Product Level Emission Intensities: Measurement and Application

Ohyun Kwon* **Hao Zhao†** **Min Qiang Zhao ‡**

November, 2024

Abstract

We propose a novel method for calculating product-level emission intensities (PLEI) at highly disaggregate level (Harmonized System 6-digit) for nine emission categories. This method effectively disentangles PLEI from the firm-level efficiency factor that varies across firms and years. Utilizing firm-level emissions data from China for the period 2000–2013, our analysis shows that: (i) there is substantial heterogeneity in PLEI across different products; (ii) the 10% most emission-intensive products contribute to nearly 75% of total emissions, while comprising only 4% of total exports; (iii) a less aggregate categorization of products markedly underestimates the variation in emission intensities; and (iv) China's export profile shows no tendency towards specialization in products with high emission intensity.

JEL classification: D22, F18, Q56

Keywords: International trade, Emission intensities, China

*School of Economics, Drexel University LeBow College of Business, ok85@drexel.edu.

†School of Environmental and Natural Resources, and Institute of Ecological Civilization, Renmin University of China, haozhao@ruc.edu.cn.

‡MOE Key Laboratory of Econometrics, the Wang Yanan Institute for Studies in Economics, Xiamen University, kent_zhao@xmu.edu.cn. Zhao acknowledges financial support from the National Natural Science Foundation of China (No. 71773102).

1 Introduction

A seminal study by Zhang et al. (2017) reveals a staggering finding: in 2007, 762,400 premature deaths worldwide were attributable to PM2.5 emissions associated with the production of goods and services for international trade. Zhang et al. (2017) further assert that "consumption in western Europe and the USA is linked to more than 108,600 premature deaths in China."

When exploring trade-related environmental impacts, researchers often need to quantify emissions embedded in trade flows, by multiplying emission intensities with trade volumes. Traditionally, this approach has been applied at sectoral levels. This paper introduces an innovative methodology for computing product-level emission intensities (PLEI) by integrating firm-level emissions data with firm-product-level output, allowing for a highly detailed analysis at the HS (Harmonized System) 6-digit export product level.

The contributions of this paper are twofold. Firstly, we introduce an innovative PLEI algorithm that iteratively reconciles firm-level emissions allocation across products with the definition of economy-wide average emission intensities at the aggregate level. This methodological advancement sets our work apart from existing literature beyond mere considerations of more products. Notable strengths of our approach include: (1) consistency with the heterogeneous firm framework, as exemplified by Shapiro and Walker (2018), allowing for the variance in emission efficiency across firms and over time, (2) the ability to distinctly separate product-specific emission intensities from firm-specific emission efficiency, and (3) broad applicability to other datasets encompassing firm-level emissions and firm-product-level output details.

Our algorithm involves an iteration process, comprising several key steps. Initially, we assume a set of starting values for PLEI, which are then used to allocate firm-level emissions across various products produced by each firm. Subsequently, emissions are aggregated for each product and normalized by its total output value, generating updated PLEI estimates. These refined PLEI values are then reintroduced into the initial

step, instigating successive iterations until convergence is attained. The robustness of our iterative algorithm has been rigorously validated through Monte Carlo simulations, demonstrating its insensitivity to fundamental assumptions, resilience to data perturbations, and superiority over alternative methodologies.

The second major contribution of this study is to unveil novel stylized facts concerning PLEI, derived from Chinese firm-level emissions data spanning from 2000 to 2013. Firstly, a significant heterogeneity in PLEI across different products is observed, evidenced by a Gini coefficient exceeding 0.95 for all types of emissions. Secondly, our findings reveal that the top 10% most emission-intensive products are responsible for approximately 75% of emissions, yet they constitute only about 4% of China's export volume. This suggests that environmental policies, when precisely targeted at these emission-intensive products, can substantially reduce emissions with minimal economic disruption. Finally, contrary to the common belief that China's lax environmental regulation leads to its specialization in the production of emission-intensive products, we do not find Chinese exports to be more emission intensive than imports or trade flows in the rest of the world.

Related Literature. Shapiro and Walker (2018) utilize US Census data to compute the PLEI for 1,440 products, which represents the most detailed analysis to date, to the best of our knowledge. However, their method allocates firm-level emissions to each product based on its output value, implicitly assuming a uniform emission intensity across all products within each firm. This assumption contradicts the primary objective of determining PLEI. Our study addresses this inconsistency by introducing an iterative algorithm that ensures a "Micro-Macro" consistency in firm-product-level emission allocation and economy-wide PLEI calculations. Furthermore, leveraging customs data enables us to ascertain PLEI for a substantially more granular spectrum of approximately 4,500 products.[1]

[1]Our methodology requires the availability of emissions data at the firm level, which necessitates allocation of emissions across the product mix. Notably, some datasets report firm emissions distinctly for each product (Barrows and Ollivier, 2018).

In a recent study, Shapiro (2021) computes industry-level CO_2 intensities using the Exiobase global input-output table across 163 industries. A notable difference with our approach is that our PLEI calculations encompass emissions linked to domestic intermediate inputs and value-added, whereas Shapiro's metrics additionally incorporate emissions from imported intermediate inputs. Put differently, our PLEI framework quantifies emissions related to China's value-added production, while Shapiro's methodology accounts for emissions in China's gross exports. It is critical to acknowledge that neither approach is strictly superior; their applicability depends on the specific analytical context. For instance, in assessing China's CO_2 contributions globally, it is more pertinent to focus on emissions tied to China's value-added activities.

Our research makes a significant contribution to the expanding body of literature on environmental emissions within an international framework (Frankel, 2009).[2] This literature includes numerous studies estimating emissions embodied in trade flows, such as examining CO_2 emissions in China's value-added exports (Dietzenbacher et al., 2012), investigating international carbon leakage post the Kyoto Protocol (Aichele and Felbermayr, 2015), and assessing the welfare consequences of trade liberalization considering CO_2 emissions (Shapiro, 2016). Our PLEI measures are poised to be invaluable for future research in this field.

The structure of the paper is as follows: Section 2 delineates our data sources and the methodology for calculating PLEI, accompanied by extensive robustness checks and external validation. Section 3 documents novel stylized facts uncovered about PLEI. The concluding section provides a synthesis of our findings and implications.

[2]For a comprehensive review of the literature, we direct readers to Copeland and Taylor (2004), Cherniwchan et al. (2017), and Copeland et al. (2022).

2 Product-Level Emission Intensities

This section outlines the primary data sources at the core of our analysis and presents our novel approach to compute PLEI.

2.1 Data Source

The primary dataset for our study is China's Environmental Statistical Database (CESD), encompassing firm-level panel data pertaining to emissions. Annually, leading enterprises responsible for 85% of emissions within each county submit detailed reports regarding their emission metrics, energy consumption, and abatement measures to the Ministry of Ecology and Environment. Within this dataset, our primary focus lies on industrial water utilization and emissions, such as COD, SO_2, NOX, CO_2, industrial exhaust, ammoniacal nitrogen, soot, and dust.[3] To facilitate integration with additional datasets, we utilize information such as the firm's name, organizational code, postal code, and administrative location code. The CESD dataset is widely used in environmental studies in China (Liu et al., 2017; Fan et al., 2019), and its rigorous data collection protocol mitigates incentives for firms to misrepresent emission records (He et al., 2020). As for the data quality, studies have cross-validated the dataset with the U.S. satellite data for SO_2 (Rodrigue et al., 2022a).

The second dataset is the Annual Survey of Industry Enterprises (ASIE) for Chinese firms. The ASIE dataset covers state-owned enterprises and private firms with annual sales above 5 million RMB (20 million RMB since 2011). It includes key firm-level information such as industry classification and total sales. The CESD and the ASIE data can be merged together based on firms' organization code and company name. Kwon et al. (2023) show that this matching process successfully captures over 75% of the output

[3] CO_2 emissions primarily result from burning fossil fuels. The CESD dataset details each firm's use of coal, natural gas, and oil. To estimate carbon emissions, we multiply the consumption of each type of energy by its corresponding carbon emission factor, as determined by Guidelines for Provincial Greenhouse Gas Inventories published by National Development and Reform commission of China.

reported in the CESD.

Subsequently, to acquire insights into the HS 6-digit products that firms export, we merge our firm-level data with the Chinese customs data. This linkage is established using firm identifying information, following a similar methodology to Roberts et al. (2012) and Yu (2015). This enables us to identify the HS 6-digit product categories associated with each individual firm's export activities in our dataset.

In line with the methodology presented in Levinson (2009), we compute both direct and total emission intensities. *Direct* emission intensity is determined by normalizing the emissions generated in the final production stage by the total production value. On the other hand, *total* emission intensity takes into account emission footprint throughout the domestic value chain.[4] For this purpose, we utilize China's input-output table from EORA, a widely utilized resource in international trade research. Finally, we rely on BACI (Gaulier and Zignago, 2010) trade data to compare China's trade flows with the rest of the world.

2.2 Methodology, Validation and PLEI Estimates

In most datasets, firms do not typically report emissions data distinctively for each product they produce.[5] Building upon widely employed assumptions in the existing literature, we present a novel approach to compute emission intensities at the product level using firm-level information.

Our underlying conceptual framework assumes that a firm's emission intensity for a

[4]The distinction between CO_2 intensity measures developed by Shapiro (2021) deserves attention. Our *total* PLEI measure focuses on emissions associated with domestic value chain in China's exports, whereas Shapiro's model accounts for emissions in gross exports, encompassing emissions footprint embedded in imported intermediates in addition to domestic value chain. As briefly discussed in introduction, these two measures serve different objectives, and neither is strictly better than the other. For example, in attributing global emissions to specific nations, our PLEI measure is more relevant; whereas Shapiro's emission intensity measure is more relevant if an importing country strategically lowers its import tariffs to reduce its domestic CO_2 emissions via import substitution.

[5]One exception is the dataset used by Barrows and Ollivier (2018), where Indian firms distinctively report emissions for different products.

specific product can be decomposed into two components. The first component encompasses the economy-wide emission intensity of the product h denoted as ζ_h ($\zeta_h \geq 0$), which is influenced by the inherent characteristics of the production process and is common across all firms. For example, producing steel inherently generates more CO_2 emissions per unit value compared to smartphones. As such, we use a set of parameters ζ_h to capture such product-level variation in emission intensity common to all firms and refer it to as PLEI.

The second component influencing emission intensity is firm-specific, denoted as λ_i ($\lambda_i > 0$). This factor captures the inherent variations in emission intensities when firms, even while producing identical products, differ in their environmental footprints. Studies show that such disparities among firms could be attributed to variations in their productivity levels (Shapiro and Walker, 2018; Forslid et al., 2018). Additionally, different environmental regulations faced by the firms can further exacerbate these variations (Fan et al., 2019; He et al., 2020).

We link the two components as below to express firm-product-level emission intensity:

$$\zeta_{hi} = \zeta_h \cdot \lambda_i. \tag{1}$$

As an intuitive example, a firm with a λ_i value of 1.2 would produce product h at an emission intensity level that is 20% higher than the benchmark level ζ_h. As such, we refer to the parameter λ_i as the firm's (emission) *inefficiency parameter*. A few discussions regarding equation (1) are in order.

First, equation (1) assumes that the firm-specific factor λ_i is uniformly applicable across all products in the firm's product mix. This is an innocuous assumption for the majority of firms in our data as 81.40% of them produce only a single product. For those firms that produce multiple products, our assumption aligns with the conceptual frameworks presented in Bernard et al. (2011) and Barrows and Ollivier (2018), which propose the existence of a "core productivity" that uniformly impacts all products. Nonetheless, as

firms diversify their product portfolios, minor deviations in emission inefficiency across different products may arise. Our Monte Carlo simulations, delineated in Section 2.2.1, demonstrate that such variations minimally affect our algorithm's performance.[6] Given the predominance of single-product firms, theoretical support from the literature, and our algorithm's quantitative robustness to firm efficiency variation across products, we utilize equation (1) in our PLEI calculation.

Secondly, leveraging the temporal dimension of our dataset, we tag each firm-year observation with an index i. This indexation offers notable flexibility, enabling us to account for potential temporal variations in firms' productivity or abatement efficiency via a time-varying λ_i. Thus, ζ_h represents the long-term average PLEI - spanning from 2000 to 2013 for each pollutant - governed by inherent product characteristics. Any temporal enhancements in the cleanliness of a product's production process manifest as the associated firms' declining λ_i trajectory.

Lastly, the emissions embodied in domestic sales remains equally important for that of HS 6-digit export commodities. For domestic sales, the PLEI is gauged at the 4-digit CIC (Chinese Industrial Classification) level, encompassing 603 categories within our sample. In the subsequent exposition, the set of h includes both HS 6-digit products and 4-digit CIC sectors.

Our PLEI calculation proceeds iteratively as below.

1. Guess a vector of emission intensities ζ_h.
2. Equation (1) allows us to express firm i's total emission associated with product h as
$$z_{hi} = z_i \frac{\frac{y_{hi}}{y_i}\frac{z_{hi}}{y_{hi}}}{\sum_{h'}\frac{y_{h'i}}{y_i}\frac{z_{h'i}}{y_{h'i}}} = z_i \frac{s_{hi}\zeta_h\lambda_i}{\sum_{h'} s_{h'i}\zeta_{h'}\lambda_i} = z_i \frac{s_{hi}\zeta_h}{\sum_{h'} s_{h'i}\zeta_{h'}} \tag{2}$$
where z_i is total emission from firm i, y_{hi} is the firm's production of product h, y_i

[6] To explain, any deviation at the product level within a firm can either exhibit correlation or remain uncorrelated across different firms. In instances where it is correlated, our ζ_h estimates will recalibrate to accommodate the correlated production deviations. When uncorrelated, these deviations essentially act as statistical noise, becoming inconsequential with an increasing number of firms producing identical products in the dataset. Particularly, as the number of firms manufacturing a shared product increases, any firm-specific product-level deviation from (1) can be perceived as benign white noise for the ζ_h calculation.

is the firm's total production, s_{hi} ($\equiv y_{hi}/y_i$) is the production share of the product. Note that the first equal sign in (2) simply reflects an identity, the second equal sign applies equation (1), and the last one simply cancels out λ_i. As such, based on firm-level emissions z_i, product output shares s_{hi} and the vector of ζ_h from step 1, we calculate firm-product level emissions z_{hi} based on equation (2).

3. The economy-wide total pollution associated with product h is $Z_h = \sum_i z_{hi}$. We can then update the guess of emission intensities by
$$\zeta_h' = \frac{Z_h}{Y_h} \tag{3}$$
where $Y_h (\equiv \sum_i y_{hi})$ is the economy-wide total production of product h.
4. If ζ_h' are sufficiently close to ζ_h, the iteration process stops; otherwise, ζ_h' replace ζ_h for the next round of iteration.

Our algorithm offers several advantages for computing PLEI. First, the solution derived from this algorithm meets the conventional definition of emission intensities, as in equation (3). Additionally, our algorithm attributes firm-level total emissions to each product, based on the respective PLEI weighted by its output share, as described in equation (2).

Second, our algorithm is particularly useful when the researcher does not observe product-level emissions within a firm. Notably, a simpler version of this algorithm was applied in Shapiro and Walker (2018), where firm-level emissions were attributed to individual products using their respective revenue shares. This method results in the expression $z_{hi} = z_i s_{hi} / \sum_{h'} s_{h'i}$, which can be viewed as a special case of our equation (2), which assumes that ζ_h is uniform across all products. From this perspective, one primary advantage of our iterative algorithm is its ability to ensure that firm-level emission allocations correspond coherently with aggregated product-level emission intensities.

Lastly, under our framework, the parameter λ_i is designed to be specific to each firm-year observation, ensuring that our PLEI metric inherently accommodates for firm-specific emissions inefficiency that may vary over time due to various factors, such as productiv-

ity (Shapiro and Walker, 2018), emissions abatement (Forslid et al., 2018; Kwon et al., 2023), market power (Rodrigue et al., 2022a), and regulatory stringency (He et al., 2020).

For our algorithm to deliver reliable results, it is essential that each product is produced by a multitude of firms, thereby neutralizing the influence of firm-specific variations. To this end, we exclude the lowermost 10% of products based on export volume.[7] Our refined sample comprises 5,144 product categories, segmented into 603 domestic industries and 4,541 HS 6-digit products. This dataset encompasses 666,213 firm-year observations with minor deviations across pollutants due to data availability. On average, each product has 440 firm-year observations, with a median of 140. Given the large number of observations associated with each product, we contend that this allows for a precise estimation of the PLEI common to all firms.

2.2.1 Monte Carlo Simulations

We use Monte Carlos simulations to demonstrate the efficacy of our algorithm in various scenarios. A succinct description of the data generating process (DGP) and the main findings are presented here, with comprehensive details deferred to Appendix A.

In essence, our DGP preserves three key dimensions in the data: the distribution of ζ_h estimates, firm-product ratio, and the distribution of product numbers across firms. To verify the robustness of our iterative algorithm against data perturbations, we introduce a uniformly distributed multiplicative measurement error, spanning between 70% and 130%. Additionally, we investigate a scenario wherein the efficiency term in the DGP, denoted by λ_{hi}, varies not only across firms but also across their individual products. Diagnostic metrics suggest that our DGP adeptly mirrors the pivotal data moments. All of our Monte Carlo results are based on an ensemble of 500 distinct simulations.

Key findings from Monte Carlo simulations can be summarized as follows. Firstly, our iterative procedure adeptly recovers the "true" (presumed) ζ_h values with high pre-

[7] For reference, the dropped products based on this criterion are produced by a median (mean) number of 5 (9.26) firms, whereas the remaining products are produced by 129 (351.82) firms.

cision. Across 500 simulations, both the median and mean pairwise correlation values between the calculated and true ζ_h consistently exceed 0.998 (with standard errors less than 0.002). Given the substantial measurement errors we employed for each simulation, these results are particularly noteworthy, implying that our algorithm remains robust to significant data disturbances. Secondly, even when the DGP deviates from our assumption so that the emissions efficiency term λ_{hi} varies by products within each firm, our iteration algorithm can still recover true ζ_h with the median and mean correlation above 0.998, albeit the standard deviation of the correlation coefficients slightly increases to 0.003. The superb performance of our algorithm can be partially explained by a large proportion of single-product firms (81.40%) and firms that produce only a few products (88.65% produce 3 products or fewer). Thus, a firm's single efficiency parameter λ_i can parsimoniously reflect PLEI variations across firms while regarding the within-firm variation across products as additional data noise.

As a supplementary analysis, we also employ an ordinary least squares (OLS) methodology to estimate ζ_h using the following equation

$$z_i = \sum_h y_{hi}\zeta_h + e_i, \tag{4}$$

where z_i represents firm-level aggregate emissions, y_{hi} indicates the firm-product level output, and e_i is the encompassed error term. Our results show that the OLS method produced *negative* $zeta_h$ estimates for approximately half of the examined products, raising concerns regarding its accuracy and interpretation. Furthermore, even when the focus is narrowed to products with positive ζ_h estimates, the correlation with the true ζ_h exhibits lower mean and considerably heightened standard deviation than our approach. In conclusion, our Monte Carlo findings underscore the superior accuracy and robustness of our iterative method.

2.2.2 Total Domestic Emission Intensities

The output from our iterative algorithm is typically referred to as *direct* emission intensity, capturing emissions associated with the final production process. This stands in contrast to *total* emission intensities, which encompass the cumulative emissions across the entire value chain in domestic production.[8]

To derive total emission intensities, we adopt the following methodology. Consider $\mathbf{C}$ as an N-by-N matrix of direct requirement coefficients. Each element, denoted by c_{ij}, specifies the monetary value of the input from industry i (row) requisite for generating a dollar's output in industry j (column). The vectors $\mathbf{x}$ and $\mathbf{y}$ respectively represent the aggregate output and the total end demand of industry i. With this construction, the relationship can be articulated as:

$$\mathbf{x} = \mathbf{C}\mathbf{x} + \mathbf{y}. \tag{5}$$

This expression manifests that the industrial output, represented by $\mathbf{x}$, is either assimilated as intermediate input, denoted by $\mathbf{C}\mathbf{x}$, or used as final goods, represented by $\mathbf{y}$.

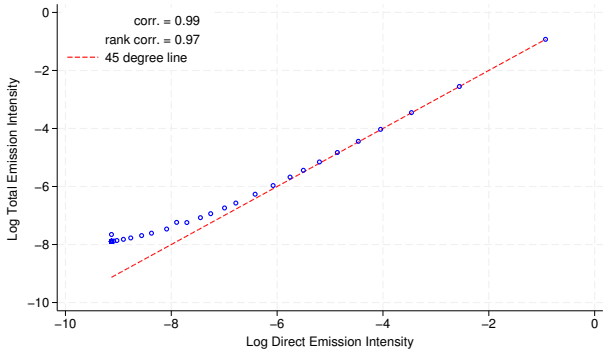
Rearranging terms, we derive: $\mathbf{x} = [\mathbf{I} - \mathbf{C}]^{-1}\mathbf{y}$, where the matrix $[\mathbf{I} - \mathbf{C}]^{-1} \equiv \mathbf{T}$ is referred to as the Leontief total requirements matrix (Leontief, 1970). In this matrix, each element, represented by t_{ij} in $\mathbf{T}$, quantifies the monetary output of industry i essential for the final production in industry j. Let $\boldsymbol{\zeta}$ and $\tilde{\boldsymbol{\zeta}}$ denote direct and total emission intensities, respectively. The relationship $\tilde{\boldsymbol{\zeta}}'\mathbf{y} = \boldsymbol{\zeta}'\mathbf{x}$ holds, wherein both sides of the equation correspond to the aggregate emissions in the economy. Consequently, the total pollution coefficients, $\tilde{\boldsymbol{\zeta}}$, can be elucidated as:

$$\tilde{\boldsymbol{\zeta}}' = \boldsymbol{\zeta}'\mathbf{T}. \tag{6}$$

To calculate the Leontif matrix, we employ the EORA dataset, which provides an

[8]Studies such as Levinson (2009), Lyubich et al. (2018), and Shapiro (2021) have employed analogous metrics for computing total emission intensities. Notably, both Levinson (2009) and Lyubich et al. (2018) limit their considerations to domestic intermediate inputs. Conversely, Shapiro (2021) extends this by accounting for emissions embedded in imported intermediate components. We align with the former methodology, primarily because direct emission intensities at the HS 6-digit level for international domains remain outside our purview. Moreover, the total *domestic* emission intensity proves more pertinent for various empirical evaluation.

Figure 1: Direct vs. Total Emission Intensities (CO_2)



Notes: These figures compare the logarithmic values of direct and total emission intensities for CO_2. We add the smallest value of total emission intensities to all values before log transformation to account for zero emission intensities. For exposition, we choose one HS 6-digit product from each of 100 percentile values of direct emission intensities and represent it as a dot. The dots presented in the figure appear to be less than 100 as many dots overlap each other toward zero (plus a small value) direct emission intensities. The displayed correlations derive from the full distribution of emission intensities rather than the 100 displayed dots. Similar figures for other emissions are presented in Appendix Figures E.3 and E.4.

input-output table encompassing 87 industries in China: 6 agricultural, 6 mining, 72 manufacturing, and 3 utility sectors.[9] Since this table aggregates products more broadly than the PLEI, we make two key assumptions. First, we assume that all products within an individual EORA industry demand the same input bundle per unit of output. Additionally, we assume that input supplies within an EORA industry are distributed across each HS 6-digit product in proportion to the product's output share.

Figure 1 plots direct emission intensities (displayed on the horizontal axis) against total emission intensities (on the vertical axis) for CO_2, which garnered attention in the context of global warming.[10] Several observations are noteworthy. Primarily, total emission intensities consistently exceed their direct counterparts. This divergence arises naturally from the definition of total emission intensities, which encapsulate emissions from both domestic intermediate inputs and the final production stage. Moreover, products with lower direct emission intensities exhibit more pronounced spikes in total emission inten-

[9]While the EORA input-output table also enumerates service industries, they are excluded from our analysis since the firms in our dataset do not pertain to these sectors.

[10]For the most part in our main text, we present results pertaining to CO_2 emissions. However, it is noteworthy that analogous findings are observed for other pollutants and are presented in Appendix E.

sities, for even a small amount of dirty intermediate input may potentially elevate the total emission measure extensively. Conversely, products characterized by elevated direct emission intensities witness only marginal hikes in total measures. This is attributed to the limited impact of integrating relatively cleaner intermediate inputs on the cumulative emission value. The correlation and rank-correlation coefficients between the two emission intensities hover around 0.90.

In essence, Figure 1 underscores that a sole reliance on direct emission intensities can lead to an understated account of total emissions, especially when the final production stage is comparably clean. Nonetheless, this discrepancy does not affect the PLEI ranking significantly.[11]

2.3 Validation

We employ two primary methodologies to validate our PLEI measures. First, we draw upon conventional wisdom of which products inherently generate more emissions during production. By ensuring that the products delineated as emission-intensive via PLEI align with our expectations, we gain confidence in the PLEI measure. Additionally, we juxtapose our PLEI values (after aggregation) against the industry-level emission intensities identified in alternative datasets.

2.3.1 Top Emission Intensive Products

In Tables 1 and 2, we enumerate the top emission-intensive product groups, as characterized by their total PLEI levels. We focus our attention primarily on CO_2 and COD, with corresponding tables for other pollutants relegated to Appendix E. Each table showcases the six most emission-intensive HS 2-digit product groups, alongside the three leading emission-intensive HS 6-digit products within each category.

[11] We also note that the conclusions drawn from our subsequent analysis are robust to employing either of emission intensities.

Both tables resonate well with conventional wisdom. For instance, Table 1 identifies the glass, cement, inorganic chemicals and metal industries as leading CO_2 emitters, stemming from their dependency on energy and coal combustion. These industries are also identified as CO_2-intensive industries according to the EU.[12] Similarly, Table 2 shows that industries such as chemical, paper, textile, and leather are significant contributors to COD emissions – which is expected due to their intensive chemical usage and resulting wastewater. As such, we view the intuitive industry patterns demonstrated in both tables (as well as those for other pollutants in Appendix E) as evidence for our PLEI adeptly reflecting intrinsic emission properties inherent in production processes.

[12] https://www.europarl.europa.eu/doceo/document/TA-9-2021-0071_EN.pdf (accessed January 26 2024).

Table 1: Top CO_2-intensive Products

HS Code	Description
700000	**Glass and glassware**
700239	Glass; unworked, in tubes, other than of glass having a linear coefficient of expansion n...
701391	Glassware; n.e.c. in heading no. 7013, of lead crystal
700210	Glass; unworked, in balls (other than microspheres of heading no. 7018)
250000	**Plastering materials, lime and cement**
250610	Quartz; other than natural sands
250870	Clays (excluding expanded clays of heading no. 6806); chamotte or dinas earths
250620	Quartzite; whether or not roughly trimmed or merely cut, by sawing or otherwise, into blo...
280000	**Inorganic chemicals**
283911	Silicates; sodium metasilicates
281520	Potassium hydroxide (caustic potash)
281830	Aluminium hydroxide
820000	**Tools, implements, cutlery, spoons and forks, of base metal; parts thereof, of base metal**
821195	Knives; with handles of base metal
820770	Tools, interchangeable; (for machine or hand tools, whether or not power-operated), for m...
821490	Cutlery; hair clippers and mincing knives
870000	**Vehicles; other than railway or tramway rolling stock, and parts and accessories thereof**
870390	Vehicles; for transport of persons (other than those of heading no. 8702) n.e.c. in headi...
870590	Vehicles; break-down lorries, road-sweepers, spraying lorries, mobile workshops, mobile r...
870990	Vehicles; parts of the vehicles of heading no. 8709
730000	**Iron or steel articles**
730439	Iron or non-alloy steel (excluding cast iron); seamless, (excluding cold-drawn or cold-ro...
732211	Radiators and parts thereof; for central heating, (not electrically heated), of cast iron
730431	Iron or non-alloy steel (excluding cast iron); seamless, cold-drawn or cold-rolled, tubes...

Table 2: Top COD-intensive Products

HS Code	Description
380000	**Chemical products n.e.c.**
380400	Lyes, residual; from the manufacture of wood pulp, whether or not concentrated, desugared...
380892	Fungicides; other than containing goods specified in Subheading Note 1 to this Chapter; p...
381190	Oxidation and gum inhibitors, viscosity improvers, anti-corrosive preparations, other pre...
280000	**Inorganic chemicals**
282732	Chlorides; of aluminium
282741	Chloride oxides and chloride hydroxides; of copper
282510	Hydrazine and hydroxylamine and their inorganic salts
480000	**Paper and paperboard**
480421	Kraft paper and paperboard; sack kraft paper, uncoated, unbleached, in rolls or sheets, o...
480530	Paper and paperboard; sulphite wrapping paper, uncoated, in rolls or sheets
480620	Paper; greaseproof papers, in rolls or sheets
540000	**Man-made filaments and textile materials**
540234	Yarn, synthetic; filament, monofilament (less than 67 decitex), textured, of polypropylen...
540331	Yarn, artificial; filament, monofilament (less than 67 decitex), of viscose rayon (not hi...
540310	Yarn, artificial; filament, monofilament (less than 67 decitex), of high tenacity viscose...
410000	**Raw hides, skins and leather**
410632	Tanned or crust hides and skins; of swine, without hair on, whether or not split, but not...
410622	Tanned or crust hides and skins; of goats or kids, without hair on, whether or not split,...
411320	Leather; further prepared after tanning or crusting, including parchment-dressed leather,...
290000	**Organic chemicals**
293391	Heterocyclic compounds; n.e.c. in headings no. 2933.1, 2933.2, 2933.3, 2933.4, 2933.5, 29...
292214	Amino-alcohols, other than those containing more than one kind of oxygen function; their ...
292243	Amino-acids, other than those containing more than one kind of oxygen function, and their...

2.3.2 Industry Level External Validation

The CESD dataset has gained widespread traction in empirical research, and its quality has been affirmed by previous investigations, as highlighted by Rodrigue et al. (2022a). A viable strategy for external validation of our PLEI measures involves aggregating them at the industry level for comparisons with emission intensities computed using alternate datasets.[13] We relegate the details to Appendix B and describe our main findings. In general, our comparisons with the Chinese Statistical Yearbook, EXIOBASE, and EORA yield mean correlation coefficients of 0.79, 0.48, and 0.46, respectively, across several pollutants. This underscores the credibility of our PLEI measures, confirming that, upon aggregation to the industry level, they align closely with industrial emission intensities derived from alternate sources.

2.4 Robustness Check and Additional Results

2.4.1 Initial Guess of ζ_h

In our baseline estimation, we uniformly designate the initial guess for ζ_h as 1.[14] To ascertain the robustness of our findings, we randomize the initial ζ_h by drawing from a uniform distribution within the range $[0.1, 5]$. This exercise yields 500 independent terminal ζ_h values, each potentially distinct due to the randomized starting point. Impressively, the median of the pairwise correlations are at 1.00 for all pollutants, underlining the robustness of our algorithm to variations in the initial ζ_h guesses.[15]

[13] While Shapiro and Walker (2018) employ the U.S. census data to calculate 1,440 product-level emission intensities at five-digit Standard Industrial Classifications level, data access restrictions hinder direct juxtaposition with their values.

[14] It is noteworthy that using any positive constant as an initial guess is equivalent.

[15] Detailed results are available upon request.

2.4.2 Time Consistency of ζ_h

We operate under the presumption that PLEI reflects the inherent time-invariant production characteristics of individual products. This subsection summarizes the key results in Appendix D, which is dedicated to carefully examining this assumption. In brief, we derive product-year specific emission intensities, denoted as ζ_{ht}, from a single iteration of ζ_h and compare these values with ζ_h. Should the actual yearly emission intensities substantially deviate from the estimated long-term average PLEI, a marked difference between ζ_{ht} and ζ_h would appear.[16] Our analyses, as laid out in Appendix D, indicate that ζ_{ht} values align closely with ζ_h over multiple years. Specifically, when regressing $\ln(\zeta_{ht})$ against $\ln(\zeta_h)$, the estimated coefficients consistently approximate 1.00, with estimated standard errors less than 0.01. This empirical evidence substantiates the hypothesis that the intrinsic emission characteristics of products evolve at a markedly slow pace over time.

2.4.3 Distribution of λ_i

Our algorithm to calculate PLEI does not require any prior knowledge of λ_i, the firm-specific emission inefficiency parameter, nor does it calculate λ_i as an intermediate step. However, once PLEI is determined, we can infer the λ_i value for each firm-year observation by applying the following equation:

$$z_i = \lambda_i y_i \sum_h s_{hi} \zeta_h, \tag{7}$$

[16] To elucidate, consider a scenario where all products are exclusively manufactured by single-product firms. In this case, a single-round iteration for ζ_{ht} using a single-year data would yield values that are independent of pre-supposed ζ_h, as they are not used for distributing emissions internally within each firm. Thus, a high correlation ζ_{ht} and ζ_h suggests that firms that produce emission intensive products according to ζ_h are also emission intensive in year t. (As mentioned earlier, single-product firms account for approximately 80% of our firm sample.) Next, consider another scenario where products are produced by multi-product firms and for some years the underlying ζ_{ht} deviate substantially from ζ_h, leading to unexpectedly lower average emission intensity for firms that produce emission-intensive products according to ζ_h. In this case, a single-iteration for ζ_{ht} would vary considerably from ζ_h. Thus, we argue that a comparison of ζ_h and ζ_{ht} provides insights into the tenability of assuming temporal consistency in PLEI.

where, as previously mentioned, s_{hi} represents the output value share of product h for firm i. The expression $y_i \sum_h s_{hi}\zeta_h$ represents the estimated emissions attributable to the firm's output, product mix, and the PLEI under the assumption that λ_i equals 1. Consequently, λ_i can be interpreted as a ratio, comparing the firm's actual emissions to the hypothetical emissions level of a representative firm with an identical product mix. Notably, λ_i serves as a comprehensive term for various determinants of firm-level emission intensities, including productivity (Shapiro and Walker, 2018), emissions abatement (Forslid et al., 2018; Kwon et al., 2023), market power (Rodrigue et al., 2022b), and regulatory stringency (He et al., 2020).

Although a detailed exploration of λ_i's nature falls outside the scope of this paper, we identify and describe key trends associated with λ_i in Figure 2. Specifically, we observe that the firm-year emission inefficiency parameter λ_i: (1) shows a consistent downward trend over time, (2) tends to be lower among firms with higher productivity, and (3) is typically higher for state-owned enterprises (SOEs). These observations might indicate an overall enhancement in productivity and emission reduction efforts among Chinese firms over time. Additionally, the disparity between SOEs and non-SOEs could be attributed to productivity differences and stricter regulatory compliance required for non-SOEs. We note that the patterns illustrated in Figure 2 align closely with the literature – reinforcing the credibility of our PLEI calculation, which is the critical input for inferring firm-level λ_i.

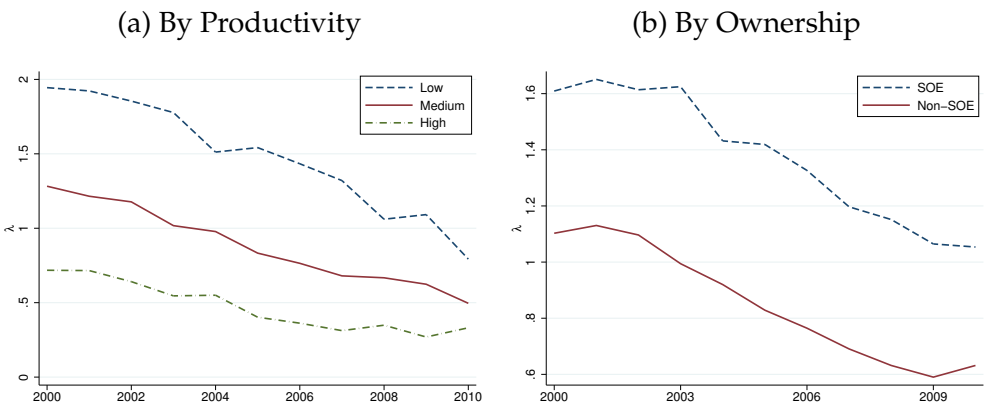
3 Stylized Facts about the PLEI

This section highlights key stylized facts regarding the PLEI. All subsequent analyses in this paper are based on *total* emission intensities.

Stylized Fact 1: PLEI feature a striking degree of heterogeneity.

The first and most notable feature of PLEI is its significant heterogeneity across prod-

Figure 2: Distribution of λ by Productivity and Ownership (CO_2)



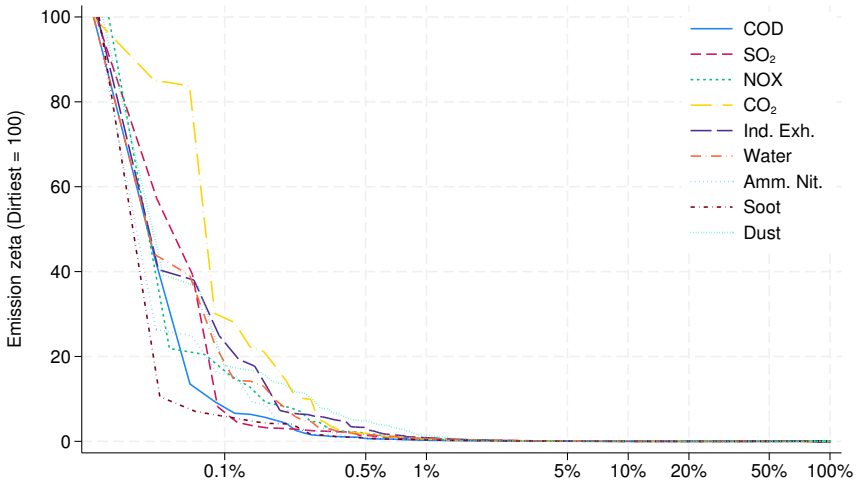
Notes: These figures show the distribution of firms' emission inefficiency parameter λ by the level of productivity (left) and ownership (right).

ucts. For illustration, we normalize the highest PLEI value for each pollutant to 100 and plot the distribution in Figure 3. Notably, emission intensities vary widely even if we focus on only the top 0.5% of emission-intensive products (approximately 20 products). Products with emission intensities below the top 1% threshold have negligible presence compared to the top emission-intensive products. This substantial discrepancy is further confirmed by the Gini coefficient for emission intensities, which exceeds 0.95 across all pollutants.

A natural implication of this on environment policy is the need for a detailed Pigouvian tax schedule that takes into account product-level variation in PLEI. While it would be first-best to tax each product individually based on specific emission contents if they are either perfectly observed (or truthfully revealed by the producers), it is rather challenging in practice especially when production and taxation take place in different countries.

For instance, the EU's CBAM to be implemented in 2026 will most likely adopt some form of *ad-valorem* tax system, which involves determining trade value and the CO_2 intensity of the corresponding products. If the CO_2 intensity is determined based on average

Figure 3: Skewness of Emission Intensities across 6-digit HS Code Products



Notes: This figure illustrates the total emission intensities of HS 6-digit products. The highest value is normalized to 100. The horizontal axis is log-scaled after multiplying 100.

emission intensities calculated at a broader product category, it will be less efficient since emission-intensive (low-emission) products would be taxed insufficiently (excessively). The striking PLEI heterogeneity presented in Figure 3 suggests that the PLEI determined at a highly granular product level can be very useful for guiding the tax system.

Stylized Fact 2: Top 10% of emission-intensive products contribute to nearly 75% of total emissions, yet these products represent only 4% of China's export value.

Table 3 provides a detailed breakdown of emissions and export volumes for different product groups. Panel A, where products are ranked and grouped based on their emission intensities, reveals a striking fact that the top 1% of the most emission-intensive products contribute, on average, to over 45% of total emissions across various pollutants. The cumulative impact of the top 10% of emission-intensive products, which encompasses approximately 440 items (summing the first three rows), is responsible for roughly 75% of total emissions. Conversely, as detailed in the lower section of Panel A, these top 1% and 10% of emission-intensive products account for merely 0.12% and 4.17% of China's total export volume, respectively. This marked imbalance further underscores the significant

potential of our PLEI measure to guide efficient tax policy. Specifically, our stylized facts 1 and 2 suggest that environment tax targeted at a small number of emission-intensive products can effectively reduce emissions while causing minimum economic distortion.

Panel B of Table 3 provides a slightly different perspective by categorizing products according to their total emissions. As anticipated, the proportion of emissions attributed to top products and their export volume both increase compared with Panel A. Still, a similar critical observation from Panel B is the concentration of emissions within a narrow product range: the top 10% of products are responsible for over 90% of the total emissions.

Based on our stylized fact 2, we introduce a binary classification for products based on their emission intensities. Specifically, products are classified as *dirty* if their emission intensity falls within the highest decile (top 10%) for any given pollutant; the rest are categorized as *clean*. Note that under this pollutant-specific classification, a product such as leather might be categorized as *dirty* with respect to its COD emissions, yet be considered *clean* in the context of its CO_2 emissions.

Table 3: Percentage of Total Emissions Accounted For by Top Emission-intensive 6-digit HS Code Products

	AVG	*CO_2*	*COD*	*SO_2*	*NOX*	*Exh.*	*Water*	*Amm.*	*Soot*	*Dust*
Panel A. Product Ranking Based on Emission Intensities										
pct. of emissions										
top 1 pct. HS	33.82	69.89	27.17	16.93	18.55	26.98	20.54	21.22	20.78	69.71
next 4 pct. HS	26.13	9.18	24.10	29.53	22.39	36.37	27.00	12.49	39.92	20.51
next 5 pct. HS	11.44	3.84	15.78	13.09	18.76	9.50	17.27	5.67	11.25	2.00
next 40 pct. HS	17.36	13.01	22.95	22.59	26.56	20.77	19.72	30.96	10.39	2.89
bottom 50 pct. HS	11.26	4.08	10.00	17.86	13.74	6.38	15.47	29.65	17.65	4.90
pct. of exports										
top 1 pct. HS	0.13	0.12	0.11	0.07	0.15	0.12	0.13	0.21	0.07	0.23
next 4 pct. HS	1.05	0.98	0.66	0.81	0.99	1.20	1.06	0.69	1.07	1.60
next 5 pct. HS	2.64	1.90	1.47	1.58	1.61	2.52	1.60	1.33	2.14	8.28
next 40 pct. HS	35.88	60.02	49.01	26.41	35.12	31.06	38.76	41.53	19.43	27.19
bottom 50 pct. HS	60.32	36.99	48.75	71.13	62.13	65.11	58.46	56.25	77.29	62.70
Panel B. Product Ranking Based on Total Emissions										
pct. of emissions										
top 1 pct. HS	58.49	78.52	50.40	44.33	52.82	57.25	50.13	47.44	53.00	81.44
next 4 pct. HS	24.55	12.49	28.13	30.41	27.78	27.87	29.40	25.61	26.77	13.58
next 5 pct. HS	7.52	3.81	9.67	10.83	8.27	7.64	9.19	10.33	8.63	2.11
next 40 pct. HS	8.82	4.83	11.20	13.45	10.28	6.86	10.50	15.26	10.77	2.64
bottom 50 pct. HS	0.63	0.35	0.60	0.98	0.86	0.38	0.78	1.35	0.83	0.24
pct. of exports										
top 1 pct. HS	11.18	15.36	15.54	15.60	16.01	4.67	15.59	21.05	6.21	0.49
next 4 pct. HS	19.78	20.85	15.58	15.30	14.62	16.76	18.21	19.85	22.79	34.10
next 5 pct. HS	15.81	15.87	12.50	16.05	14.92	17.25	15.98	12.81	17.12	16.80
next 40 pct. HS	41.87	38.70	45.76	42.06	39.97	48.81	40.44	36.67	41.64	37.56
bottom 50 pct. HS	11.35	9.21	10.60	10.98	14.47	12.51	9.78	9.62	12.23	11.05
Num. of HS products	.	4482	4460	4336	3757	4262	4473	4383	4195	4036

Notes: The table presents a breakdown of emissions associated with the most emission-intensive products classified under 6-digit HS codes. Specifically, Panel A of the table details the proportion of total emissions accounted for by the top emission-intensive HS codes, while Panel B focuses on HS codes with the highest absolute emission levels. The "AVG" column represents the average values of each row across nine pollutants.

Stylized Fact 3: Broad product categories significantly underestimate the heterogeneity in PLEI.

The heterogeneity in emission intensities may have been underappreciated when aggregated for broader product categories, such as at the sectoral level or using HS 2-digit classifications. This oversight becomes particularly evident in contexts where PLEI within a broad category exhibit significant variation.

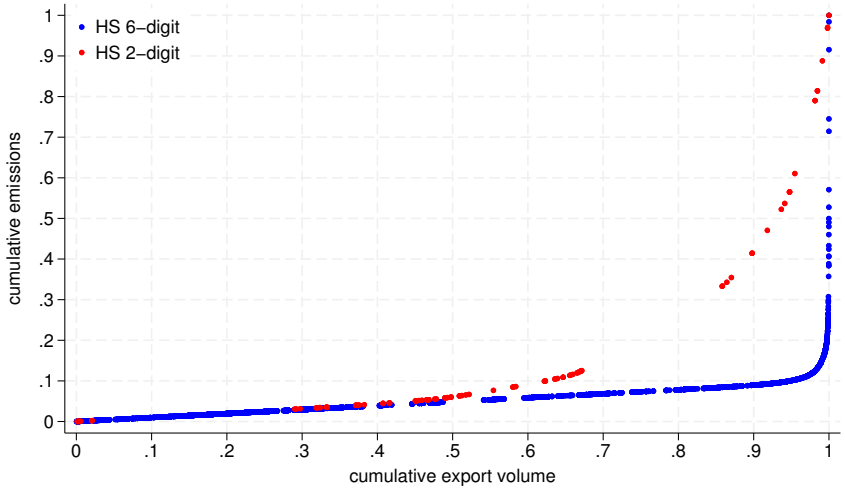
In Figure 4, we order products by ascending CO_2 emission intensities and present the relationship between cumulative emissions and export volume for both HS 2-digit and 6-digit levels.[17] This figure illustrates a marked discrepancy in cumulative emissions between the two levels of classification. Notably, the HS 6-digit level graph exhibits a steep curve towards its right end, suggesting that approximately 80% of emissions can be traced back to a relatively minor portion of the trade volume (less than 3%, as quantified in Table 3). In contrast, the HS 2-digit level analysis, which features much flatter distribution, indicates that a similar proportion of emissions is associated with a much larger segment of trade, approximately 30% of the total volume.

This pronounced discrepancy highlights the critical importance of granular PLEI data for effective policy formulation. Consider, for example, the European Union's recent policy "Fit for 55", which aims to reduce net greenhouse gas emissions by at least 55% by 2030. If this goal is applied to the forthcoming Carbon Border Adjustment Mechanism, which imposes taxes on imported goods based on their carbon content, a policy evaluation based on HS 2-digit level PLEI would suggest that at least 10% of Chinese imports would be affected according to Figure 4. Yet, if the tax rates can be set at a more granular level with the help of 6-digit level PLEI, only a negligible share (less than 1%) of trade would be impacted.

Stylized Fact 4: Surprisingly, Chinese exports are not dirtier than its imports or trade flows in the rest of the world. From 2000 to 2013, the growth of China's dirty-product

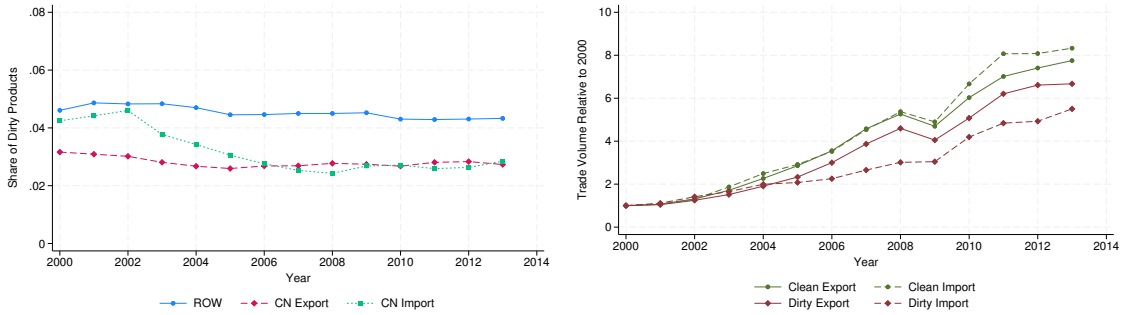
[17] For data on other pollutants, see the additional figures in Appendix E.

Figure 4: Cumulative Total Emissions and Trade Volume (CO_2)



Notes: This figure illustrates the cumulative distributions of total emissions and export volumes. Products are sorted in increasing order based on their total emission intensities.

Figure 5: Emissions in Chinese Trade, Imports vs. Exports (CO_2)



Notes: The left figure presents the trade share of dirty products, defined as those in the top 10% of emission intensities, within China's exports, imports, and global trade flows. The right figure plots the growth trajectory of both clean and dirty product imports and exports since 2000.

exports lagged behind its clean-product exports.

Utilizing our PLEI measures, we analyze China's trade patterns concerning environmental emissions. In the left panel of Figure 5, we examine the proportion of dirty products within China's exports, imports, and international trade flows. Under the Pollution Haven Hypothesis (PHH) (Copeland and Taylor, 1994), one might anticipate China's increasing specialization in dirty products, given its comparatively lax environmental regulations relative to major trading partners (e.g., US, Hong Kong, Korea, and Japan) and steadily declining trade costs since 2001. However, contrary to this hypothesis, our observations do *not* reveal an increasing concentration in dirty-product exports from China during this period.

Perhaps more surprisingly, the proportion of dirty products in China's imports, which was expected to be cleaner, was actually higher than that in exports prior to 2006, and comparable thereafter. This observation is reminiscent of the well-known Leontief paradox (Leontief, 1953).[18]

Additionally, it is pertinent to highlight that the trade share of dirty products among countries, excluding China, surpasses the trade share of both China's imports and exports of these goods. This observation further challenges the notion that China specializes in producing environmentally harmful products. While a comprehensive analysis of the minimal role of environmental policy in shaping trade patterns falls outside the purview of this study, it is noteworthy to mention that similar research on China corroborates this finding. For instance, Duan et al. (2021) also conclude that environmental policy plays a relatively insignificant role in influencing trade specialization.

In the right panel of Figure 5, we depict the evolution of China's imports and exports for dirty and clean products, respectively, since 2000. Most significantly, the growth of

[18]The classification of dirty products is based on their production process. Therefore, the set of dirty products for China might differ from that for its trading partners. Ideally, with equivalent data for all of China's trading partners, dirty products for imports and exports should be separately defined, but data limitations prevent this. We assume for convenience that dirty products in terms of China's production are also likely to be dirty in other countries. In Appendix B, we show that emission intensities are highly correlated across countries at the sectoral level, with correlation coefficients generally above 0.6.

both imports and exports of clean products surpasses dirty counterparts, suggesting that product-level trade shocks in favor of clean products, which affect both imports and exports (e.g., worldwide increased demand for semiconductors), play a pivotal role. As a result, this leads to the observation that the growth of China's clean exports exceeds that of dirty exports, contradicting the predictions made by the PHH. Meanwhile, the growth of clean imports significantly exceeds that of dirty imports. Thus, the difference-in-differences inferred from this figure, comparing dirty products' net growth (export growth minus import growth) with that of clean products, appears to be consistent with the PHH. The caveat is that one has to take into account worldwide faster increase in clean products' trade in order to identify the force of PHH from data. This message is consistent with the previous studies, which show that concurrent confounding factors are a major challenge for empirically verifying the PHH (Grossman and Krueger, 1995; Copeland et al., 2022).

4 Conclusion

This paper introduces a novel methodology for calculating product-level emission intensities. Key advantages of our method include: (1) the capability to compute PLEI at a highly detailed product level; (2) flexibility to incorporate firm-level heterogeneity; (3) an iterative approach ensuring theoretical consistency of PLEI at both firm and aggregate levels, particularly in determining product-level emissions for multi-product firms; and (4) broad applicability, compatible with firm-level emissions data and firm-product level output data.

Applying this methodology to Chinese data from 2000 to 2013, we uncover significant insights about PLEI. We observe pronounced heterogeneity across products, with the top 10% of the most emission-intensive product categories accounting for approximately 75% of total emissions in Chinese exports, yet constituting only about 4% of trade volume.

Additionally, our findings indicate that using broad categories to estimate emission intensities can severely underestimate the true extent of PLEI heterogeneity. Therefore, we argue that policies related to emission intensities, such as the European Union's Carbon Border Adjustment Mechanism, would greatly benefit from a detailed measurement of PLEI.

References

AICHELE, R. AND G. FELBERMAYR (2015): "Kyoto and Carbon Leakage: An Empirical Analysis of the Carbon Content of Bilateral Trade," *The Review of Economics and Statistics*, 97, 104–115.

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*, 88, 134–158.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2011): "Multiproduct Firms and Trade Liberalization," *The Quarterly Journal of Economics*, 126, 1271–1318.

CHERNIWCHAN, J., B. R. COPELAND, AND M. S. TAYLOR (2017): "Trade and the Environment: New Methods, Measurements, and Results," *Annual Review of Economics*, 9, 59–85.

COPELAND, B. R., J. S. SHAPIRO, AND M. SCOTT TAYLOR (2022): "Globalization and the Environment," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 5 of *Handbook of International Economics: International Trade, Volume 5*, 61–146.

COPELAND, B. R. AND M. S. TAYLOR (1994): "North-South Trade and the Environment," *The Quarterly Journal of Economics*, 109, 755–787.

——— (2004): "Trade, Growth, and the Environment," *Journal of Economic Literature*, 42, 7–71.

DIETZENBACHER, E., J. PEI, AND C. YANG (2012): "Trade, Production Fragmentation, and China's Carbon Dioxide Emissions," *Journal of Environmental Economics and Management*, 64, 88–101.

DUAN, Y., T. JI, Y. LU, AND S. WANG (2021): "Environmental Regulations and International Trade: A Quantitative Economic Analysis of World Pollution Emissions," *Journal of Public Economics*, 203, 104521.

FAN, H., J. S. GRAFF ZIVIN, Z. KOU, X. LIU, AND H. WANG (2019): "Going Green in China: Firms' Responses to Stricter Environmental Regulations," .

FORSLID, R., T. OKUBO, AND K. H. ULLTVEIT-MOE (2018): "Why Are Firms That Export Cleaner? International Trade, Abatement and Environmental Emissions," *Journal of Environmental Economics and Management*, 91, 166–183.

FRANKEL, J. (2009): "Environmental Effects of International Trade," Scholarly Articles, Harvard Kennedy School of Government.

GAULIER, G. AND S. ZIGNAGO (2010): "BACI: International Trade Database at the Product-Level (the 1994-2007 Version)," .

GROSSMAN, G. M. AND A. B. KRUEGER (1995): "Economic Growth and the Environment*," *The Quarterly Journal of Economics*, 110, 353–377.

HE, G., S. WANG, AND B. ZHANG (2020): "Watering Down Environmental Regulation in China," *The Quarterly Journal of Economics*, 135, 2135–2185.

KWON, O., H. ZHAO, AND M. Q. ZHAO (2023): "Global Firms and Emissions: Investigating the Dual Channels of Emissions Abatement," *Journal of Environmental Economics and Management*, 118, 102772.

LEONTIEF, W. (1953): "Domestic Production and Foreign Trade; The American Capital Position Re-Examined," *Proceedings of the American Philosophical Society*, 97, 332–349.

——— (1970): "Environmental Repercussions and the Economic Structure: An Input-Output Approach," *The Review of Economics and Statistics*, 52, 262–271.

LEVINSON, A. (2009): "Technology, International Trade, and Pollution from US Manufacturing," *American Economic Review*, 99, 2177–2192.

LIU, M., R. SHADBEGIAN, AND B. ZHANG (2017): "Does Environmental Regulation Affect Labor Demand in China? Evidence from the Textile Printing and Dyeing Industry," *Journal of Environmental Economics and Management*, 86, 277–294.

LYUBICH, E., J. SHAPIRO, AND R. WALKER (2018): "Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Environmental Policy," *AEA Papers and Proceedings*, 108, 136–142.

ROBERTS, M. J., D. Y. XU, X. FAN, AND S. ZHANG (2012): "The Role of Firm Factors in Demand, Cost, and Export Market Selection for Chinese Footwear Producers," .

RODRIGUE, J., D. SHENG, AND Y. TAN (2022a): "The Curious Case of the Missing Chinese Emissions," *Journal of the Association of Environmental and Resource Economists*, 9, 755–805.

——— (2022b): "Exporting, Abatement, and Firm-Level Emissions: Evidence from China's Accession to the WTO," *The Review of Economics and Statistics*, 1–45.

SHAPIRO, J. S. (2016): "Trade Costs, CO_2, and the Environment," *American Economic Journal: Economic Policy*, 8, 220–254.

——— (2021): "The Environmental Bias of Trade Policy," *The Quarterly Journal of Economics*, 136, 831–886.

SHAPIRO, J. S. AND R. WALKER (2018): "Why Is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade," *American Economic Review*, 108, 3814–3854.

YU, M. (2015): "Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms," *The Economic Journal*, 125, 943–988.

ZHANG, Q., X. JIANG, D. TONG, S. J. DAVIS, H. ZHAO, G. GENG, T. FENG, B. ZHENG, Z. LU, D. G. STREETS, R. NI, M. BRAUER, A. VAN DONKELAAR, R. V. MARTIN, H. HUO, Z. LIU, D. PAN, H. KAN, Y. YAN, J. LIN, K. HE, AND D. GUAN (2017): "Transboundary Health Impacts of Transported Global Air Pollution and International Trade," *Nature*, 543, 705–709.

Online Appendix: Product Level Emission Intensities: Measurement and Application

by

Ohyun Kwon

Hao Zhao

Min Qiang Zhao

A Monte Carlo Simulation for PLEI

To assess the performance of our iterative algorithm, we employ a Monte Carlo (MC) simulation approach. For simplicity yet without compromising generality, we choose a single emission category, COD, as our reference pollutant for simulation. The data generating process (DGP) is carefully designed to reflect the key moments of our firm-product level data.

Our simulation serves two core purposes. First, we investigate our algorithm's performance when the DGP deviates from the assumption of a uniform emission inefficiency parameter λ_i across all products within a given firm. This evaluation is achieved by comparing the outcome under two distinct DGPs. The former, wherein the inefficiency parameter λ_i is solely variant across firms, aligns perfectly with our algorithm's assumption. The latter DGP permits λ_{hi} to vary across both firms and their respective products. This design offers insights into the robustness of our algorithm in the face of product-level emission inefficiency variation within firms. Second, we rely on MC simulations to demonstrate that ordinary least squares (OLS) do not perform as well as our iteration method in terms of estimating PLEI.

A.1 Data Generating Process

As mentioned in Section 2.1, our firm-product level data combines three sources: the China's Environmental Statistical Database, the Annual Survey of Industry, and Chinese customs data. This final dataset encompasses 569,624 firms and spans 5,084 unique products. To streamline simulations, we constrain the sample of each simulated dataset to 10,000 firms and cap the product count at 90, ensuring a comparable firm-product ratio (calculated as $90 \approx 10\,000 \times 5084/569\,624$).

The benchmark estimates from Section 2.2 serve as the true ζ_h values in our Monte Carlo simulation. We note that (i) a substantial proportion (50.8%) of our ζ_h estimates are zero, (ii) the distribution of ζ_h is densely packed around positive yet tiny values, and (iii) a handful of sizable ζ_h values skew the distribution rightward. Due to the first two observations, in our Monte Carlo simulations, we assume that 70% of products possess zero emission intensities since, in addition to 50.8% products with zero emissions, around 20% of products' ζ_h is can be regarded as zero up to a tolerance ($\zeta_h < 1 \times 10^{-7}$).

To start, we categorize products with ζ_h values over the 70th percentile value into 27 evenly distributed product groups based on their emission intensity. This product group count is derived by multiplying the number of products in our simulation (90) by the proportion of products with positive emission intensity (30%). For products not exceeding the 70th percentile threshold, we shuffle them randomly and segment them into 63 uniform groups. This correspondence between the simulated 90 product groups and the original 5,084 products remains invariant across each simulation dataset. As detailed in Appendix A.2, this approach aptly preserves the distribution of ζ_h.

In the Monte Carlo section, we label the 27 product groups with positive ζ_h as "dirty," while the rest are termed "clean". This definition of being "dirty" is slightly different from the top 10% emission-intensive products as the main text. For each of the "dirty" product groups, the same median value from the ζ_h estimates is adopted across all simulations. Clean products' emission intensities are zero.

Next, we detail the methodology for selecting a sample of 10,000 firms from the pool of 569,624 firms. The aim is to preserve the quantity distribution of products produced each firm. From our raw data, 94.47% of firms produce 10 or fewer products. After aggregating the original products into 90 distinct groups, we observe that 94.74% of firms still produce 10 or fewer such product groups. By construction, the latter percentage is higher than the former, but the striking similarity in percentages can be attributed to the highly skewed distribution of products per firm. Namely, firms that produce hundreds of HS 6-digit products are also involved with more than 10 product groups, hence aligning the two percentages. For simplicity, if firms are engaged in the production of more than 10 product groups, we prioritize their top 10 based on the largest output shares. In the ensuing discussions of our Monte Carlo analysis, we interchangeably use "product" for "product group" and, as we detail later, designate the ζ_h values of these groups as the "true" ζ_h values.

For firm sampling, while preserving the product number distribution, we randomly choose firms with replacement from our raw dataset. Upon selecting a firm, a simulated counterpart is created, replicating its product numbers and corresponding output values (y_{hi}). Thus, firms in our Monte Carlo simulation, although fewer than that in our raw data, have the same count and quantity distribution of products.

For our initial Monte Carlo simulations, referred to as DGP-1, the parameter λ_i is designed to exhibit variation across firms yet remain invariant across all products within a firm. This ensures that the emissions at the product level (h) for a specific firm i can be expressed as $z_{hi} = y_{hi}\zeta_h\lambda_i$. For dirty products where $\zeta_h > 0$, the parameter λ_i must implicitly conform to the following equation[19]

$$\sum_i y_{hi}\lambda_i = \sum_i y_{hi}, \quad \forall h. \tag{A.1}$$

Given the substantially larger number of firms relative to products, multiple potential

[19]To see this, express emission intensity is as $\left(\zeta_h = \frac{\sum_i y_{hi}\zeta_h\lambda_i}{\sum_i y_{hi}}\right)$, where the numerator and the denominator respectively indicate total emissions and output associated with product h. By realigning the denominator to the left and negating mutual ζ_h values, equation (A.1) is derived.

solutions for λ_i can satisfy equation (A.1). Consequently, we utilize the empirical distribution to set constraints on the λ_i values during simulations. In particular, we curtail both the top and bottom 20% of the empirical λ_i distribution to eliminate outliers and impose that the set of λ_i solution to equation (A.1) has the same minimum, maximum, and mean values.

Having determined, ζ_h, y_{hi} and λ_i, the firm-product emissions are calculated as $z_{hi} = y_{hi}\zeta_h\lambda_i$. We then posit that emission data are observed with firm-specific noise, represented as ϵ_i. Thus, the observed emission at the firm level is articulated as $z_i = (1 + \epsilon_i)\sum_h z_{hi}$. Taking a conservative stance, we propose that ϵ_i is sourced from a uniform distribution spanning -30% to 30%.

In the second data generating process, referred to as DGP-2, we let λ_{hi} to be more flexible to exhibit variability not just across firms, but also among the products within each firm. The counterpart to equation (A.1) can be formulated as:

$$\sum_i y_{hi}\lambda_{hi} = \sum_i y_{hi}, \quad \forall h. \tag{A.2}$$

For each simulation under DGP-2, we determine a set of λ_{hi} satisfying equation (A.2). This process retains the minimum, maximum, and mean values associated with λ_i, highlighting the nuances of within-firm and between-product emission intensity variations. It is noteworthy that a uniform λ_{hi} across all products within a firm simplifies to the specific case of λ_i elucidated in equation (A.1).

Given this, the emission value of product h from firm i can be expressed as $z_{hi} = y_{hi}\zeta_h\lambda_{hi}$. Incorporating the measurement error, we introduce a random perturbation, ranging uniformly between -30% and 30%, to the firm's emission data z_i à la DGP-1.

A total of 500 simulated datasets were generated for both DGPs. The iterative algorithm initializes the value of ζ_h at 1. For robustness, we explored other starting values for ζ_h, randomly chosen between 0.1 and 5. All of the results remain highly robust.

Table A.1: Firm-Product Level Output Values

Percentiles	Data	Simulated Data	
		Mean	SD
1%	8.637E-07	8.730E-07	1.390E-07
5%	1.819E-05	1.840E-05	2.130E-06
10%	9.123E-05	9.170E-05	8.700E-06
25%	1.368E-03	1.378E-03	1.212E-04
50%	0.080	0.080	0.005
75%	0.586	0.587	0.016
90%	2.223	2.223	0.059
95%	4.804	4.812	0.155
99%	22.430	22.592	1.505

Notes: In the table, the second column displays the distribution of output at the firm-product level in our dataset. The last two columns present the means and standard deviations of the output values for the corresponding variables across 500 independent simulated datasets.

A.2 Comparison Between Actual and Simulated Datasets

For empirical relevance, we generate our simulated data that reflect the empirical distributions of firm-product level output and product-specific emission intensities. The simulated data's stochastic nature arises from firm sampling, solutions for λ_i, and measurement errors. In this section, we compare crucial moments from the actual and simulated datasets to affirm their similarity.

Table A.1 presents nine distinct percentiles of firm-product level output values from both actual and simulated datasets. Given that DGP-1 and DGP-2 employ the same method for output distribution creation, only results under DGP-1 are show. In Table A.1, close alignment of the percentile values attest to our method to obtain simulated firm-product level output data. Specifically, the actual data value and the average of the simulated values remain within the distance of one standard deviation for every percentile level.

To provide a comprehensive comparison, we compare the actual data with our simulated datasets, focusing particularly on firm-product level emissions. Given the absence of direct observations for emissions at this granularity for multi-product firms in the raw data, we infer them using the empirical estimates of ζ_h and λ_i. The same methodology

Table A.2: Firm-Product Level Emissions

Percentiles	Actual Data	Simulated Data: DGP-1		Simulated Data: DGP-2	
		Mean	SD	Mean	SD
1%	0	0	0	0	0
5%	0	0	0	0	0
10%	0	0	0	0	0
25%	0	8.680E-23	8.600E-22	2.360E-22	3.160E-21
50%	3.920E-07	3.500E-08	2.830E-08	3.500E-08	2.820E-08
75%	3.440E-05	4.610E-05	2.020E-06	4.600E-05	2.020E-06
90%	2.700E-04	2.871E-04	1.080E-05	2.859E-04	1.080E-05
95%	7.472E-04	7.591E-04	3.260E-05	7.561E-04	3.250E-05
99%	5.451E-03	4.677E-03	3.357E-04	4.670E-03	3.348E-04

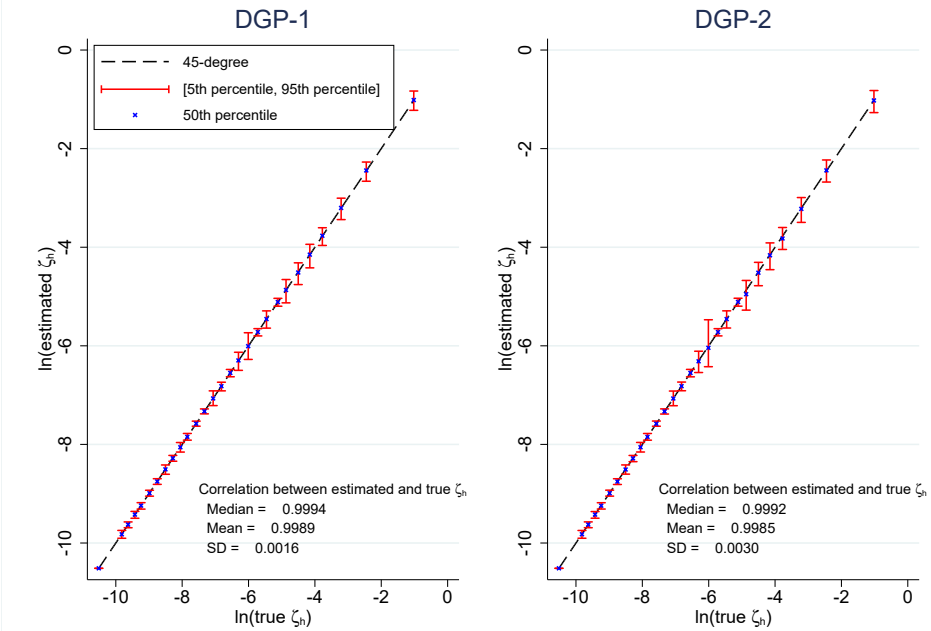
Notes: The second column of the table presents the distribution of COD emissions at the firm-product level in our dataset. The subsequent four columns display the means and standard deviations of the corresponding emission values across 500 independent simulated datasets, using two different data generating processes. DGP-1 refers to the first data generating process where λ_i varies across firms, while DGP-2 refers to the second data generating process where λ_{hi} varies across both firms and products.

is applied to generate z_{hi} values in the simulated datasets. As presented in Table A.2, there is a notable alignment between emissions derived from simulated datasets and the actual emissions across both DGPs. This alignment underscores the efficacy of our DGPs in generating datasets that represent the authentic emission profiles.

A.3 DGP-1: λ_i Varies Across Firms

Utilizing our iterative algorithm, we compute ζ_h for each of the 500 simulated datasets. In the left panel of Figure A.1, we display a scatter plot comparing estimated and true ζ_h values, both in logarithmic scale. To enhance clarity, we adjust both values of clean products (those with true ζ_h values of zero) by adding a small constant - half the smallest positive value of the true ζ_h - before taking their logarithm. Across 500 independent Monte Carlo simulations, we depict the median as well as the 5th and 95th percentile values of estimated ζ_h for each product. With the estimated values predominantly aligning with the 45-degree line, this suggests our algorithm consistently and accurately retrieves the true ζ_h values. Notably, both the mean and median correlation coefficients (prior to

Figure A.1: Monte Carlo Results



Notes: The depicted figures represent the ζ_h values computed via our iterative algorithm for both DGP-1 and DGP-2. In each plot, the logarithm of median values are illustrated for each product, along with the span between the 5th and 95th percentile logarithmic values, derived from 500 distinct Monte Carlo simulations. For clean products, we augment both the estimated and true ζ_h values by adding half the smallest positive value from the true ζ_h of dirty products prior to logarithm conversion. Each figure's bottom-right corner provides the mean, median, and standard deviation of the correlation coefficients between the estimated and true ζ_h values.

logarithm conversion), presented in the left panel of Figure A.1, exceed 0.998.

A.4 DGP-2: λ_{hi} Varies Across Firms and Products

In DGP-2, we introduce a nuanced layer of complexity by permitting the true λ_{hi} to differ across products within a given firm. This added variation, which is not explicitly addressed by our iteration algorithm, introduces further intricacies to our PLEI estimates beyond the inherent measurement error. Under this DGP, we still perform our iteration under the maintained assumption that the inefficiency parameter only varies at the firm level. The purpose is to assess the ramification of our assumption when it does not fully align with the true data generating process.

The outcomes of DGP-2 are depicted in the right panel of Figure A.1. In contrast with DGP-1, there is a marginal decline in the correlation coefficients between the estimated and true ζ_h values. However, their mean and median remain prominently above 0.998. An observable, albeit slight, widening in the 5th to 95th percentile range is discernible for certain dirty products, which suggests that estimation is less efficient. Yet, the alignment of estimated values with the 45-degree line underscores the robustness of our PLEI calculations to potential intra-firm product variations in λ_{hi}.

A.5 Regression Approach

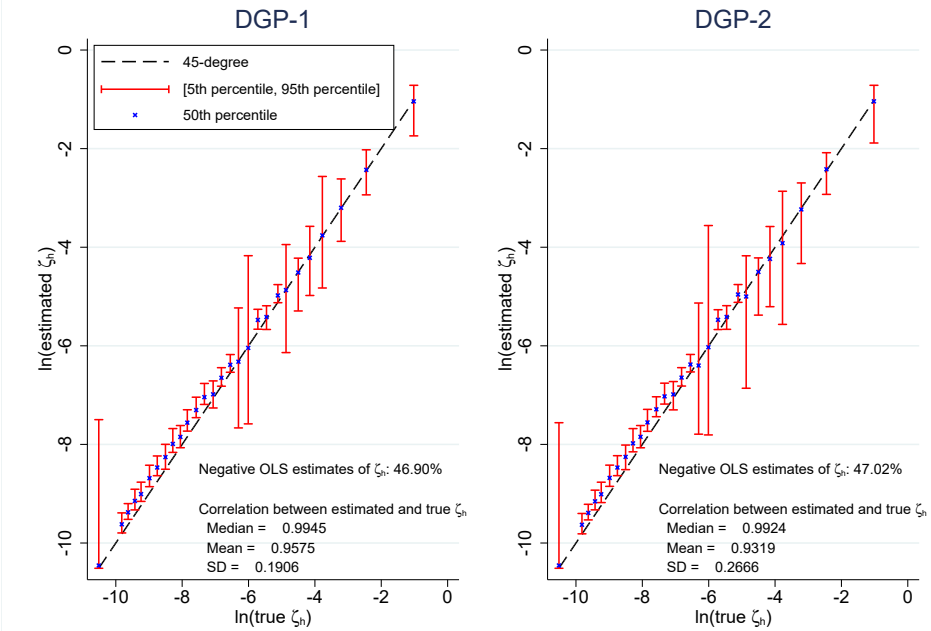
A potential strategy to estimate ζ_h involves employing ordinary least squares. Consider the subsequent model specification:

$$z_i = \sum_h y_{hi}\zeta_h + e_i. \tag{A.3}$$

In this equation, e_i represents the error term, and ζ_h can be interpreted as the PLEI. While the computational efficiency of OLS might be an advantage given its speed relative to our iterative approach, it is inherent with certain limitations. Primarily, equation (A.3) relies on e_i to encompass the firm efficiency parameter in an additive fashion. This suggests that a firm's emission level would equate to e_i even with zero production – inconsistent with most assumptions in the literature. Furthermore, OLS can yield negative ζ_h estimates, suggesting that production reduces emissions paradoxically.

We assess the limitations of employing OLS for estimating ζ_h by using it on the simulated datasets from DGP-1 and DGP-2. The outcomes are illustrated in Figure A.2, where the left and right panels depict the findings from DGP-1 and DGP-2, respectively. It is evident from the left panel of Figure A.2 that both mean and median correlation coefficients between OLS-estimates and actual ζ_h values are substantially lower than those from our iteration method. The most telling metric of OLS's inefficacy is the surge in the standard deviation of the correlation coefficients: a jump from 0.0016 (iteration method) to 0.1906 (OLS). This highlights our iteration method's superior accuracy in calculating PLEI close

Figure A.2: Monte Carlo Results (Ordinary Least Squares)



Notes: The depicted figures represent the ζ_h values estimated using OLS for both DGP-1 and DGP-2. In each plot, the logarithm of median values are illustrated for each product, along with the span between the 5th and 95th percentile logarithmic values, derived from 500 distinct Monte Carlo simulations. For clean products, we augment both the estimated and true ζ_h values by adding half the smallest positive value from the true ζ_h of dirty products prior to logarithm conversion. Negative estimates of ζ_h from the OLS are excluded in the visualizations. Each figure's bottom-right corner provides the mean, median, and standard deviation of the correlation coefficients between the estimated and true ζ_h values.

to the true ζ_h values.

Moreover, a significant fraction of OLS estimates (46.90%) are negative. As a result, we can not plot the estimates on the log-log scale employed in Figure A.1. For clarity, only non-negative OLS estimates are plotted in Figure A.2. Compared with the left panel of Figure A.1, we observe certain median values of ζ_h estimates deviate noticeably from the 45-degree line. Additionally, the variation - quantified by the span between the 5th and 95th percentiles - in the OLS method is considerably greater than the one registered by the iterative approach. This underscores the iterative approach's enhanced precision in median ζ_h estimates, particularly within the non-negative domain. A similar examination under DGP-2, showcased in the right panel of Figure A.2, echoes these observations.

Table A.3: Mean Error Comparison between Iteration and OLS

Clean Products		
	MAE^{ITER}	MAE^{OLS}
DGP-1	4.045E-06	2.657E-04
DGP-2	4.077E-06	2.651E-04
Dirty Products		
	$MAPE^{ITER}$	$MAPE^{OLS}$
DGP-1	0.0608	0.3748
DGP-2	0.0744	0.4361

Notes: For clean and dirty products, we report the mean absolute errors (MAE) and the mean absolute percentage errors ($MAPE$) between the estimated and true ζ_h, respectively. $ITER$ and OLS indicate our iteration algorithm and OLS, respectively.

To contrast the efficacy of our iteration algorithm with OLS, we compute the mean absolute errors (MAE) for *clean* products and mean absolute percentage errors ($MAPE$) for *dirty* products. The MAE^j is given by:

$$MAE^j = \frac{1}{N \cdot H_c} \sum_n \sum_h \left| \zeta_h^{true} - \zeta_h^{j,n} \right|,$$

where n indexes a simulated dataset, N denotes the total number of simulations, H_c represents the number of clean products, j denotes the estimation method ($ITER$ or OLS), and $\zeta_h^{true} = 0$ for clean products. For *dirty* products, $MAPE^j$ is described as:

$$MAPE^j = \frac{1}{N \cdot H_d} \sum_n \sum_h \left| \frac{\zeta_h^{true} - \zeta_h^{j,n}}{\zeta_h^{true}} \right|,$$

where H_d denotes the number of dirty products. While $MAPE$ is more meaningful compared to MAE, the computation of $MAPE$ necessitates a non-zero ζ_h^{true} and thus it is not applicable to clean products.

Table A.3 presents the MAE and $MAPE$ outcomes. The iteration method yields an MAE for clean products nearly two orders of magnitude lower than its OLS counterpart. For dirty products, the iteration technique's $MAPE$ is approximately 16% of what the OLS delivers. These findings underline the superiority of our iteration approach in determining ζ_h.

B Verifying PLEI with External Datasets

In addition to the CESD, various datasets report on China's industrial emissions and their intensities. Primary sources include China Statistical Yearbook and China Statistical Yearbook on Environment from the National Bureau of Statistics (SYB). These yearbooks disclose annual emissions for various pollutants, which include NOX, SO_2, soot and dust, ammoniacal nitrogen, and COD, classified under the CIC 2-digit level. Additionally, the Eora Global Supply Chain Database (EORA) and EXIOBASE (EXIO), both of which are multi-region environmentally extended input-output (IO) tables, also include sectoral level emissions and output data, although each dataset has its own distinct sectoral classifications.

Comparing our PLEI calculations to emission intensities from these alternative sources is not straightforward. One major complication is the inconsistent emission categories reported in different datasets. For our pollutant types, we associate them with the same or nearest emissions accounts, ensuring a reasonable comparison.[20] A further challenge emerges from the distinct industrial identifiers each dataset deploys. To ensure consistency during cross-comparison, we aggregate emission intensity measures across all sources to the CIC 2-digit level, the broadest classification shared by the datasets.

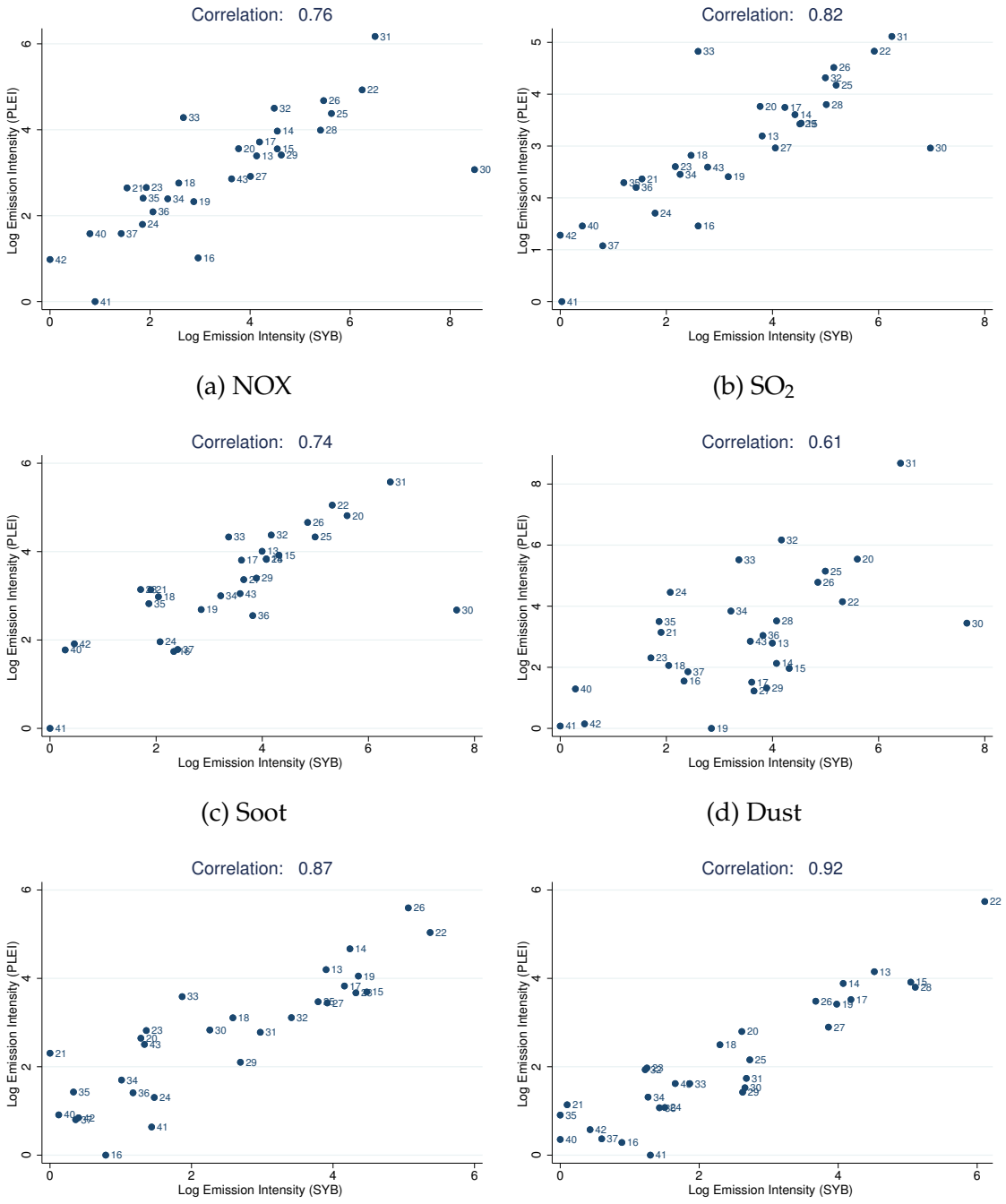
Figures B.1 through B.3 offer a bilateral comparison of emission intensity estimates across datasets. Broadly speaking, our PLEI-derived industry-level emission intensities exhibit strong correlations with those derived from other sources. All observed correlations are positive, with the majority surpassing 0.5. The strongest correlation emerges between PLEI and SYB estimates. For comparisons involving EORA and EXIO, discrepancies are more pronounced, potentially attributable to discrepancies in pollutant classifi-

[20]For example, CO_2 emissions used in our PLEI computation stem from reported corporate fossil fuel utilization. In contrast, the closest measure EORA provides, termed "Fuel Combustion Activities (PRIMAP—CO2—IPC1A—Gg)", spans only a segment of sectors. This leads us to adopt data tagged as "National Total excluding LULUCF (PRIMAP—CO2—TOTALexcludingLULUCF—Gg)", not ideally aligned with our PLEI metric.

cations, industry identifiers, or the intrinsic quality of the EORA/EXIO datasets.[21] Nevertheless, the robust correlation observed for specific emission categories (like Ammoniacal Nitrogen, Dust, and SO_2 in the PLEI vs. EORA comparison, and Dust in the PLEI vs. EXIO analysis) reinforces the efficacy of our PLEI estimates in gauging the environmental impact of industrial production.

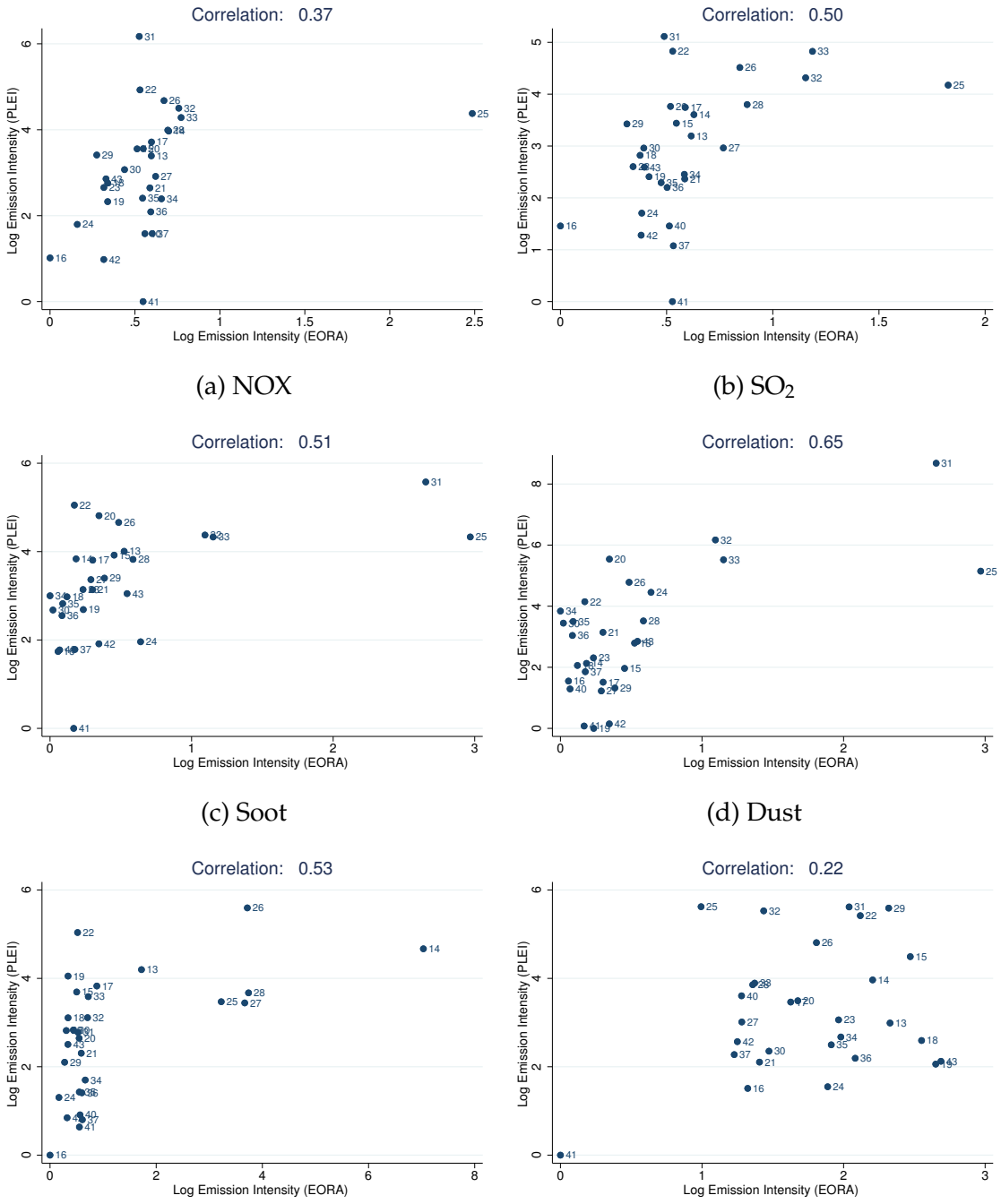
[21] The creation of the multi-region input-output table necessitates reliance on sectorial or regional growth rates to infer certain emission account data, such as those related to land-use changes and deforestation.

Figure B.1: The Comparison of Emission Intensity Estimates: PLEI vs. SYB



Notes: These figures illustrate the bilateral comparison of emission intensity estimates between PLEI and SYB, standardized to the CIC 2-digit categorization. Each subfigure's title shows the specific pollutant under analysis. Each dot represents the paired estimates of emission intensities for a specific CIC 2-digit industry, with the industry code displayed next to the dot. The x-axis showcases the logarithmic values of emission intensities sourced from SYB data, while the y-axis depicts the logarithmic values derived from PLEI assessments. The correlation coefficient between these two estimation sets is displayed at the top of each subfigure.

Figure B.2: The Comparison of Emission Intensity Estimates: PLEI vs. EORA



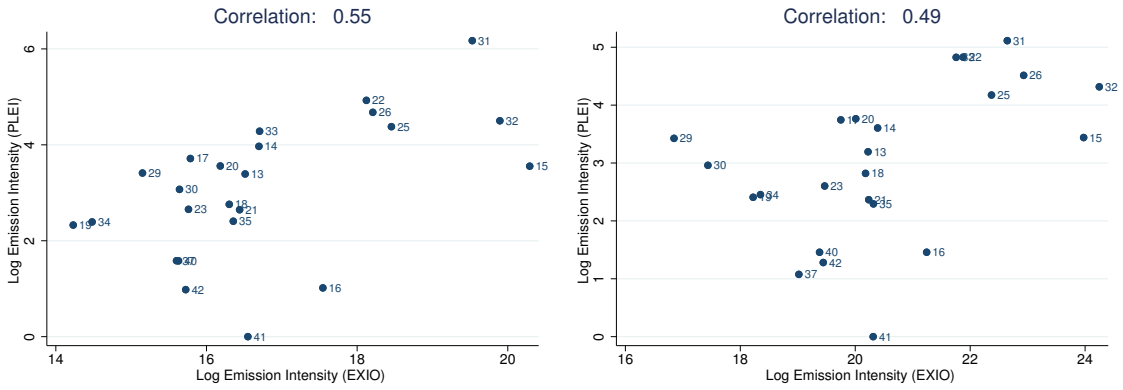
(a) NOX (b) SO_2

(c) Soot (d) Dust

(e) Amm. Nit. (f) CO_2

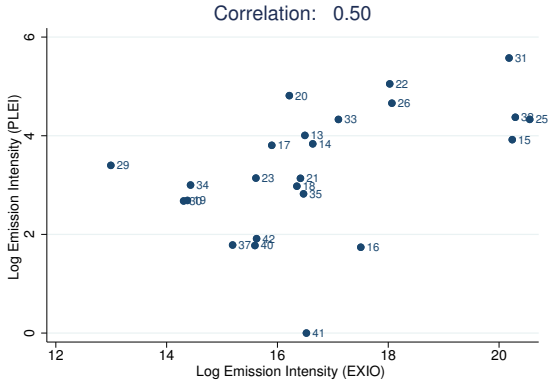
Notes: These figures illustrate the bilateral comparison of emission intensity estimates between PLEI and EORA, standardized to the CIC 2-digit categorization. Each subfigure's title shows the specific pollutant under analysis. Each dot represents the paired estimates of emission intensities for a specific CIC 2-digit industry, with the industry code displayed next to the dot. The x-axis showcases the logarithmic values of emission intensities sourced from EORA data, while the y-axis depicts the logarithmic values derived from PLEI assessments. The correlation coefficient between these two estimation sets is displayed at the top of each subfigure.

Figure B.3: The Comparison of Emission Intensity Estimates: PLEI vs. EXIO

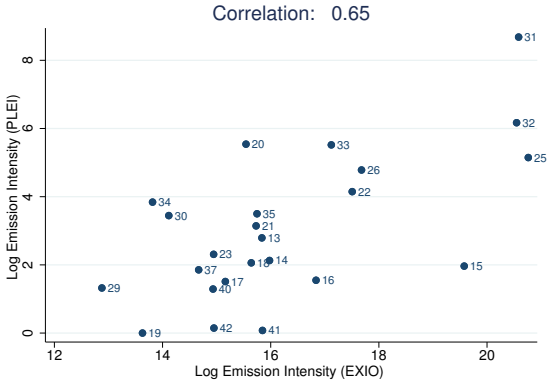


(a) NOX

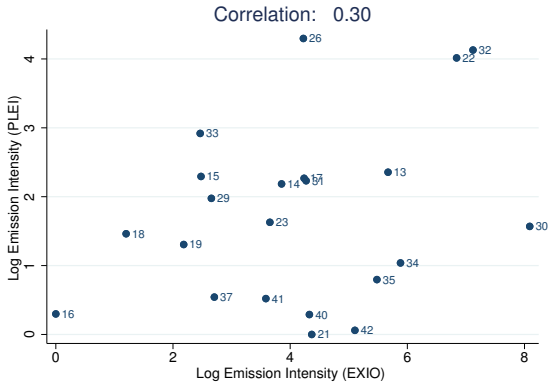
(b) SO_2



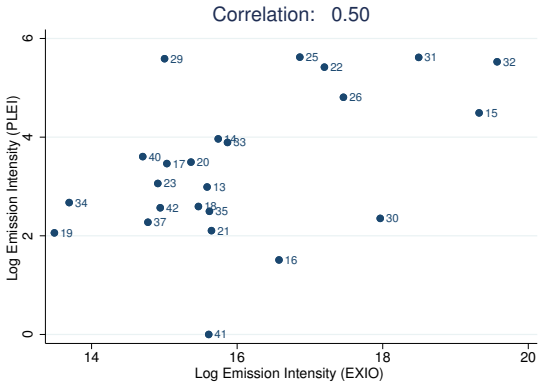
(c) Soot



(d) Dust



(e) Water



(f) CO_2

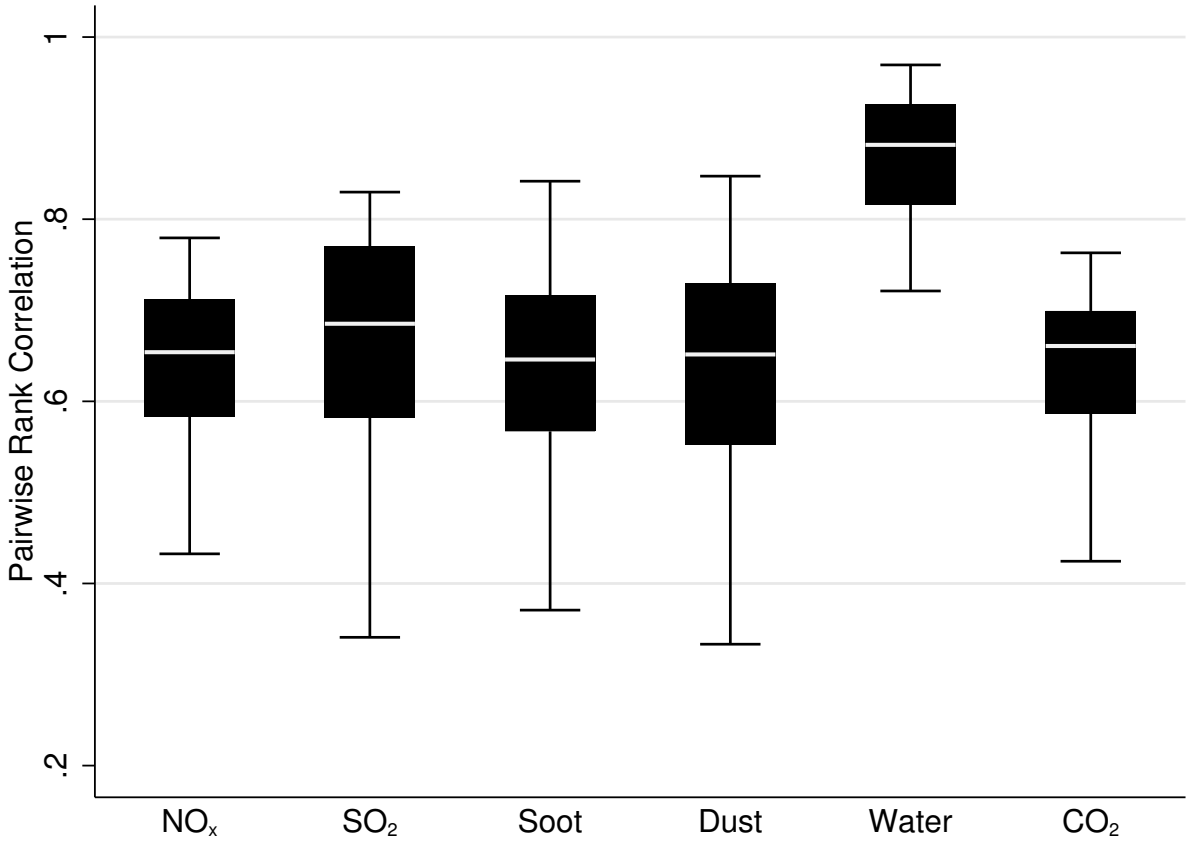
Notes: These figures illustrate the bilateral comparison of emission intensity estimates between PLEI and EXIO, standardized to the CIC 2-digit categorization. Each subfigure's title shows the specific pollutant under analysis. Each dot represents the paired estimates of emission intensities for a specific CIC 2-digit industry, with the industry code displayed next to the dot. The x-axis showcases the logarithmic values of emission intensities sourced from EXIO data, while the y-axis depicts the logarithmic values derived from PLEI assessments. The correlation coefficient between these two estimation sets is displayed at the top of each subfigure.

C Cross-Country Comparison of Industrial-Level Emission Intensities

In this subsection, we employ EXIO to investigate the correlation of industrial-level emission intensities between China and its trade partners. Our empirical analysis utilizes the PLEI measure, derived from Chinese manufacturing data, to identify products with high pollution levels in both export and domestic sectors. This methodology inherently presumes a high degree of consistency in the ranking of emission intensities across different countries. Leveraging the multi-regional IO table of EXIO, which provides output and emissions data for 162 industries across 48 countries, we calculate the pairwise rank correlation of emission intensities at the industrial level between China and each of the other 47 countries, focusing on the pollutants measured by the PLEI.

Figure C.1 presents the box plots illustrating the pairwise rank correlations for the six pollutants. The box plots reveal that the majority of these correlations surpass 0.6, indicating a strong level of consistency in the ranking of industrial emission intensities between China and other countries. This finding underscores the reliability of our empirical examination using PLEI.

Figure C.1: Cross-National Comparison of Industrial-Level Emission Intensities



Notes: This figure displays box plots of the pairwise rank correlation of emission intensities between China and other countries for six pollutants calculated using EXIO. The horizontal axis indicates the type of pollutant. The vertical axis shows the value of rank correlation.

D Time Consistency of ζ_h

We operate under the presumption that PLEI reflect the time-invariant production attributes of each product. This subsection rigorously probes this hypothesis. A naïve method involves calculating emission intensities, denoted as $\{\zeta_{ht}\}$, separately for each year. Yet, this strategy encounters two problems. First, a large amount of firm-level observation for individual products is needed for a precise PLEI calculation. Utilizing a single year's data exacerbates the likelihood of encountering products manufactured predominantly by few firms, thereby making it difficult to disentangle the differentiation between product-specific and firm-specific attributes. The subsequent, and arguably more pressing issue, arises when $\{\lambda_{it}\}$ calculated from this strategy loses inter-temporal comparability. Specifically, within our algorithm, if a firm witnesses a 10% increase in λ_{it} in a given year, it signals a commensurate increase in emission intensities for its entire product portfolio, both when compared across firms and compared with other years. Such interpretation becomes untenable when $\{\zeta_{ht}\}$ is calculated distinctively for each year. When λ_{it} cannot fully absorb time-varying firm-specific inefficiency factor, $\{\zeta_{ht}\}$ then should not be compared acoss years directly, since in this case a high ζ_{ht} may simply reflect a high residual of firm inefficiency conditional on $\{\lambda_{it}\}$.

In light of these challenges, we propose an alternative methodology, leveraging the iterative structure of our algorithm. Starting from our null hypothesis, equation (2) can be reformulated as:

$$z_{hit} = z_{it}\frac{s_{hit}\zeta_h}{\sum_{h'} s_{h'it}\zeta_{h'}}, \tag{D.1}$$

with subscript i representing a firm, rather than a firm-year observation. Equation (D.1) allocates firm-specific emissions across products, and remains valid under our assumption that $\{\zeta_{ht}\} = \{\zeta_h\}$. Subsequently, we can derive the cumulative emissions and output pertinent to the product (Z_{ht} and Y_{ht}) to infer the emission intensities for year t:

Table D.1: Time Consistency of ζ_h

Emssion Type	β_1	S.E.	R^2	Rank Corr.
COD	0.94	0.00	0.96	0.94
SO_2	0.97	0.00	0.95	0.92
NOX	0.42	0.00	0.80	0.93
CO_2	0.95	0.00	0.91	0.47
Ind. Exh.	0.93	0.00	0.90	0.92
Water	1.01	0.00	0.90	0.95
Amm. Nit.	0.78	0.00	0.99	0.92
Soot	1.07	0.00	0.96	0.91
Dust	1.14	0.00	0.98	0.84

Notes: The table elucidates a pronounced correlation between the mean product-level emission intensities, denoted as ζ_h, and their annual fluctuations, represented by ζ'_{ht}. Displayed columns encompass the β_1 coefficients, accompanied by their respective standard errors and the adjusted R-squared values, all derived from estimating equation (D.3). The inclusion of year fixed effects does not alter these results. The last column displays the rank correlation between ζ_h and ζ'_{ht}.

$$\zeta'_{ht} = \frac{Z_{ht}}{Y_{ht}}. \tag{D.2}$$

It is essential to highlight that our baseline iterative algorithm terminates once the predefined $\{\zeta_h\}$ (as per equation (D.1)) and the derived $\{\zeta'_{ht}\}$ (from equation (D.2)) converge within an acceptable tolerance. Using this logic, the proximity between ζ'_{ht} and ζ_h can serve as a diagnostic tool, assessing the initial appropriateness of assuming ζ_h in equation (D.1).

Loosely speaking, let $\Phi_t(\cdot)$ denote the operations (D.1) and (D.2). The convergence of our algorithm at the annual level can be succinctly described as a fixed point criterion $\zeta_{ht} = \Phi_t(\zeta_{ht})$. Using our long-run average PLEI as an argument, we can obtain $\zeta'_{ht} = \Phi(\zeta_h)$. By observing how similar ζ'_{ht} is with ζ_h, we can get a sense of how close ζ_h is to the fixed point of $\Phi_t(\cdot)$. In an extreme case, if ζ_h is the fixed point of $\Phi_t(\cdot)$, we expect $\zeta'_{ht} = \zeta_h$; instead, if ζ_h is far from the fixed point (i.e., rather different from ζ_{ht}), we expect ζ'_{ht} to be quite different from ζ_h.

Upon computing ζ'_{ht}, we estimate the following equation to compare two emission

intensities:

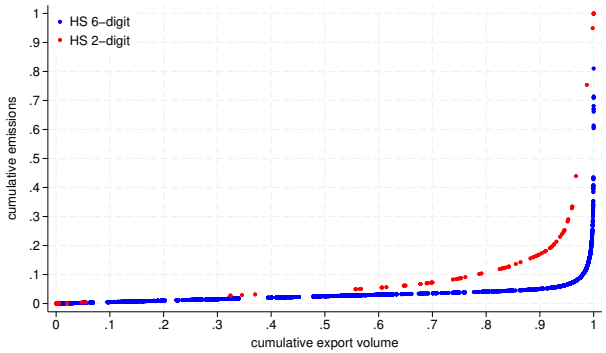
$$\ln(\zeta'_{ht}) = \beta_0 + \beta_1 \ln(\zeta_h) + \epsilon_{ht}. \tag{D.3}$$

The derived values for β_1 are exhibited in Table D.1. Remarkably, the regression coefficients uniformly equal 1.00 with remarkable precision across all emission categories. The R-squared, likewise, stands at 1.00, implying that ζ_h accounts for almost all variations in ζ'_{ht}. In the last column, the rank correlation is presented. Despite its overall high magnitude, discernible deviations are observed, particularly for CO_2. Such discrepancies are attributed to the dense distribution of emission intensities, with values slightly exceeding, or precisely at, 0. Minimal numerical variations can precipitate substantial rank shifts. Collectively, Table D.1 corroborates the notion that emission intensities can be parsimoniously parameterized with a long-term average value.

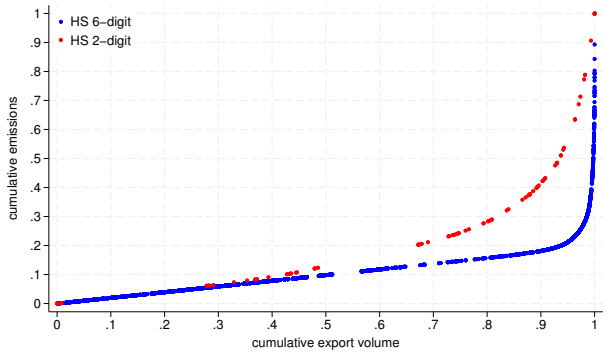
E Additional Tables and Figures

Figure E.1: Cumulative Total Emissions and Trade Volume

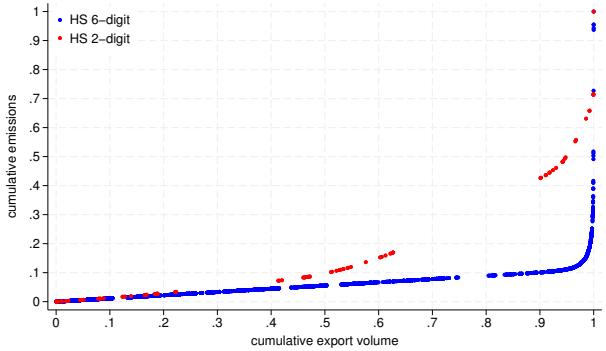
(a) COD



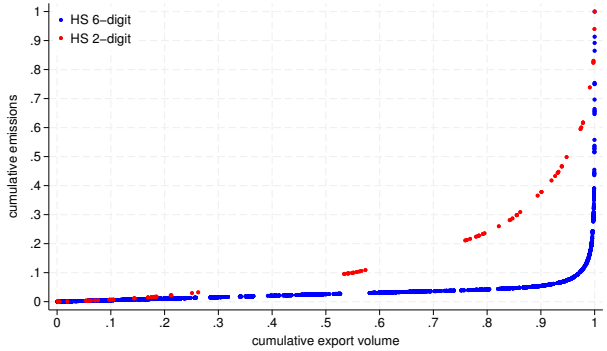
(b) SO_2



(c) NOX

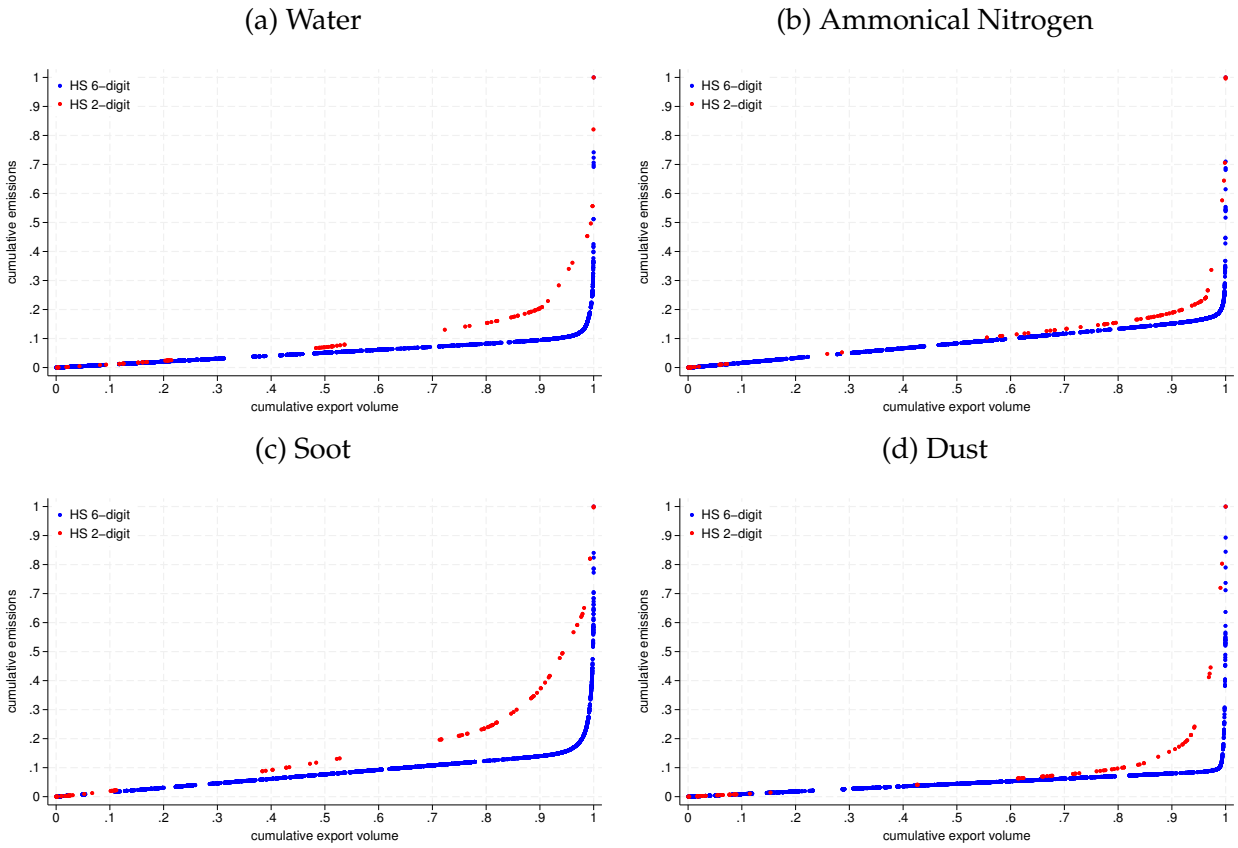


(d) Industrial Exhaust



Notes: This figure illustrates the cumulative distributions of total emissions and export volumes. Products are sorted in increasing order based on their total emission intensities.

Figure E.2



Notes: This figure illustrates the cumulative distributions of total emissions and export volumes. Products are sorted in increasing order based on their total emission intensities.

Table E.1: Top SO_2-intensive Products

HS Code	Description
810000	**Metals; n.e.c., cermets and articles thereof**
810990	Zirconium; other than unwrought, n.e.c. in heading no. 8109
811292	Gallium, germanium, hafnium, indium, niobium (columbium), rhenium and vanadium; articles ...
811299	Gallium, germanium, hafnium, indium, niobium (columbium), rhenium and vanadium; articles ...
280000	**Inorganic chemicals**
282732	Chlorides; of aluminium
283319	Sodium sulphates; other than disodium sulphate
282510	Hydrazine and hydroxylamine and their inorganic salts
710000	**Natural, cultured pearls; precious, semi-precious stones; precious metals, metals clad wi...**
710122	Pearls; cultured, worked, whether or not graded (but not strung, mounted or set), tempora...
710691	Metals; silver, unwrought, (but not powder)
711620	Stones; precious or semi-precious stones (natural, synthetic or reconstructed) articles o...
230000	**Food industries, residues and wastes thereof; prepared animal fodder**
230610	Oil-cake and other solid residues; whether or not ground or in the form of pellets, resul...
230210	Bran, sharps and other residues; of maize (corn), whether or not in the form of pellets, ...
230649	Oil-cake and other solid residues; whether or not ground or in the form of pellets, resul...
720000	**Iron and steel**
720690	Iron or non-alloy steel; primary forms (excluding ingots and iron of heading no. 7203)
721790	Iron or non-alloy steel; wire, n.e.c. in heading no. 7217
720925	Iron or non-alloy steel; (not in coils), flat-rolled, width 600mm or more, cold-rolled, o...
540000	**Man-made filaments and textile materials**
540234	Yarn, synthetic; filament, monofilament (less than 67 decitex), textured, of polypropylen...
540824	Fabrics, woven; of artificial filament yarn, containing 85% or more by weight of artific...
540831	Fabrics, woven; of artificial filament yarn, of artificial monofilament, strip and the li...

Table E.2: Top Industrial Exhaust-intensive Products

HS Code	Description
280000	**Inorganic chemicals**
282739	Chlorides; other than of ammonium, calcium, magnesium, aluminium and nickel
283911	Silicates; sodium metasilicates
283230	Thiosulphates
720000	**Iron and steel**
720825	Iron or non-alloy steel; in coils, without patterns in relief, flat-rolled, of a width 60...
720690	Iron or non-alloy steel; primary forms (excluding ingots and iron of heading no. 7203)
720719	Iron or non-alloy steel; semi-finished products of iron or non-alloy steel, containing by...
690000	**Ceramic products**
690310	Refractory ceramic goods; containing by weight more than 50% of graphite or other forms ...
690220	Refractory bricks, blocks, tiles and similar refractory ceramic constructional goods; con...
690100	Bricks, blocks, tiles and other ceramic goods of siliceous fossil meals (e.g. kieselguhr,...
250000	**Plastering materials, lime and cement**
252100	Limestone flux; limestone and other calcareous stone, of a kind used for the manufacture ...
251749	Stones; of heading no. 2515 or 2516 (excluding marble), in granules, chippings and powder...
253010	Vermiculite, perlite and chlorites; unexpanded
810000	**Metals; n.e.c., cermets and articles thereof**
810210	Molybdenum; articles thereof, including waste and scrap, powders
810390	Tantalum; articles n.e.c. in heading no. 8103
811221	Chromium and articles thereof; unwrought chromium, powders
760000	**Aluminium and articles thereof**
760521	Aluminium; alloys, wire, maximum cross-sectional dimension exceeding 7mm
760511	Aluminium; (not alloyed), wire, maximum cross-sectional dimension exceeds 7mm
760421	Aluminium; alloys, hollow profiles

Table E.3: Top Water-consuming Products

HS Code	Description
250000	**Plastering materials, lime and cement**
251749	Stones; of heading no. 2515 or 2516 (excluding marble), in granules, chippings and powder...
251741	Stones; of marble, in granules, chippings and powder, whether or not heat-treated
251320	Emery, natural corundum, natural garnet and other natural abrasives, whether or not heat-...
720000	**Iron and steel**
720825	Iron or non-alloy steel; in coils, without patterns in relief, flat-rolled, of a width 60...
721633	Iron or non-alloy steel; H sections, hot-rolled, hot-drawn or extruded, of a height of 80...
720837	Iron or non-alloy steel; in coils, without patterns in relief, flat-rolled, of a width 60...
280000	**Inorganic chemicals**
282732	Chlorides; of aluminium
281410	Ammonia; anhydrous
282710	Chlorides; of ammonium
440000	**Wood and articles of wood; wood charcoal**
440122	Wood; for fuel, in chips or particles, non-coniferous, whether or not agglomerated
441300	Wood; densified wood, in blocks, plates, strips or profile shapes
440839	Wood, tropical; as in Subheading note 2 to this Chapter, n.e.c. in item no. 4408.31, shee...
350000	**Albuminoidal substances; modified starches; glues; enzymes**
350790	Enzymes and prepared enzymes; other than rennet and concentrates thereof
350290	Albumins, albuminates and other albumin derivatives; other than egg or milk albumin, incl...
350510	Dextrins and other modified starches
290000	**Organic chemicals**
292111	Amine-function compounds; acyclic monoamines and their derivatives, methylamine, di- or t...
292243	Amino-acids, other than those containing more than one kind of oxygen function, and their...
290382	Halogenated derivatives of cyclanic, cyclenic or cycloterpenic hydrocarbons; aldrin (ISO)...

Table E.4: Top Ammoniacal Nitrogen-intensive Products

HS Code	Description
280000	**Inorganic chemicals**
283220	Sulphites; other than of sodium
282732	Chlorides; of aluminium
280530	Earth-metals, rare; scandium and yttrium, whether or not intermixed or interalloyed
810000	**Metals; n.e.c., cermets and articles thereof**
810210	Molybdenum; articles thereof, including waste and scrap, powders
810194	Tungsten (wolfram); unwrought, including bars and rods obtained simply by sintering
810294	Molybdenum; unwrought, including bars and rods obtained simply by sintering
290000	**Organic chemicals**
292411	Acyclic amides (including acyclic carbamates) and their derivatives; salts thereof; mepro...
293491	Other heterocyclic compounds, n.e.c. in 2934.1, 2934.2 and 2934.3
292424	Cyclic amides (including cyclic carbamates) and their derivatives; ethinamate and i... salts
250000	**Plastering materials, lime and cement**
251749	Stones; of heading no. 2515 or 2516 (excluding marble), in granules, chippings and powder...
250100	Salt (including table salt and denatured salt); pure sodium chloride whether or not in aq...
250850	Clays (excluding expanded clays of heading no. 6806); andalusite, kyanite and sillimanite...
50000	**Animal originated products; not elsewhere specified or included**
50210	Animal products; hair and bristles, of pigs, hogs or boars, and waste thereof
51191	Animal products; of fish or crustaceans, molluscs or other aquatic invertebrates; dead an...
50590	Animal products; skins and other parts of birds, feathers and down (not for stuffing), po...
230000	**Food industries, residues and wastes thereof; prepared animal fodder**
230250	Bran, sharps and other residues; of leguminous plants, whether or not in the form of pell...
230320	Beet-pulp, bagasse and other waste of sugar manufacture; whether or not in the form of pe...
230990	Dog or cat food; (not put up for retail sale), used in animal feeding

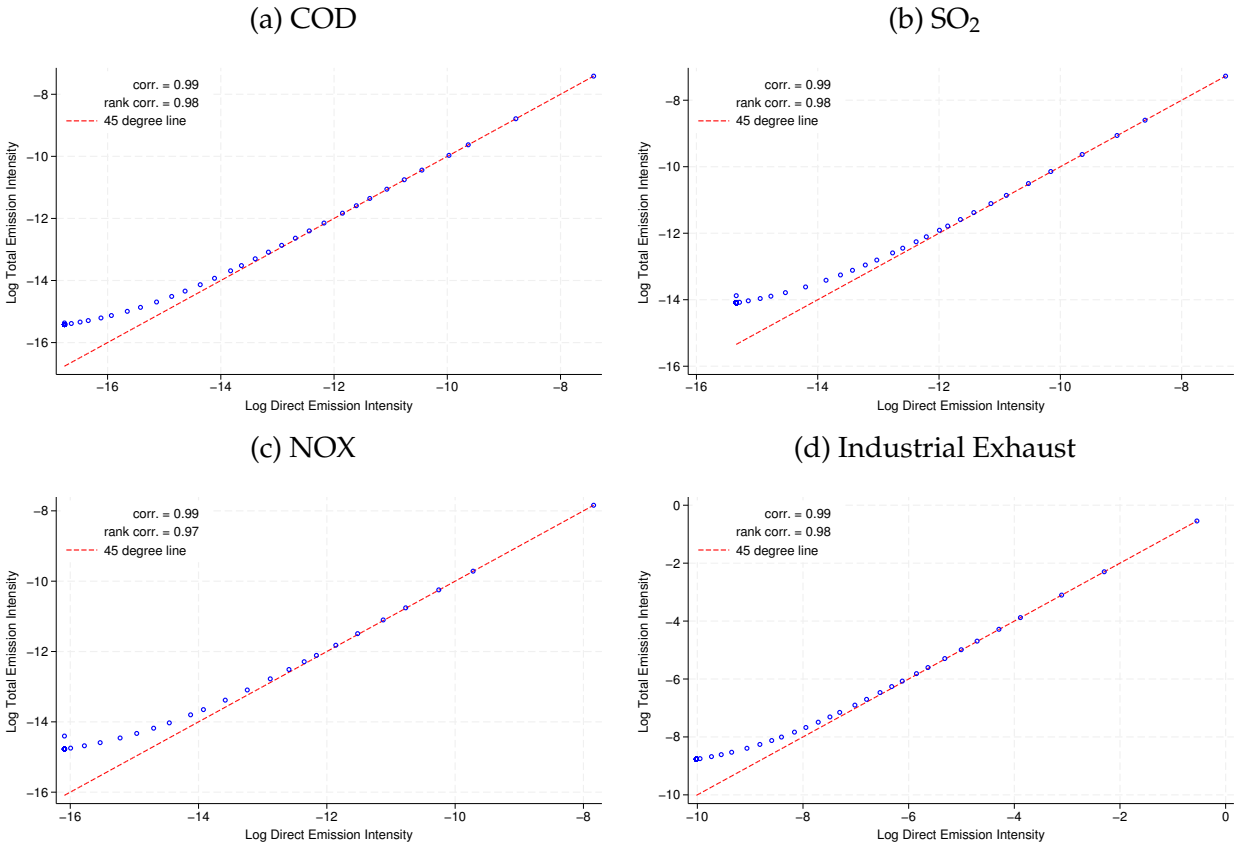
Table E.5: Top Soot-intensive Products

HS Code	Description
790000	**Zinc and articles thereof**
790111	Zinc; unwrought, (not alloyed), containing by weight 99.99% or more of zinc
790390	Zinc; powders and flakes
790400	Zinc; bars, rods, profiles and wire
310000	**Fertilizers**
310230	Fertilizers, mineral or chemical; nitrogenous, ammonium nitrate, whether or not in aqueou...
310540	Fertilizers, mineral or chemical; ammonium dihydrogenorthophosphate (monoammonium phospha...
310100	Fertilizers, animal or vegetable; whether or not mixed together or chemically treated; fe...
280000	**Inorganic chemicals**
282731	Chlorides; of magnesium
281830	Aluminium hydroxide
282510	Hydrazine and hydroxylamine and their inorganic salts
690000	**Ceramic products**
690290	Refractory bricks, blocks, tiles and similar refractory ceramic constructional goods; n.e...
690100	Bricks, blocks, tiles and other ceramic goods of siliceous fossil meals (e.g. kieselguhr,...
691110	Tableware and kitchenware; of porcelain or china
710000	**Natural, cultured pearls; precious, semi-precious stones; precious metals, metals clad wi...**
710691	Metals; silver, unwrought, (but not powder)
710122	Pearls; cultured, worked, whether or not graded (but not strung, mounted or set), tempora...
711620	Stones; precious or semi-precious stones (natural, synthetic or reconstructed) articles o...
480000	**Paper and paperboard**
480269	Uncoated paper and paperboard (not 4801 or 4803); over 10% by weight of mechanical or ch...
480421	Kraft paper and paperboard; sack kraft paper, uncoated, unbleached, in rolls or sheets, o...
481110	Paper and paperboard; tarred, bituminised or asphalted, in rolls or sheets, other than go...

Table E.6: Top Dust-intensive Products

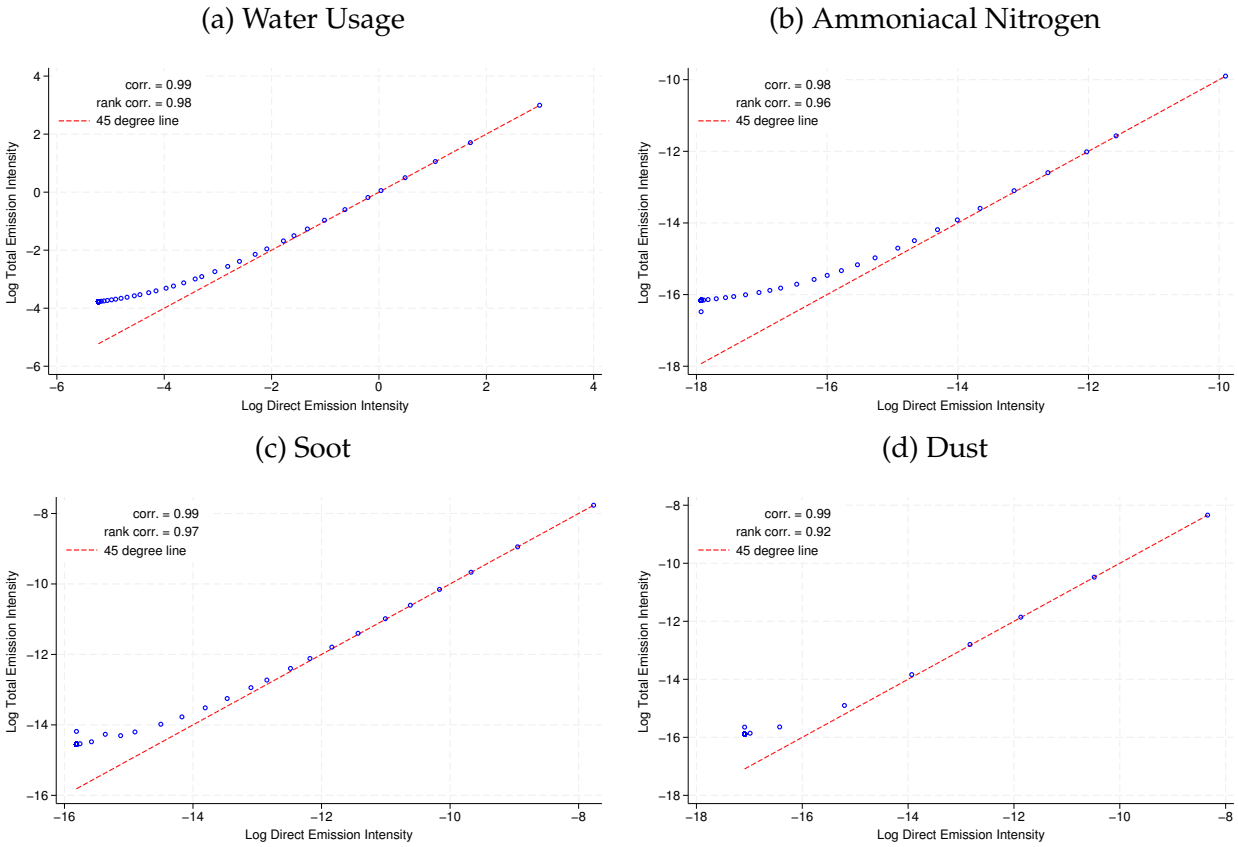
HS Code	Description
310000	**Fertilizers**
310230	Fertilizers, mineral or chemical; nitrogenous, ammonium nitrate, whether or not in aqueou...
310430	Fertilizers, mineral or chemical; potassic, potassium sulphate
310540	Fertilizers, mineral or chemical; ammonium dihydrogenorthophosphate (monoammonium phospha...
280000	**Inorganic chemicals**
281830	Aluminium hydroxide
282919	Chlorates; other than sodium
284130	Salts; sodium dichromate
720000	**Iron and steel**
720690	Iron or non-alloy steel; primary forms (excluding ingots and iron of heading no. 7203)
721631	Iron or non-alloy steel; U sections, hot-rolled, hot-drawn or extruded, of a height of 80...
721621	Iron or non-alloy steel; L sections, hot-rolled, hot-drawn or extruded, of a height of le...
790000	**Zinc and articles thereof**
790111	Zinc; unwrought, (not alloyed), containing by weight 99.99% or more of zinc
790400	Zinc; bars, rods, profiles and wire
790700	Zinc; articles n.e.c. in chapter 79
250000	**Plastering materials, lime and cement**
253010	Vermiculite, perlite and chlorites; unexpanded
251990	Magnesia, fused or dead-burned (sintered); whether or not containing small quantities of ...
251512	Marble and travertine; merely cut, by sawing or otherwise, into blocks or slabs of a rect...
710000	**Natural, cultured pearls; precious, semi-precious stones; precious metals, metals clad wi...**
710691	Metals; silver, unwrought, (but not powder)
711590	Metal; precious or metal clad with precious metal, other than that of item no. 7115.10
710122	Pearls; cultured, worked, whether or not graded (but not strung, mounted or set), tempora...

Figure E.3: Direct vs. Total Emission Intensities



Notes: These figures compare the logarithmic values of direct and total emission intensities for various emissions. We add the smallest value of total emission intensities to all values before log transformation to account for zero emission intensities. For exposition, we choose one HS 6-digit product from each of 100 percentile values of direct emission intensities and represent it as a dot. The dots presented in the figure appear to be less than 100 as many dots overlap each other toward zero (plus a small value) direct emission intensities. The displayed correlations derive from the full distribution of emission intensities rather than the 100 displayed dots.

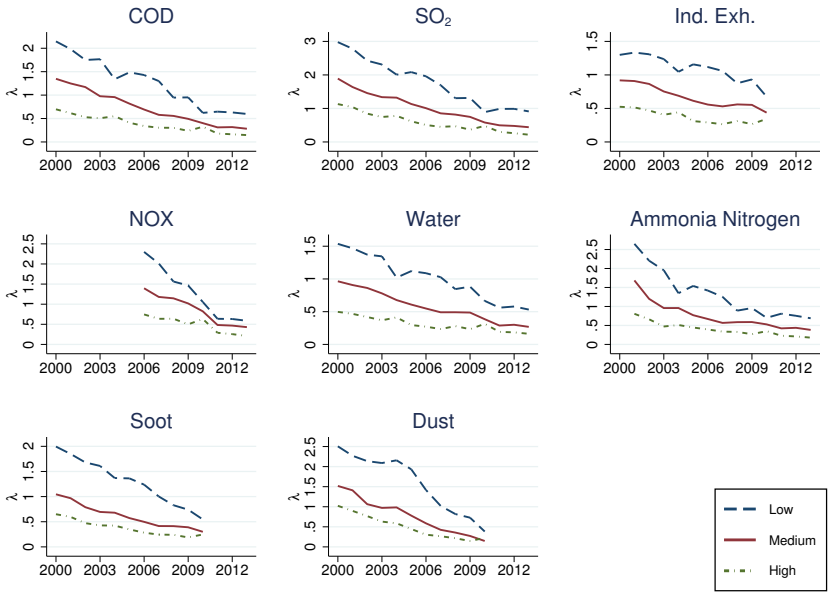
Figure E.4: Direct vs. Total Emission Intensities



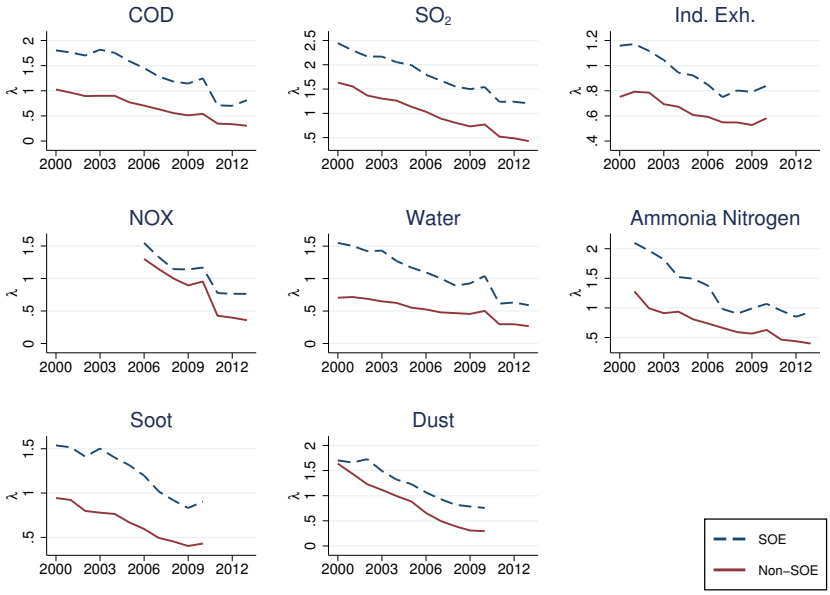
Notes: These figures compare the logarithmic values of direct and total emission intensities for various emissions. We add the smallest value of total emission intensities to all values before log transformation to account for zero emission intensities. For exposition, we choose one HS 6-digit product from each of 100 percentile values of direct emission intensities and represent it as a dot. The dots presented in the figure appear to be less than 100 as many dots overlap each other toward zero (plus a small value) direct emission intensities. The displayed correlations derive from the full distribution of emission intensities rather than the 100 displayed dots.

Figure E.5: Distribution of λ by Productivity and Ownership

(a) By Productivity



(b) By Ownership



Notes: These figures show the distribution of firms' emission inefficiency parameter λ by the level of productivity (upper) and ownership (lower).