290
CO_2 emissions (tonnes)	3 948	117 753	12	123	1 962	122 290
Log Value Added / CO_2	10.06	1.67	8.00	10.11	12.03	121 155
Log Value Added / labour	10.87	0.76	10.15	10.91	11.60	121 154
Firm age (years)	27.47	19.51	7	24	51	122 210
Energy spending / intermediate inputs	0.11	5.49	0	0.02	0.09	122 055
Energy spending / total costs	0.08	5.44	0	0.01	0.05	122 066
Intermediate input costs / total costs	0.68	0.15	0.5	0.7	0.86	122 066
Labour costs / total costs	0.32	0.15	0.14	0.3	0.5	122 066
Electricity quantity share (share of joules)	0.67	0.3	0.25	0.69	1	122 290

Note: The table reports summary statistics of firm-level data for manufacturing firms with at least 20 employees in France. The statistics are weighted by the sample weight.
Source: authors' calculations

Table 2.3. Firm-year-level summary statistics for Croatia

	mean	sd	p10	p50	p90	count
Employment	127	330	22	48	258	29 124
Value Added (millions)	5.4	33.9	0.3	1.3	8.5	29 124
CO_2 emissions (tonnes)	2 168	20 048	32	203	2 832	28 549
Log Value Added / CO_2	9.10	1.27	7.66	9.06	10.59	28 332
Log Value Added / labour	10.48	0.80	9.58	10.46	11.45	28 906
Firm age	14.17	8.6	3	13	25	22 765
Energy spending / intermediate inputs	0.1	0.13	0.02	0.06	0.25	29 124
Energy spending / total costs	0.05	0.05	0.01	0.04	0.1	29 124
Intermediate input costs / total costs	0.65	0.23	0.27	0.71	0.89	29 124
Labour costs / total costs	0.35	0.23	0.11	0.29	0.73	29 124

Note: The table reports summary statistics of firm-level data for manufacturing firms with at least 20 employees in Croatia. The statistics are unweighted and cover close to the population of firms.
Source: authors' calculations

[16] Note that these statistics are computed on firm-level data whilst the statistics on dispersion earlier in the section are the average of industry-country-year dispersions. Therefore, the two statistics effectively have different weighting.

Table 2.4. Firm-year level summary statistics for Lithuania

	mean	sd	p10	p50	p90	count
Employment	127	205	26	65	270	15 593
Value Added (millions)	2.8	14.8	0.2	1.0	5.4	15 593
CO_2 emissions (tonnes)	1 896	23 383	21	138	1426	15 002
Log Value Added / CO_2	9.36	1.38	7.66	9.35	11.09	14 854
Log Value Added / labour	10.12	0.85	9.06	10.13	11.16	15 443
Firm age	13.03	6.44	5	13	22	15 593
Energy spending / intermediate inputs	0.08	0.16	0.01	0.04	0.18	15 587
Energy spending / total costs	0.05	0.06	0.01	0.03	0.1	15 593
Intermediate input costs / total costs	0.72	0.18	0.47	0.76	0.9	15 593
Labour costs / total costs	0.28	0.18	0.1	0.24	0.53	15 593

Note: The table reports summary statistics of firm-level data for manufacturing firms with at least 20 employees in Lithuania. The statistics are weighted by the sample weight.
Source: authors' calculations

Table 2.5. Establishment-year level summary statistics for Indonesia

	mean	sd	p10	p50	p90	count
Employment	195	720	21	45	406	370 723
Value Added (millions)	4.4	33.5	0.0	0.2	6.6	370 723
CO_2 emissions (tonnes)	3 766	48 223	6	130	4112	370 723
Log Value Added / CO_2	9.34	1.74	7.23	9.28	11.57	370 203
Log Value Added / labour	10.06	1.43	8.30	10.01	11.89	370 206
Firm age	10.33	8.48	1	8	23	370 723
Energy spending / intermediate inputs	0.53	0.3	0.11	0.54	0.94	370 723
Energy spending / total costs	0.21	0.2	0.02	0.14	0.5	370 723
Intermediate input costs / total costs	0.41	0.26	0.09	0.38	0.8	370 723
Labour costs / total costs	0.59	0.26	0.21	0.63	0.91	370 723
Electricity quantity share (share of joules)	0.75	0.43	0	1	1	350 294

Note: The table reports summary statistics of establishment level data for manufacturing establishments with at least 20 employees in Indonesia.
Source: authors' calculations

Table 2.6. Firm-year level summary statistics for Sweden

	mean	sd	p10	p50	p90	count
Employment	137	610	22	45	227	51188
Value Added (millions)	16.4	123.0	1.3	3.5	21.9	51188
CO2 emissions	2482	37515	9	87	1381	51188
Log Value Added / CO2	10.71	1.63	8.71	10.69	12.77	50789
Log Value Added / labour	11.24	0.49	10.75	11.22	11.80	50788
Firm age	22.36	8.89	9	23	33	51188
Energy spending / intermediate inputs	0.03	0.05	0	0.02	0.07	51125
Energy spending / total costs	0.02	0.03	0	0.01	0.05	51166
Intermediate input costs / total costs	0.76	0.12	0.61	0.78	0.9	51176
Labour costs / total costs	0.24	0.12	0.1	0.22	0.39	51176
Electricity quantity share	0.7	0.26	0.34	0.74	1	48591

Note: The table reports summary statistics of firm-level data for manufacturing firms with at least 20 employees in Sweden.
Source: authors' calculations

3 Dispersion in carbon productivity and labour productivity

This section explores the patterns of within-industry dispersion in carbon and labour productivity. It first compares the extent of dispersion in each measure across countries. It then carries out a counterfactual exercise showing how much CO_2 emissions could be reduced by increasing the efficiency of the least productive firms. Next, it explores the relationship between changes in dispersion and aggregate outcomes. Trends in percentiles of the distribution of carbon productivity and its dispersion are plotted, as well as a detailed portray of carbon productivity distributions in 2-digit manufacturing industries. Finally, both levels and changes of dispersion are correlated with other industry characteristics.

Dispersion in carbon productivity between firms

There is more within-industry dispersion in carbon productivity than labour productivity in all countries examined. On average across industries and countries, the 90th percentile firm is 24 times more carbon productive than the 10th percentile firm in the same industry. In comparison, for labour productivity the equivalent difference is five times.[17] Therefore, on average, there is five times more dispersion in carbon productivity than for labour productivity. The differences in the extent of dispersion are most pronounced in France, and least pronounced in Indonesia.

These findings are in line with Lyubich, Shapiro and Walker (2018[4]) who make this comparison for US firms, finding that the 90th percentile firm is 10 times more carbon productive than the 10th percentile firm in the US. Possible explanations for this extensive dispersion and the greater dispersion in carbon productivity are examined and discussed later in this section.

Figure 3.1 shows this dispersion by plotting the distribution of firms' carbon and labour productivity within 4-digit industries for each of the five countries in 2015 (the latest year for which data is available for all countries).[18] Each distribution is centred around 0.

This finding of high dispersion in carbon productivity – greater than in labour productivity – is highly robust. Figure A B.1 plots the distributions within even more finely defined cells, which additionally split firms into five employment bands. That is, it controls for broad differences in firm size even within industries. Similarly, Figure A B.2 plots the distributions within-firm age bands, finding similar patterns. The key finding

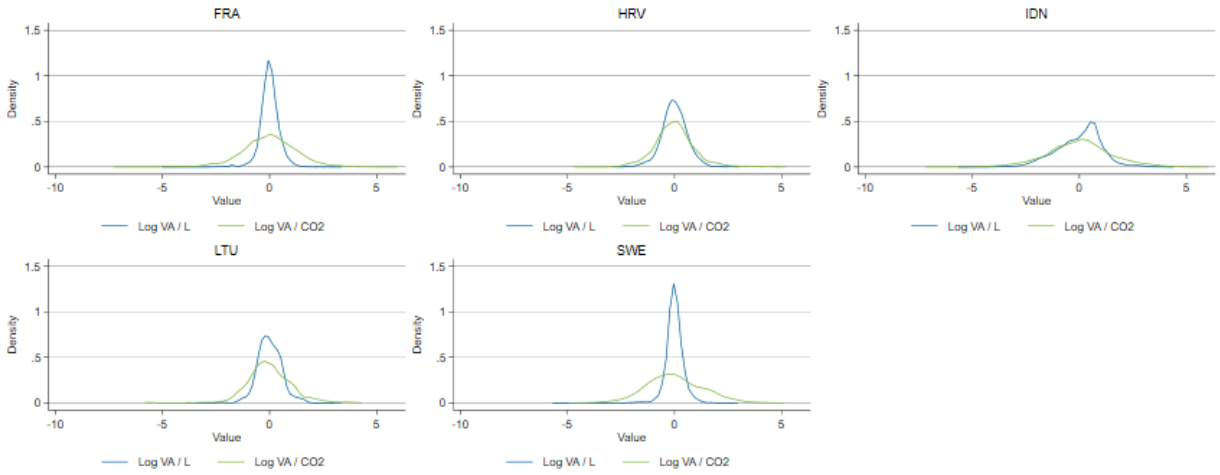
[17] This statistic is estimated by taking the average of log dispersion across industries and then computing the implied average difference in productivities. This use of logs is preferred to reduce the influence of large outliers in dispersion in the industry level data. The same methodology is used in Lyubich, Shapiro and Walker (2018[4]). Simply taking the raw average difference between the 90th percentile and 10th percentile firms carbon productivity yields an average ratio of 40 (compared with 19 for labour productivity).

[18] Note that it is possible to use 4-digit industries for the density plots as they can be computed at the firm level and produced for each country separately, i.e. they do not require the use of cross-country data. The sample is restricted to manufacturing sector firms with at least 20 employees.

that there is more dispersion in carbon productivity than labour productivity also holds when including firms with fewer than 20 employees, and when looking at the services sector (these data are only available for Croatia and Lithuania, and the results are omitted for brevity). It also holds for 2- and 3-digit industries, addressing any concerns that this may be due to small sample sizes.

Figure 3.1. There is more dispersion in carbon productivity than labour productivity

Distributions of carbon productivity and labour productivity across firms within 4-digit industries in 2015



Note: The charts plot the within-industry-year distributions of carbon and labour productivity across firms in France, Croatia, Indonesia, Lithuania, and Sweden. Industries are defined at the 4-digit level and distributions are centred at 0. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations.

One possible explanation for this dispersion could be that firms in the same industry are producing different products. To investigate this, detailed data on firms' sales by product for French firms is obtained from the Enquête annuelle de production (EAP), covering 2009 to 2020. This is matched with firm-level measures of carbon and labour productivity. For this exercise, firms are assigned to the product in which they have the highest share of sales. As robustness checks, firms with their main product accounting for less than 50% or 80% of their total sales are excluded to ensure that firms producing the same products are being compared. Measures of dispersion – the difference between the 90th and 10th percentile firm – are then computed within a product or industry at different levels of granularity.

The results show that, even comparing firms producing the same 8 digit product, there is huge dispersion in carbon productivity. Table 3.1 shows the average dispersion (across years) in carbon and labour productivity within 2, 3, and 4 digit industries, and within 6 and 8 digit products.[19] As expected, when the industry or product is more narrowly defined, the extent of dispersion decreases. For carbon productivity, the 90th percentile firm is 20 times (3.0 log points) more productive than the 10th percentile firm in 4 digit industries, compared with 15-16 times (2.7-2.8 log points) in 6 digit products, and 14 times (2.6 log points) in 8 digit products. In other words, even when comparing firms producing the same 8 digit product in the same country (France), there are huge differences between firms' carbon productivity and this difference

[19] The results are similar when looking at specific years and when taking the weighted average across products or industries, with value added used as the weight.

is far greater than the dispersion in labour productivity, for which the 90th percentile firm is only 3 times more productive than the 10th percentile firm.

Table 3.1. Firms producing the same product have very different carbon productivity

Average carbon and labour productivity in France within different definitions of products and industries

Product or industry-level	Dispersion in log carbon productivity	Dispersion in log labour productivity
2 digit industry	3.42	1.16
3 digit industry	3.17	1.16
4 digit industry	3.02	1.13
6 digit product (highest firm share)	2.79	1.09
6 digit product (firm share >50%)	2.80	1.09
6 digit product (firm share >80%)	2.73	1.08
8 digit product (highest firm share)	2.64	1.05
8 digit product (firm share >50%)	2.65	1.06
8 digit product (firm share >80%)	2.61	1.07

Note: The table shows dispersion in carbon and labour productivity within different definitions of products and industries. Carbon productivity is defined as value added per CO_2 and labour productivity is defined as value added per worker. Dispersion is measured as the log difference between the 90th and 10th percentile firms within the product or industry. Industries are measured at either the 2, 3, or 4 digit levels. Products are measured at either the 6 or 8 digit level. Firms are assigned to a single product depending on the one that accounts for the largest share of sales. Firms for which the most important product accounts for less than 50% or 80% of sales are excluded in some rows (as indicated in the first column). The average dispersion is taken across industries or products and across years.
Source: authors' calculations

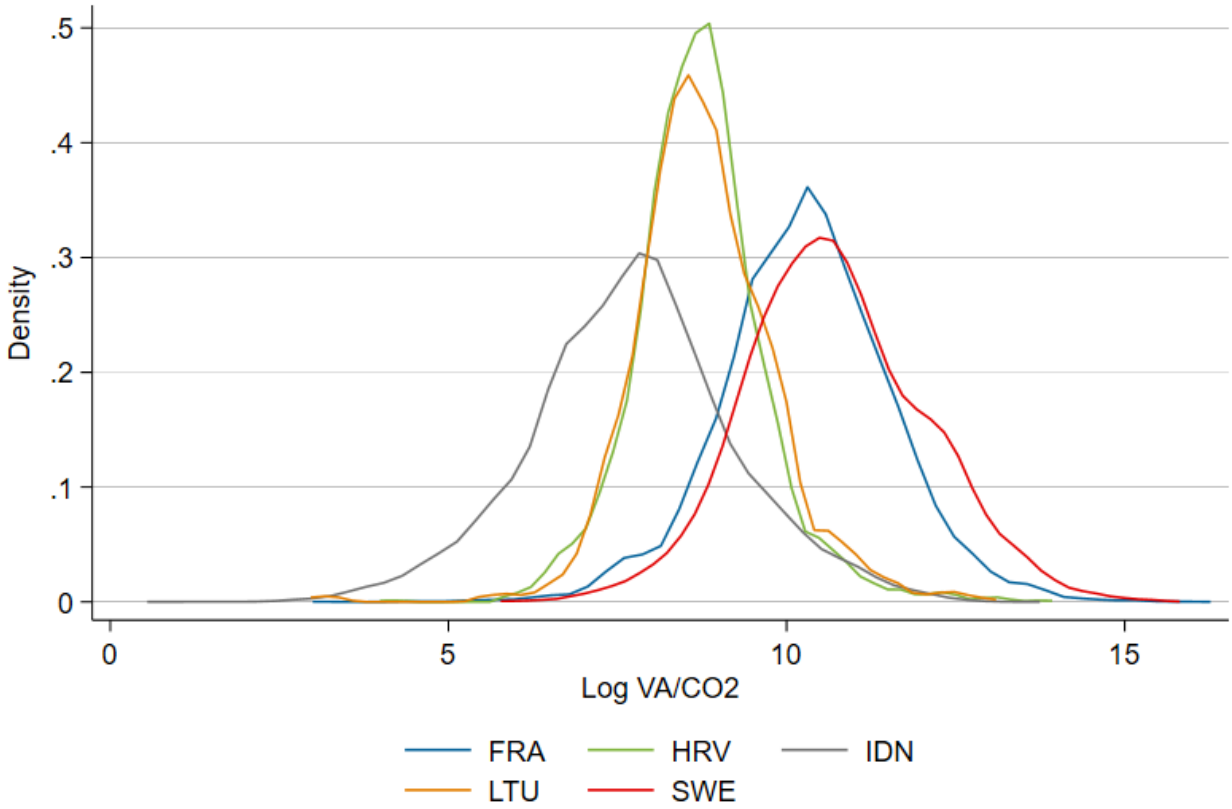
There are also substantial differences in the distributions of carbon productivity across countries. Figure 3.2 plots the distributions of value added per CO_2 for each country in 2015. The within-industry density is taken to examine the differences of firms compared with other firms in their industry, then the population mean of carbon productivity is added to the plot to show differences in absolute magnitudes of carbon productivity across countries.

The chart reveals interesting patterns of the carbon productivity distribution across countries. French and Swedish firms tend to be the most carbon productive while Indonesian firms tend to be least productive. Despite this, the distributions are overlapping, meaning that some of the most productive Indonesian firms are still more productive than the least productive French and Swedish firms, although these firms may of course be in different sectors so this comparison should be interpreted with some caution.

These results are qualitatively robust to using only total energy spending to estimate CO_2 emissions for all countries (instead of detailed data on energy quantities by type for France, Indonesia, and Sweden). This is shown in Figure A B.3. As discussed in the previous section, using only total energy spending leads to an underestimate of dispersion in CO_2 and CO_2 productivity. Despite this, the overall pattern observed in Figure 3.2 is preserved when using total energy spending in all countries.

Figure 3.2. Swedish and French firms are more productive on average but there are many highly productive Indonesian firms

Distributions of carbon productivity across countries in 2015



Note: The charts plot the within-industry-year distributions of carbon productivity across firms in France, Croatia, Indonesia, Lithuania, and Sweden. Industries are defined at the 4-digit level and distributions are recentred by adding the population mean of each country. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations.

Dispersion over time and across industries

This section examines trends in carbon productivity at various percentiles of the within-industry distributions over time. Averaging across industries, there have been general improvements in carbon productivity over time at various key percentiles of the distribution (10th, 50th, and 90th), notably in Lithuania, Indonesia, and Sweden.[20] This is shown in Figure 3.3, where industries are defined at the 3 digit level due to data availability.[21]

Figure 3.4 presents changes in dispersion over time more clearly, by plotting the average ratio between the carbon productivity of the 90th and 10th percentile firms within an industry. To allow for an interpretation of differences between firms in ratios, the exponential of average log dispersion is taken and plotted. There

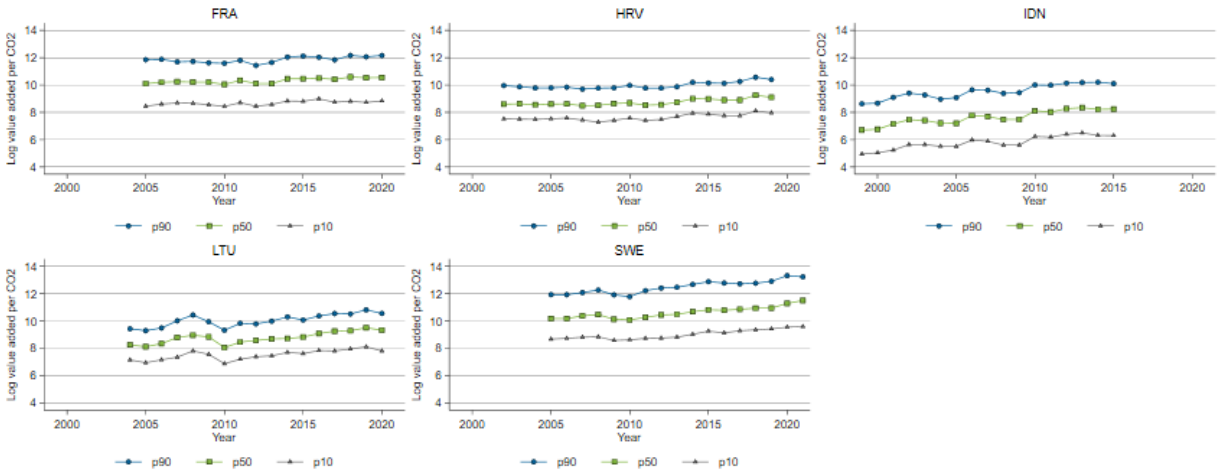
[20] Figure A B.5 and Figure A B.6 in the annex show that these results are robust to weighting across industries.

[21] The equivalent trends in labour productivity are presented in Figure A B.4. It is again clear that there is less dispersion in labour productivity than in carbon productivity. There has been little change over time in labour productivity at each percentile in France and Croatia, while it has generally improved in Indonesia and Lithuania.

has been no reduction in within-industry dispersion in carbon productivity across countries. In Sweden, there has been a significant rise in dispersion throughout the period, and in France and Lithuania there has been a rise in dispersion towards the end of the sample period. In Indonesia and Croatia, dispersion has been relatively flat over the entire period.

Figure 3.3. Carbon productivity has improved across all percentiles of the distribution in each country

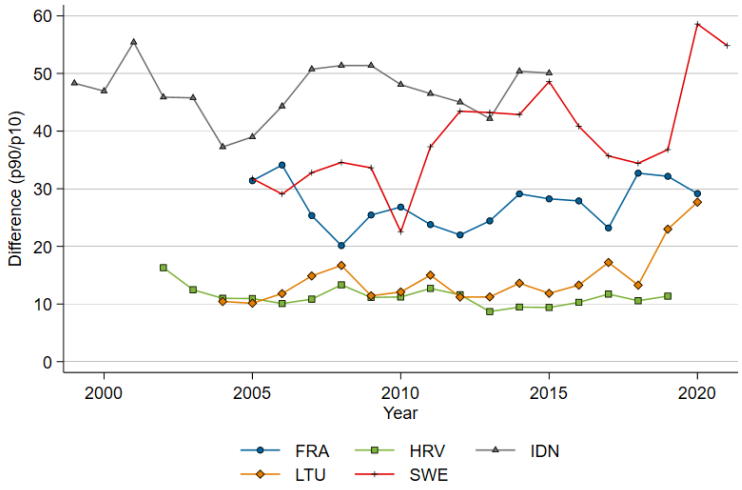
Carbon productivity at the 10th, 50th, and 90th percentiles within 3-digit industries over time



Note: The charts plot the 10th, 50th, and 90th percentile of within-industry carbon productivity over time. Industries are defined at the 3-digit level and the unweighted average across industries is taken. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations

Figure 3.4. Dispersion in carbon productivity has been stable or rising over time

Difference between carbon productivity at the 90th and 10th percentile within industries



Note: The charts plot the difference between carbon productivity at the 90th and 10th percentile within industries. The average of log dispersion is taken, then the exponent of this is computed so that dispersion can be interpreted as the ratio difference between the 90th and 10th percentile firm. Industries are defined at the 3-digit level and the unweighted average across industries is taken. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations.

Looking at aggregates may hide interesting information on the distributions in specific sectors. Figure 3.5 plots percentiles of carbon productivity for 2-digit manufacturing sectors in 2005 (the dots) and 2015 (the arrows). The percentiles are computed at the 3-digit level to account for more detailed differences in production and then averaged to the 2-digit level. The sample is restricted to industries for which at least 10 firms are observed in each of the five countries in the sample in each year. The unweighted average is then taken across the countries.

The carbon productivity of firms at different percentiles of the distribution within industries differs across industries. The extent of dispersion also differs significantly across industries. Some industries, such as rubber and plastic products, have a smaller difference between the carbon productivity of the 10th and 90th percentile firms than others, such as non-metallic minerals. Later in this section, the characteristics of industries with more dispersion are explored.

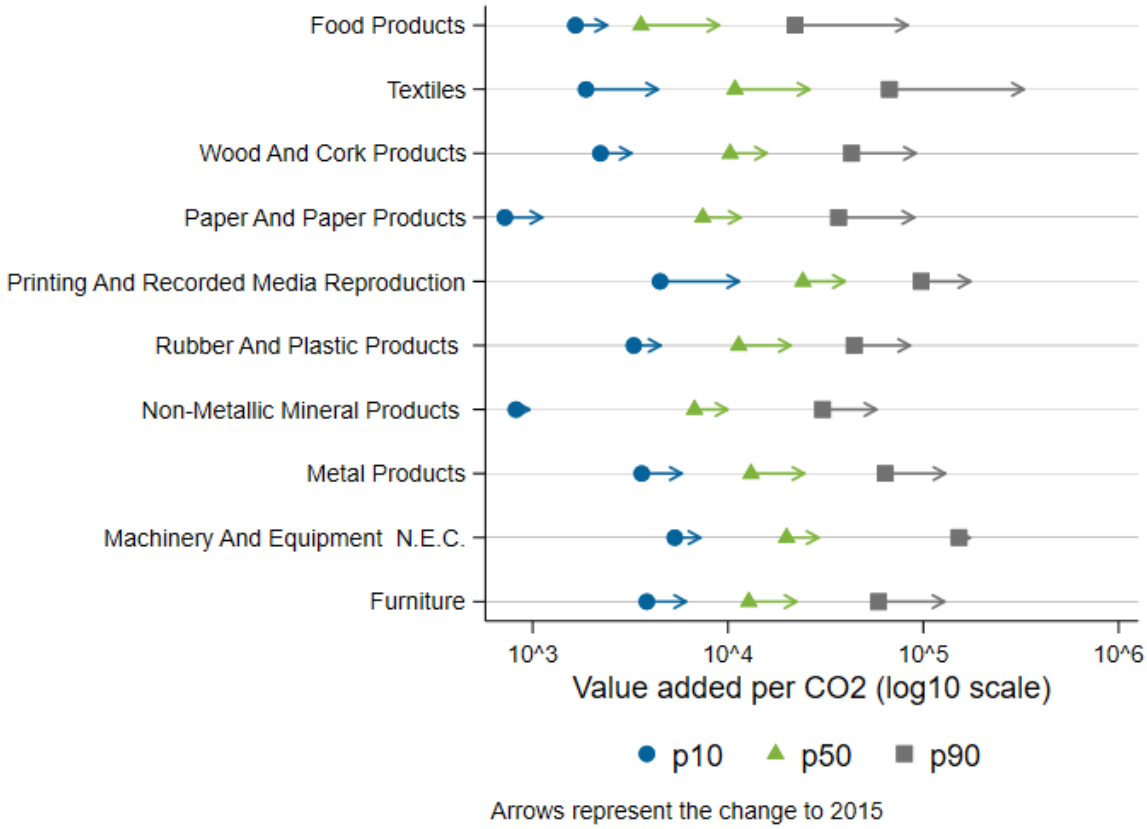
There are also substantial differences across industries in the change in carbon productivity at different percentiles of the distribution over the sample period. Some industries have experienced an improvement in carbon productivity across the distribution, such as metal products, while some have improved more at the top than at the bottom of the distribution, including food products, resulting in an increase in dispersion. The characteristics of industries that are correlated with changes in dispersion are examined later in this section.

Figure A B.7, Figure A B.8, Figure A B.9, Figure A B.10, and Figure A B.11 show the equivalent charts for individual countries instead of taking the average across countries. There is huge variation across industries in both the extent of dispersion and changes in dispersion when looking at individual countries. There are also clear differences across countries: Indonesia has experienced a substantial improvement in carbon productivity across all percentiles of all industries; in Lithuania and Croatia, there is general improvement in many industries and seems to be some catching up of the 10th percentile in many industries; while in France and Sweden, the top has generally outperformed the bottom in many industries.

Figure A B.12 shows the equivalent analysis but for labour productivity. It plots the 10th, 50th, and 90th percentiles of labour productivity in the same 2 digit industries, averaged across countries. There is noticeably less dispersion in labour productivity than in carbon productivity for all industries in the sample. Furthermore, industries that tend to have more dispersion in carbon productivity are also more dispersed in labour productivity. The relationship between levels and changes in carbon productivity and labour productivity is explored in the following sub-section.

Figure 3.5. Evolution of carbon productivity across the distribution between 2005 and 2015, by industry

Carbon productivity at the 10th, 50th, and 90th percentiles within 2-digit industries and their change over time on a log scale



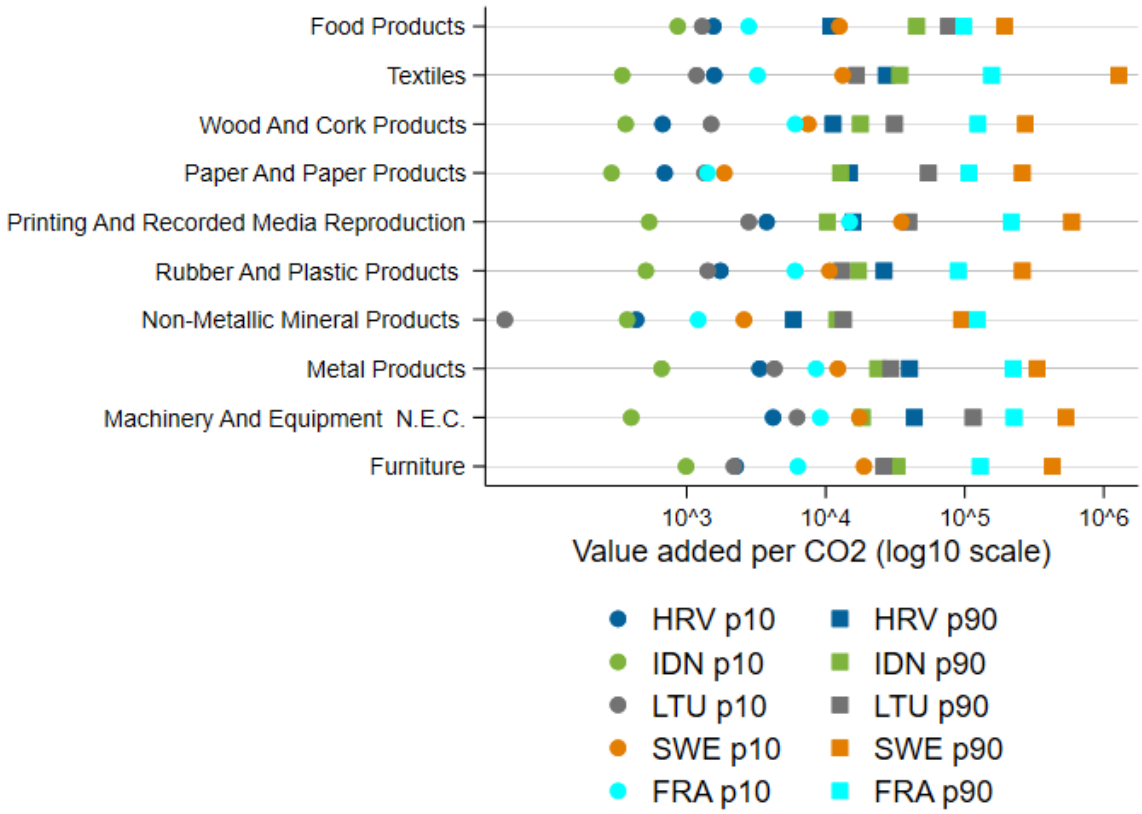
Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2005 to 2015. The values are plotted on a base 10 log scale. The unweighted average across countries is taken. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations.

There is also variation in the performance of firms across countries among firms operating in the same industry. Figure 3.6 plots the average carbon productivity of the top (90th percentile) and bottom (10th percentile) firms across countries. As before, the percentiles are computed for 3-digit industries and averaged to the 2-digit level.

At both the top and the bottom of the distribution, there are often large differences in carbon productivity between countries. Sweden has the most productive frontier firms in most industries, followed by France, while the ranking of other countries differs by sector. In some industries, the 10th percentile French or Swedish firms have similar carbon productivity to the 90th percentile Indonesian firms. Averaging across 3-digit industries, the average carbon productivity of the most productive firms is 15 times higher in Sweden than in Indonesia, and the least productive firms are also 15 times more productive. Figure A B.13 shows that a similar pattern emerges for labour productivity.

Figure 3.6. Variation across countries and industries in frontier and lagging firms' carbon productivity

10th and 90th percentile value added per CO_2 across countries in each manufacturing industry



Note: The chart shows the 10th and 90th percentiles of carbon productivity within industries for each country in 2015. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations.

Table A B.1 shows all 3-digit industries in the sample ranked by their average carbon productivity dispersion across countries. The most dispersed industries tend to be heavy industries, with energy costs comprising a higher share of total costs and firms having more carbon emissions on average. For example, manufacture of coke oven products and manufacture of basic iron and steel are two of the three most dispersed industries (the other being manufacture of refined petroleum products). The production of steel is an industry with a relatively homogenous output but two distinct production technologies with differing levels of emissions intensity (blast furnace or electric arc). This industry therefore illustrates one mechanism through which firms may have such different carbon productivities, even when producing a homogenous output.

The aggregate implications of dispersion

Counterfactual exercise: improving the carbon productivity of unproductive firms

To what extent could reducing dispersion affect the carbon footprint of production? To provide evidence on this, a counterfactual exercise is carried out whereby the carbon productivity of the least productive firms is raised to a certain percentile – either the 25th, 50th, or 75th percentile of the distribution in their 4-digit industry (the most detailed level available in the firm-level data) in their country in 2015 (the latest year with all countries in the sample). Then, assuming all the firms had at least that carbon productivity, the implied amount of carbon that would be produced to achieve the actual level of production observed is computed. Put differently, if all firms below the Nth percentile of carbon productivity in their sector had a carbon productivity equal to the firm at the Nth percentile, how much less carbon would be emitted to produce the same level of value added?

The results from this counterfactual exercise are presented in Figure 3.7 for each country in the sample. Focusing on the exercise of raising all firms to the 50th percentile within their industry, this would reduce carbon emissions by 42% in Croatia, 43% in Lithuania, 68% in Indonesia, 72% in Sweden, and 81% in France. The greater improvements reflect a higher degree of dispersion at the bottom of the distribution. The bars at the bottom of the graph represent the aggregate effect in all countries together (note that this is the aggregate from summing all values so is heavily affected by the largest countries – France and Indonesia).

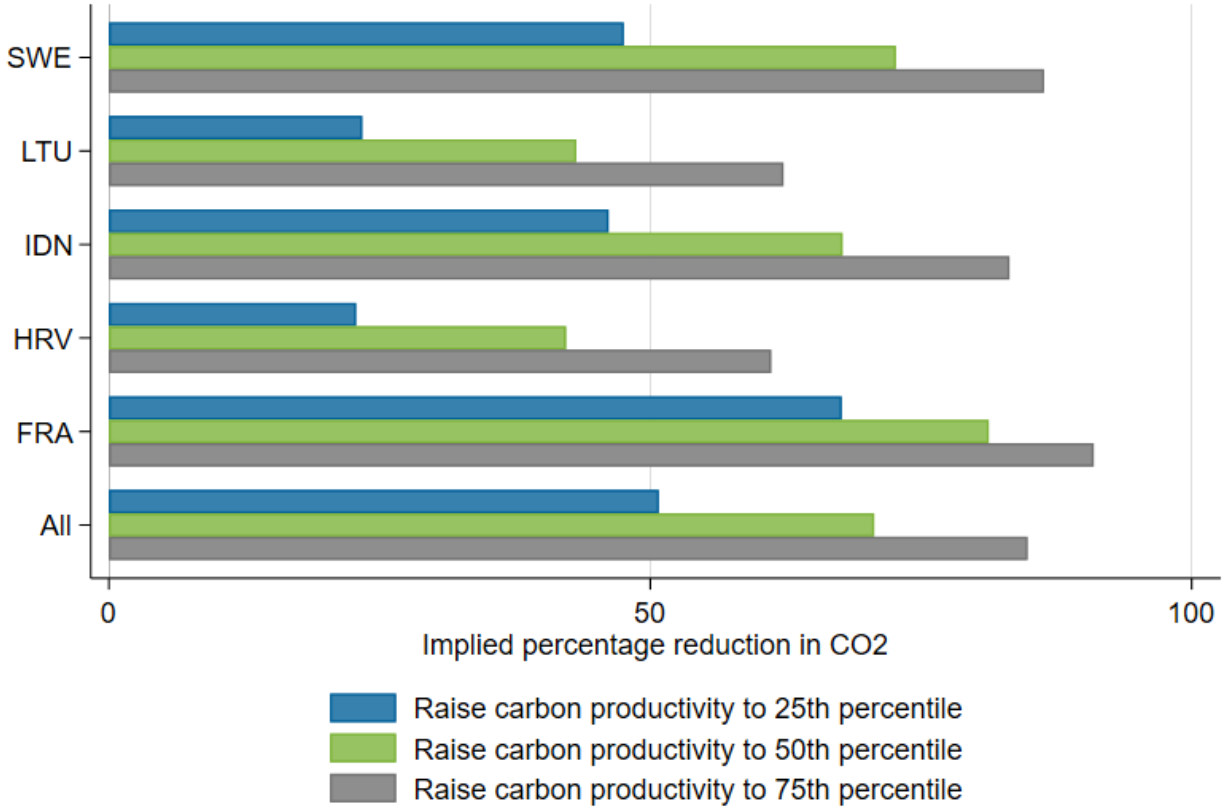
The results are similar when using 3 digit industries. Furthermore, they are not driven by specific sectors. Looking at the carbon-weighted distribution of the counterfactual exercise across 4-digit industries and countries, the 10th percentile industry experiences a 55% reduction while the 90th percentile industry experiences an 81% reduction in total emissions.

The implied reduction in CO2 appears large, and this is driven by two factors. First, the huge dispersion in carbon productivity means that the improvement that would be experienced by improving the carbon productivity of the least productive firms to the median is large. Second is the empirical observation that carbon emissions are strongly negatively correlated with carbon productivity (this is documented in Section 4 within industries but also holds across industries). Hence, large emitters are more likely to have below-median carbon productivity, and improving the carbon productivity of large emitters will have a greater impact on aggregate carbon emissions than improving that of smaller emitters.

To provide a benchmark comparison, Figure A B.14 presents results from the equivalent exercise for the number of workers that could provide the same level of output if all firms operated at least at the Nth percentile of labour productivity. Reflecting that there is much less dispersion in labour productivity than in carbon productivity, the percentage difference in workers necessary to achieve the same output is much smaller than the percentage reduction in carbon. Looking at the exercise of bringing all firms to the 50th percentile in their industry and aggregating across countries, for carbon productivity the percentage reduction in carbon is 71% while for labour productivity the percentage reduction in labour is 19%.

Figure 3.7. Improving the carbon productivity of the least efficient firms can substantially reduce CO_2 emitted to produce the same output

Counterfactual exercises showing implied reduction in CO_2 to produce the same aggregate value added if all firms were at least as productive as the 25th, 50th, and 75th percentile firms in their industry



Note: The chart shows the implied percentage reduction in CO_2 that would be achieved if all firms had at least a carbon productivity equal to the 25th, 50th, or 75th percentile firm in their 4-digit industry. The data is for 2015 and is restricted to manufacturing sector firms with at least 20 employees.
Source: authors' calculations

Relationship between changes in dispersion and changes in macro performance

Increases in within-industry dispersion in carbon productivity may be associated with either increases or decreases in aggregate outcomes of CO_2, value added, and carbon productivity. Increases in dispersion could be due to top firms becoming more productive, unproductive firms becoming less productive, or just relative differences between the two. More dispersion can also come through entry and exit of firms at different points of the distribution.[22]

To examine the relationship between dispersion and aggregate outcomes, regressions in both levels and changes are estimated. The regressions in levels mainly exploit cross-sectional variation and measures whether industries that have more dispersion have systematically different aggregate outcomes. The specification can be written:

[22] The role of entry and exit will be explicitly investigated in future work.

$$\log(y_{ckt}) = \beta\ DispCP_{ckt} + \varepsilon_{ckt}$$ Equation 1

The outcome variable $\log y_{ckt}$ is the log of the aggregate value of either value added, CO_2 emissions, employment, carbon productivity, or labour productivity, in industry k in country c in year t. The explanatory variable of interest is dispersion in carbon productivity, measured as the difference between the 90th and 10th percentile firms within a 3-digit industry and year:

$$DispCP_{ckt} = \log\left(\frac{VA_{ckt}}{CO2_{ckt}}\right)^{p90} - \log\left(\frac{VA_{ckt}}{CO2_{ckt}}\right)^{p10}$$ **Equation 2**

The regressions are weighted by CO_2 to be representative of total manufacturing sector emissions.[23] The relationship of interest is the pure correlation between dispersion and aggregate outcomes. Therefore, a simple bivariate regression specification is chosen, in line with the specification of Andrews, Criscuolo, and Gal (2019[2]). The results are robust to including country-year fixed effects.

The results are presented in Table 3.2, showing that industries with more dispersion in carbon productivity have worse aggregate carbon productivity but higher aggregate labour productivity. They also tend to be significantly larger in terms of labour, value added and, especially, CO_2 emissions.

To examine how changes in dispersion are correlated with changes in aggregate outcomes, the estimated regressions are instead estimated in 5-year differences:

$$\Delta_5 \log(y_{ckt}) = \beta\ \Delta_5 DispCP_{ckt} + \varepsilon_{ckt}$$ **Equation 3**

The results are presented in Table 3.3. Increases in dispersion are negatively correlated with both carbon and labour productivity, statistically significantly different from 0 at the 5% level. This suggests that increases in dispersion may be driven by less productive firms becoming even less productive, or net entry resulting in more dispersion and lower overall productivity.24 It suggests that dispersion in carbon productivity may be causing a drag on aggregate carbon productivity.

[23] Note that cross-country data is required here so 3-digit industries are the most detailed aggregation possible.

[24] The equivalent specifications with the explanatory variable replaced as dispersion in labour productivity are presented in Table A B.2 and Table A B.3 of the Annex. The results suggest that, as industries become more dispersed in labour productivity over time, their aggregate carbon and labour productivity increase on average.

Table 3.2. Industries with more dispersion in carbon productivity have worse aggregate carbon productivity

Regressions of aggregate variables on within-industry dispersion in carbon productivity

	(1)	(2)	(3)	(4)	(5)
	log(VA)	log(CO_2)	log(L)	log(VA/CO_2)	log(VA/L)
log(VA/CO_2) p90-p10	0.474***	1.211***	0.157**	-0.736***	0.326***
	(0.0379)	(0.0516)	(0.0725)	(0.0549)	(0.0606)
Observations	3026	3151	3137	3026	3024

Note: The table presents results from regressions of aggregate outcomes on within-industry-country-year dispersion in carbon productivity, measured as the log difference between the 90th and 10th percentile firm. The outcome variables of interest, in logs, are: value added (VA), CO_2, labour (L), carbon productivity (VA/CO_2), and labour productivity (VA/L). The data is at the 3-digit industry-country-year level and is restricted to firms with at least 20 employees in manufacturing industries. The regressions are weighted by total CO_2 to be representative of aggregate emissions. Standard errors are robust to heteroskedasticity. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

Table 3.3. Increasing dispersion in carbon productivity is correlated with a fall in aggregate carbon productivity

Regressions of changes in aggregate variables on changes in within-industry dispersion in carbon productivity

	(1)	(2)	(3)	(4)	(5)
	$\Delta_5 \log(VA)$	$\Delta_5 \log(CO_2)$	$\Delta_5 \log(L)$	$\Delta_5 \log(VA/CO_2)$	$\Delta_5 \log(VA/L)$
$\Delta_5 DispCP$	-0.147***	-0.0384	-0.0545***	-0.103**	-0.0841***
	(0.0302)	(0.0393)	(0.0211)	(0.0453)	(0.0297)
Observations	1957	2019	2119	1916	1955

Note: The table presents results from regressions of changes in aggregate outcomes on changes in within-industry-country-year dispersion in carbon productivity, measured as the log difference between the 90th and 10th percentile firm. The outcome variables of interest, in logs, are value added (VA), CO_2, labour (L), carbon productivity (VA/CO_2), and labour productivity (VA/L). The data is at the 3-digit industry-country-year level and is restricted to firms with at least 20 employees in manufacturing industries. The regressions are weighted by total CO_2 to be representative of aggregate emissions. Standard errors are robust to heteroskedasticity.* p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

Explanators of dispersion in carbon productivity

There are many factors which may explain why some industries have more dispersion in carbon productivity than others. This section relates dispersion in carbon productivity to other industry-level characteristics both in terms of the cross-sectional variation and within-industry changes over time.

The following regression specification is estimated:

$$DispCP_{ckt} = \beta\, x_{ckt} + \alpha_{ct} + \varepsilon_{ckt}$$

Equation 4

The variable $DispCP_{ckt}$ is within-industry dispersion in carbon productivity, as defined above. The explanatory variable of interest x_{ckt} is changed in each regression specification, but a multivariate specification with all possible explanatory variables is also estimated. The remaining terms are a country-

year fixed effect α_{ct} to control for country-specific characteristics and an error term ε_{ckt}.[25] Note that sector-year fixed effects are not included in the regressions because the specification is designed to capture the characteristics of industries with more dispersion. Therefore it is not desirable to control for averages of industries across countries for this purpose. The regressions are estimated by OLS on the cross-country micro-distributive dataset and standard errors are robust to heteroskedasticity.[26] The estimated coefficients and confidence intervals are presented in Figure 3.8, while Table A B.4 presents the results in a table.[27]

First, dispersion in carbon productivity is positively correlated with dispersion in labour productivity. There is an extensive literature that document the factors that may explain dispersion in labour productivity (see for example Andrews, Criscuolo, and Gal (2019[2])); the strong positive correlation here may indicate that these same factors may also be related to dispersion in carbon productivity. The relationship between dispersion in carbon and labour productivity is explored and discussed in more detail below.

Second, industries that are more dispersed in terms of capital productivity (value added per capital) or capital intensity (capital per worker) tend to also be more dispersed in terms of carbon productivity. Capital and energy usage are likely to be complementary to production, so industries with more variation in capital usage across firms are also more dispersed in terms of carbon productivity. Industries that are more capital intensive on average are also more dispersed in carbon productivity.

Third, dispersion in carbon productivity is related to cost shares of several types of inputs. It is positively correlated with dispersion in intermediate input cost shares and dispersion in the labour cost share, but uncorrelated with the import cost share. Industries with a higher average labour share are less dispersed in carbon productivity on average, somewhat in line with the positive correlation found with capital intensity. Unsurprisingly, carbon productivity dispersion is positively correlated with the average energy cost share, and, unsurprisingly, with dispersion in the energy cost share between firms.

Fourth, industries with a lower average share of electricity in total energy costs are more dispersed. This may reflect that there is more scope for emissions dispersion through choice of energy inputs in these industries or reflect the nature of production in these industries (as heavy industries are likely to rely less on electricity). Relatedly, dispersion in the electricity share is, unsurprisingly, positively correlated with dispersion in carbon productivity.

Fifth, there is no relationship between average firm age and carbon productivity dispersion. However, industries with more dispersion in firm age have more dispersion in carbon productivity on average, suggesting some role for technology vintages.

These results show simple bivariate regressions, without controlling for other factors, and should be interpreted as simple correlations. There may be structural factors – either observed or unobserved – that drive the observed correlations.

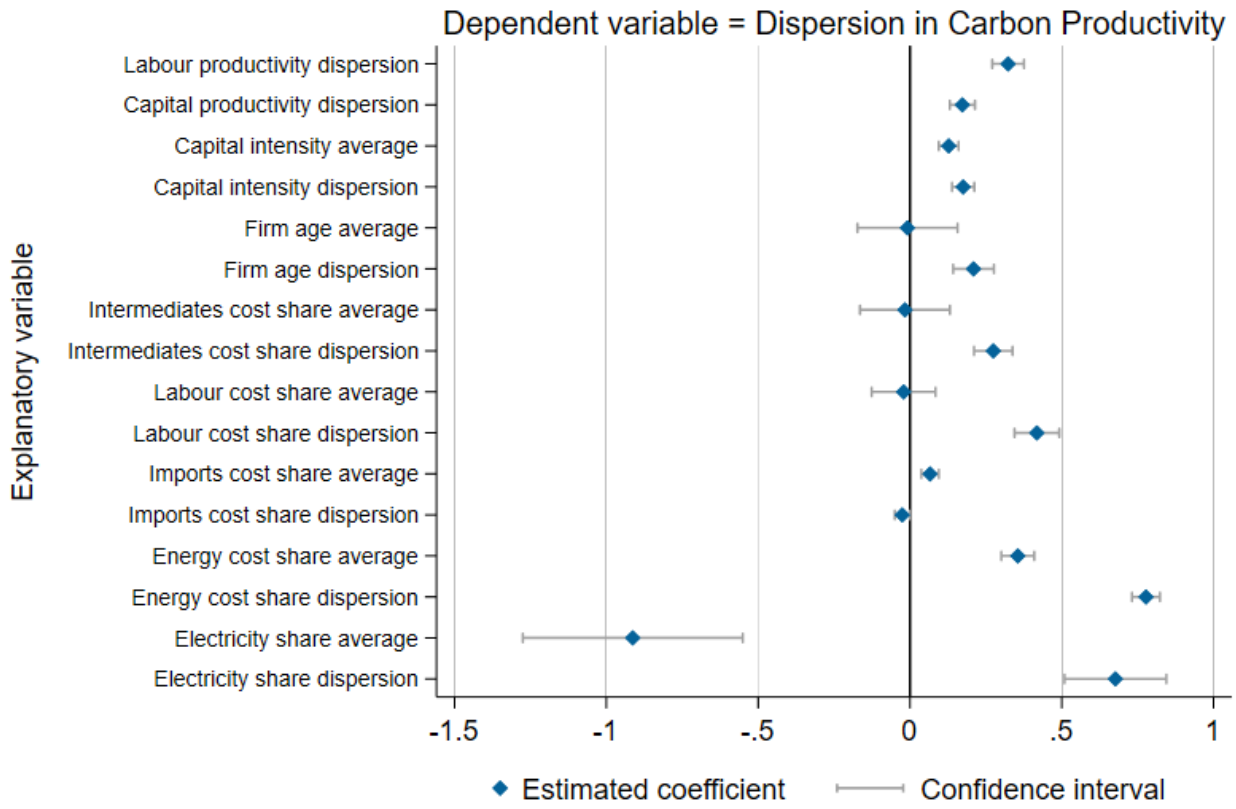
[25] All regressions including high-dimensional fixed effects are estimated using Correia (2017[97]).

[26] Note that some countries are excluded from certain regressions as not all variables are available in all countries.

[27] The regressions included in the table do not exactly correspond to the figure: to save space in the table, instead of two bivariate regressions on the average of a variable and its dispersion, a single multivariate regression with both the average and dispersion is presented.

Figure 3.8. Factors correlated with carbon productivity dispersion

Estimated coefficients from bivariate OLS regressions of carbon productivity dispersion on explanatory variables



Note: The chart shows results from regression estimates of within-industry dispersion in carbon productivity on various explanatory variables. The regressions are estimated by OLS on industry-country-year level data, with industries defined at the 3-digit level. Each row is estimated as a separate bivariate regression. Country-year fixed effects are included as controls. The sample is restricted to manufacturing industries and firms with at least 20 employees. Standard errors are robust to heteroskedasticity and 95% confidence intervals are plotted.
Source: authors' calculations.

It is therefore perhaps more informative to look at correlations between changes in dispersion in carbon productivity and changes in the explanatory variables. In particular, by looking at changes over time, the role of any unobserved time-invariant industry characteristics which may be correlated with both carbon productivity dispersion and the explanatory variable can be removed. The following regression is estimated:

$$\Delta_5 DispCP_{ckt} = \beta\, \Delta_5 x_{ckt} + \alpha_{ct} + \varepsilon_{ckt}$$

Equation 5

Where Δ_5 represents 5-year changes in a country-industry unit. The coefficients and confidence intervals are presented in Figure 3.9, with a full table of regression estimates in Table A B.5.[28]

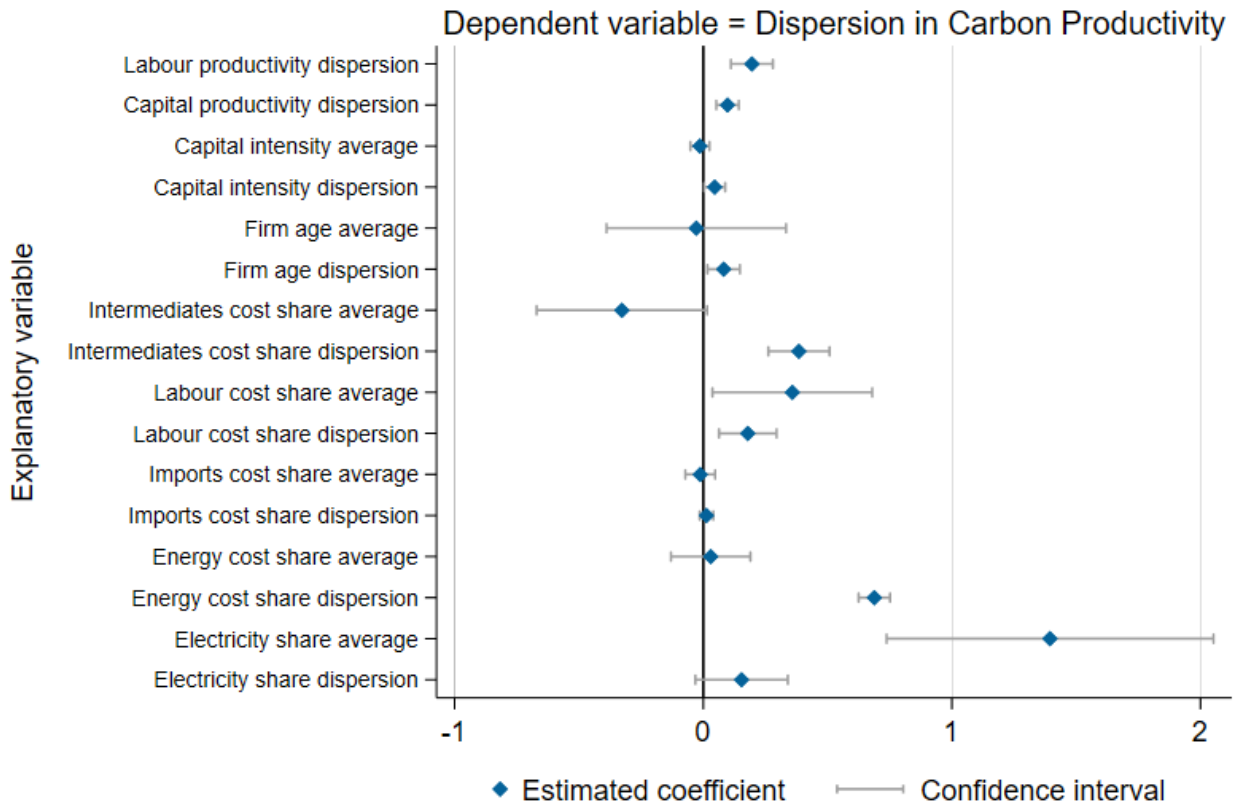
[28] As before, the regressions included in the table do not exactly correspond to the figure: to save space in the table, instead of two bivariate regressions on the average of a variable and its dispersion, a single multivariate regression with both the average and dispersion is presented.

Compared with the cross-sectional regressions above, many coefficients are no longer statistically significantly different from zero. The results must, of course, be interpreted with some caution when looking at many regressions with different explanatory variables.

There is still a strong positive correlation between labour productivity dispersion and carbon productivity dispersion. Furthermore, there is a significant positive correlation with capital productivity dispersion. There is also a positive correlation with firm age dispersion, energy cost share dispersion, intermediate input cost share dispersion, and labour cost share dispersion.

Figure 3.9. Factors correlated with changes in carbon productivity dispersion

Estimated coefficients from bivariate OLS regressions of carbon productivity dispersion on explanatory variables estimated in 5 year changes



Note: The chart shows results from regression estimates of 5-year differences of within-industry dispersion in carbon productivity on 5 year differences of various explanatory variables. The regressions are estimated by OLS on industry-country-year level data, with industries defined at the 3-digit level. Each row is estimated as a separate bivariate regression. Country-year fixed effects are included as controls. The sample is restricted to manufacturing industries and firms with at least 20 employees. Standard errors are robust to heteroskedasticity and 95% confidence intervals are plotted.
Source: authors' calculations.

Correlation between dispersion in carbon productivity and in labour productivity

One of the key aims of this paper is to understand the relationships between environmental and economic performance, especially between carbon productivity and labour productivity. The previous sub-section revealed a strong and statistically significant relationship between dispersion in carbon productivity and

labour productivity, both in levels and changes. This sub-section examines the robustness of this result by adding control variables to the regressions. Specifically, the following relationship is estimated:

$$DispCP_{ckt} = \beta\ DispLP_{ckt} + \gamma X_{ckt} + \alpha_{ct} + \varepsilon_{ckt}$$

Equation 6

Dispersion in labour productivity is defined as before and in line with the definition of dispersion in carbon productivity:

$$DispLP_{ckt} = \log\left(\frac{VA_{ckt}}{L_{ckt}}\right)^{p90} - \log\left(\frac{VA_{ckt}}{L_{ckt}}\right)^{p10}$$

Equation 7

Additional controls, X_{ckt}, are included and vary across specifications. The regressions are also then estimated in 5-year differences, with controls also specified in differences, written as:

$$\Delta_5 DispCP_{ckt} = \beta\ \Delta_5 DispLP_{ckt} + \gamma \Delta_5 X_{ckt} + \alpha_{ct} + \varepsilon_{ckt}$$

Equation 8

The results from the regression estimated in levels are presented in Table 3.4 and the results from the regressions in changes are presented in Table 3.5. There is a robust positive relationship between dispersion in carbon productivity and dispersion in labour productivity. This suggests that there may be structural industry characteristics that explain both these dispersions. There is an extensive literature documenting the extent and explanations of dispersion in labour productivity (see for example Berlingieri and Criscuolo (2017[17]), Andrews, Criscuolo, and Gal (2019[2]), Syverson (2011[19])), as well as possible solutions to help improve the productivity of laggard firms (Berlingieri et al., 2020[18]). Several possible explanations of dispersion in labour productivity have been proposed. These include a lack of technology transfer between frontier and laggard firms, barriers to entry or scale-up, or a lack of competition, among others. The results presented here - along with the evidence in the next section that there is a strong relationship between labour and carbon productivity at the firm-level - suggest that such remedies may also help to improve the environmental performance of firms if designed appropriately, and this should be investigated in more detail.

Table 3.4. Industries with more dispersion in carbon productivity also have more dispersion in labour productivity

Regressions of dispersion in carbon productivity on dispersion in labour productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10
log(VA/L) p90-p10	0.323***	0.224***	0.216***	0.201***	0.141***	0.0564*	0.186***	0.162***
	(0.0267)	(0.0343)	(0.0348)	(0.0354)	(0.0346)	(0.0305)	(0.0361)	(0.0350)
log(L) p90-p10			0.122***					
			(0.0462)					
log(Age) p90-p10				0.0924***				
				(0.0350)				
log(Intermediates/total cost) p90-p10					0.481***			
					(0.0329)			
log(Energy spending/total cost) p90-p10						0.718***		
						(0.0206)		
log(Labour costs/total cost) p90-p10							0.173***	
							(0.0408)	
log(K/L) p90-p10								0.153***
								(0.0183)
Observations	3505	3024	3024	2954	3024	3024	3024	2950
Basic Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents results from OLS regressions on industry-country-year level data. The dependent variable is dispersion in carbon productivity and dispersion in labour productivity, where dispersion is measured as the log difference between the 90th and 10th percentile firms in a cell. All regressions include country-year fixed effects. The basic controls: log employment, log value added, log CO_2, log average age, log firm count. The additional explanatory variables are: log dispersion in employment, log dispersion in age, log dispersion in intermediate inputs costs share, log dispersion in energy cost share, log dispersion in labour cost share, and log dispersion in capital per worker. The sample is restricted to manufacturing firms with at least 20 employees. Standard errors are robust to heteroskedasticity. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

Table 3.5. Changes in dispersion in carbon productivity are correlated with changes in labour productivity dispersion

Regressions of changes in dispersion in carbon productivity on changes in dispersion in labour productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta_5$log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10	Δ_5.log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10	$\Delta_5$log(VA/CO$_2$) p90-p10
Δ_5 log(VA/L) p90-p10	0.196***	0.196***	0.193***	0.189***	0.156***	0.0450	0.169***	0.197***
	(0.0430)	(0.0451)	(0.0446)	(0.0472)	(0.0460)	(0.0430)	(0.0456)	(0.0480)
Δ_5 log(L) p90-p10			0.0733					
			(0.0658)					
Δ_5 log(AGE) p90-p10				0.0117				
				(0.0363)				
Δ_5 log(M/TC) p90-p10					0.337***			
					(0.0617)			
Δ_5 log(EN/TC) p90-p10						0.669***		
						(0.0349)		
Δ_5 log(LC/TC) p90-p10							0.169***	
							(0.0642)	
Δ_5 log(K/L) p90-p10								0.0278
								(0.0218)
Observations	2353	1915	1915	1857	1915	1915	1915	1866
Basic Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents results from OLS regressions estimated in 5-year differences on industry-country-year level data. The dependent variable is dispersion in carbon productivity and dispersion in labour productivity, where dispersion is measured as the log difference between the 90th and 10th percentile firms in a cell. All regressions include country-year fixed effects. The controls, in order of appearance in the table, are: log employment, log value added, log CO_2, log average age, log firm count, log dispersion in employment, log dispersion in age, and log dispersion in capital per worker. The sample is restricted to manufacturing firms with at least 20 employees. Standard errors are robust to heteroskedasticity. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

4 Factors explaining firms' carbon productivity

This section identifies the factors associated with firms' carbon productivity. It directly uses the firm-level data to estimate regressions of (log) carbon productivity on various firm characteristics. Specifically, the following relationship is estimated:

$$\log\left(\frac{VA_{ict}}{CO2_{ict}}\right) = \beta_c\, x_{\mathrm{ict}} + \alpha_{k(i)t} + \varepsilon_{ict} \qquad \text{Equation 9}$$

for firm i in country c operating in industry $k(i)$ at time t, where $\alpha_{k(i)t}$ is a 4-digit industry-year fixed effect and ε_{ict} is an error term. The relationships are estimated as separate bivariate regressions, meaning that the explanatory variable x_{ict} is adapted across specifications. The regression is necessarily estimated separately for each country because firm-level data cannot be pooled across countries due to access regulations. Therefore, separate estimates of β are presented for each country.

The results are presented in Figure 4.1. The results from each explanatory variable are usually similar across countries in terms of their magnitudes and, in almost all cases, the estimated coefficients have the same sign in each country.[29]

First, carbon productivity is strongly positively correlated with labour productivity. In line with this, Barrows and Ollivier (2018[16]) find that more efficient firms are less emission intensive. This relationship is explored and discussed in more detail later in the section.

Second, the relationship between carbon productivity and the share of non-energy intermediate inputs costs in total costs varies by country. In France, Lithuania, and Sweden, the relationship is positive, while in Croatia and Indonesia, the relationship is negative. Note, however, that when all the explanatory variables are included together, the relationship between non-energy intermediates and carbon productivity is negative in every country. This is likely to reflect that firms with a higher share of intermediates have more complex production procedures requiring more parts to be assembled. A counteracting force would be that firms with high shares of intermediates may be outsourcing a higher

[29] Most of the results are similar if the explanatory variables are included together in a multivariate regression, shown in Figure A B.15, although these results should be interpreted with caution as the explanatory variables are likely to be highly collinear. The coefficient on labour productivity remains positive and significant, while the coefficients on the intermediate share become negative. The coefficients on employment, and CO_2 remain negative. The coefficient on the capital-labour ratio becomes positive in Indonesia and Sweden (only), while the coefficients on energy cost share move closer to 0 and are either positive, negative, or insignificant in each country, as do the coefficients on firm age. Finally, the import and export shares are negatively associated with carbon productivity when controlling for other variables.

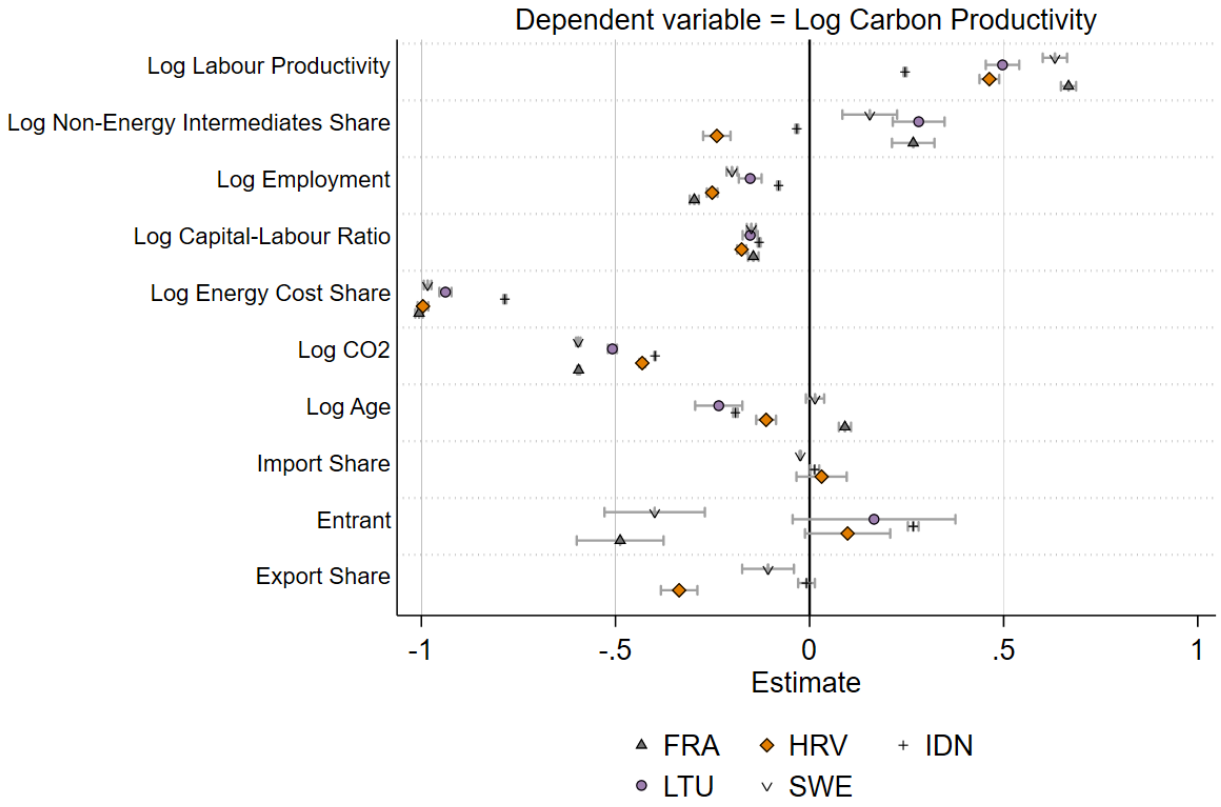
share of production or be in a more downstream industry and may therefore be more energy efficient, but it appears that the first explanation dominates as the observed correlations are strongly negative.

Third, larger firms, measured by employment or CO_2, are less carbon productive, as are more capital intensive firms. Younger firms are more carbon productive in three of the five countries, perhaps reflecting that they use newer vintages of capital which are more environmentally efficient. However, in France, younger firms tend to be less carbon productive. In both France and Sweden, firms that are not new entrants have higher carbon productivity, but note that entrants with at least 20 employees (those included in the sample) may not be representative of all entrants.

Finally, firms that export more tend to be less carbon productive, which may reflect that firms that export tend to be larger and more capital intensive. There is no statistically significant relationship between imports and carbon productivity in Indonesia and Croatia, while there is a negative relationship in Sweden – this is further lack of evidence of outsourcing carbon intensive products.

Figure 4.1. Relationships between carbon productivity and other firm characteristics

Estimates from bivariate OLS regressions of log value added per CO_2 on various explanatory variables



Note: The chart shows estimates from bivariate OLS regressions of log value added per CO_2 on various explanatory variables. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The regressions all include industry-year fixed effects. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations

In the Annex, results are presented for two linear probability model specifications. The first examines characteristics of firms at the top of the carbon productivity distribution in its industry, by regressing a

dummy variable for whether the firm is in the top 10% of firms in its industry for value added per CO2. Specifically, it estimates:

$$\left\{\frac{VA_{ict}}{CO2_{ict}}\right\}^{top10\%} = \beta_c \mathrm{X}_{\mathrm{ict}} + \alpha_{k(i)t} + \varepsilon_{ict} \qquad \textbf{Equation 10}$$

Where $\left\{\frac{VA_{ict}}{CO2_{ict}}\right\}^{top10\%}$ is a dummy for whether the firm is in the top decile of the distribution for carbon productivity within an industry-year.

Second, to examine the characteristics of firms with lagging carbon productivity performance, a dummy variable for whether the firm is in the bottom tercile of the carbon productivity distribution in its industry. Specifically, it estimates:

$$\left\{\frac{VA_{ict}}{CO2_{ict}}\right\}^{bottom33\%} = \beta_c \mathrm{X}_{\mathrm{ict}} + \alpha_{k(i)t} + \varepsilon_{ict} \qquad \textbf{Equation 11}$$

Where $\left\{\frac{VA_{ict}}{CO2_{ict}}\right\}^{bottom33\%}$ is a dummy for whether the firm is in the bottom tercile.

The results from these specifications are presented in Figure A B.17 and Figure A B.18, respectively. The economic implications of the results are broadly in line with those of the baseline continuous specification of Figure 4.1. Frontier firms in carbon productivity are more labour productive, have lower employment, are less capital intensive, and are younger. The reverse is true for unproductive firms, except that the relationship with age is less clear across countries.

Given the strong correlation between labour productivity and carbon productivity, it might be expected that equivalent regressions to the above but with labour productivity as the dependent variable might yield similar results. Figure A B.16 presents results for bivariate regressions with log labour productivity as the outcome and the same explanatory variables on the right-hand side. The results are often contrasting to those for carbon productivity as the outcome. Labour productivity is positively correlated with the intermediate input share, employment, capital-labour ratio, and carbon emissions. There is also a clearer relationship with age and entry: older firms tend to be more labour productive.

Finally, the relationship between changes in labour productivity and changes in carbon productivity within firms over time is explored. To do so, the following regression specification is estimated:

$$\Delta_l \log\left(\frac{VA_{ict}}{CO2_{ict}}\right) = \beta \Delta_l \log\left(\frac{VA_{ict}}{L_{ict}}\right) + \alpha_{k(i)t} + \Delta_l X_{ict} + X_{ic,t=0} + \alpha_1 FirstYear_{ict} + \alpha_2 FinalYear_{ict} + \alpha_3 YearSpan_{ict} + \varepsilon_{ict} \qquad \textbf{Equation 12}$$

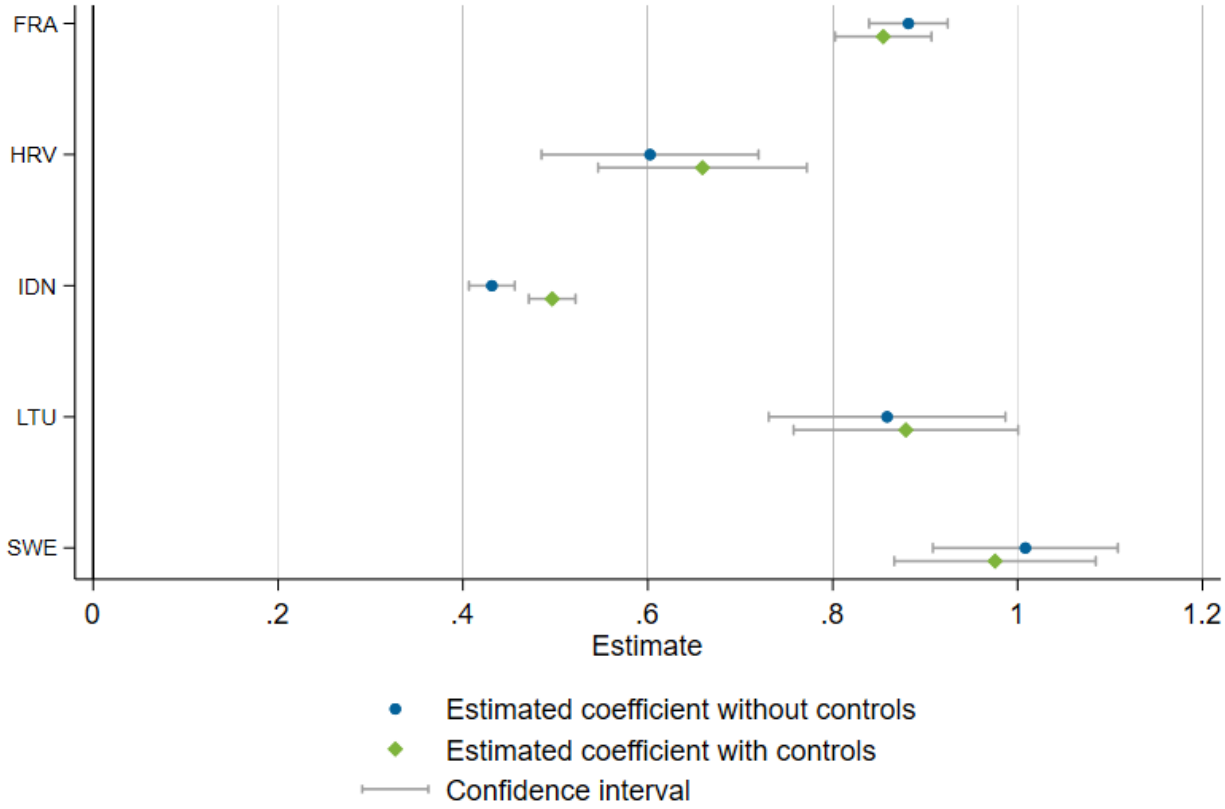
The regression is designed to be robust to repeated cross-sectional sampled data, where firms might not be sampled in all years. It is estimated in long changes – taking the difference between the first and last observation of the firm. The variables $\left(\frac{VA_{ict}}{CO2_{ict}}\right)$, $\left(\frac{VA_{ict}}{L_{ict}}\right)$, and $\alpha_{k(i)t}$ are defined as before (carbon productivity, labour productivity, and industry-year fixed effects). X_{ict} is a set of firm controls, including (log) capital, (log)intermediate inputs, and (log) age (set to 1 in the first year and included in levels). To account for differences in the time periods where each firm is first observed and finally observed, first year (the first year the firm is observed), final year (the final year the firm is observed), and year span (the number of years between the first and the last year of observation) dummy variables are included. The results are robust to the exclusion of controls, to recalibrating sample weights to account for the requirement of two

periods, and to alternative specifications such as shorter period changes or a specification in levels with firm fixed effects.

The results, with and without controls, are presented in Figure 4.2. There is a strong positive relationship between changes in labour productivity and changes in carbon productivity. Put differently, as firms improve their labour productivity, they also improve their carbon productivity, even controlling for changes in their overall size. The estimated coefficients are between 0.4 and 0.9 depending on the country, suggesting that a 10% increase in labour productivity is associated with between a 4% and 9% increase carbon productivity.

Figure 4.2. As firms become more labour productive, they also become more carbon productive

Estimated coefficients from OLS regressions of change in carbon productivity on change in labour productivity and controls



Note: The chart shows estimates from OLS regressions of the long change in log value added per CO_2 on the long change in log labour productivity. Controls for (log) capital, (log) intermediate inputs, and (log) age are included in initial levels and, for the first two, in changes. Additional fixed effects are included for the entry and exit year of the firm, and the length of that period. The regressions also include industry-year fixed effects. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations.

5 The effects of energy prices

One of the key policy tools to help reduce emissions is carbon taxes and trading schemes. These raise the price of emitting carbon to internalise the social costs of pollution and to provide an incentive to reduce emissions. They may also have economic implications by affecting the overall competitiveness of firms as well as through their complementarity or substitutability with other factors of production such as labour, capital, and intermediate inputs.

This section examines the effects of energy prices on the emissions and economic performance of firms in a cross-country setting. It then looks at how different types of firms – in terms of their carbon productivity, age, or size – respond to changes in energy prices. Future work will aim to estimate the effects of broader economic and environmental policies on firms' economic outcomes.

The baseline specification estimates the change in the log of firm outcomes on the one period lagged change in the firm's energy price index:

$$\Delta_3 \log y_{ict} = \beta \Delta_3 \log EnergyPrice_{ic,t-1} + \alpha_{k(i)t} + \varepsilon_{ict}$$

Equation 13

for firm i in country c in year t. The outcomes variables y_{ict} are CO_2 emissions from energy use, value added, employment, capital, carbon productivity (value added per CO_2), and labour productivity (value added per worker). All regressions include an industry-year fixed effect $\alpha_{k(i)t}$ to compare firms with similar production functions and control for industry-specific and macro shocks.[30] The regressions are estimated in three-year differences to allow for medium term firm adjustments.[31] Standard errors are clustered by firm to account for potential serial correlation.

To compute firms' changing exposure to energy price shocks, the weighted average log change is computed over their different energy types n as:

$$\Delta_3 \log EnergyPrice_{ic,t-1} = \sum_n EnergyShare_{nic,t-4} \times (\log Price_{nic,t-1} - \log Price_{nic,t-4})$$

Equation 14

The energy price changes that firms face may be endogenously determined with respect to their economic and environmental performance. For example, larger firms are likely to pay lower prices for energy, so changes in price and firm size would be correlated. Therefore, in addition to reporting OLS results, an instrumental variables strategy is used following an established literature for comparability (see Sato et al. (2019[21]) for a discussion). The instrumental variable is a shift-share which uses the firm's shares of each energy, held fixed, as exposure weights to national changes in the prices of energy types. The exclusion restriction requires that the share-weighted changes in these national prices are uncorrelated with the error

[30] The baseline specifications use 2 digit industries but the results are very similar when using 4 digit industries.

[31] Future work will also investigate 5 year differences.

term. The literature relies on the argument that the shocks (energy prices) are exogenous because, for example, national energy prices should be driven by international energy markets and national policies, rather than firm-specific characteristics.[32] The instrumental variable for the change in energy price is written:

$$\Delta_3 \log EnergyPrice^{IV}_{ic,t-1} = \sum_n EnergyShare_{nic,t-4} \times (\log Price_{nc,t-1} - \log Price_{nc,t-4})$$

Equation 15

The regressions must be estimated separately for each country as they require access to firm-level data which cannot be pooled across countries. Only countries that have data on the quantity of different energy types used can be included in the regressions because energy prices must be directly used to estimate CO_2 for countries where only values are available, which means that CO_2 would be mechanically related to prices. Therefore, the results are presented for France and Indonesia.[33]

The lagged structure of the specifications require that firms are observed in operation in periods t-4, t-3, t-1, and t. Therefore, the results should be interpreted as the effects of energy prices conditional on firm survival. Additionally in the case of France, where the data is collected through a sample, the results may be biased towards firms that are repeatedly sampled, which tend to be larger. To check whether this affects the results, future work will examine the robustness of results to sample weights that are adjusted to account for the probability that firms appear in the repeated sample. The results are also robust to a specification in contemporaneous changes, which only requires that the firm appears in t-3 and t.

Baseline results

The results from the instrumental variables regressions, estimated by two-stage least squares (2SLS), are presented in Table 5.1. The top panel presents results for France and the bottom panel for Indonesia. Under the assumptions discussed above, the estimates can be interpreted as the causal effects of energy prices on firm outcomes. The first stage Kleibergen-Paap (KP) F statistics are large (around 170 in France and above 18 000 in Indonesia), implying that the first stages are strong (i.e. the instrument is informative). The "reduced form" regression results – the firm outcomes regressed against the instrument (national energy prices) – are presented in Table A B.6 of the Annex.

Increases in energy prices cause a fall in CO_2 emissions and a rise in carbon productivity in both countries. With the regression specified in log differences, the estimated coefficients can be approximately interpreted as elasticities, meaning that a 10% increase in the energy price is associated with a 3.7% increase in carbon productivity in France or a 1.8% increase in Indonesia. This suggests that energy pricing incentivises firms to reduce their emissions, either by using less energy overall or by switching to greener sources of energy.

[32] Recent research shows that identification that relies on exogeneity of the shocks in a shift-share IV setting requires a large number of shocks (Borusyak, Hull and Jaravel, 2021[94]). In this setting, that would require a large number of energy types, but there are at most 18 energy types. Therefore, alternative identification may choose to rely on the exogeneity of the shares (Goldsmith-Pinkham, Sorkin and Swift, 2020[95]). The baseline results of this paper are robust to using an instrument with industry energy type shares as and firm prices, relying on the assumption that the types of energies firms use are uncorrelated with changes in their outcomes, except through prices.

[33] Results for Sweden are not yet available.

In France, there is no statistically significant relationship between energy prices and value added or employment, while in Indonesia higher energy prices cause a relative rise in value added and employment. This may suggest that, in Indonesia, higher energy prices stimulate firms to become more efficient by, for example, investing in new technologies or improving their human capital. In France, there is a negative effect of energy prices on the capital stock, perhaps reflecting that energy and capital are complementary in production.

The estimates from the OLS regressions using firms' own energy price index are presented in Table A B.7. Increases in energy prices are associated with improvements in carbon productivity and a decline in CO_2 emissions in both countries. As discussed above, the OLS should be interpreted with caution. The estimates are likely to be biased because energy price changes that firms face may be endogenously determined with respect to their economic and environmental performance, for example, if larger firms pay lower prices for energy.

Table 5.1. Increasing energy prices cause a reduction in CO_2 and an increase in CO_2 productivity

IV regression estimates of changes in firm outcomes on change in energy prices

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO_2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO_2
FRA						
Energy Price	-.007	-.37***	-.047	-.098**	.043	.369***
	(.063)	(.113)	(.05)	(.046)	(.073)	(.125)
Observations	40 080	40 649	40 649	40 153	40 080	40 080
KP F Stat	164.8	175.8	174.2	163.9	164.8	166.4
IDN						
Energy Price	.101***	-.11***	.023***	-.005	.075***	.18***
	(.017)	(.023)	(.008)	(.029)	(.016)	(.022)
Observations	179 305	207 338	207 338	105 919	179 305	179 282
KP F Stat	18448.4	19175.5	19175.5	11171.2	18448.4	18458.1
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	IV	IV	IV	IV	IV	IV

Note: The table presents results of 2SLS regression estimates of various outcomes of firms on their energy price index. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

Heterogeneity

This paper has documented that there is huge dispersion across firms in their carbon productivity. This section explores whether different types of firms also respond differently to energy price shocks. Three margins of heterogeneity are explored: tercile of carbon productivity; firm age band; and firm employment band. Each category of the heterogeneity variable is interacted with the energy price change, meaning that the marginal effect of each group is simply the coefficient of that interaction (i.e. there is no base category).

In each case, the instrumental variables results are presented, where the instruments are interacted with the relevant heterogeneity band. The heterogeneity results should be interpreted with the caveat that the samples are restricted to firms with employment of at least 20, and in France larger firms are more likely to be sampled.

Carbon productivity

Table 5.2 presents the results on the effects of energy prices at different terciles of the carbon productivity distribution. In both countries, firms across the distribution of initial carbon productivity have a positive and statistically significant effect of energy prices on their carbon productivity. In France, the strongest marginal effects are among those with initially higher carbon productivity, while in Indonesia the marginal effect is similar across all types of firms.

Table 5.2. Higher energy prices cause a rise in carbon productivity across the distribution of initial carbon productivity

IV regression estimates of changes in firm outcomes on change in energy prices by firms' initial carbon productivity

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO_2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO_2
FRA						
Low CP x Energy Price	-0.010	-0.244**	-0.035	-0.045	0.028	0.238**
	(0.062)	(0.102)	(0.047)	(0.038)	(0.072)	(0.114)
Mid CP x Energy Price	0.058	-0.274*	-0.039	-0.035	0.090	0.335*
	(0.082)	(0.158)	(0.068)	(0.055)	(0.095)	(0.174)
High CP x Energy Price	0.067	-0.449**	0.155	-0.122	-0.065	0.555**
	(0.108)	(0.199)	(0.097)	(0.076)	(0.125)	(0.217)
R-squared	-0.001	0.016	0.002	0.002	-0.001	0.006
Observations	39 724	40 242	40 242	39 793	39 724	39 724
KP F Stat	29.2	30.5	30.2	28.8	29.2	29.5
IDN						
Low CP x Energy Price	0.072***	-0.057**	0.028***	-0.029	0.042**	0.133***
	(0.020)	(0.026)	(0.009)	(0.033)	(0.018)	(0.025)
Mid CP x Energy Price	0.073***	-0.105***	0.021**	0.022	0.051**	0.178***
	(0.022)	(0.029)	(0.010)	(0.036)	(0.020)	(0.026)
High CP x Energy Price	0.048**	-0.113***	0.024**	0.002	0.028	0.149***
	(0.025)	(0.032)	(0.012)	(0.041)	(0.023)	(0.031)
R-squared	0.006	0.006	0.000	0.000	0.007	0.023
Observations	163 441	164 948	164 948	96 967	163 441	163 437
KP F Stat	5424.4	5485.8	5485.8	3340.3	5424.4	5430.4
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	IV	IV	IV	IV	IV	IV

Note: The table presents results of 2SLS regression estimates of various outcomes of firms on their energy price index interacted with the firm's within-industry carbon productivity tercile in t-4. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations.

Age

Table 5.3 reports the impacts of energy prices on outcomes for firms of different ages. The effect on carbon productivity appears to be similar for firms of different ages. In France, the negative effect on capital appears to be driven by younger firms, perhaps because they have smaller sunk costs on capital investments so can more easily adapt their production processes to higher energy prices. In Indonesia, the positive effect on employment growth is driven by smaller firms.

Table 5.3. Firms of all ages have higher relative carbon productivity with energy price rises

IV regression estimates of changes in firm outcomes on change in energy prices by firms' age band

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO_2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO_2
FRA						
Entrant x Energy Price	0.074	-0.493	-0.215	-0.336**	0.269	0.589
	(0.161)	(0.331)	(0.192)	(0.140)	(0.257)	(0.379)
Young x Energy Price	-0.115	-0.256	-0.052	-0.212*	-0.062	0.155
	(0.158)	(0.217)	(0.114)	(0.108)	(0.141)	(0.254)
Mid age x Energy Price	-0.026	-0.339**	-0.047	-0.032	0.029	0.317**
	(0.084)	(0.135)	(0.057)	(0.055)	(0.093)	(0.157)
Old x Energy Price	0.016	-0.368***	-0.041	-0.067*	0.058	0.392***
	(0.063)	(0.115)	(0.051)	(0.040)	(0.074)	(0.125)
R-squared	0.000	0.000	-0.001	0.024	-0.002	-0.004
Observations	39 724	40 242	40 242	39 793	39 724	39 724
KP F Stat	20.8	22.9	22.6	19.8	20.9	21.1
IDN						
Entrant x Energy Price	0.114***	-0.127***	0.027**	0.006	0.090***	0.253***
	(0.027)	(0.035)	(0.012)	(0.040)	(0.025)	(0.033)
Young x Energy Price	0.117***	-0.050	0.023**	0.040	0.092***	0.161***
	(0.025)	(0.032)	(0.011)	(0.038)	(0.023)	(0.031)
Mid age x Energy Price	0.069***	-0.158***	0.016*	-0.003	0.051***	0.221***
	(0.020)	(0.026)	(0.009)	(0.034)	(0.018)	(0.025)
Old x Energy Price	0.083***	-0.193***	-0.001	-0.114*	0.084***	0.293***
	(0.032)	(0.043)	(0.014)	(0.061)	(0.030)	(0.039)
R-squared	0.003	-0.002	0.002	0.000	0.002	-0.001
Observations	163 442	164 988	164 988	96 968	163 442	163 437
KP F Stat	3402.9	3393.7	3393.7	2309.3	3402.9	3402.7
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	IV	IV	IV	IV	IV	IV

Note: The table presents results of 2SLS regression estimates of various outcomes of firms on their energy price index interacted with the firm's age band in t-4. Entrants are defined as younger than 3 years; young firms are defined as 3-5 years; mid age firms are defined as 6 to 19 years; and old firms are at least 20 years. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm.
Source: authors' calculations

Employment

Finally, Table 5.4 reports the results across different employment bands. Note that the KP F statistics for France are small in these regressions so the results should be interpreted with caution. In France, elasticity between energy prices and carbon productivity are highest for small firms, while in Indonesia they are similar across the distribution.

Table 5.4. In France, the smallest firms have the highest elasticity of energy prices to carbon productivity

IV regression estimates of changes in firm outcomes on change in energy prices by firms' initial employment band

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO_2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO_2
FRA						
Small x Energy Price	-0.024	-0.829**	-0.155	-0.067	0.128	0.797**
	(0.141)	(0.342)	(0.099)	(0.094)	(0.168)	(0.386)
Medium x Energy Price	0.006	-0.304***	-0.059	-0.097**	0.072	0.331**
	(0.065)	(0.119)	(0.051)	(0.043)	(0.075)	(0.130)
Large x Energy Price	-0.007	-0.304***	-0.001	-0.030	-0.007	0.298**
	(0.064)	(0.109)	(0.054)	(0.041)	(0.076)	(0.123)
R-squared	0.001	-0.009	0.002	0.003	-0.003	-0.012
Observations	39 726	40 244	40 244	39 795	39 726	39 726
KP F Stat	4.2	4.6	4.6	4.3	4.2	4.2
IDN						
Small x Energy Price	0.077***	-0.117***	0.031***	-0.007	0.046**	0.197***
	(0.020)	(0.027)	(0.009)	(0.032)	(0.018)	(0.025)
Medium x Energy Price	0.075***	-0.172***	-0.015	-0.013	0.090***	0.244***
	(0.022)	(0.028)	(0.011)	(0.037)	(0.020)	(0.026)
Large x Energy Price	0.080***	-0.173***	0.006	0.001	0.074***	0.247***
	(0.026)	(0.035)	(0.012)	(0.047)	(0.024)	(0.034)
R-squared	0.006	-0.001	0.011	0.000	0.002	-0.001
Observations	163 442	164 988	164 988	96 968	163 442	163 437
KP F Stat	6208.2	6263.1	6263.1	3818.3	6208.2	6211.8
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	IV	IV	IV	IV	IV	IV

Note: The table presents results of 2SLS regression estimates of various outcomes of firms on their energy price index interacted with the firm's employment band in t-4. Small firms are defined as those with fewer than 50 employees, while medium size firms have between 50 and 249 employees, and large firms have at least 250 employees. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm.
Source: authors' calculations.

6 Conclusions and policy implications

This paper develops unique cross-country data on the environmental and economic performance of firms. It shows that there is huge dispersion across firms in their carbon productivity within narrowly defined industries, and that helping less carbon productive firms to improve their environmental performance could substantially help achieve the green transition.

Improving the carbon productivity of firms in the industrial sector is a key component of the transition to net zero. In an increasingly carbon-constrained world, carbon efficiency is also set to become a key driver of firms' competitiveness. Understanding the drivers of firms' carbon productivity and documenting the extent of its dispersion across firms can help design more effective decarbonisation policies that target different components of the firms' population.

A holistic policy framework should aim to push the green technological frontier while also boosting catch-up of unproductive firms and providing incentives to switch away from carbon-emitting production processes.

First, policies should keep promoting innovation in new technologies to push the frontier of carbon efficiency. Effective green innovation policies include supply-side instruments (R&D tax credits, R&D grants) and demand-side instruments such as carbon pricing, which has been shown to incentivise patenting activity in low carbon technologies (Calel and Dechezleprêtre, 2016[56]).

Second, in light of the huge dispersion between firms in carbon productivity, policies should help improve the environmental performance of unproductive firms while increasing long term allocative efficiency. The evidence of this paper suggests that improving the carbon productivity of the least productive firms to that of the median firm in their industry would reduce carbon emissions by 71% to achieve the same output. Such policies could include those which promote technological diffusion and upgrading, provide investment incentives, access to finance to enable scale-up, and engagement in international trade and efficient global value chains. These policies can be specifically tailored to promote environmental performance. Evidence shows that only 4% of green recovery packages adopted in reaction to the Covid-19 crisis targeted the industrial sector, which accounts for 23% of global emissions (Aulie et al., 2023[1]). More ambitious technology adoption policies seem necessary in the manufacturing sector.

Third, the strong relationship between environmental and economic performance of firms, as well as their dispersion at the industry-level, suggest that there may be synergies rather than trade-offs between environmental and economic policies, especially when aiming to reduce dispersion between firms by helping unproductive firms improve. Policies to improve labour productivity may have co-benefits for carbon efficiency and aggregate emissions. This also provides encouragement that the green transition can be achieved without being detrimental to aggregate economic performance.

Finally, carbon pricing should continue to be used to internalise the negative externalities associated with pollution. Carbon pricing also provides a direct incentive to reduce carbon usage in the short term, by using cleaner fuel types or less energy, and in the long term through directed technical change. Carbon pricing could also be extended to cover more firms, not just the largest ones: the evidence presented in this paper shows that all types of firms respond to higher energy prices by improving their carbon productivity without significant economic effects. Carbon pricing also generates revenues that can be used in additional economic and environmental policies.

Future work on this project will focus on extending coverage to new countries and analysing the effects of additional environmental and economic policies.

References

Ahmadi, Y., A. Yamazaki and P. Kabore (2022), "How Do Carbon Taxes Affect Emissions? Plant-Level Evidence from Manufacturing", *Environmental and Resource Economics*, Vol. 82/2, pp. 285-325, https://doi.org/10.1007/s10640-022-00678-x. [38]

Albrizio, S., T. Kozluk and V. Zipperer (2017), "Environmental policies and productivity growth: Evidence across industries and firms", *Journal of Environmental Economics and Management*, Vol. 81, pp. 209-226, https://doi.org/10.1016/j.jeem.2016.06.002. [49]

Alpay, E., S. Buccola and J. Kerkvliet (2002), "Productivity growth and environmental regulation in Mexican and U.S. food manufacturing", *American Journal of Agricultural Economics*, Vol. 84/4, pp. 887-901, https://doi.org/10.1111/1467-8276.00041. [80]

Alpino, M., L. Citino and A. Frigo (2023), *The effects of the 2021 energy crisis on medium-sized and large industrial firms: evidence from Italy 1 Introduction1*, https://twitter.com/ben_moll. [59]

Amann, J. et al. (2021), "Switching it up: The effect of energy price reforms in Oman", *World Development*, Vol. 142, https://doi.org/10.1016/j.worlddev.2020.105252. [28]

André, C. et al. (2023), "Rising energy prices and productivity: short-run pain, long-term gain?", *OECD Economics Department Working Papers*, No. 1755, OECD Publishing, Paris, https://doi.org/10.1787/2ce493f0-en. [27]

Andrews, D., C. Criscuolo and P. Gal (2019), "The Best versus the Rest: Divergence across Firms during the Global Productivity Slowdown", *CEP Discussion Paper* 1645. [2]

Appelt, S. et al. (2022), *Micro-data based insights on trends in business R&D performance and funding: Findings from the OECD microBeRD+ project*, OECD Science, Technology and Industry Working Papers. [54]

Arquié, A. and M. Thie (2023), *Energy, Inflation and Market Power: Excess Pass-Through in France Working Paper*, http://www.cepii.fr. [60]

Aulie, F. et al. (2023), "Did Covid-19 accelerate the green transition?", *OECD SCIENCE, TECHNOLOGY AND INDUSTRY POLICY PAPERS*. [1]

Barrows, G. and H. Ollivier (2018), "Cleaner firms or cleaner products? How product mix shapes emission intensity from manufacturing", *Journal of Environmental Economics and Management*, Vol. 88, pp. 134-158, https://doi.org/10.1016/j.jeem.2017.10.008. [16]

Berlingieri, G., P. Blanchenay and C. Criscuolo (2017), "The Great Divergence(s)", *OECD Science, Technology, and Innovation Policy Papers*. [17]

Berlingieri, G. et al. (2017), *The MultiProd Project: A Comprehensive Overview*. [5]

Berlingieri, G. et al. (2020), "Laggard firms, technology diffusion and its structural and policy determinants", *OECD Science, Technology and Industry* 86. [18]

Bernard, J. and M. Kichian (2021), "The Impact of a Revenue-Neutral Carbon Tax on GDP Dynamics: The Case of British Columbia", *Energy Journal*, Vol. 42/3, pp. 205-223, https://doi.org/10.5547/01956574.42.3.jber. [72]

Borghesi, S. et al. (2012), *The European Emission Trading Scheme and environmental innovation diffusion: Empirical analyses using Italian CIS data The European Emission Trading Scheme and environmental innovation diffusion: Empirical analyses using Italian CIS data* *, http://www.unife.it/dipartimento/economia/pubblicazioni/quaderni-del-dipartimento/. [68]

Borusyak, K., P. Hull and X. Jaravel (2021), "Quasi-Experimental Shift-Share Research Designs", *Review of Economic Studies*. [94]

Boyd, G. and J. Lee (2019), "Measuring plant level energy efficiency and technical change in the U.S. metal-based durable manufacturing sector using stochastic frontier analysis", *Energy Economics*, Vol. 81, pp. 159-174, https://doi.org/10.1016/j.eneco.2019.03.021. [22]

Broner, F., P. Bustos and V. Carvalho (2019), *Sources of comparative advantage in polluting industries Sources of Comparative Advantage in Polluting Industries*. [76]

Brucal, A. and A. Dechezleprêtre (2021), "Assessing the impact of energy prices on plant-level environmental and economic performance: Evidence from Indonesian manufacturers"*, OECD Environment Working Papers*, No. 170, OECD Publishing, Paris, https://doi.org/10.1787/9ec54222-en. [23]

Brunel, C. (2017), "Pollution Offshoring and Emission Reductions in EU and US Manufacturing", *Environmental and Resource Economics*, Vol. 68/3, pp. 621-641, https://doi.org/10.1007/s10640-016-0035-1. [77]

Calel, R. (2020), "Adopt or innovate: Understanding technological responses to cap-and-trade", *American Economic Journal: Economic Policy*, Vol. 12/3, pp. 170-201, https://doi.org/10.1257/POL.20180135. [69]

Calel, R. and A. Dechezleprêtre (2016), "Environmental policy and directed technological change: Evidence from the european carbon market", *Review of Economics and Statistics*, Vol. 98/1, pp. 173-191, https://doi.org/10.1162/REST_a_00470. [56]

Calligaris, S. et al. (2023), *Employment dynamics across firms during COVID-19*. [55]

Calvino, F., C. Criscuolo and R. Verlhac (2020), "Declining business dynamism: Structural and policy determinants", *OECD Publishing*, Vol. November/94, pp. 1-74. [52]

Calvino, F. and L. Fontanelli (2023), *A portrait of AI adopters across countries*. [53]

Cameron, A. and M. Garrone (2024), "Carbon Intensity and Corporate Performance: A Micro-level Study of EU ETS Industrial Firms", https://doi.org/10.2873/188206. [65]

Capelle, D. et al. (2023), *Mitigating Climate Change at the Firm Level: Mind the Laggards*. [6]

Chen, H. and Y. Xu (2021), "Environmental regulation and exports: Evidence from the comprehensive air pollution policy in China", *International Journal of Environmental Research and Public Health*, Vol. 18/3, pp. 1-12, https://doi.org/10.3390/ijerph18031316. [75]

Colmer, J. et al. (2024), "Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence from the European Union Emissions Trading System", *Review of Economic Studies*, https://doi.org/10.1093/restud/rdae055. [32]

Colmer, J. et al. (2020), *Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence From the European Union Emissions Trading Scheme Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence from the European Union Emissions Trading Scheme* *, https://ssrn.com/abstract=3917883. [70]

Correia, S. (2017), "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator". [97]

Cox, M. et al. (2014), "Labor demand effects of rising electricity prices: Evidence for Germany", *Energy Policy*, Vol. 75, pp. 266-277, https://doi.org/10.1016/j.enpol.2014.10.021. [30]

Criscuolo, C. et al. (2024), "Digital and green spillovers across firm-level production networks", Forthcoming OECD Working Paper. [7]

Curtis, M. (2018), "Who loses under cap-and-trade programs? The labor market effects of the Nox budget trading program", *Review of Economics and Statistics*, Vol. 100/1, pp. 151-166, https://doi.org/10.1162/REST_a_00680. [88]

Curuk, M. and R. Rozendaal (2022), "LABOR SHARE, INDUSTRY CONCENTRATION AND ENERGY PRICES: EVIDENCE FROM EUROPE", https://ssrn.com/abstract=4212962. [62]

D'Arcangelo, F. et al. (2022), "Estimating the CO2 emission and revenue effects of carbon pricing: New evidence from a large cross-country dataset", *OECD Economics Department Working Papers*, No. 1732, OECD Publishing, Paris, https://doi.org/10.1787/39aa16d4-en. [46]

Dechezleprêtre, A., D. Nachtigall and F. Venmans (2023), "The joint impact of the European Union emissions trading system on carbon emissions and economic performance", *Journal of Environmental Economics and Management*, Vol. 118, https://doi.org/10.1016/j.jeem.2022.102758. [31]

Dechezleprêtre, A. and M. Sato (2017), "The impacts of environmental regulations on competitiveness", *Review of Environmental Economics and Policy*, Vol. 11/2, pp. 183-206, https://doi.org/10.1093/reep/rex013. [50]

Dussaux, D. (2020), "The joint effects of energy prices and carbon taxes on environmental and economic performance: Evidence from the French manufacturing sector", *OECD Environment Working Papers*, No. 154, OECD Publishing, Paris, https://doi.org/10.1787/b84b1b7d-en. [24]

Dussaux, D. and S. Monjon (2023), "Selling under other skies when energy prices skyrocket: How do the companies adapt their export strategy when energy prices rise?", *Energy Policy*, Vol. 183, https://doi.org/10.1016/j.enpol.2023.113777. [64]

Fabrizi, A. et al. (2024), "The impact of environmental regulation on innovation and international competitiveness", *Journal of Evolutionary Economics*, Vol. 34/1, pp. 169-204, https://doi.org/10.1007/s00191-024-00852-y. [90]

Fetzer, T., C. Palmou and Schneebacher (2024), *How do firms cope with economic shocks in real time?*, https://osf.io/5entz/. [29]

Ganapati, S., J. Shapiro and R. Walker (2020), "Energy cost pass-through in US manufacturing: Estimates and implications for carbon taxes", *American Economic Journal: Applied Economics*, Vol. 12/2, pp. 303-342, https://doi.org/10.1257/app.20180474. [61]

Goldsmith-Pinkham, P., I. Sorkin and H. Swift (2020), "Bartik instruments: What, when, why, and how", *American Economic Review*, Vol. 110/8, pp. 2586-2624, https://doi.org/10.1257/AER.20181047. [95]

Gray, W. and R. Shadbegian (2003), "Plant vintage, technology, and environmental regulation", *Journal of Environmental Economics and Management*, Vol. 46/3, pp. 384-402, https://doi.org/10.1016/S0095-0696(03)00031-7. [82]

Greenstone, M., J. List and C. Syverson (2012), "The Effects of Environmental Regulation on the Competitiveness of U.S. Manufacturing". [48]

Grolleau, G., N. Mzoughi and S. Pekovic (2012), "Green not (only) for profit: An empirical examination of the effect of environmental-related standards on employees' recruitment", *Resource and Energy Economics*, Vol. 34/1, pp. 74-92, https://doi.org/10.1016/j.reseneeco.2011.10.002. [85]

Grover, A. and V. Karplus (2020), *The Energy-Management Nexus in Firms Which Practices Matter, How Much and for Whom?*, http://www.worldbank.org/prwp. [58]

Gu, W., J. Hussain and M. Willox (2019), *Environmentally adjusted multifactor productivity growth for the Canadian manufacturing sector*, Statistics Canada = Statistique Canada. [73]

Hamamoto, M. (2006), "Environmental regulation and the productivity of Japanese manufacturing industries", *Resource and Energy Economics*, Vol. 28/4, pp. 299-312, https://doi.org/10.1016/j.reseneeco.2005.11.001. [81]

Harrison, K. and W. Antweiler (2003), *Incentives for Pollution Abatement: Regulation, Regulatory Threats, and Non-Governmental Pressures*, https://doi.org/10.1002/pam.10137. [78]

Hintermann, B. et al. (2024), *The effect of climate policy on productivity and cost pass-through in the German manufacturing sector* *. [71]

Horbach, J. (2010), *The impact of innovation activities on employment in the environmental sector: Empirical results for Germany at the firm level*, https://www.researchgate.net/publication/5101855. [86]

Horbach, J. and K. Rennings (2013), "Environmental innovation and employment dynamics in different technology fields - An analysis based on the German Community Innovation Survey 2009", *Journal of Cleaner Production*, Vol. 57, pp. 158-165, https://doi.org/10.1016/j.jclepro.2013.05.034. [87]

Hottenrott, H., S. Rexhäuser and R. Veugelers (2016), "Organisational change and the productivity effects of green technology adoption", *Resource and Energy Economics*, Vol. 43, pp. 172-194, https://doi.org/10.1016/j.reseneeco.2016.01.004. [91]

Hsieh, C. and P. Klenow (2009), *Misallocation and Manufacturing TFP in China and India*, https://academic.oup.com/qje/article/124/4/1403/1917179. [20]

Jaffe, A. and K. Palmer (1996), "Environmental regulation and innovation: A panel data study". [92]

Jaraite, J. and C. Di Maria (2016), "Did the EU ETS make a difference? An empirical assessment using Lithuanian firm-level data", *Energy Journal*, Vol. 37/1, pp. 1-23, https://doi.org/10.5547/01956574.37.1.jjar. [36]

Känzig, D. (2023), *NBER WORKING PAPER SERIES THE UNEQUAL ECONOMIC CONSEQUENCES OF CARBON PRICING*, http://www.nber.org/papers/w31221. [67]

Känzig, D. and M. Konradt (2023), *NBER WORKING PAPER SERIES CLIMATE POLICY AND THE ECONOMY: EVIDENCE FROM EUROPE'S CARBON PRICING INITIATIVES*, http://www.nber.org/papers/w31260. [66]

Klemetsen, M., K. Rosendahl and A. Jakobsen (2020), *THE IMPACTS of the EU ETS on NORWEGIAN PLANTS' ENVIRONMENTAL and ECONOMIC PERFORMANCE*, World Scientific Publishing Co. Pte Ltd, https://doi.org/10.1142/S2010007820500062. [37]

Klenow, P., E. Pastén and C. Ruane (2024), *Carbon Taxes and Misallocation in Chile*. [10]

Koźluk, T. and V. Zipperer (2013), "Environmental Policies and Productivity Growth: A Critical Review of Empirical Findings"*, OECD Economics Department Working Papers*, No. 1096, OECD Publishing, Paris, https://doi.org/10.1787/5k3w725lhgf6-en. [51]

Kumbhakar, S., O. Badunenko and M. Willox (2022), "Do carbon taxes affect economic and environmental efficiency? The case of British Columbia's manufacturing plants", *Energy Economics*, Vol. 115, https://doi.org/10.1016/j.eneco.2022.106359. [39]

Kuosmanen, N. and T. Maczulskij (2024), "Going green while getting lean: Decomposing carbon and green total factor productivity", *Journal of Environmental Management*, Vol. 352, https://doi.org/10.1016/j.jenvman.2024.120046. [15]

Lanoie, P. et al. (2011), "Environmental policy, innovation and performance: New insights on the porter hypothesis", *Journal of Economics and Management Strategy*, Vol. 20/3, pp. 803-842, https://doi.org/10.1111/j.1530-9134.2011.00301.x. [79]

Levinson, A. (2015), "A direct estimate of the technique effect: Changes in the pollution intensity of US manufacturing, 1990–2008", *Journal of the Association of Environmental and Resource Economists*, Vol. 2/1, pp. 43-56, https://doi.org/10.1086/680039. [12]

Lutz, B. (2016), *Emissions Trading and Productivity: Firm-level Evidence from German Manufacturing*, http://ftp.zew.de/pub/zew-docs/dp/dp16067.pdf. [34]

Lu, Y. and J. Pless (2022), *Growing the Economy with Green Regulation: Evidence from Industrial Firm Productivity in China*. [83]

Lyubich, E., J. Shapiro and R. Walker (2018), "Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Environmental Policy", *AEA Papers and Proceedings*, Vol. 108, pp. 136-142, https://doi.org/10.1257/pandp.20181089. [4]

Maria D'arcangelo, F., G. Pavan and S. Calligaris (2022), *The Impact of the European Carbon Market on Firm Productivity: Evidence from Italian Manufacturing Firms The Impact of the European Carbon Market on Firm Productivity: Evidence from Italian Manufacturing Firms The Impact of the European Carbon Market on Firm Productivity: Evidence from Italian Manufacturing Firms. *.* [35]

Marin, G. and F. Vona (2021), "The impact of energy prices on socioeconomic and environmental performance: Evidence from French manufacturing establishments, 1997–2015", *European Economic Review*, Vol. 135, https://doi.org/10.1016/j.euroecorev.2021.103739. [25]

Martin, R., L. de Preux and U. Wagner (2014), "The impact of a carbon tax on manufacturing: Evidence from microdata", *Journal of Public Economics*, Vol. 117, pp. 1-14, https://doi.org/10.1016/j.jpubeco.2014.04.016. [43]

Martinsson, G. et al. (2022), *Carbon Pricing and Firm-Level CO 2 Abatement: Evidence from a Quarter of a Century-Long Panel*. [44]

Mertens, M., S. Müller and G. Neuschäffer (2022), *Identifying rent-sharing using firms' energy input mix Standard-Nutzungsbedingungen*, IWH Discussion Papers, http://www.iwh-halle.de. [63]

Murray, B. and N. Rivers (2015), "British Columbia's revenue-neutral carbon tax: A review of the latest "grand experiment" in environmental policy", *Energy Policy*, Vol. 86, pp. 674-683, https://doi.org/10.1016/j.enpol.2015.08.011. [40]

Najjar, N. and J. Cherniwchan (2021), "Environmental regulations and the cleanup of manufacturing: Plant-level evidence", *Review of Economics and Statistics*, Vol. 103/3, pp. 476-491, https://doi.org/10.1162/rest_a_00904. [47]

Petrick, S. (2013), *Carbon efficiency, technology, and the role of innovation patterns: Evidence from German plant-level microdata*, Kiel Working Paper, http://hdl.handle.net/10419/70254. [13]

Petrick, S., K. Rehdanz and U. Wagner (2010), *Energy Use Patterns in German Industry: Evidence from Plant-level Data*, http://ssrn.com/abstract=1689044https://ssrn.com/abstract=1689044Electroniccopyavailableat:http://ssrn.com/abstract=1689044. [9]

Petrick, S. and U. Wagner (2014), *The impact of carbon trading on industry: Evidence from German manufacturing firms*, http://hdl.handle.net/10419/94357. [33]

Rexhäuser, S. and C. Rammer (2014), "Environmental Innovations and Firm Profitability: Unmasking the Porter Hypothesis", *Environmental and Resource Economics*, Vol. 57/1, pp. 145-167, https://doi.org/10.1007/s10640-013-9671-x. [93]

Rottner, E. and K. Von Graevenitz (2022), *What drives Carbon Emissions in German Manufacturing: Scale, Technique or Composition?*. [14]

Sato, M. et al. (2019), "International and sectoral variation in industrial energy prices 1995–2015", *Energy Economics*, Vol. 78, pp. 235-258, https://doi.org/10.1016/j.eneco.2018.11.008. [21]

Shapiro, J. and R. Walker (2018), *Why is pollution from us manufacturing declining? the roles of environmental regulation, productivity, and trade*, American Economic Association, https://doi.org/10.1257/aer.20151272. [11]

Steinbuks, J. and K. Neuhoff (2014), "Assessing energy price induced improvements in efficiency of capital in OECD manufacturing industries", *Journal of Environmental Economics and Management*, Vol. 68/2, pp. 340-356, https://doi.org/10.1016/j.jeem.2014.07.003. [26]

Syverson, C. (2011), *What determines productivity*, https://doi.org/10.1257/jel.49.2.326. [19]

Tian, R. (2020), “Emissions Trading Schemes and Directed Technological Change: Evidence from China”. [74]

Von Graevenitz, K. and E. Rottner (2020), *Energy use patterns in German manufacturing since 2003; Energy use patterns in German manufacturing since 2003*, http://hdl.handle.net/10419/214229. [8]

Vona, F. et al. (2018), “Environmental regulation and green skills: An empirical exploration”, *Journal of the Association of Environmental and Resource Economists*, Vol. 5/4, pp. 713-753, https://doi.org/10.1086/698859. [84]

Wagner, U. et al. (2020), *Carbon Footprints of European Manufacturing Jobs: Stylized Facts and Implications for Climate Policy Carbon Footprints of European Manufacturing Jobs: Stylized Facts and Implications for Climate Policy* *, https://ssrn.com/abstract=3577550. [3]

Walker, W. (2013), “The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce”, *Quarterly Journal of Economics*, Vol. 128/4, pp. 1787-1835, https://doi.org/10.1093/qje/qjt022. [89]

Werner, S. (2016), “European District Heating Price Series”, Energiforsk. [57]

Yamano, N. and J. Guilhoto (2020), “CO2 emissions embodied in international trade and domestic final demand: Methodology and results using the OECD Inter-Country Input-Output Database”*, OECD Science, Technology and Industry Working Papers*, No. 2020/11, OECD Publishing, Paris, https://doi.org/10.1787/8f2963b8-en. [96]

Yamazaki, A. (2022), “Environmental taxes and productivity: Lessons from Canadian manufacturing”, *Journal of Public Economics*, Vol. 205, https://doi.org/10.1016/j.jpubeco.2021.104560. [41]

Yamazaki, A. (2017), “Jobs and climate policy: Evidence from British Columbia’s revenue-neutral carbon tax”, *Journal of Environmental Economics and Management*, Vol. 83, pp. 197-216, https://doi.org/10.1016/j.jeem.2017.03.003. [42]

Zhang, H. and J. Wu (2022), “The Energy Saving and Emission Reduction Effect of Carbon Trading Pilot Policy in China: Evidence from a Quasi-Natural Experiment”, *International Journal of Environmental Research and Public Health*, Vol. 19/15, https://doi.org/10.3390/ijerph19159272. [45]

Annex A. Data collection procedure

Input data requirements

The module requires at least one dataset containing information on energy use. The data may be at the facility, establishment, or firm-level, but it must be possible to convert the data to the firm-level (as discussed above). All data should be at the unit-year level. The data should contain a unit identifier (and a firm identifier if the unit is not the firm), year, industry, and employment, as well as the environmental data. The data must be representative of industries, either by covering the full census of firms or by providing sampling weights.

Additionally, it is optional to provide production or balance sheet data, and/or a firm or establishment level business register. It should be possible to match these to the environmental data source at the firm-level. Information on production that can be matched at the firm-level, ideally for all firms in the environmental survey, is highly desirable for the project as it ensures that representativeness can be maintained.

All monetary units should be in single units of local currency. If the country adopted the euro, it should transform the series prior to the introduction of the euro into euros using the fixed official exchange rate at the time of accession.

Employment must also be provided. Ideally, the employment variable would be the headcount of employment, as is the baseline in MultiProd, and should be consistently defined in each of the provided sources.

Energy use data

Energy use data can cover businesses' total expenditure on energy, their expenditure by type of energy, and quantities of energy consumed in the production process. Using this information, combined with emissions factors, it is possible to obtain an estimate of the total amount of emissions produced from firms' use of energy.

The following types of energy could be included and have available data on prices:

- Electricity
- Heat (or steam)
- Natural gas
- Coking coal
- Other bituminous coal
- Petroleum coke
- Fuel oil
- Gas and diesel oil
- Kerosene-type jet fuel
- Other kerosene

Additionally, the module has been updated to include additional energy types for countries that have data on energy quantity by type. The additional energy types included are:

- Hard coal
- Liquified petroleum gas
- Lignite
- Coke oven gas.

Production survey data

Data on economic variables at the firm-level can be obtained from production survey data. It must be possible to match the data at the firm-level to the environmental data source. For all the countries included in the sample, it is possible to match the data to all firms in the environmental survey: either the production survey covers the census of firms or the environmental questions are included in the production survey.

Variables that could be obtained from the production data are:

- Birth year
- Revenue (or gross output)
- Value added
- Total fixed assets
- Tangible fixed assets
- Intangible fixed assets
- Labour costs (wage bill)
- Intermediate input costs
- Research and development (total)
- Exports value
- Imports value
- Investment
- Investment in tangible assets
- Investment in intangible assets
- Total assets
- Net worth
- Total debt
- Cash holdings
- EBITDA (earnings before interest, taxes, depreciation, and amortization)

Business register data

Business register data has numerous uses in the module. A business register covers the population of firms and typically contains information on firm size (measured by employment), industry, and structure. It could be at the establishment or the firm-level.

If the environmental survey is at the establishment level, it is highly desirable to also have a business register at the establishment level. The module ultimately conducts all analysis at the firm-level, so it is necessary to convert establishment level data to the firm-level (see below for further discussion). In this case, an establishment level register is used to detect what share of a firm's total employment is covered in the survey and this is used to scale-up the values when converting to the firm-level.

The business register, at either establishment or firm-level, is also used to construct an industry concordance from ISIC revision 3.1 to ISIC revision 4, where there is a many-to-many correspondence. The business register is helpful here as the full panel structure of firms can be exploited.

Finally, the business register may be used to retrieve a firm identifier, firm-level employment, or the firm's industry, when these variables are not available in the environmental survey.

External datasets used to estimate emissions

Quantity of energy by industry

Data on quantities of energy by fuel type and industry are collected from two sources. First, Eurostat provides data on the quantity of energy used by type for each industry. The data is only available for manufacturing industries and has been collected since 2018, although many countries do not have data until later years. All quantities are converted into TJ using conversion factors provided by Eurostat. This data has the advantage of having very detailed energy types available but is not available for services industries, non-European countries, or for many years.[34]

Therefore, this is supplemented with data from the IEA, which includes the services sector (as a whole) and countries outside Europe, as well as being available for many years. The data collects information on six main energy types. These data are used as the main source of information on energy, and are complemented with the Eurostat data to split out the more aggregated energy types to more detailed energy types (e.g. gas to natural gas and refinery gas) using the shares of each sub-energy type from Eurostat to distribute the IEA values.[35] Specifically, in the Eurostat data, the sum of detailed energy types is taken within each match to a more aggregated energy type in the IEA data. Then the shares are computed of each detailed energy type in the Eurostat data. These shares are used in the IEA data to split out the more aggregated energy types.

Data on energy use by industry is not available from IEA or Eurostat for Indonesia but can instead be constructed from the microdata (and is anyway not necessary for the construction of the baseline CO2 measure because firm-level information is available).

Energy prices

Energy price data is available by country, energy type, and year. Four sources of data are used. First, IEA provides energy prices of many types of energy.[36] Second, Eurostat provides data on prices of electricity and gas faced by industry, with a breakdown of price according to the amount of energy consumed.[37] Third, Eurostat HICP provides a heat price index, but no indication of the level of the price for heat.[38] Fourth, data from Werner (2016[57]) is used to provide the level of heat price in a specific year, which can

[34] https://ec.europa.eu/eurostat/cache/metadata/en/nrg_quant_esms.htm

[35] https://www.iea.org/data-and-statistics/data-product/energy-efficiency-indicators

[36] https://www.iea.org/data-and-statistics/data-product/energy-prices

[37] https://ec.europa.eu/eurostat/cache/metadata/en/nrg_pc_202_sims.htm and https://ec.europa.eu/eurostat/cache/metadata/en/nrg_pc_204_sims.htm

[38] https://ec.europa.eu/eurostat/cache/metadata/en/cult_exp_esms.htm

be combined with the Eurostat HICP heat price index to produce the level of heat prices over the full sample period.[39]

Prices include taxes and levies, as is desirable for the analysis, and are converted into units of national currency per TJ using conversion factors from Eurostat and OECD exchange rate data. Eurostat data is available on a half-yearly basis: the average is taken over each year. The datasets are collated to provide panel data on energy prices by type and country.

Imputation method

There are missing values in the collated dataset. To impute values to fill these gaps, a weighted average across other countries' energy price is used. The inference is conducted for each energy type separately. The main rationale for using the prices of other countries to infer missing prices is that many energy prices are determined on a world market, meaning that the values of other countries is likely to give a better estimate of the true price than by, for example, simply making a linear interpolation.

The procedure uses the average value from a group of similar countries to impute missing values. To define this group of countries, optimal weightings are determined by comparing the prices of the target country with prices of other countries for the specific energy type of interest in the years where the target country has an available price. An algorithm (the Stata package "synth") is used to assign weights to all other countries based on their similarity to the target country. These weights are then applied to the years with missing values to impute as value as the weighted average across similar countries. If no data is available for the energy type-country in any year, the average weights constructed using other energy types are used.

For example, Croatia has data on electricity prices from 2005 onwards but not before 2004. For 2005 onwards, the values of Croatia's electricity price is compared with all other available countries' prices. The countries with the biggest weights for Croatia's electricity price are Germany (28.5%), Slovakia (19%), and Türkiye (16.7%).

The procedure is tested by setting some values to missing, applying the procedure to impute these values, and then comparing the imputed values to the true values. This testing suggests that the imputation procedure works well and provides a relatively accurate prediction of missing values.

Estimating price as a function of quantity consumed using available data

The price paid for each fuel type may depend on the quantity of that fuel type purchased. In particular, firms that consume larger quantities may get preferential rates. Therefore, using average prices that do not vary by quantity purchased may underestimate the true dispersion in emissions across firms.

Eurostat provides prices that vary by quantity consumed for electricity and natural gas – the two main energy types for most countries. For each year and country, the price is provided for different consumption bands. For example, in France in Q1 2019, the price per TJ of electricity was 51111 EUR if consuming less than 0.072 TJ of electricity, but was 15 916 EUR if consuming more than 540 TJ.

In practice, it is likely that, on average, the price that is paid for electricity or gas is a continuous function of quantity (instead of being fixed within a quantity band). Furthermore, to compute emissions, it is preferable to work with price as a continuous function of quantity. Therefore, prices are estimated as a continuous function of quantity, taking the mid-point of each quantity band. The price data is available on a half-yearly basis, so the average over each year is taken to be in line with other data used in the project.

[39] https://energiforskmedia.blob.core.windows.net/media/21926/european-district-heating-price-series-energiforskrapport-2016-316.pdf

Testing suggests that a log-log specification provides the best fit. Specifically, linear regression is used to estimate:

$$logP_{nct} = \alpha + \beta logQ_{nct} + \varepsilon_{nct}$$

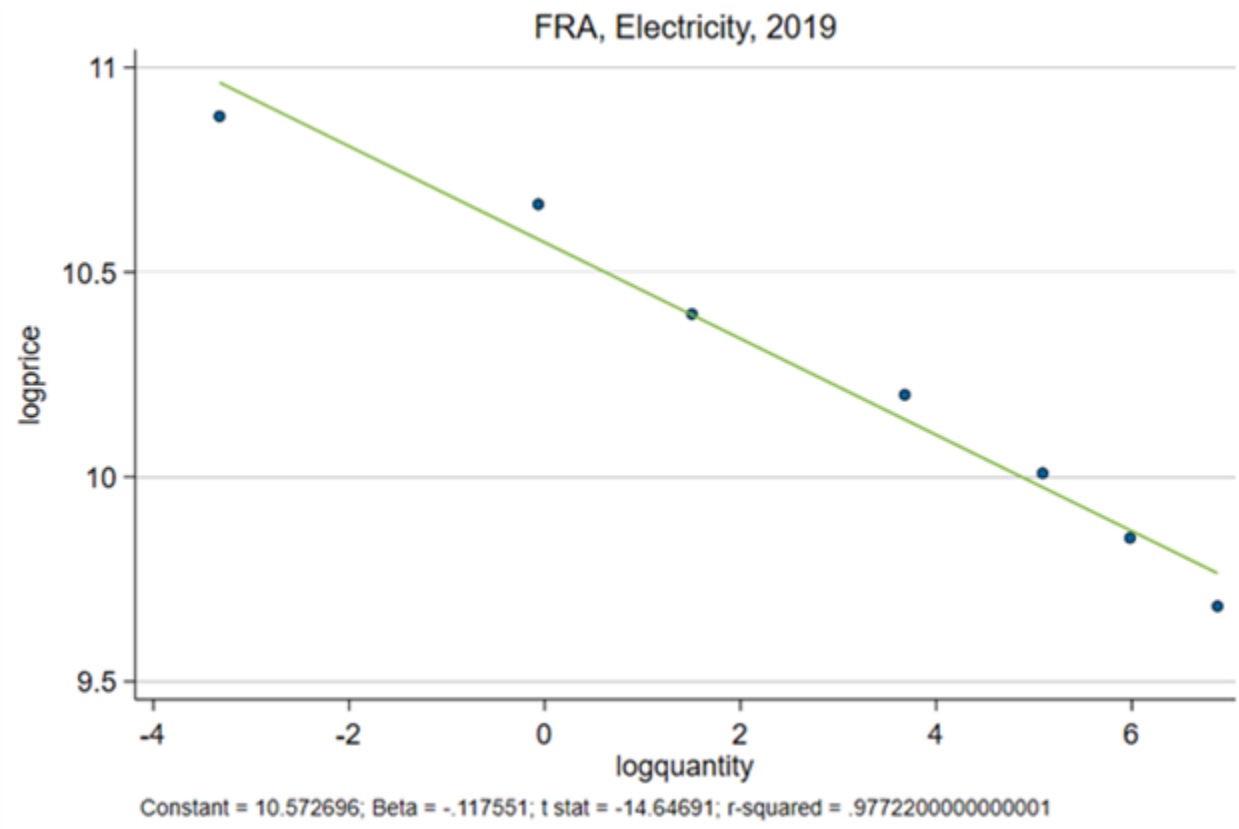
Equation 16

For each country, energy type, and year. In each regression, there are only between 5 and 9 data points because prices are only provided for this many quantity bands. Despite this, the regressions fit well. Finally, to avoid extrapolating extreme prices far outside of the observed price range, the price is restricted to be within the range of 0.1 times the minimum and 2 times the maximum actually observed in the data (i.e. even if a firm consumes a huge quantity of energy, it is not assumed to pay a price lower than the minimum price). These bounds are chosen for the minimum and maximum based on observations of prices in the French data, which were often below the minimum price observed in the Eurostat data, but less often observed to be greater than the maximum. Note that this characteristic of energy prices is further motivation for using the estimated monotonic relationship between price and quantity instead of simply using the bands, as it allows for lower prices than the minimum, which is commonly observed in the data.

The chart below shows this procedure for electricity in France in 2019. The R-squared is 0.977, suggesting that the model in logs provides a very tight fit.

Figure A A.1. Estimating price as a continuous function of quantity fits well

Log price of electricity against the log of the quantity band midpoints and associated regression line for electricity in 2019 for France



Note: The chart the relationship between log price and the log of the quantity band mid points. It also shows the associated OLS regression line and coefficients. The chart is for electricity in 2019 in France.
Source: Author's calculations using data from Eurostat

For countries where variable price data is missing, values are imputed using the same procedure described above. Furthermore, in cases where prices are variable in year t-h and t+l (for h,l>0), but not variable in period t, the coefficients are linearly interpolated (this applies to relatively few cases).

Finally, prices that vary with quantity are only available for electricity and natural gas, but the prices of all energy types are likely to vary with quantity. Therefore, the price gradients for electricity and gas are applied to other energy types within a country-year. First, the minimum price is compared with the average price for electricity and/or gas. If gradients are available for both then the average of this ratio is taken, otherwise it is taken for the only energy type with an available gradient. This ratio is then used to compute a minimum price for the new energy type using the available average price. The same is then done for the maximum. The gradient can then be applied to the new energy type, again taking the average slope across both electricity and gas if both are available. Doing this improves the estimation of true energy quantity and emissions in the French data. The same procedure can be applied to other countries in the sample.

Emissions factors

Emissions factors provide estimates of the amount of carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O) emitted from the consumption of a unit of each energy type. The emissions factors vary by country and over time. The data is provided by IEA.[40] The units are converted to tonnes of CO2 equivalent per TJ of energy, for each energy type. For electricity, the emissions factor represents the emissions from electricity produced from all sources, according to the production methods of that country. For fossil fuels, the emissions factor is from the direct combustion of that energy type.

As with the energy price data, there are some missing values. An equivalent method of imputation is used to fill these values, taking the weighted average across similar countries.

Industry code conversion

Most OECD Member countries changed from using the ISIC revision 3.1 to ISIC revision 4 industry classification system around 2008. The correspondence between ISIC rev. 3.1 and ISIC rev 4 is many-to-many (m:m), so it is not straightforward to obtain a consistent industry classification going back before 2008. Therefore, a conversion procedure in the data is used. The industry conversion procedure uses firm-level panel data to identify the relevant correspondences between the two systems.

First, changeover proportions are produced by taking firms observed in the data in both the year before the changeover and the year of the changeover. This produces a m:m crosswalk. If there are any industries for which a crosswalk cannot be created (because there are no panel firms in those industries), the crosswalk is supplemented with an external crosswalk taken from the MultiProd project (Berlingieri et al., 2017[5]).

Then an industry code is assigned according to the following procedure:

1. If the firm already has the new industry code, this is the baseline industry code.
2. If the proposed match from the crosswalk is unique, then the uniquely proposed industry code is used.

If panel data is available for a firm and it is observed to transition to a stable industry code in the ISIC rev 4 classification, assign this as the industry code for the firm in earlier years so long as the industry code in the earlier years has a possible correspondence to the later industry code.

[40] https://www.iea.org/data-and-statistics/data-product/emissions-factors-2

a. For example, suppose the ISIC rev 3.1 industry A maps to an ISIC rev 4 industry Y with 90% probability and Z with 10% probability. A firm is observed with industry A in 2007 and industry Z from 2008 onwards. Then this firm is assigned to have industry Z in 2007. However, if in 2005 the same firm is observed with an ISIC rev 3.1 industry code B, which only maps to industry X, the firm would be assigned industry X.

b. This procedure aims to use information about the firm. For example, publishing from ISIC rev 3.1 splits into a manufacturing sector (Printing and reproduction of recorded media) and a services sector industry (publishing activities). If a firm is observed to be in the services sector after 2008 in ISIC rev 4, it makes sense to assume that they were in the services sector before, rather than going straight for probabilistic matching.

3. Finally, probabilistic matching is applied using the crosswalk constructed. The firm is assigned a new industry code probabilistically in the first year in which it is observed without the new industry code. Then the remaining years of the firm are filled, keeping the industry code constant. A seed is set to ensure that the same probabilistic mapping is applied in each run of the code.

The most common industry observed for each firm is assigned to the firm throughout the entire timespan. This ensures that industry trends are not driven by firms switching in and out of the industry classification. Of course, the disadvantage is that there may be true industry switches that are held constant.

Industry variables may be included in each data source. Therefore, a preferred industry code must be chosen. The first choice is the business register, followed by the production survey, and finally the environmental survey.

Estimating emissions

Estimating CO2 emissions when only total energy spending is available

To outline the procedure, the following variables are defined:

S_{it} = total spending on energy by firm i, at time t. Each firm operates in an industry k and in a country c.

Q_{nkct} = quantity of fuel type n used by industry k in country c at time t. All quantities are in tera joules (TJ).

P_{nct} = price of fuel type n in country c at time t. This can also vary according to the quantity of the fuel type consumed, in which case it is denoted $P_{nct}(Q_n)$. All prices are for a tera joule of energy.

F_{nct} = emissions factor for fuel type fuel type n in country c at time t. Emissions can be computed as F×Q.

In all cases, the energy types included in the analysis are restricted to be those for which data is available on all variables. In practice, price data is available for the fewest energy types, so only the energy types with price data are included (these are coking coal, electricity, fuel oil, gas oil and diesel oil, heat, kerosene-type jet fuel, natural gas, other bituminous coal, other kerosene, and petroleum coke). Where possible, energy types that are included in the energy use data are matched to the most similar energy type in the price data (e.g. lignite to coking coal), making the assumption that they follow similar price trends. The shares of quantity consumed always sum to 1, so proportionality assumptions ensure that emissions are not underestimated due to these restricted energy types.

First, the simple procedure used to estimate emissions using constant prices is outlined before the preferred procedure which implements an algorithm to estimate emissions using prices that vary by quantity consumed is described.

Using consumption-invariant prices

Basic data on energy prices are available at the energy type-country-year level, meaning that they do not depend on the volume of energy consumed. When these are used, it is still relatively straightforward to compute emissions. Total quantity consumed by the firm is estimated first. Note that the following identity for spending holds:

$$S = \sum_n Q_n \times P_n = Q \sum_n \frac{Q_n}{Q} \times P_n$$

Equation 17

Where the second equality holds because quantity and price are independent in this setting (i.e., this requires prices that do not depend on the quantity of energy consumed).

Applying this to the available data, the total quantity of energy used by the firm is estimated as:

$$\widehat{Q_{it}} = \frac{S_{it}}{\sum_n \frac{Q_{nkc}}{Q_{nc}} \times P_{nct}}$$

Equation 18

Note that all estimated values are denoted with a hat. The denominator is the price index for a TJ of energy. This requires two assumptions:

Assumption (1): the firm's quantity shares of each energy type are assumed to be equal to the industry-country quantity shares.

Assumption (2): prices are determined at the national level (there is no variation in prices across firms).

Now that firm-level quantity has been estimated, total emissions $\widehat{M_{it}}$ can be computed as:

$$\widehat{M_{it}} = \widehat{Q_{it}} \sum_n \frac{Q_{nkc}}{Q_{nc}} \times F_{nct}$$

Equation 19

Using prices that vary with quantity consumed

The preferred estimate of emissions uses information on prices as a function of quantity for each energy type. Intuitively and in the data, firms that consume more energy pay a lower average price per unit. Therefore, ignoring the variation in prices would lead to an underestimate of energy consumed and of emissions for large firms and an overestimate for small firms.

The firm's quantity of each type using observed firm total spending, industry quantity shares by type, and prices must be solved. Recall that the equation for spending and quantity by type can be written as:

$$S = \sum_n Q_n \times P_n$$

Equation 20

The problem now is that the price of each energy type is a function of the quantity of that energy type, while the quantity consumed of that energy type is also a function of the price. Furthermore, the functions are non-linear and cannot be easily solved analytically.

To solve for the total quantity of energy used by each firm, the following maximisation problem is solved:

$$\max_{q} Q$$
$$s.t.$$
$$logP_n = \alpha_n + \beta_n logQ_n$$
$$S = Q \sum_n P_n \frac{Q_n}{Q}$$

Where the share of each quantity is fixed and is observed in the external data (but is updated ex-ante based on firm size) and the price parameters are determined by this estimation procedure Annex A with limits on the minimum and maximum price also imposed. The parameters in blue are fixed and observed in the data, while the variables in red are to be determined.

This maximisation problem finds the largest quantity for which the implied prices of each energy type are consistent with the total spending of the firm. Note that by maximising Q instead of minimising Q, it is assumed that firms behave optimally in the sense that they consume the maximum quantity of energy possible, given their spending. Put differently, there may be smaller values of Q that satisfy the constraints, but it is assumed that firms would not choose these combinations as they could consume more energy for the same total spending.

The parameters vary with each industry, country, and year. Therefore, solutions to the optimisation must be found for all values of total spending in each industry-country-year. The algorithm to estimate the quantity variable is as follows:

1. Order observations in decreasing order of total spending within an industry-country-year (biggest spending first).
2. Beginning with the firm with the largest expenditure.
    a. Propose that the firm pays the smallest possible prices (with bounds imposed on the minimum and maximum possible price based on the energy price data).
    b. Compute the largest possible quantity that could be consumed, given the firm's expenditure.
    c. If this quantity implies an expenditure larger than the firm's actual expenditure, reduce the value of the observed quantity.
    d. Continue to reduce the proposed quantity until implied expenditures equal observed expenditures.
3. Use the final quantity of the previous firm as the first proposed quantity of the firm with the next biggest expenditure.
    a. Reduce proposed quantity until implied spending equals observed spending.
4. Repeat this procedure for all firms.

To improve efficiency of the algorithm by minimising the number of computations, quantities are updated downwards in large jumps initially until the difference between implied and proposed expenditure becomes negative, then is updated upwards again until it is positive, and so on until convergence to a difference of 0.

In this setting, assumption (2) from before can be relaxed to:

Assumption (2*): Prices do not vary across firms, conditional on their quantity consumed.

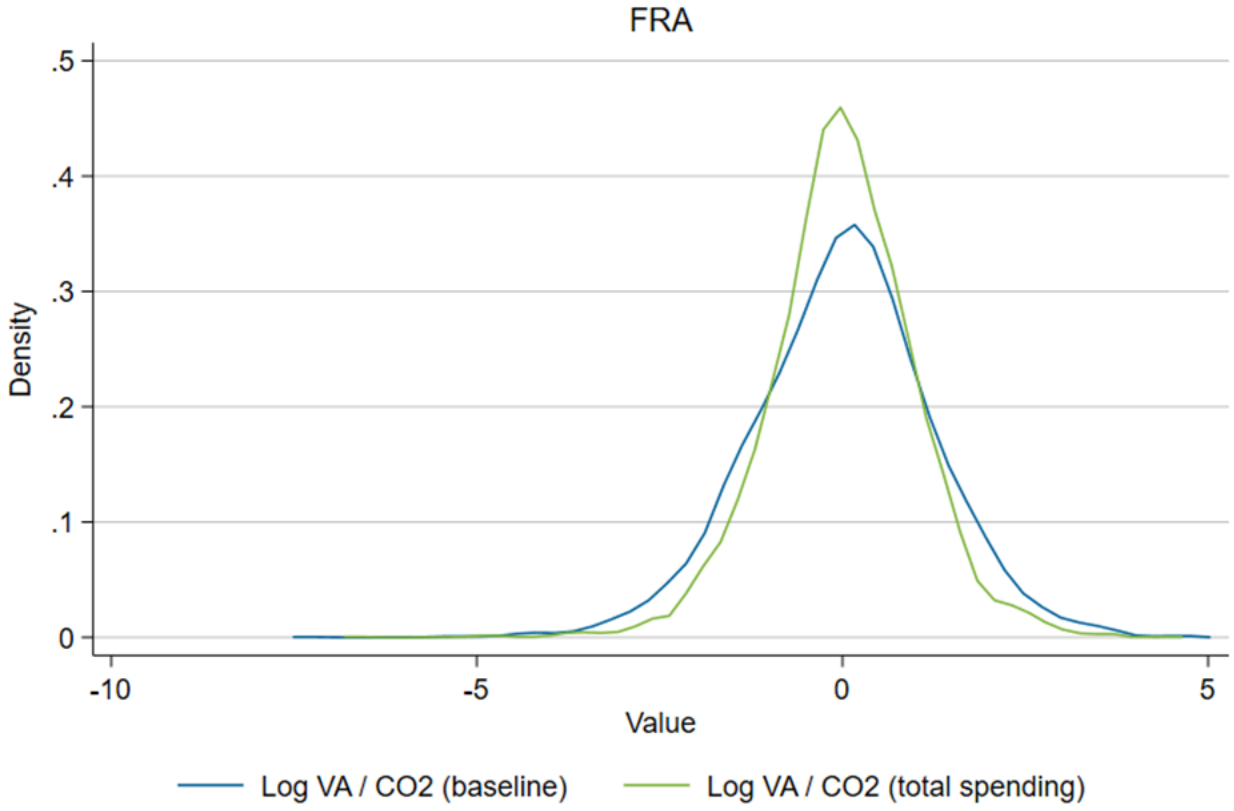
Two additional adjustments are also made to the quantity shares of each energy type $\frac{Q_n}{Q}$ using detailed French data. First, in some cases, quantity shares are only available at the 2 or 3 digit level. The French data is used to compare the average share of electricity at the 2 (or 3, accordingly) digit level with the average share at the 4 digit level. This ratio is then used to adjust the 2 (or 3) digit share data for electricity

in line with the ratio, and the inverse for fossil fuels. Second, even within 4 digit industries, there is some systematic variation in the types of energy that firms use, in particular with firm size. Therefore, a firm-level regression is estimated in France for each 2 digit industry, with the log of the electricity share as the dependent variable and the log of employment as the explanatory variable. The estimated coefficient is then applied to other countries to adjust up or down the electricity share, and the inverse for fossil fuels.[41]

Once total quantity by type has been computed, it is straightforward to estimate emissions using the same procedure as described above.

Figure A A.2. Using only total energy spending to estimate emissions slightly underestimates dispersion in carbon productivity

Distribution of carbon productivity when using data on quantity of energy use by type of fuel and when using total energy spending only



Note: The charts plot the within-industry-year distributions of carbon productivity across firms in France. CO2 is estimated using data on energy use by type (in blue) or just using total energy spending (in green). Industries are defined at the 4 digit level. The distributions are centred at 0. The sample is restricted to manufacturing sectors and firms with at least 20 employees.
Source: authors' calculations

[41] Alternative regression specifications were explored with more explanatory variables but this simple version seemed to be most accurate. Only parameters on the electricity share are used to reduce issues with adding up constraints after the adjustments.

Converting establishment to firm-level data

The analysis of this project is conducted at the firm-level while some countries (including France) collect environmental data at the establishment level. It is therefore necessary to aggregate establishment level data to the firm-level.

If all the establishments within a firm are included in the sample, then it is straightforward to aggregate variables to the firm-level by simply taking the sum (because all variables considered by the module are additive). However, not all establishments within a firm may be included in the environmental survey. In this case, simply aggregating the observed values from the establishment level data would lead to an underestimate of firm-level values.

To overcome the issue of incomplete coverage of establishments within a firm, the employment covered by surveyed establishments is compared to the employment of the firm. There are numerous possible methods to obtain a value for the share of employment covered in the environmental survey, depending on the data sources available. If an establishment level business register is available, then the environmental survey is matched to the register at the establishment level and the share of employment covered it computed. Otherwise, the total employment of establishments in the environmental survey is computed and compared to the best available data source for firm-level employment (production survey or firm-level business register).

The total values (e.g. in energy quantity or emissions) of the establishments in the environmental survey are then scaled up by the ratio of the firm's employment to the sum of establishment employment. For example, if the total employment of establishments is 50 while the firm's employment is 100, the sum of CO2 from the environmental survey would be scaled up by a factor of 2 for this firm. The scaling factor is bounded to be at least 1 (so that the sum of establishment values within a firm is never scaled down) and at most 100 (so that the values are not scaled by a huge value). Note that if establishment level values are scaled by an implausibly large factor, the outlier detection methodology described below will smooth or remove implausible outlier values by comparing with other variables available at the firm-level.

The key assumption of this method is that all environmental variables can be scaled proportionally to employment shares. Testing by randomly dropping establishments within a firm for which all establishments are observed (checked in the business register data) and conducting the methodology, then comparing results against the firm's true value, suggests that this performs well.

Method for converting sample weights to the firm-level

While most variables, such as CO2, can simply be summed across establishments, it is not accurate to simply sum weights across establishments. Furthermore, it would not be accurate to take the average of weights. To see this, consider the example below of a firm with two establishments, where the sample weight is an inverse probability weight, and suppose this is the only firm in the sample which is representative of the full economy. To obtain the total weighted CO2 of the aggregate economy from the sampled establishment data, the computation would be 1 * 100 + 10 * 10 = 200, where the first number in each multiplication is the weight and the second is the observed value of CO2. Once the data is aggregated to the firm-level, if the average or sum of the sample weights is taken, the estimate of the total economy CO2 would be wrong (it would be estimated as 5.5 * 110 = 605 or 11 * 110 = 1210, respectively).

The proposed solution to aggregate weights from the establishment to the firm-level is the following. Suppose a sample weight w_s is observed for establishment s. To obtain an accurate sample weight at the firm-level i for a target variable x_s (in the above example the target variable is CO2), the following formula is used:

$$w_{x,i} = \sum_{i \in f} \frac{x_s}{\sum_{s \in i} x_s} w_s$$

Equation 21

This formula corresponds to the within-firm target variable weighted average of the establishment weights. Returning to the example from above, the firm's weight for CO2 would be (100/110)*1 + (10/110)*10 = 20/11. Then, it is possible to obtain the total CO2 of the economy using this weight: 110*20/11 = 200.

Table A A.1. Illustrative example for aggregating establishment weights to the firm-level

Establishment ID	Firm ID	Sample weight	CO2 (establishment)	CO2 (firm)
1	1	1	100	110
2	1	10	10	110

This method can also be used consistently with the scaling described above. Under the assumption required for the scaling to be accurate – that the target variables increase proportionally with employment – the weight scaling method still obtains macro aggregates. To see this, assume that only establishment 2 is observed in the environmental survey data so that establishment 1 has a sample weight of 0 (because it is not in the sample) while the sample weight of establishment 2 is increased to 20. This is shown in the example below. The true total CO2 in the economy is still 200. Assume that the perfect scaling factor of 11 is observed so that CO2 can be accurately scaled to the firm-level. Then, the sample weight of the firm is 20 * 10 / 110 = 20/11, as before.

Table A A.2. Illustrative example for aggregating establishment weights to the firm-level with scaling

Establishment ID	Firm ID	Sample weight	CO2 (establishment)	CO2 (firm)
1	1	0	100	110
2	1	20	10	110

Detection of extreme outliers to reduce data entry errors

For some countries, the raw data cannot be directly accessed. Therefore, it is important to include procedures to detect and clean extreme outliers on the firm-level data as these may be data entry errors, and these may be particularly common in energy use surveys where the units of measurement are variable, and conversions can be complex.

Significant errors in data entry, such as entering a value in single units when it is supposed to be entered in thousands, can lead to major incorrect outliers that significantly affect results. To try to detect such errors, two methods are used. First, and most reliably, the within-firm time series is used to detect major jumps in specific variables. Second, extreme cross-sectional outliers are detected by comparing observations with other firms within an industry in each year. The procedures are described in detail in this sub-section.

The aim of these procedures is to detect data entry errors while not affecting genuine outliers. Indeed, outliers may be some of the most important observations economically and should not be dropped. Yet, given the nature of the survey data and the decentralised manner in which the codes are run, it is important to impose some provisions for errors.

Within-firm detection

With sufficient firm-level panel data, it is possible to identify implausible jumps in variables within a firm over time. Numerous measures are taken to try to avoid incorrectly identifying actual changes in a firm's activities as an error. The within-firm detection is carried out on a firm-by-firm basis, meaning that detection for one firm is independent from that of another.

Since this method relies on identifying sharp changes within a firm, it is necessary to have a sufficiently long and consistent panel of observations for the firm's potential errors to be identified. This requires two conditions. First, there must be a minimum number of periods surrounding an observation. On this, it is required that there is at least one period before and three periods after an observation (adjustable), or vice versa, to identify whether that observation is a significant error (therefore, at least four comparison periods total). The reasoning for requiring periods both before and after the observation in question is that firms are likely to have extreme outliers in their first and last year of operation, so this should not be incorrectly identified as an error. There are two exceptions to this. First, it is not necessary to observe a period before the observation is if the firm is observed to be older than one year, so it is not a new firm, in which case the first observation is still checked, so long as there are at least three observations after. The second exception is that, in the last period of the entire sample, outliers are still checked. The second requirement is that the observed periods must not be too many years apart. Specifically, if consecutive observations are more than three years apart (adjustable), they are not used to identify errors, as it is plausible that the firm may have changed significantly in this time.

Figure A A.3 provides four illustrative examples. In the top left panel, firm 1 is only observed in two periods, so neither period is checked using within-firm detection. In the top right panel, the firm's first two observations are many years apart, so cannot be checked. The observation in 2015 can be checked as the firm is at least two years old, and there are three periods observed after. Beyond that, it is not possible to check as there are not at least three periods either side of the observations. In the bottom left panel, the firm is known to have been born in the first period, so it is not checked. After that, it is checked in every period except the last, when it may have ended existence. In the bottom right panel, the firm is known to be 25 years old in the first period so it is checked, along with all others until the final period (where the sample of data covers periods later than 2008).

Figure A A.3. Illustrative examples of sample requirements for within-firm outlier detection

Firm ID	Year	Firm Age	Identify?
1	2007	1	No
1	2008	2	No

Firm ID	Year	Firm Age	Identify?
2	2000	1	No
2	2005	6	No
2	2015	16	Yes
2	2016	17	No
2	2017	18	No
2	2018	19	No

Firm ID	Year	Firm Age	Identify?
3	2002	1	No
3	2003	2	Yes
3	2005	4	Yes
3	2006	5	Yes
3	2007	6	Yes
3	2008	7	No

Firm ID	Year	Firm Age	Identify?
4	2002	25	Yes
4	2003	26	Yes
4	2005	28	Yes
4	2006	29	Yes
4	2007	30	Yes
4	2008	31	No

For the observations that are identified as candidates for checking within-firm errors, the procedure used relies on comparing the value of a variable with the moving average of that variable in the consecutive periods (excluding the contemporary observation). Furthermore, to identify a variable as an error, it must not only be an extreme outlier in its level but must also generate extreme outliers in an associated ratio (for example, if there is a sharp decrease in energy used that happens at the same time as a sharp decrease in the number of employees or revenue, the sharp drop in energy used would not be considered as an error).

Three more conditions are required for an observation to be considered an outlier. First, there must not be a clear within-firm trend break downwards. This aims to avoid incorrectly identifying firm observations as outliers when they have simply downsized. This was identified in the data as relatively common, while corresponding sudden jumps upwards were not generating issues for the outlier detection. Second, outliers are not detected if there is a persistent negative growth rate of the variable. Third, if more than half of the observations are zero, then it is not inspected for outliers.

Figure A A.4 provides an illustrative example. For firm 5 in the top panel, value added (VA) is relatively constant over time. Focusing on 2005, if VA is compared with the moving average (MA) of value added in neighbouring periods, the values are similar. However, its CO2 in 2005 is less than 1% of the size of the associated moving average. This imbalance is identified by looking at the ratio of VA/CO2, which jumps sharply upwards in 2005. Therefore, it is likely that the value for CO2 was entered with error.

In contrast, in the second panel, the entry for value added also dropped sharply in 2005. Therefore, even though there is the same sharp drop in CO2 as before, the ratio of value added to CO2 is relatively similar to the moving average. It is feasible that, in this year, the firm simply downsized its operations, so this entry would not be considered as an error.

Figure A A.4. Illustrative examples of within-firm outlier detection

Firm ID	Year	VA	MA of VA	CO2	MA of CO2	VA/ CO2	MA of VA/ CO2
5	2002	10000		2000		5.00	
5	2003	11000		2500		4.40	
5	2004	10500		2200		4.77	
5	2005	12000	11000	22	2400	545.45	4.615545
5	2006	11000		2300		4.78	
5	2007	11500		2500		4.60	
5	2008	12000		2900		4.14	

Firm ID	Year	VA	MA of VA	CO2	MA of CO2	VA/ CO2	MA of VA/ CO2
6	2002	10000		2000		5.00	
6	2003	11000		2500		4.40	
6	2004	10500		2200		4.77	
6	2005	150	11000	22	2400	6.82	4.615545
6	2006	11000		2300		4.78	
6	2007	11500		2500		4.60	
6	2008	12000		2900		4.14	

It is necessary to specify a parameter beyond which the variable is considered to be sufficiently different from the within-firm moving average to consider it a potential error (for both the level and the ratio). After testing and eyeballing the data, a parameter of 50 is chosen. If a variable x is at least a factor of 50 different from the moving average, and a key ratio involving X is also 50 times different from its respective moving average, then the variable x is considered to be an error. Then, x is adjusted to be equal to its moving average and all corresponding ratios and log values are also updated.

Specifically, the following two conditions are used to identify the distance of variable x from its moving average, ma, for the distance parameter α. The observation x_{it} is considered an error if either:

$$abs\left\{\frac{(x_{it} - ma(x_{it}))}{ma(x_{it})}\right\} > \alpha$$

Equation 22

$$abs\left\{\frac{ma(x_{it})}{x_{it}}\right\} > \alpha \text{ if } sign(x_{it}) = sign\big(ma(x_{it})\big) \text{ and } abs(x_{it}) < abs\big(ma(x_{it})\big)$$

Equation 23

The first condition detects large absolute differences of x from its moving average and is robust to the two variables having different signs. The second condition identifies cases where the value of x is very small relative to the moving average, for example if it had been entered in millions instead of single units.

The procedure is implemented on just one observation per firm at a time. That is, only the largest outlier for each firm is identified and smoothed, if necessary. Then the procedure is repeated, identifying just one outlier at a time using the updated dataset, until no more outliers remain.

There are, of course, limitations to this approach. It is only possible to carry out the procedure on firms with panel data. In surveys with stratified samples, it is more likely to observe a sufficient panel for larger firms which, in turn, may be less likely to submit data with errors.

Cross-sectional outlier detection

Implausible values can also be identified by comparing values across firms within an industry. This is particularly useful if firm-level panel data is not available, as is often the case with environmental surveys. To detect data entry errors, the value of a variable is compared with the difference between the 10th and 90th percentile value of that variable within the firm's industry in each year. This is similar to the Tukey method, but with the 10th and 90th percentiles instead of the 25th and 75th.

Like with the within-firm detection, it is important to avoid incorrectly identifying genuine outliers as data entry errors. In particular, when trying to identify outliers, it is important not to incorrectly identify particularly large firms as outliers. Therefore, to identify a variable as a data entry error and not just an outlier, it must be that the value of the level of the variable and the ratio of the variable with respect to the firm's employment are sufficiently different from the respective within-industry 90-10 percentile ranges.

Firms may have plausible values of two variables relative to other firms but the comparison of these two variables within the firm may be implausible. For example, value added of hundreds of millions is plausible, as is employment of just five, but a firm with both of these together may be implausible. Therefore, ratios of key variables relative to both labour and carbon are checked. Firm-year observations are dropped if these ratios exceed the specified plausibility threshold for the industry. In practice, this mainly drops observations with implausibly small values of CO2 which, from examination of the data, most likely to be data entry errors.

If a variable is detected as being a data entry error, the observation is dropped. The variables on which this procedure is implemented, if available, are gross output; value added; energy expenditure; carbon dioxide emissions; total environmental investments; total environmental spending; and total pollution. The parameter which determines whether a variable is sufficiently different from the within-industry 90-10 percentile range to be considered an error is 50 times larger or smaller.

Confidentiality

The program can automatically apply three forms of confidentiality. The user can specify that any or all of these should be applied.

First, the program can apply a threshold rule based on the number of units per cell. If the minimum elements per cell option is specified, the code blanks cells for statistics based on fewer than X units, where X is specified by the user, for example, automatically blanking cells with fewer than five firms. If the raw data is at the establishment level, then the count of establishments is used for the cell.

Secondly, the program can implement a dominance rule. If the user chooses to apply dominance, the program blanks cells in which the largest firm or the largest two firms account for more than Y% of the cell's total, in terms of employment (L), output (GO), or on a variable-by-variable basis. The code program can also implement blanking based on the P-percent rule. To allow for the possibility of negative value added, the p-percent rule is applied after converting to absolute values.

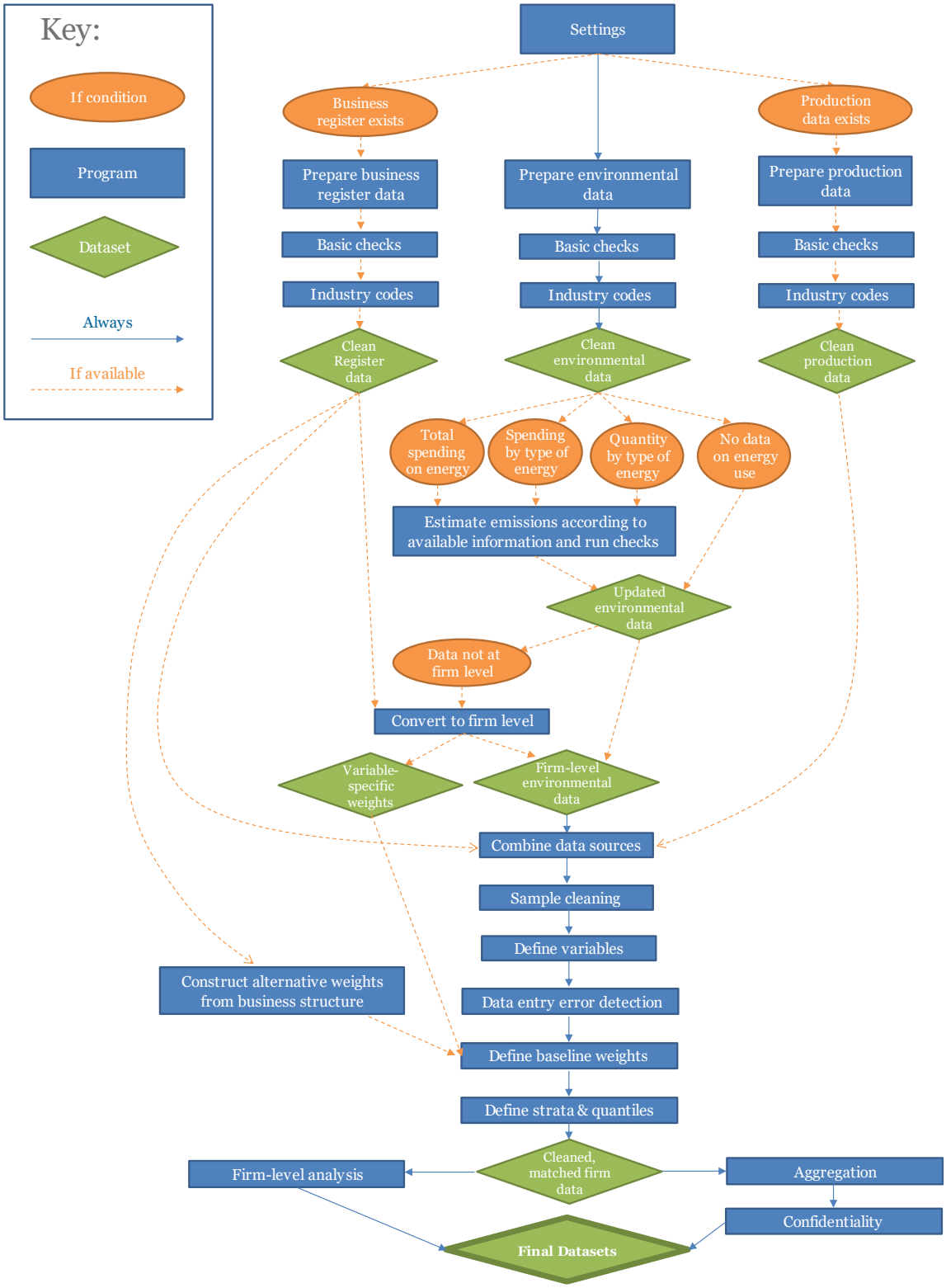
Third, some countries, though not any of those currently included in the project, do not allow specific percentiles of the data to be released. In these cases, percentiles are replaced with the average over the nearest X observations, where X is a number to be determined by the country.

Table A A.3. Confidentiality requirements by country

Country	Min units per cell	Dominance	Rules
Croatia	-	-	-
France	3 (establishments)	1 unit, 85%.	_N variables also blanked
Indonesia	-	-	-
Lithuania	-	-	-
Sweden	3	1 unit, 50%. 2 units, 69%	-

Diagrammatic representation of the procedure

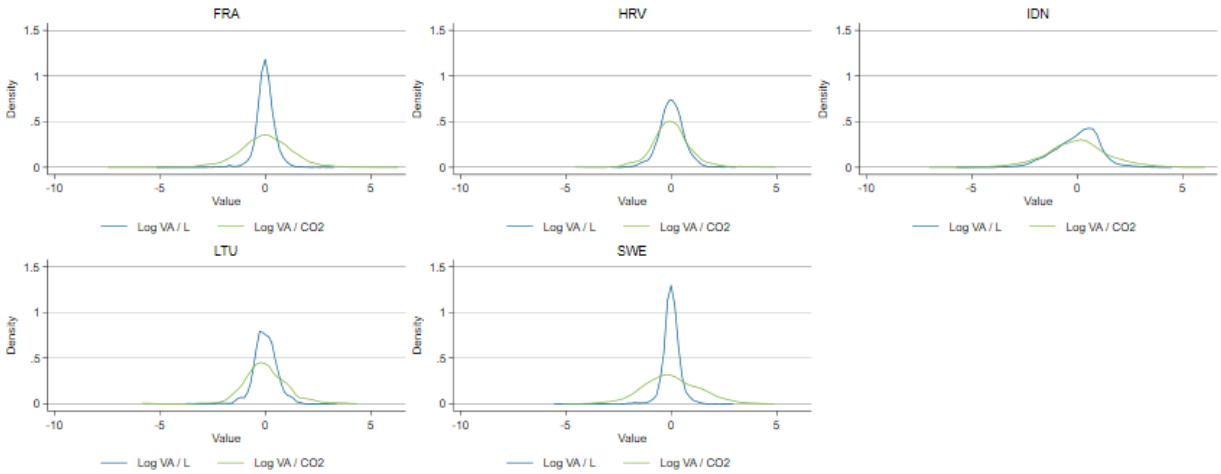
Figure A A.5. Diagrammatic representation of the data collection process



Annex B. Additional analysis

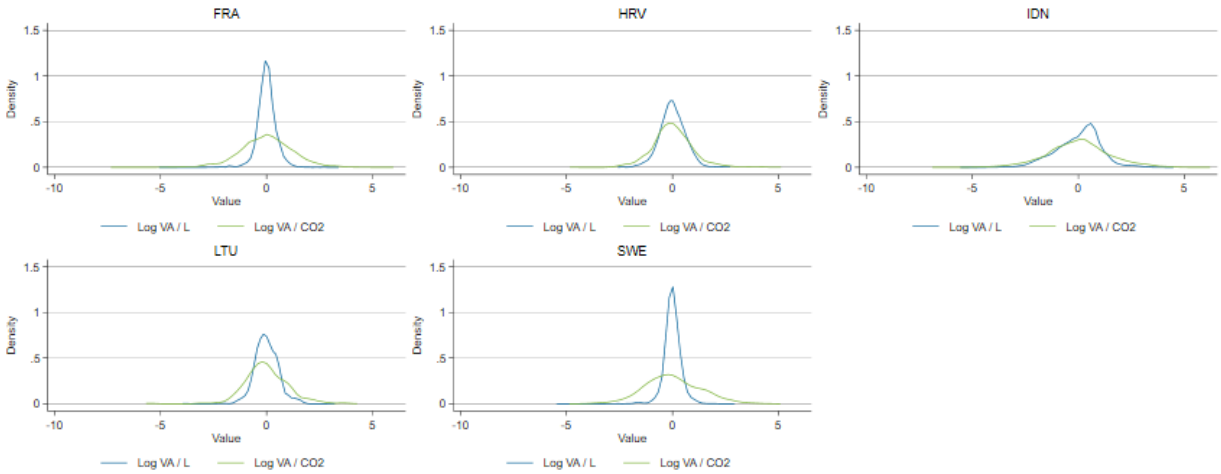
Density plots

Figure A B.1. Within-industry and size class distributions of carbon and labour productivity



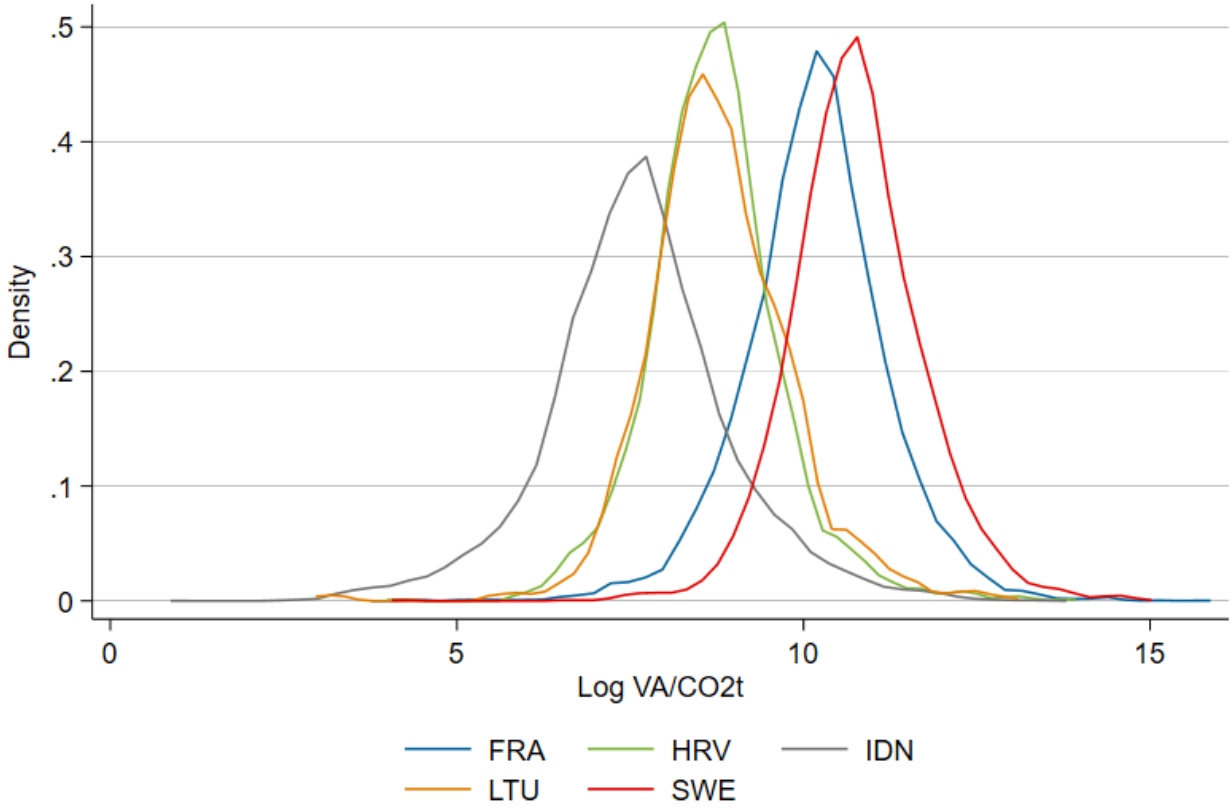
Note: The charts plot the within-industry-employment band-year distributions of carbon and labour productivity across firms in France, Croatia, Indonesia, Lithuania, and Sweden. Industries are defined at the 4 digit level and employment bands are defined as: 20-49, 50-249, and 250+. The distributions are centred at 0. The sample is restricted to manufacturing sectors and firms with at least 20 employees.
Source: authors' calculations

Figure A B.2. Within-industry and age band distributions of carbon and labour productivity



Note: The charts plot the within-industry-employment band-year distributions of carbon and labour productivity across firms in France, Croatia, Indonesia, Lithuania, and Sweden. Industries are defined at the 4 digit level and age bands are defined as: 1-2, 3-5,6-19, 20+. The distributions are centred at 0. The sample is restricted to manufacturing sectors and firms with at least 20 employees.
Source: authors' calculations

Figure A B.3. Distributions of carbon productivity using data on total energy spending only

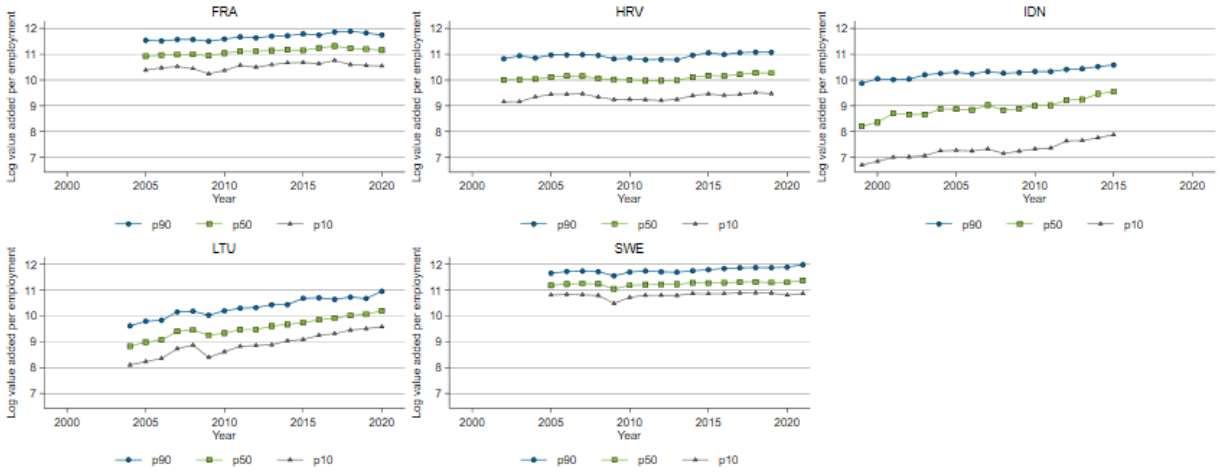


Note: The charts plot the within-industry-year distributions of carbon productivity across firms in France, Croatia, Indonesia, Lithuania, and Sweden. Industries are defined at the 4 digit level and distributions are recentred by adding the population mean of each country. CO2 is computed only using data on total energy spending for all countries. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations.

Time series

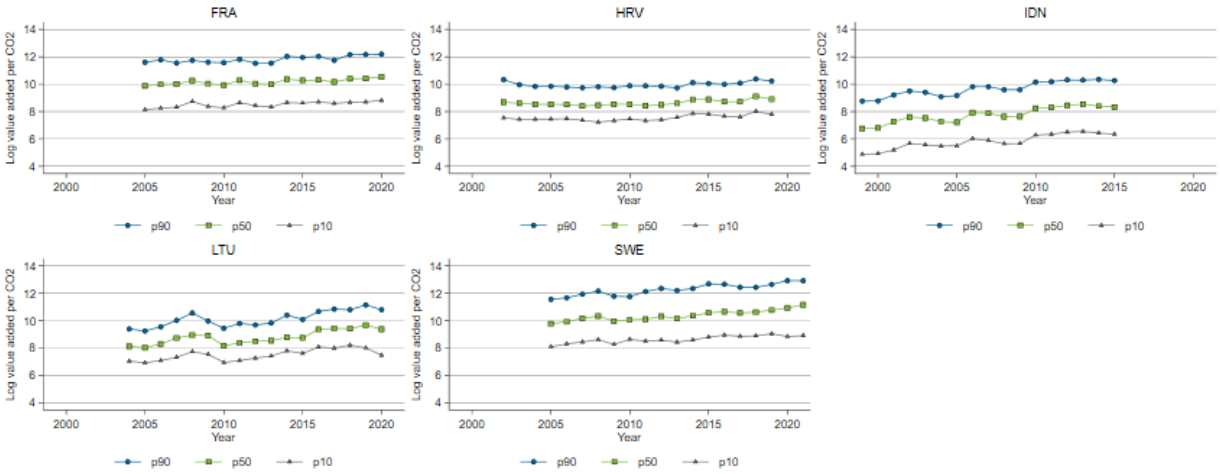
Figure A B.4. Labour productivity has typically evolved similarly across different percentiles of the distribution

Labour productivity at the 10th, 50th, and 90th percentiles within 3 digit industries over time



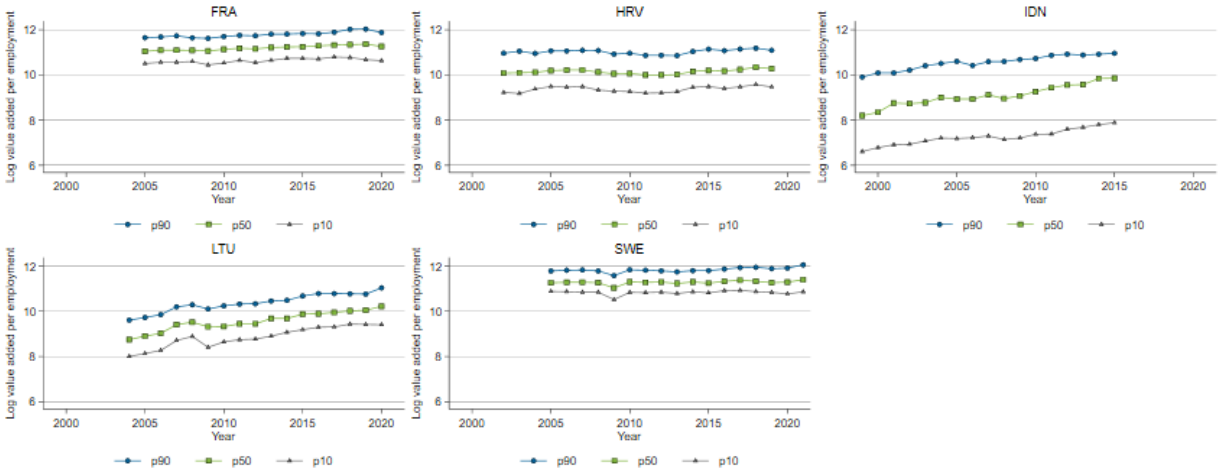
Note: The charts plot the 10th, 50th, and 90th percentile of within-industry labour productivity over time. Industries are defined at the 3 digit level and the unweighted average across industries is taken. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations

Figure A B.5. Carbon productivity at the 10th, 50th, and 90th percentiles within 3 digit industries over time, weighted by output



Note: The charts plot the 10th, 50th, and 90th percentile of within-industry carbon productivity over time. Industries are defined at the 3 digit level and the output-weighted average across industries is taken. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations

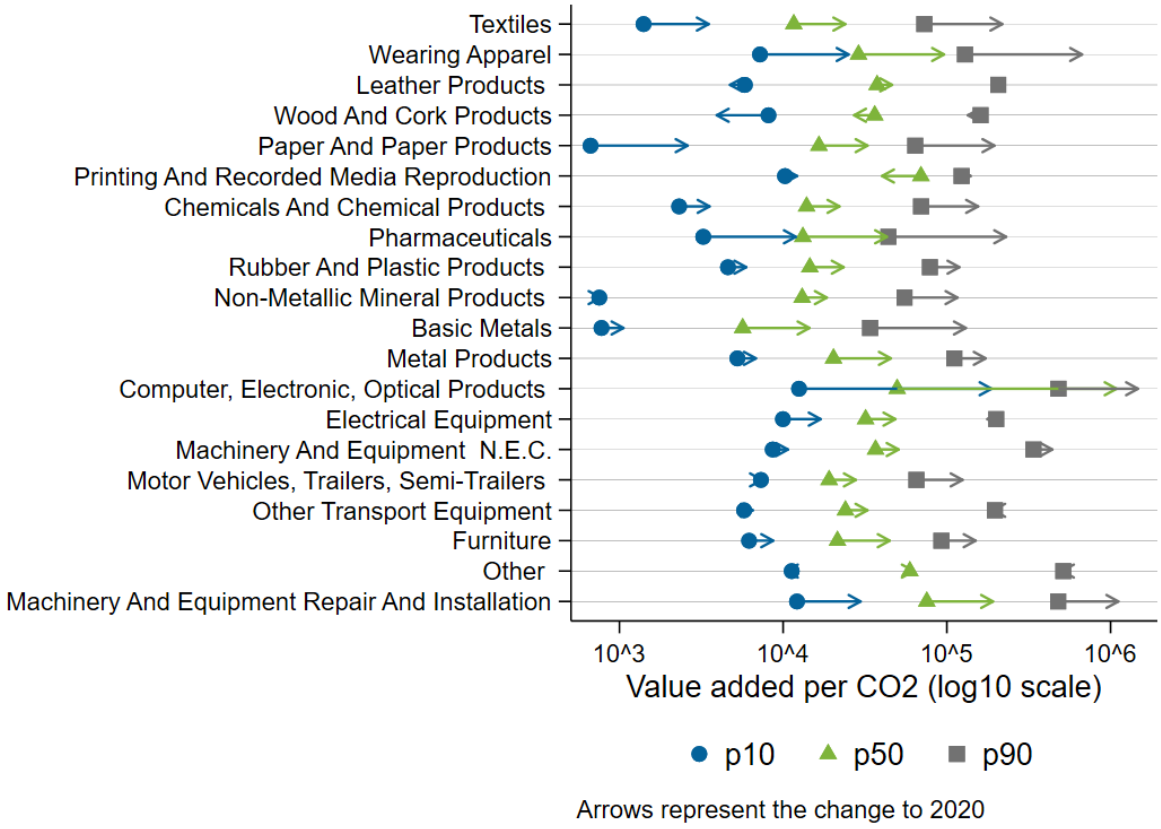
Figure A B.6. Labour productivity at the 10th, 50th, and 90th percentiles within 3 digit industries over time, weighted by output



Note: The charts plot the 10th, 50th, and 90th percentile of within-industry labour productivity over time. Industries are defined at the 3 digit level and the output-weighted average across industries is taken. The sample is restricted to manufacturing industries with at least 10 firms and firms with at least 20 employees.
Source: authors' calculations

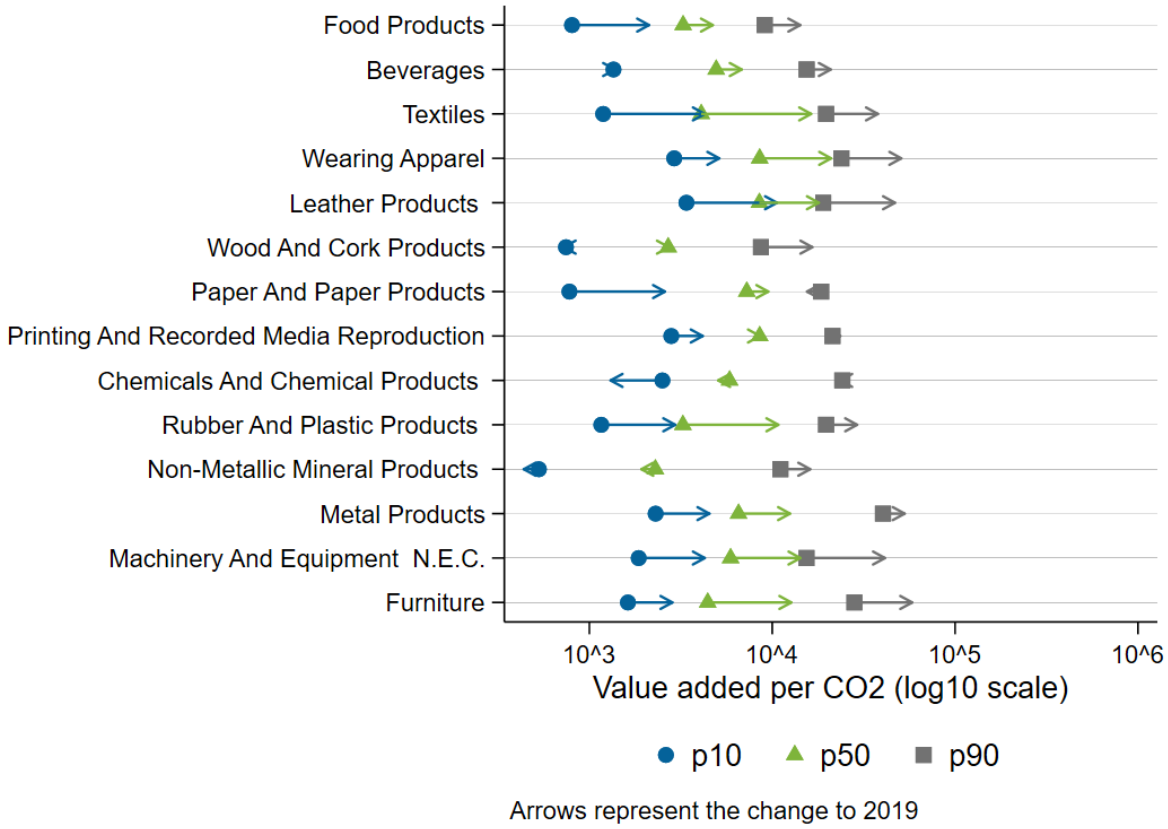
Distributions by industry

Figure A B.7. Carbon productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale in France



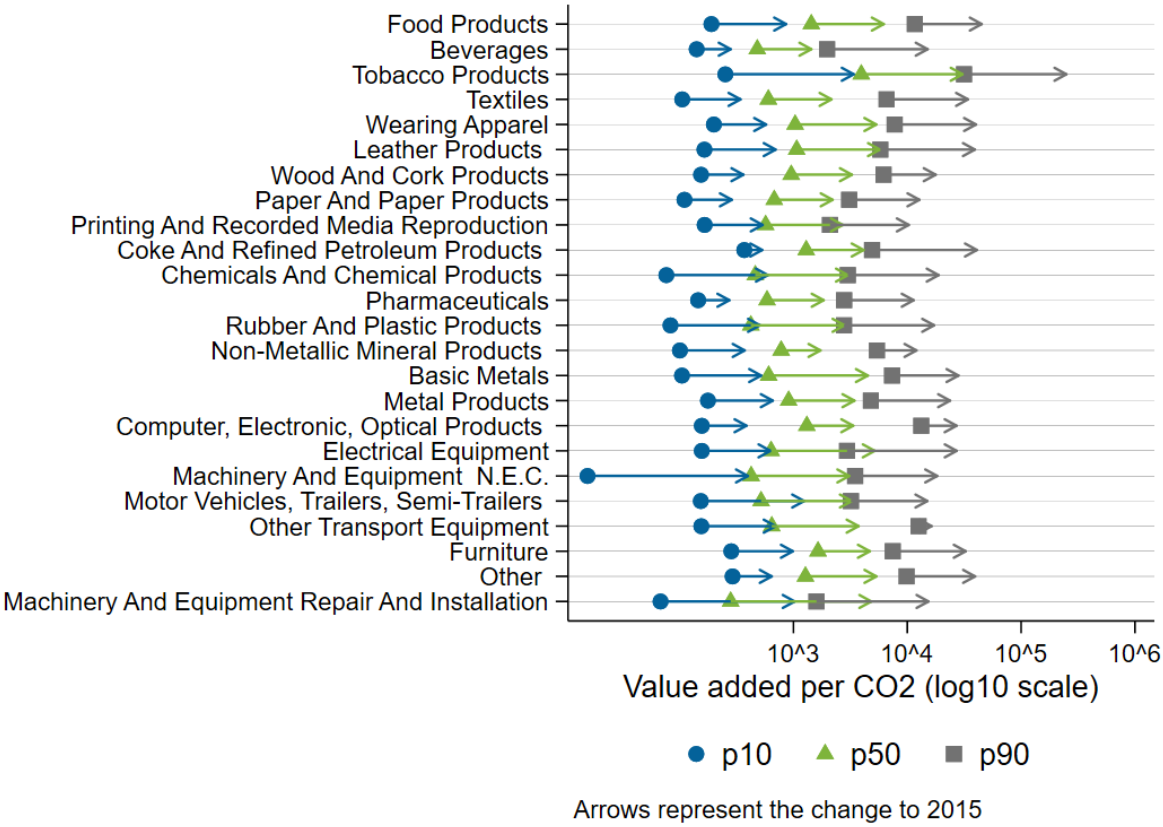
Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2005 to 2020 in France. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

Figure A B.8. Carbon productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale in Croatia



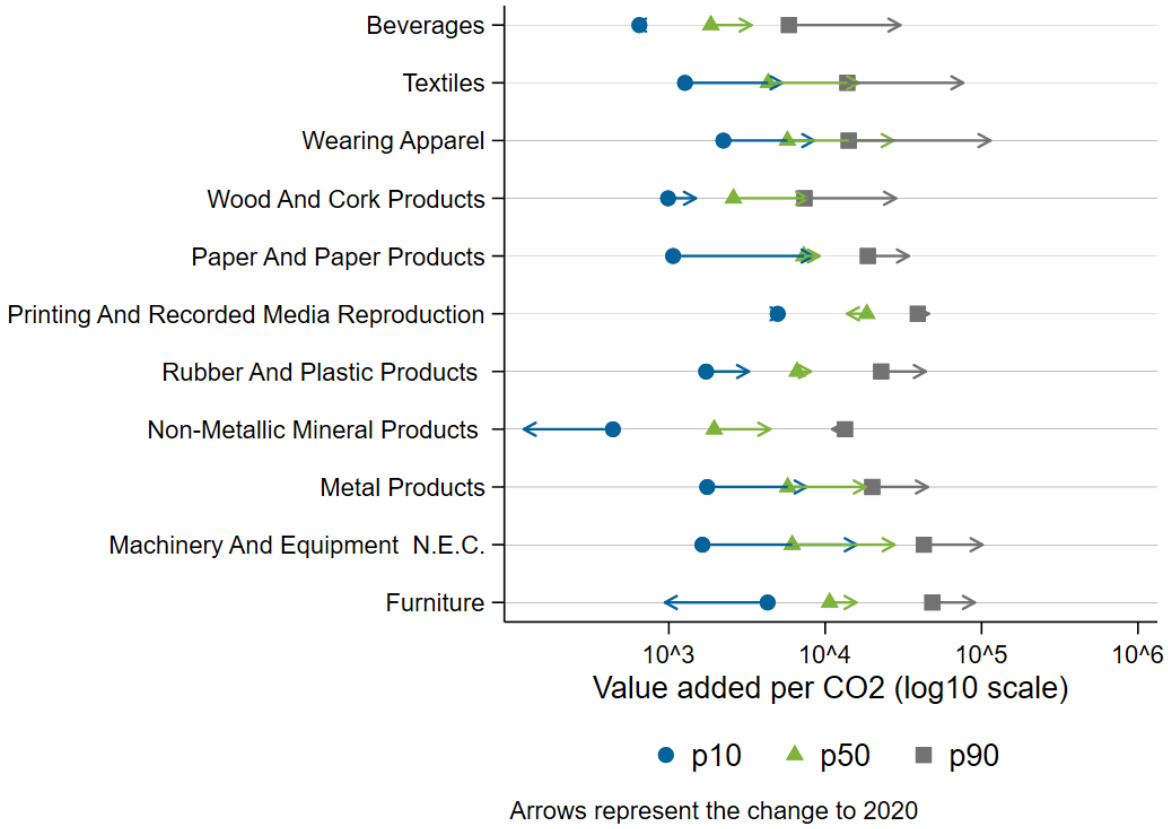
Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2002 to 2020 in Croatia. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

Figure A B.9. Carbon productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale for Indonesia



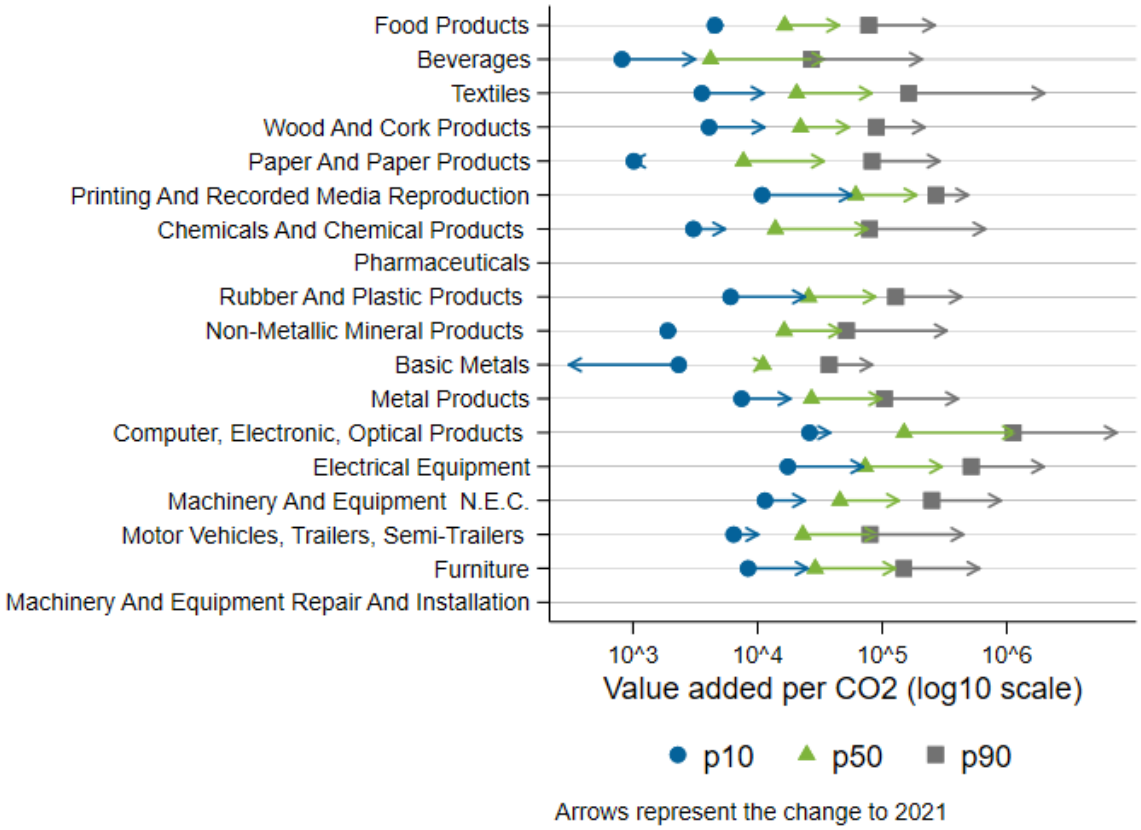
Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2000 to 2020 in Indonesia. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

Figure A B.10. Carbon productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale for Lithuania



Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2004 to 2020 in Lithuania. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

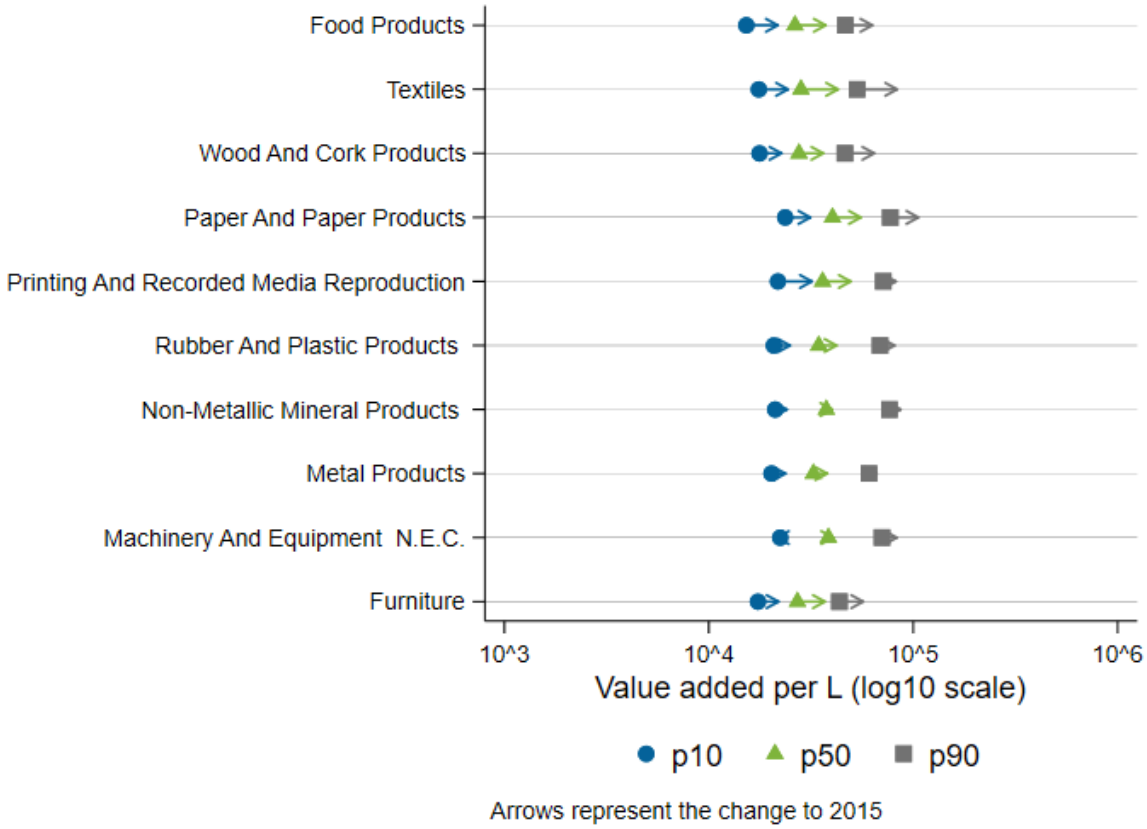
Figure A B.11. Carbon productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale for Sweden



Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2004 to 2020 in Sweden. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

Figure A B.12. Variation across industries in labour productivity dispersion

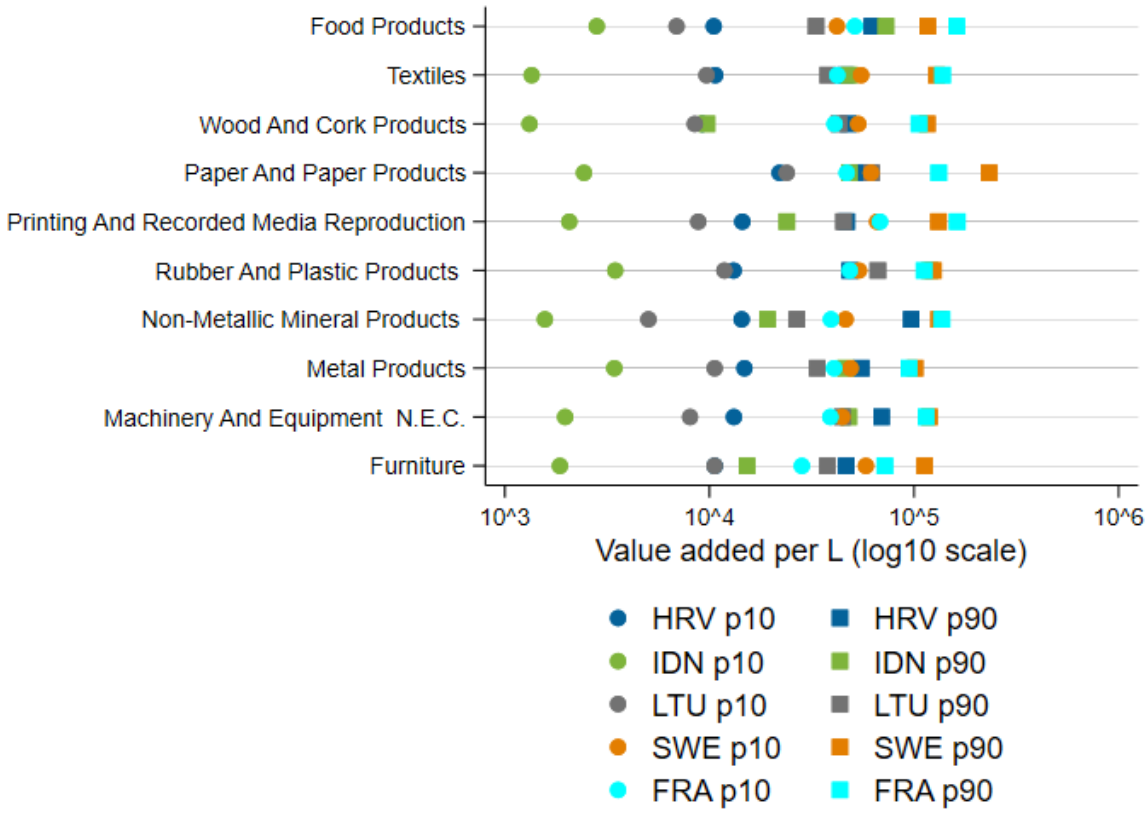
Labour productivity at the 10th, 50th, and 90th percentiles within 2 digit industries and their change over time on a log scale



Note: The chart shows the 10th, 50th, and 90th percentiles of carbon productivity within industries and their changes from 2004 to 2015. The values are plotted on a base 10 log scale. The unweighted average across countries is taken. The sample is restricted to manufacturing industries with at least 15 firms in every country-year, and firms with at least 20 employees. The industries included are the same as in the chart for carbon productivity and are in the same order.
Source: authors' calculations

Figure A B.13. Variation across countries and industries in frontier and lagging firms' labour productivity

10th and 90th percentile value added per worker across countries in each manufacturing industry



Note: The chart shows the 10th and 90th percentiles of labour productivity within industries for each country in 2015. The values are plotted on a base 10 log scale. The sample is restricted to manufacturing industries with at least 10 firms in every country-year included in the plot, and firms with at least 20 employees.
Source: authors' calculations

Table A B.1. Ranking of industries by carbon productivity dispersion

Industry (3 digit)	Dispersion in VA/CO2	Dispersion in VA/Worker	Difference in dispersion of VA/CO2 and VA/Worker	Average of energy cost share	Average of CO2 emissions
Manufacture of coke oven products	6.31	2.77	3.54	0.34	4356
Manufacture of refined petroleum products	4.82	2.83	1.98	0.13	78014
Manufacture of basic iron and steel	4.65	2.02	2.63	0.11	52475
Manufacture of consumer electronics	4.32	3.24	1.08	0.05	2577
Manufacture of tobacco products	4.30	3.99	0.31	0.07	4668
Vegetable and animal oils and fats	4.29	3.10	1.18	0.11	3255

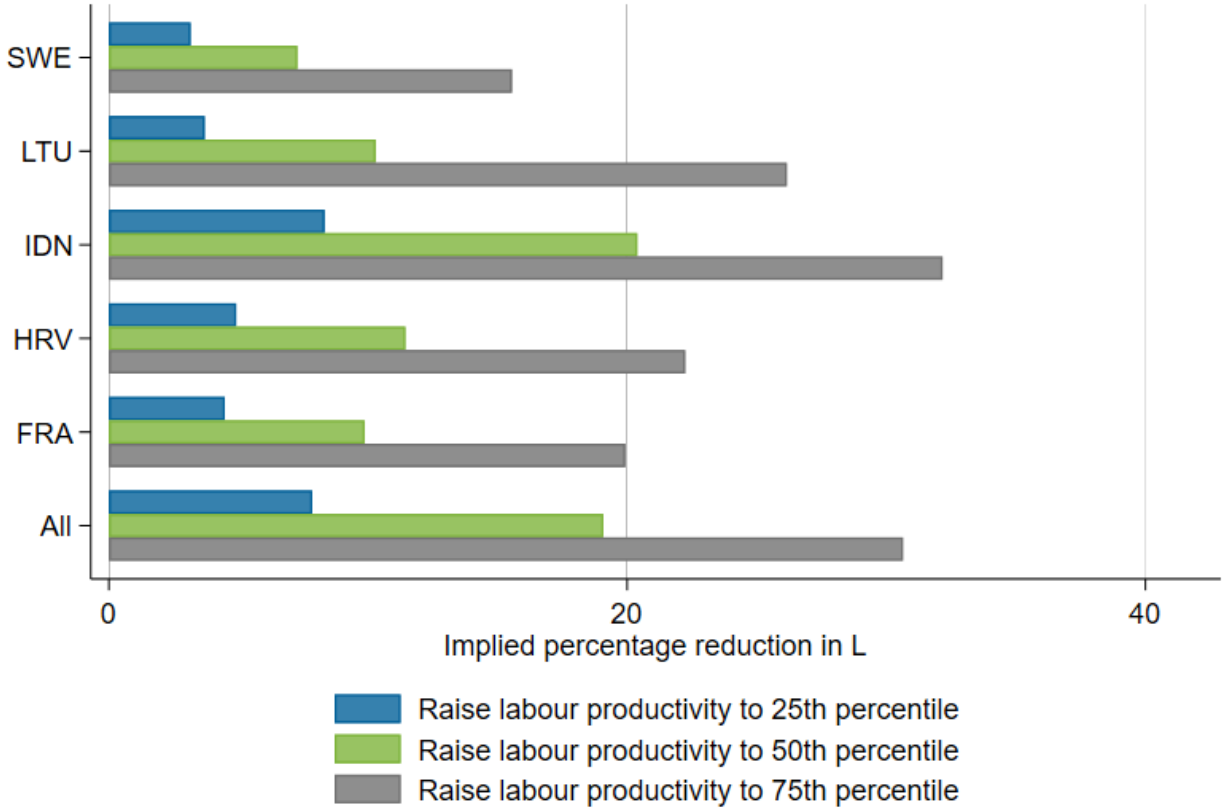
manufacturing					
Manufacture of basic chemicals and plastics	4.21	2.17	2.04	0.13	36477
Industrial machinery and equipment installation	3.79	0.91	2.89	0.02	127
Basic precious and non-ferrous metals manufacturing	3.72	2.15	1.57	0.13	9623
Batteries and accumulators manufacturing	3.69	2.72	0.97	0.10	6313
Non-metallic mineral products manufacturing n.e.c.	3.68	1.89	1.79	0.13	8386
Manufacture of paper and paper products	3.66	1.63	2.03	0.09	7651
Manufacture of glass and glass products	3.66	1.63	2.03	0.10	11666
Manufacture of games and toys	3.62	2.01	1.61	0.07	407
Repair of metal products and machinery	3.55	1.44	2.11	0.06	1006
Building of ships and boats	3.53	1.92	1.61	0.06	1605
Manufacture of other textiles	3.41	1.57	1.84	0.06	697
Jewellery and related articles manufacturing	3.40	1.89	1.51	0.04	248
Textiles spinning, weaving, finishing	3.39	1.95	1.44	0.13	3763
Other manufacturing n.e.c.	3.38	1.63	1.75	0.05	378
Leather tanning and luggage manufacturing	3.36	1.92	1.44	0.05	417
Transport equipment manufacturing n.e.c.	3.35	1.93	1.41	0.06	1404
Manufacture of lighting equipment	3.34	1.59	1.75	0.06	3077
Electronic components and boards manufacturing	3.33	1.66	1.68	0.06	2747
Manufacture of measuring and control equipment	3.32	1.35	1.98	0.04	268
Manufacture of pharmaceuticals	3.32	2.12	1.20	0.05	3463
Manufacture of prepared animal feeds	3.30	2.57	0.74	0.12	5184
Manufacture of wiring and wiring devices	3.25	1.72	1.53	0.05	1736
Electrical equipment manufacturing	3.24	1.53	1.70	0.05	1459
Manufacture of rubber products	3.24	1.98	1.26	0.09	1932
Manufacture of sports goods	3.21	1.60	1.60	0.04	185
Air and spacecraft manufacturing	3.19	1.38	1.81	0.05	15542
Fruit and vegetables	3.18	1.57	1.61	0.08	1982

processing and preserving					
Knitted and crocheted apparel manufacturing	3.17	1.55	1.63	0.06	1882
Manufacture of other chemical products	3.14	1.88	1.26	0.06	1494
Manufacture of wood products	3.11	1.54	1.57	0.07	1007
Medical and dental instruments and supplies manufacturing	3.10	1.36	1.74	0.05	1028
Manufacture of dairy products	3.10	2.16	0.94	0.09	7416
Manufacture of communication equipment	3.10	1.49	1.61	0.06	1514
Motor vehicle parts and accessories manufacturing	3.10	1.73	1.36	0.06	3243
Manufacture of man-made fibres	3.09	3.04	0.05	0.18	19712
Manufacture of musical instruments	3.09	2.64	0.45	0.10	4864
Casting of metals	3.06	1.53	1.52	0.12	3374
Manufacture of beverages	3.05	2.25	0.81	0.08	2419
Fish and seafood processing and preserving	3.03	1.57	1.46	0.06	1193
Manufacture of electric motors and transformers	3.02	1.68	1.35	0.04	1788
Railway locomotives and rolling stock manufacturing	3.02	1.10	1.92	0.03	2639
Manufacture of special-purpose machinery	2.93	1.68	1.25	0.06	745
Manufacture of general-purpose machinery	2.91	1.59	1.31	0.06	914
Manufacture of other food products	2.91	1.59	1.32	0.09	1668
Manufacture of domestic appliances	2.86	1.52	1.34	0.05	1698
Manufacture of motor vehicle bodies	2.86	1.25	1.61	0.04	348
Manufacture of furniture	2.86	1.34	1.51	0.05	490
Manufacture of metal products	2.84	1.45	1.40	0.07	657
Sawmilling and planing of wood	2.83	1.55	1.28	0.10	914
Manufacture of structural metal products	2.83	1.52	1.31	0.06	708
Manufacture of grain mill products	2.77	2.00	0.77	0.10	5211
Manufacture of footwear	2.74	1.61	1.13	0.06	915
Manufacture of plastics products	2.72	1.53	1.20	0.08	1024
Wearing apparel manufacturing, except fur	2.71	1.53	1.18	0.05	402

Computers and peripheral equipment manufacturing	2.69	1.29	1.40	0.04	511
Printing and related services	2.53	1.40	1.13	0.05	513
Manufacture of motor vehicles	2.48	1.48	1.00	0.05	16670
Processing and preserving of meat	2.44	1.46	0.98	0.07	1560
Manufacture of weapons and ammunition	2.14	1.26	0.89	0.03	2760
Manufacture of optical instruments	2.08	1.09	0.99	0.07	4046
Reproduction of recorded media	1.46	0.65	0.81	0.06	374
Manufacture of medical equipment	1.20	0.92	0.27	0.07	456

Note: The table shows 3 digit industries, ranked by their carbon productivity dispersion – largest to smallest. Dispersion in carbon and labour productivity are defined as the difference between the 90th percentile and 10th percentile firms in their 3 digit industry. The fourth column shows the difference between carbon and labour productivity dispersions. The remaining two columns show the average share of energy in total costs and average CO2 emissions per firm. The measures are averaged across countries and years.
Source: authors' calculations

Figure A B.14. Counterfactual exercises showing implied reduction in employment to produce the same aggregate value added if all firms were at least as productive as the 25th, 50th, and 75th percentile firms in their industry



Note: The chart shows the implied percentage reduction in employment that would be achieved if all firms had at least a labour productivity equal to the 25th, 50th, or 75th percentile firm in their 3 digit industry. The data is for 2015 and is restricted to manufacturing sector firms with at least 20 employees.
Source: authors' calculations

Table A B.2. Greater dispersion in labour productivity is associated with lower aggregate labour productivity

Regressions of aggregate variables on within-industry dispersion in labour productivity

	(1)	(2)	(3)	(4)	(5)
	log(VA)	log(CO2)	log(L)	log(VA/CO2)	log(VA/L)
$DispLP$	-0.0399	0.0947	0.296***	-0.116*	-0.327***
	(0.0343)	(0.0866)	(0.0425)	(0.0690)	(0.0489)
Observations	3043	3168	3154	3043	3041

Note: The table presents results from regressions of aggregate outcomes on within-industry-country-year dispersion in labour productivity, measured as the log difference between the 90th and 10th percentile firm. The outcome variables of interest, in logs, are: value added (VA), CO2, labour (L), carbon productivity (VA/CO2), and labour productivity (VA/L). The data is at the 3 digit industry-country-year level and is restricted to firms with at least 20 employees in manufacturing industries. The regressions are weighted by total CO2 to be representative of aggregate emissions. Standard errors are robust to heteroskedasticity. Country-year fixed effects are included as controls. * p<0.10 ** p<0.05 *** p<0.01.

Table A B.3. Increasing dispersion in labour productivity is correlated with increases in aggregate labour productivity but falling carbon productivity

Regressions of changes in aggregate variables on changes in within-industry dispersion in labour productivity

	(1)	(2)	(3)	(4)	(5)
	$\Delta_5 \log(VA)$	$\Delta_5 \log(CO2)$	$\Delta_5 \log(L)$	$\Delta_5 \log(VA/CO2)$	$\Delta_5 \log(VA/L)$
$\Delta_5 DispLP$	0.243***	0.141**	0.00705	0.102*	0.234***
	(0.0467)	(0.0706)	(0.0238)	(0.0605)	(0.0367)
Observations	1969	2031	2131	1928	1967

Note: The table presents results from regressions of changes in aggregate outcomes on changes in within-industry-country-year dispersion in labour productivity, measured as the log difference between the 90th and 10th percentile firm. The outcome variables of interest, in logs, are: value added (VA), CO2, labour (L), carbon productivity (VA/CO2), and labour productivity (VA/L). The data is at the 3 digit industry-country-year level and is restricted to firms with at least 20 employees in manufacturing industries. The regressions are weighted by total CO2 to be representative of aggregate emissions. Standard errors are robust to heteroskedasticity. Country-year fixed effects are included as controls. * p<0.10 ** p<0.05 *** p<0.01.

Factors correlated with dispersion

Table A B.4. Factors correlated with dispersion in carbon productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10	log(VA/CO2) p90-p10
log(VA/L) p90-p10	0.323***								
	(0.0267)								
log(VA/K) p90-p10		0.173***							
		(0.0211)							
log(mean(K/L))			0.114***						

			(0.0177)						
log(K/L) p90-p10			0.166***						
			(0.0193)						
log(mean(Age))				0.0760					
				(0.0889)					
log(Age) p90-p10				0.210***					
				(0.0356)					
log(mean(Intermediates/total cost))					1.144***				
					(0.108)				
log(Intermediates/total cost) p90-p10					0.629***				
					(0.0453)				
log(mean(Labour costs/total cost))						0.192***			
						(0.0539)			
log(Labour costs/total cost) p90-p10						0.454***			
						(0.0385)			
log(mean(Imports/total cost))							0.0506***		
							(0.0164)		
log(Imports/total cost) p90-p10							-0.0142		
							(0.0132)		
log(mean(Energy spending/total cost))								0.284***	
								(0.0194)	
log(Energy spending/total cost) p90-p10								0.759***	
								(0.0204)	
mean(Electricity share)									0.0344
									(0.191)
Electricity share p90-p10									0.685***
									(0.0955)
Observations	3505	3271	3376	3434	3504	3503	2076	3467	2462
Excluded countries							FRA LTU		HRV LTU

Note: The chart shows results from regression estimates of within-industry dispersion in carbon productivity on various explanatory variables. The regressions are estimated by OLS on industry-country-year level data, with industries defined at the 3 digit level. Country-year fixed effects are included as controls. The sample is restricted to manufacturing industries and firms with at least 20 employees. Standard errors are robust to heteroskedasticity. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations

Table A B.5. Factors correlated with changes in dispersion in carbon productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10	d5logVA_ _CO2_disp_90_10

d5logVA__L_disp_90_10	0.196***								
	(0.0430)								
d5logVA__K_disp_90_10		0.0976***							
		(0.0231)							
d5LOG_K__L_av			-0.0207						
			(0.0211)						
d5logK__L_disp_90_10			0.0570**						
			(0.0226)						
d5LOG_AGE_av				0.0958					
				(0.200)					
d5logAGE_disp_90_10				0.0888**					
				(0.0366)					
d5LOG_M__TC_av					0.132				
					(0.176)				
d5logM__TC_disp_90_10					0.401***				
					(0.0666)				
d5LOG_LC__TC_av						0.476***			
						(0.172)			
d5logLC__TC_disp_90_10						0.209***			
						(0.0603)			
d5LOG_IM__TC_av							-0.0194		
							(0.0321)		
d5logIM__TC_disp_90_10							0.0187		
							(0.0142)		

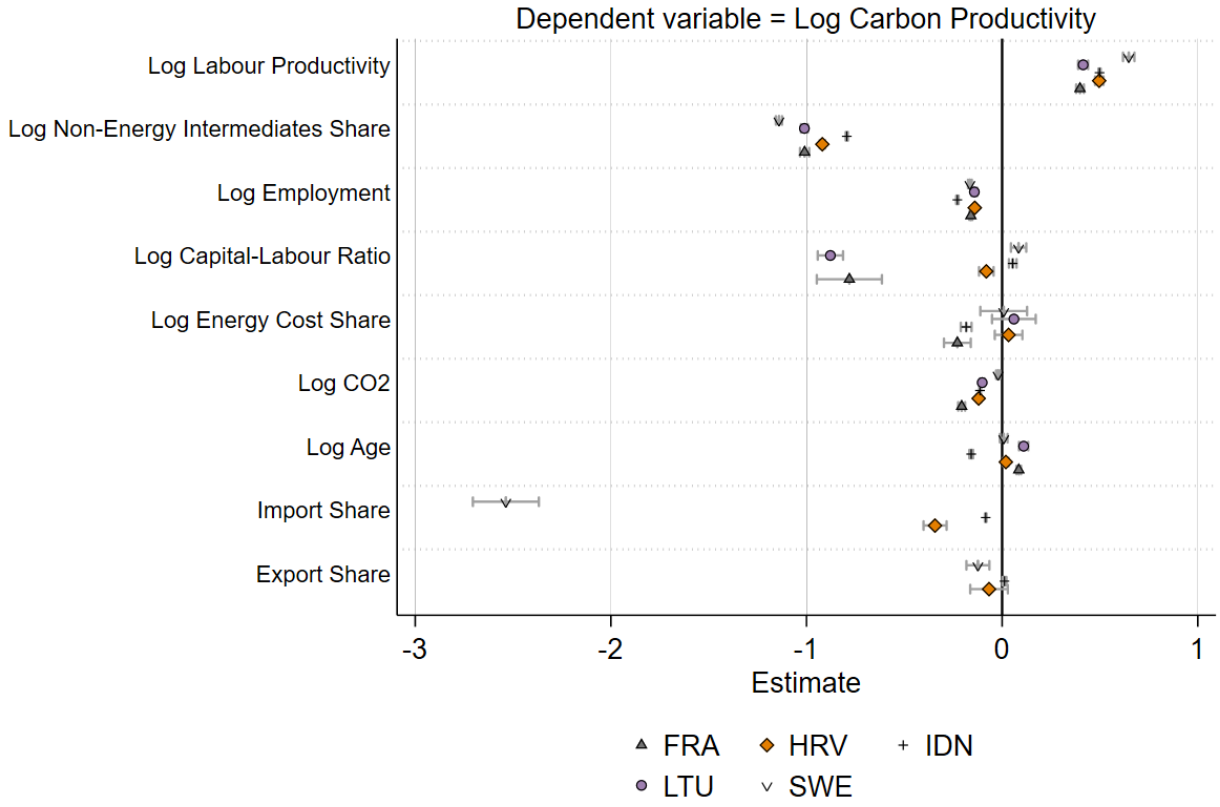
d5LOG_EN__TC_av								-0.151**	
								(0.0662)	
d5logEN__TC_disp_90_10								0.683***	
								(0.0326)	
d5shEL_av									1.949***
									(0.331)
d5shEL_disp_90_10									0.339***
									(0.101)
Observations	2353	2116	2238	2295	2353	2352	1398	2310	1652
Excluded countries							FRA LTU		HRV LTU

Note: The chart shows results from regression estimates of 5 year differences of within-industry dispersion in carbon productivity on 5 year differences of various explanatory variables. The regressions are estimated by OLS on industry-country-year level data, with industries defined at the 3 digit level. Each row is estimated as a separate bivariate regression. Country-year fixed effects are included as controls. The sample is restricted to manufacturing industries and firms with at least 20 employees. Standard errors are robust to heteroskedasticity. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations

Factors explaining firms' carbon productivity

Figure A B.15. Factors correlated with firms' carbon productivity

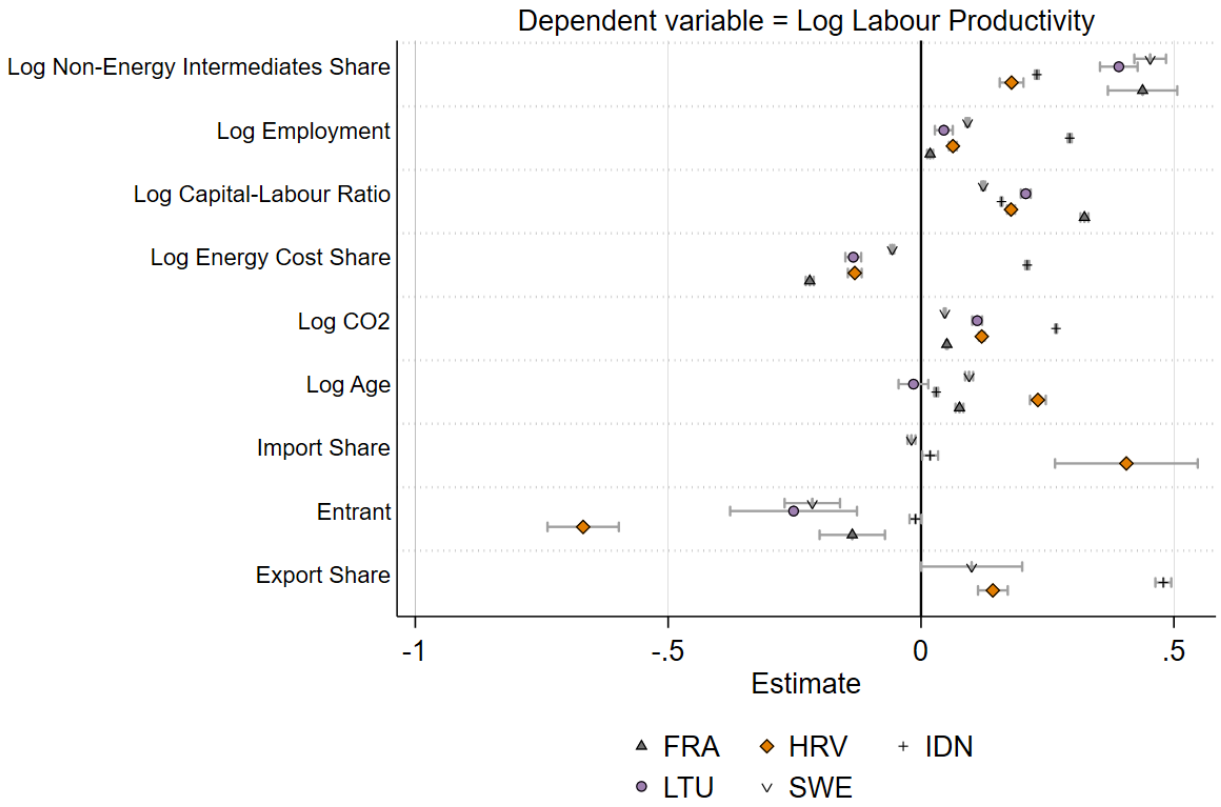
Estimates from multivariate OLS regressions of log value added per CO2 on various explanatory variables



Note: The chart shows estimates from multivariate OLS regressions of log value added per CO2 on all the explanatory variables listed simultaneously. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The regressions all include industry-year fixed effects. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations

Figure A B.16. Factors correlated with firms' labour productivity

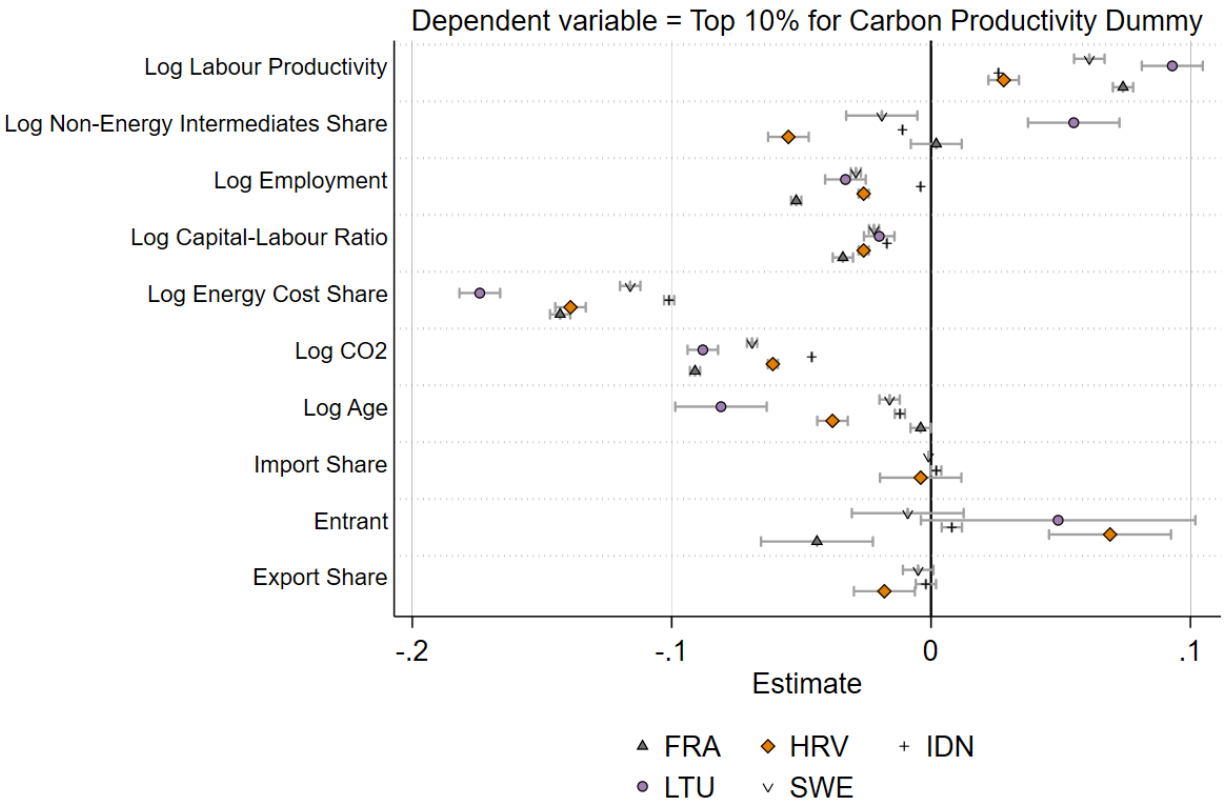
Estimates from bivariate OLS regressions of log value added per worker on various explanatory variables



Note: The chart shows estimates from bivariate OLS regressions of log value added per worker on various explanatory variables. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The regressions all include industry-year fixed effects. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations

Figure A B.17. Characteristics of frontier carbon productivity firms

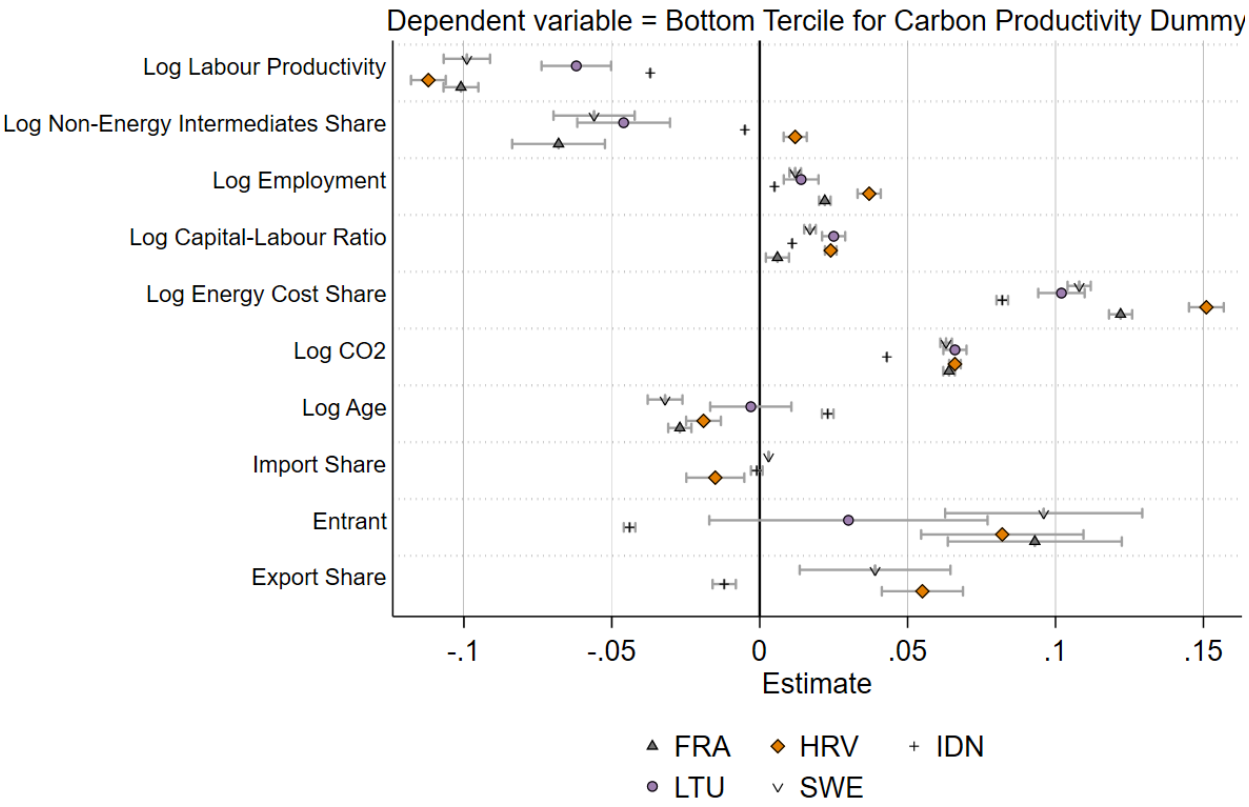
Estimates from bivariate OLS regressions of a dummy for high carbon productivity on various explanatory variables



Note: The chart shows estimates from bivariate OLS regressions of a dummy for whether the firm is in the top decile of the distribution for carbon productivity within an industry-year on various explanatory variables. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The regressions all include industry-year fixed effects. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations

Figure A B.18. Characteristics of low carbon productivity firms

Estimates from bivariate OLS regressions of a dummy for low carbon productivity on various explanatory variables



Note: The chart shows estimates from bivariate OLS regressions of a dummy for whether the firm is in the bottom tercile of the distribution for carbon productivity within an industry-year on various explanatory variables. They are estimated separately for each country, with country-specific coefficients reported in the plot. The y axis corresponds to the explanatory variable in each regression. The bars indicate 95% confidence intervals with standard errors that are robust to heteroskedasticity. The regressions all include industry-year fixed effects. The sample is restricted to firms with at least 20 employees in manufacturing industries.
Source: authors' calculations

The effects of energy prices

Table A B.6. Reduced form regression estimates of changes in firm outcomes on change in national energy prices

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO2
FRA						
Energy Price	-.007	-.365***	-.047	-.096**	.042	.36***
	(.061)	(.111)	(.049)	(.045)	(.071)	(.119)
R-squared	0.090	0.071	0.129	0.038	0.118	0.073
Observations	40080	40649	40649	40153	40080	40080
IDN						

Energy Price	.063***	-.067***	.014***	-.003	.047***	.112***
	(.011)	(.014)	(.005)	(.018)	(.01)	(.014)
R-squared	0.040	0.071	0.010	0.239	0.040	0.082
Observations	180042	208494	208494	106316	180042	180019
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	OLS	OLS	OLS	OLS	OLS	OLS

Note: The table presents results of OLS regression estimates of various outcomes of firms on national energy price index. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations

Table A B.7. Increasing energy prices are associated with increasing carbon productivity

OLS regression estimates of changes in firm outcomes on change in energy prices

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Value Added	Log CO2	Log Employment	Log Capital	Log VA per Worker	Log VA per CO2
FRA						
Energy Price	-.049***	-.187***	-.012	-.022**	-.039***	.134***
	(.013)	(.031)	(.01)	(.01)	(.014)	(.035)
R-squared	0.090	0.074	0.129	0.038	0.119	0.074
Observations	40080	40649	40649	40153	40080	40080
IDN						
Energy Price	.089***	-.031***	.003*	.004	.085***	.112***
	(.005)	(.006)	(.002)	(.009)	(.005)	(.006)
R-squared	0.042	0.068	0.009	0.236	0.042	0.081
Observations	188553	216771	216771	110561	188553	188529
Controls	none	none	none	none	none	none
Fixed Effects	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year	Ind-year
Estimation	OLS	OLS	OLS	OLS	OLS	OLS

Note: The table presents results of OLS regression estimates of various outcomes of firms on their energy price index. The regressions are estimated separately for each country using firm-level data. All specifications include an industry-year fixed effect. Standard errors are clustered by firm. * p<0.10 ** p<0.05 *** p<0.01.
Source: authors' calculations