REGULATING MISMEASURED POLLUTION:
IMPLICATIONS OF FIRM HETEROGENEITY
FOR ENVIRONMENTAL POLICY

By

Eva Lyubich, Joseph S. Shapiro, and Reed Walker

January 2018

Regulating Mismeasured Pollution:
Implications of Firm Heterogeneity for Environmental Policy[*]

Eva Lyubich	Joseph S. Shapiro	Reed Walker
UC Berkeley	Yale University and NBER	UC Berkeley and NBER
elyubich@berkeley.edu	joseph.shapiro@yale.edu	rwalker@berkeley.edu

January 2018

Abstract

This paper provides the first estimates of within-industry heterogeneity in energy and CO_2 productivity for the entire U.S. manufacturing sector. We measure energy and CO_2 productivity as output per dollar energy input or per ton CO_2 emitted. Three findings emerge. First, within narrowly defined industries, heterogeneity in energy and CO_2 productivity across plants is enormous. Second, heterogeneity in energy and CO_2 productivity exceeds heterogeneity in most other productivity measures, like labor or total factor productivity. Third, heterogeneity in energy and CO_2 productivity has important implications for environmental policies targeting industries rather than plants, including technology standards and carbon border adjustments.

JEL: F18, F64, H23, Q56

Regulating Mismeasured Pollution:
Implications of Firm Heterogeneity for Environmental Policy

Eva Lyubich
UC Berkeley
elyubich@berkeley.edu

Joseph S. Shapiro
Yale University
joseph.shapiro@yale.edu

Reed Walker
UC Berkeley
rwalker@berkeley.edu

This paper provides the first estimates of within-industry heterogeneity in energy and CO_2 productivity for industries spanning the entire U.S. manufacturing sector. We define energy and CO_2 productivity as log dollars of output per dollar of energy input or per ton of CO_2 emitted.[1] Three key takeaways emerge. First, within narrowly defined industries, heterogeneity across plants in energy and CO_2 productivity is enormous. For example, given one dollar of energy inputs, a plant at the 90th percentile of a typical industry's energy productivity distribution produces 580 percent more output than a plant at the 10th percentile of the same industry. Second, these values significantly exceed heterogeneity in most other measures of productivity. For example, the corresponding 90-10 differences for labor and total factor productivity are 400 percent and 150 percent, respectively. Third, heterogeneity in energy and CO_2 productivity has important implications for industry-based environmental regulations. Many countries have considered pairing a carbon tax on domestic output with a tariff on imports which is proportional to the carbon content of the imports. We show that an industry-based carbon tariff, which abstracts from within-industry heterogeneity, will substantially differ from the correct (plant-level) Pigouvian tax for many plants.

There are many existing environmental regulations are standards that apply uniformly across plants within an industry. For example, the U.S. Clean Air Act requires plants in regulated industries and regions to meet an industry-level technology standard by installing "Best Available Control Technologies." Similarly, the Clean Water Act's Industrial Effluent Guidelines require plants to meet an industry-level technology standard. Several tradable

---

[1]Energy and CO_2 productivity are the inverse of energy and CO_2 intensity. We use the former metric to facilitate comparisons to other single-factor and total-factor productivity measures.

permit markets use industry-level rebates to compensate firms.[2] Due to substantial data requirements, researchers and policy makers have a limited understanding of the extent of producer heterogeneity in energy and CO_2 productivity. We show that plants within an industry have very different pollution emissions rates, and therefore, such industry-level regulations will be too stringent for some plants and too lenient for others.

We use confidential, plant-level data from the U.S. Census of Manufacturers (CM) and the Manufacturing Energy Consumption Survey (MECS) to explore this heterogeneity. We distinguish about 375 6-digit NAICS industries. One industry, for example, manufactures carbon black; another makes ethyl alcohol. Our main results calculate plant-level energy expenditures on raw fuels and electricity as reported in CM and MECS. We also calculate plant-level CO_2 emissions through converting each fuel consumption choice to CO_2 equivalents using emissions factors (e.g., tons CO_2 emitted per ton of coal burned).

A few estimates near the paper's end analyze carbon tariffs. These estimates also account for energy consumption and emissions required to produce intermediate inputs that are used for final good production, sometimes called "indirect emissions." For example, in most of the paper, emissions for the cookware industry include coal, gas, oil, and electricity used to shape a pan. Indirect emissions for the cookware industry also include fossil fuels used to make aluminum, which is then purchased as an intermediate input to make a pan. We calculate indirect emissions in two separate ways. The first is standard: we invert the U.S. input-output table to compute the dollars of coal, oil, and natural gas inputs required to produce a dollar of output in each industry. This accounts for energy used to produce inputs, and energy used to produce the inputs to inputs, etc. Our second measure of indirect energy is non-standard: we use plant-level data on the dollar value of each individual material input the plant uses, along with associated industry codes for each material, which are all part of the CM Materials Trailer. We combine this information with the inverted input-output

---

[2]California's AB-32 cap-and-trade distributes additional permit allocations to energy intensive, trade-exposed industries using an industry-level assistance factor to help combat against regulatory leakage. These assistance factors are applied at the industry-level when determining permit allocations for a facility.

table to calculate indirect energy and emissions separately for each plant.

This paper builds on several literatures. One explores the implications of firm heterogeneity for environmental policy and either argues for market-based instruments like pollution taxes or cap-and-trade markets (Carlson, Burtraw, Cropper, and Palmer, 2000; Goulder and Parry, 2008) or analyzes industry-based regulation in Melitz-type settings when firms are heterogeneous (Shapiro and Walker, 2016). Several papers within this literature specifically analyze border adjustments (Cosbey, Droege, Fischer, and Munnnings, 2017; Kortum and Weisbach, 2017). This paper also relates to work analyzing the efficiency of imperfectly targeted environmental policies (Jacobsen, Knittel, Sallee, and van Benthem, 2017). A related literature shows that total factor productivity is heterogeneous within narrowly-defined or homogenous industries (Syverson, 2011); other work interprets heterogeneity as factor misallocation (Hsieh and Klenow, 2009). Existing analysis of heterogeneity in energy productivity is limited, though includes studies of a few specific industries like glass (Boyd and Pang, 2000). The remainder of the paper discusses data, methodology, and results.

1 Data and Methodology

We measure plant-level energy inputs using data from the 2007 CM and the 2006 MECS. The CM includes about 350,000 U.S. manufacturing plants operating in 2007, while MECS includes a probabilistic sample of around 15,000 plants. We join MECS and the CM at the plant level, using a unique plant identifier. Our MECS estimates use survey weights to make statistics represent the broader manufacturing sector. The CM reports each plant's value of shipments, capital stock, production hours, and expenditure on electricity, fuels, and materials. We exclude "administrative records" since many of their values are imputed. We also exclude records where output, fuel expenditures, or electricity expenditures are imputed.

The CM and MECS both report plant-level expenditure on fuels and on electricity, which we use to compute CO_2 emissions. MECS further reports physical quantities and expendi-

tures for each fuel, which we convert to CO_2 (see Appendix for details). Since the CM does not report expenditures by fuel type, we use MECS to calculate industry-level averages of CO_2 per dollar of fuel expenditure, and we multiply each CM establishment's fuel expenditure by these averages. For electricity inputs, we use the EPA's eGrid database—which assigns annual total output emissions rates (CO_2 per KWh) to 26 regions of the country—to calculate the mean marginal emissions based on plant location of electricity consumption.

We account for indirect emissions only in our estimates of carbon tariffs. We do this in two separate ways. First, we use the 2007 U.S. benchmark input-output data of the Bureau of Economic Analysis. We invert the input-output table to compute the total dollars of coal, oil, and natural gas inputs required to produce a dollar of output in each industry. We apply emissions coefficients from the Energy Information Agency and Environmental Protection Agency (EPA) to calculate the total CO_2 emitted per dollar of output in an industry. Our second measure of indirect emissions comes from the CM Materials Trailer, which provides plant-level detail on the dollar value of each material input, along with associated input industry codes. We multiply these expenditures by the corresponding industry emissions rate from the inverted input-output table. Thus, while emissions rates are constant across intermediate input industries, plant-level variation in intermediate input intensity generates additional heterogeneity in energy and emissions productivity.

We use all these data to construct multiple measures of energy and emissions productivity. For comparability with common productivity measures, we construct productivity measures as the log of the value of shipments per dollar of direct energy input, or per metric ton of CO_2 emitted. We also discuss estimates that define productivity as log dollars of value added per unit of energy input or CO_2 emissions. We calculate value added by subtracting expenditures on materials, parts, and energy from the plant's total value of shipments.

For each industry, we measure productivity heterogeneity by calculating the 90th and 10th percentile of energy and CO_2 productivity across plants within the same industry.[3] We

---

[3]To respect confidentiality requirements for 90-10 statistics, we use each industry's mean and standard deviation of the respective productivity measure to simulate the 90th and 10th percentile using a normal

also compute the within-industry standard deviation of all productivity measures for each of the 375 industries. Lastly, we summarize these industry-level dispersion measures by taking the unweighted mean across all industries. This latter statistic provides some insight on within-industry heterogeneity in productivity for the mean industry.

2   Results

Table 1 shows the mean and dispersion of seven different productivity measures. This table accounts for only "direct" CO_2 or other inputs at a plant. Columns 1-6 calculate productivity as the log of the plant's value of shipments divided by some measure of a plant's factor demand, CO_2 emissions, or intermediate inputs. Columns 2 and 3 report value of shipments per ton of CO_2 produced, where CO_2 is calculated using the CM and MECS samples, respectively. Columns 4-6 report other single factor productivity measures, as indicated in the column headings. Column 7 presents statistics from a total factor productivity index.[4]

Panel A of Table 1 shows mean productivity levels. For example, column 1 implies that energy costs are roughly 1.5 percent of output value $(0.015 = 1/\exp(4.16))$ for the mean plant in our sample, since the log of output per dollar energy input is 4.16. Panel B of Table 1 summarizes the industry-level dispersion measures. The first row presents the mean of the within-industry 90-10 ratio, taken across all industries in our sample. The second row of Panel B shows the standard deviation of the within-industry 90-10 ratio, taken across all industries. The third row shows the difference between the 90th percentile industry and the 10th percentile industry of this within-industry 90-10 dispersion measure. Panel C shows similar values, but using within-industry standard deviations.

Panels B and C of Table 1 show substantial heterogeneity in output per dollar of energy expenditure or per ton of CO_2 emitted, which is the paper's first main finding. The top-left entry in Panel B, for example, shows that given a dollar of energy inputs in the industry with

---

distribution. Estimates using the simulated data are nearly identical to those from the underlying microdata.

[4]This index uses a Cobb-Douglas production technology with three inputs: labor, capital, and materials. Output elasticities for each input are constructed from industry-level revenue shares (Syverson, 2011).

Table 1: Single and Total Factor Productivity Statistics

	Direct Energy (1)	CO$_2$ [CM] (2)	CO$_2$ [MECS] (3)	Labor (4)	Capital (5)	Materials (6)	TFP (7)
			Panel A: Industry-Wide Statistics				
Mean	4.16	8.42	8.79	4.51	1.01	0.95	1.81
SD	0.94	1.16	1.16	0.83	0.95	0.66	0.56
		Panel B: Within-Industry 90-10 Difference in Productivity					
Mean	1.92	2.27	2.27	1.63	2.22	1.34	0.92
SD	0.47	0.57	1.17	0.45	0.50	0.62	0.39
p90-10	1.21	1.46	3.01	1.16	1.27	1.58	1.00
		Panel C: Within-Industry Standard Deviation of Productivity					
Mean	0.75	0.89	0.89	0.64	0.87	0.52	0.36
SD	0.19	0.22	0.46	0.18	0.19	0.24	0.15
p90-10	0.47	0.50	1.14	0.44	0.43	0.55	0.33

NOTES: Panel A means and SD are computed from plant-level CM and MECS observations. Panel B statistics are calculated using the 375 within-industry 90-10 dispersion measures. Panel C statistics are calculated using the 375 within-industry standard deviation measures. See text for details.

the mean energy productivity dispersion, a plant at the 90th percentile of the within-industry energy productivity distribution produces 580 percent more output than a plant at the 10th percentile of that within-industry distribution does. Dispersion in CO$_2$ productivity is even wider, at 2.27 log points (870 percent difference). The standard deviation of energy and of CO$_2$ productivity within the average industry is 0.75 to 0.89 log points, respectively.

Panels B and C also show the paper's second main finding—dispersion in CO$_2$ and energy productivity is larger than dispersion in most other productivity measures. Both panels show that dispersion in energy and CO$_2$ productivity is more than twice as large as dispersion in total factor productivity (TFP). Typically single-factor productivity measures are more dispersed than TFP, but Table 1 shows that dispersion in energy and CO$_2$ productivity exceeds dispersion in other single-factor productivity measures like labor or material productivity.[5]

---

[5]The greater dispersion of single factor productivity compared to TFP stems from cross-plant differences in factor intensities. For example, if one plant has a greater labor share than another plant due to lower local wages, the two plants may have the same TFP but different labor productivity. Differences across plants in

Dispersion in energy and capital productivity is more similar, though worth interpreting cautiously since the durability of capital investments makes the value of the capital stock difficult to measure. All pairwise t-tests (not shown for space) reject the hypothesis that dispersion in energy and CO_2 productivity equals dispersion in the other productivity levels. Appendix Table A1 shows similar conclusions from value added productivity measures.

It may be unsurprising that CO_2 productivity varies so much, since differences in fuel inputs, variation across the grid in the CO_2 intensity of electricity generation, and related forces make CO_2 more variable than energy expenditure. It is more surprising that energy productivity varies more than other single-factor productivity measures, since even though some fuels are dirtier than others, one might expect plants to use similar amounts of energy to produce a single unit of output. Panel A of Table 1 shows that mean productivity for energy and labor are similar, so the difference in dispersion is not driven by scale effects.

Figure 1 demonstrates the paper's first two conclusions. This graph plots the distribution of industry-level 90-10 dispersion measures. Each of the roughly 375 observations underlying one of these lines is an industry; the value of each observation equals the within-industry 90-10 productivity ratio. The mean of the CO_2 distribution (dark solid line) lies above the mean of all other productivity dispersion measures, demonstrating that CO_2 dispersion for the average industry is greater than dispersion in the other productivity measures. The greater width of the CO_2 distribution relative to the TFP and labor distributions shows that within-industry dispersion in CO_2 productivity is more variable across industries than within-industry dispersion in TFP or labor productivity.

Implications for Carbon Tariffs

In many countries, policymakers have proposed import tariffs proportional to the carbon content of imported goods in order to guard against emissions leakage.[6] These are often

---

factor prices (e.g., wages) generally affect single-factor productivity but not TFP (Syverson, 2011).

[6]This type of policy was in the Waxman-Markey bill that passed the U.S. House but not the Senate in 2009. In 2017, France, Mexico, and Canada discussed imposing one on the U.S. after the Trump Administration announced it was withdrawing from the Paris Treaty on Climate Change. California has just implemented
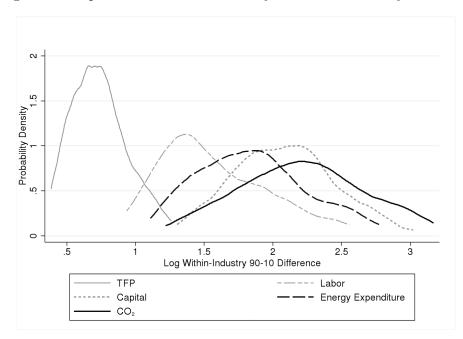
Figure 1: Dispersion of Within-Industry 90-10 Productivity Measures

referred to as carbon border adjustments or carbon tariffs.

Table 2 reports the level and distribution of the external cost of CO_2 emissions per dollar of output. If another country imposed a carbon tariff on imports from the U.S., the social cost of carbon (SCC) per dollar output provides one measure of the relevant tariff. We assume a standard SCC of \$40 per metric ton of CO_2. Each column represents a different method of calculating CO_2 per dollar of output. Column 1 presents direct emissions from fuels plus electricity per dollar of output using CM data. Column 2 adds indirect emissions to the direct emissions estimates from column 1, where indirect emissions are calculated by inverting the industry-level input-output table.[7] Adding industry-level indirect emissions changes the mean externality (Panel A) but not the within-industry dispersion (Panels B-

---

such a measure for government purchase decisions (the Buy Clean California Act).

[7]Total emissions are the sum of direct and indirect emissions. Direct emissions come from plant-level data. Indirect emissions come from the input-output table. Note that the input-output table provides both direct and indirect emissions for an industry. We subtract the industry-level direct emissions from total industry-level emissions to get our measure of indirect emissions.

C). Column 3 shows the same direct plus indirect emissions estimates from column 2 but uses MECS rather than CM to measure direct plant-level emissions. While MECS is a smaller sample than CM, it contains plant-level information on the types of fossil fuels used. Column 4 replaces the industry-level indirect emissions estimates used in columns 2 and 3 with indirect emissions calculated using the CM Materials Trailer. Column 4 uses plant-level information on input purchases to calculate indirect emissions. For each input material, however, it only accounts for the industry average of direct emissions of that input material and not its indirect emissions. Column 5 is similar to column 4, but for each material input, it calculates total (not just direct) emissions of each input using the inverted input-output table. Column 6 is similar to column 5 but uses MECS to measure direct emissions.

Table 2: Social Costs of Carbon Per Dollar of Output

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Industry-Wide Statistics					
Mean	0.019	0.029	0.023	0.036	0.051	0.046
SD	0.048	0.049	0.032	0.056	0.074	0.051
	Panel B: Within-Industry 90-10 Diff. in SCC/$					
Mean	0.060	0.060	0.051	0.076	0.098	0.084
SD	0.204	0.204	0.102	0.203	0.208	0.109
p90-10	0.523	0.523	0.261	0.521	0.533	0.280
	Panel C: Within-Industry Std. Dev. of SCC/$					
Mean	0.023	0.023	0.020	0.030	0.038	0.033
SD	0.080	0.080	0.040	0.079	0.081	0.043
p90-10	0.035	0.035	0.045	0.040	0.053	0.068
Direct Source	CM	CM	MECS	CM	CM	MECS
Indirect Source		BEA	BEA	CM	CM	CM
Leontief Inverse		X	X		X	X

NOTES: Panel A means and SD are computed from plant-level CM and MECS observations. Panel B statistics are calculated using the 375 within-industry 90-10 dispersion measures. Panel C statistics are calculated using the 375 within-industry standard deviation measures. Each column computes SCC per dollar of output using different inputs, as indicated in the column headings and table footers. A column represents either direct or total emissions, where direct emissions come from either the CM or MECS, and indirect emissions come from either the BEA I-O table or the CM Material trailer. See text for details.

Panel A of Table 2 shows that the external cost of CO_2 emissions for the mean plant is 2 to 5 percent of product value. Column 2 suggests that a uniform Pigouvian carbon tariff imposed on imports from U.S. manufacturers should be around 3 percent. In the first row of Panels B-C, column 2 shows that the mean industry has a 90-10 SCC difference of 0.06. This implies that even if a carbon tariff were imposed based on industry-specific means instead of the economy-wide 3 percent, many plants would have a carbon tariff which is well below the appropriate plant-level tax, whereas others would face a tax rate that is far too high. Comparing columns 3 and 5 of Panel B shows that using plant-level records of intermediate good purchases from CM, rather than industry-level records from the input-output table, approximately doubles both the 90-10 and standard deviation measures of dispersion.
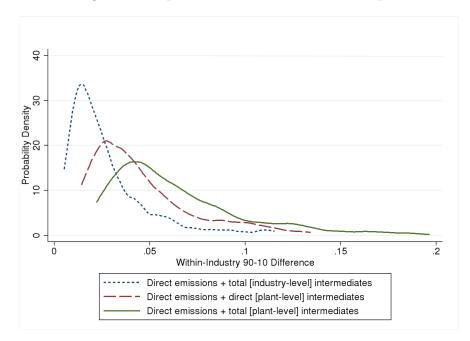
Figure 2: Dispersion in SCC Per Dollar Output

NOTES: Each kernel density plot was created using the approximately 375 6-digit NAICS dispersion measures for the corresponding emissions intensity measure. Densities are censored at the 5th and 95th percentiles.

Figure 2 plots the distribution of industry-level 90-10 differences in SCC per dollar output. This shows the main conclusions from Table 2 visually. Many industries have high 90-10 differences, and this distribution of dispersions has a long right tail which is understated by our censoring at the 5th and 95th percentile. Thus, a Pigouvian tax based on industry

averages would still miss significant heterogeneity in true SCC per dollar of output.

Detailed analyses of carbon tariffs have noted many challenges, ranging from legal ambiguity to information burdens. This paper uses plant-level data to highlight another tradeoff—while a plant-specific tariff would impose a large information burden, an industry-level tariff would have substantial targeting errors stemming from firm heterogeneity.

3   Discussion and Conclusions

The records used for this paper are the most detailed data we are aware of that cover the entire U.S. manufacturing sector. The plant-level granularity and detailed information on plant-level input purchases reveal significant heterogeneity in energy and CO_2 productivity, which exceeds heterogeneity for most other measures of single-factor and total factor productivity.

However, there are at least three reasons why our approach may understate the true extent of heterogeneity. First, we do not observe the full upstream set of plants that contribute to final output for a given plant in our data. Instead, we assign industry-level emission and energy intensities to construct our indirect emission and energy measures.[8] If supplying plants are significantly different in terms of emissions or CO_2 productivity, then we would understate heterogeneity. Second, our productivity estimates are based on revenues and not quantities. This should lead to underestimates of dispersion since more productive plants tend to have lower prices. Lastly, by excluding "administrative records" and other imputes from the CM, we are missing many of the smallest manufacturing establishments which might contribute to even more within-industry heterogeneity.[9]

How large are the welfare consequences of this heterogeneity for policies like technology standards or carbon tariffs that target industries and not plants? What are the economic

---

[8]Even when we observe plant-level input purchases, we only observe the industry of those inputs and not the specific plant.

[9]It is worth noting that while these reasons suggest we are understating true heterogeneity, any remaining measurement error after excluding imputed observations could lead to overstatement of true heterogeneity.

reasons why energy productivity is more widely dispersed than labor or total factor productivity? How would decreasing factor misallocation across firms affect CO_2 emissions? Finally, what does heterogeneity in CO_2 productivity imply about heterogeneity in marginal abatement costs? We leave these important questions for future work.

References

BOYD, G. A., AND J. X. PANG (2000): "Estimating the linkage between energy efficiency and productivity," *Energy Policy*, 28(5), 289–296.

CARLSON, C., D. BURTRAW, M. CROPPER, AND K. L. PALMER (2000): "Sulfur Dioxide Control by Electric Utilities: What are the Gains from Trade?," *Journal of Political Economy*, 108(6), 1292–1326.

COSBEY, A., S. DROEGE, C. FISCHER, AND C. MUNNNINGS (2017): "Developing guidance for implementing border carbon adjustments: Lessons, cautions, and research needs from the literature," Mimeo, RFF.

GOULDER, L. H., AND I. W. H. PARRY (2008): "Instrument Choice in Environmental Policy," *Review of Environmental Economics & Policy*, 2(2), 152–174.

HSIEH, C.-T., AND P. J. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

JACOBSEN, M. R., C. R. KNITTEL, J. M. SALLEE, AND A. A. VAN BENTHEM (2017): "Sufficient statistics for imperfect externality-correcting policies," NBER Working Paper 22063.

KORTUM, S., AND D. WEISBACH (2017): "The Design of Border Adjustments for Carbon Prices," *National Tax Journal*, 70(2).

SHAPIRO, J. S., AND R. WALKER (2016): "Why is Pollution from U.S. Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade," Mimeo, Yale University.

SYVERSON, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–365.

Appendix A   For Online Appendix: Additional Tables

Table A1: Descriptive Statistics of Industry-Level Characteristics - Value Added

	Direct Energy (1)	CO_2 [CM] (2)	CO_2 [MECS] (3)	Labor (4)	Capital Stock (5)	Materials (6)
			Panel A. Industry-Wide Statistics			
Mean across all plants	3.63	7.89	8.24	3.97	0.48	0.42
SD across all plants	1.07	1.28	1.31	0.81	1.03	1.07
			Panel B. Within-Industry 90-10 Productivity Diff.			
Mean	2.20	2.53	2.47	1.76	2.43	2.33
SD	0.54	0.59	1.28	0.56	0.53	0.76
p90-10	1.39	1.51	3.27	1.43	1.37	1.95
			Panel C. Within-Industry Productivity Standard Dev.			
Mean	0.86	0.99	0.96	0.69	0.95	0.91
SD	0.21	0.23	0.50	0.22	0.21	0.30
p90-10	0.50	0.54	1.28	0.51	0.45	0.63

NOTES: Panel A means and SD are computed from plant-level CM and MECS observations. Panel B statistics are calculated using the 375 within-industry 90-10 dispersion measures. Panel C statistics are calculated using the 375 within-industry standard deviation measures. See text for details.

Table A2: Social Costs of Carbon Per Dollar of Value Added

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Summary Stats, CM					
Mean across all plants	0.06	0.09	0.04	0.23	0.41	0.13
SD across all plants	2.72	2.72	0.31	3.3	4.69	1.01
	Panel B. Within-Industry 90/10 Productivity Diff.					
Mean	0.55	0.55	0.24	1.99	3.76	0.64
SD	4.22	4.22	1.49	12.19	26.54	3.12
p90-10	10.81	10.81	3.82	31.26	68.05	8.00
	Panel C. Within-Industry Productivity Std. Dev.					
Mean	0.21	0.21	0.10	0.78	1.47	0.25
SD	1.65	1.65	0.58	4.76	10.35	1.22
p90-10	0.18	0.18	0.14	0.92	1.71	0.36
Direct Source	CM	CM	MECS	CM	CM	MECS
Indirect Source		BEA	BEA	CM	CM	CM
Leontief Inverse		X	X		X	X

NOTES: Panel A means and SD are computed from plant-level CM and MECS observations. Panel B statistics are calculated using the 375 within-industry 90-10 dispersion measures. Panel C statistics are calculated using the 375 within-industry standard deviation measures. Each column computes SCC per dollar of output using different inputs, as indicated in the column headings and table footers. A column represents either direct or total emissions, where direct emissions come from either the CM or MECS, and indirect emissions come from either the BEA I-O table or the CM Material trailer. See text for details.

Appendix B   Data

Appendix B.1   Direct Energy and Emissions

The main text describes how we construct the analysis sample. Here we describe a few additional sample restrictions designed to limit measurement error.

For each variable in the raw data, the final sample excludes observations that are more than 100 times larger than the 99th percentile value. We do not apply this rule to ratios, e.g., this restriction is applied to CO_2 and to output but not to CO_2 productivity. The final sample also excludes plants that report zero or have missing values for any of our variables,[10] and plants that do not report positive values for at least one material in the materials trailer. Finally, the sample excludes establishments that are unique in their industry after all the

---

[10]In cases where electricity kWh variables are missing in MECS but not CM, we use CM values to calculate total emissions in MECS.

above restrictions, since we cannot compute 90-10 dispersions or standard deviations for these industries.

We calculate emissions from fuel use by multiplying each establishment's consumption by fuel-specific emissions factors. We assign these emissions factors using data from the EPA when possible and from the EIA otherwise. We treat acetylene, hydrogen, and waste and byproduct gases as zero emissions. For emissions from electricity, we assign CO_2 per MWh using the EPA's eGRID dataset. We match eGRID regions to counties and compute emissions from electricity at the establishment level by multiplying each establishment's electricity consumption with the corresponding emissions factor from eGRID. In cases where a county overlaps with several eGRID regions, we take an unweighted mean of emissions intensities across the relevant eGRID regions. For observations in the CM that are missing the county variable, we take the unweighted mean of emissions factors across counties within the state. We do not account for process emissions.

Appendix B.2 Indirect Emissions

We use the BEA's 2007 benchmark Make Table, Use Table, and Import Table to construct an industry-level input-output (I-O) table. The BEA tables distinguish between industries and commodities to reflect the fact that some industries produce commodities other than the primary product of that industry (known as secondary commodities). We use tables after redefinitions, which in certain casesreallocate secondary commodity outputs to the industry in which they are the primary product, because this makes industries more homogenous.[11] In practice most I-O codes in the benchmark analysis represent both a commodity and an industry. Exceptions to this are four commodities which are not industries (scrap goods, non-comparable imports, used and second-hand goods, and rest-of-world adjustment), and nine industries corresponding to different types of government enterprises. In cases where a government industry has an analog in private industry – for example federal electric utilities – the BEA assigns both the public and private industries' commodity outputs to the private industry's commodity code.

The make table is an industry-by-commodity table, with each element m_{ij} representing industry i's output of commodity j, in nominal dollars. The use (and import) tables are commodity-by-industry tables, with each element u_{ij} representing the total (imported) amount of commodity i used in industry j's production, also in nominal dollars. In addition to the commodity-by-industry pairs, the use table contains three value added rows (compensation of employees; taxes on production and imports less subsidies; and gross operating surplus) and 20 final demand columns. These additional rows and columns play an important role in ensuring that total inputs equal total outputs, but they are not rows or columns of the final I-O table. The use and import tables are available from the BEA at producer values and purchaser values – we use producer values throughout to maintain consistency with the make table. We construct a domestic use table by subtracting import values from the use table.

The BEA combines crude oil and natural gas extraction into one industry (code 211000).

---

[11]The BEA reallocates secondary output from an industry to the industry in which it is the primary product when the two industries' input structures differ significantly

We split this industry into two, in order to treat oil and natural gas extraction separately. We assign all of the petroleum refineries commodity produced by the original industry to the new crude oil industry, and we assign all of the industrial gas manufacturing commodity produced by the original industry to the new natural gas industry. The rest of commodity output is assigned such that total production of gas and crude oil are proportional to their overall production according to the EIA. We assume that the commodity input mix for each of the two new industries is the same, with levels proportional to industry output, and we maintain oil and gas extraction as one commodity.

We normalize elements of the make table by commodity totals to generate a "market shares" table, in which each element s_{ij} is the proportion of commodity j produced by industry i. Analogously, we normalize elements of the domestic use table by industry totals to generate a direct requirements table, in which each element d_{ij} is the proportion of industry j's production made up by commodity i. Because we are interested only in combustible fuel use, we adjust direct requirements values by proportions of fuel used for combustion using EIA values.[12]

We generate the industry level I-O matrix by multiplying the market share matrix by the direct requirements matrix. The elements of this matrix are how much of each input an industry uses to produce one dollar of output. Thus equilibrium is defined by:

$$X = AX + Y$$

where X is an industry-length vector of gross production, Y is an industry-length vector of final demand, and A is the I-O matrix. We can rearrange to get

$$X = (I - A)^{-1}Y$$

$(I - A)^{-1}$ is referred to as the Leontief inverse. Using the Leontief inverse, we can calculate how much output is necessary in total from every industry to meet a given vector of final demand.

Thus, we calculate total emissions embedded in the production necessary to meet a unit of demand for goods from a given industry by left multiplying the Leontief Inverse by a row vector of the raw emissions intensities for coal, crude oil, and natural gas, which we get from the EPA. Since we are using CM data to calculate a more granular measure of direct emissions from production, we calculate indirect embedded emissions by subtracting emissions from the direct requirements from the total emissions:

$$\text{IndirectEmissions}_j = \text{TotalEmissions}_j - \sum_i (\text{DirectEmissions}_j \times \text{InputOutput}_{ji})$$

where the direct emissions vector is calculated from the total emissions vector, resetting all values to 0 except those corresponding to utilities and fuel industries.

After creating the BEA-level emissions intensities, we convert from BEA industry definitions to NAICS industry definitions using the concordance provided by the BEA. If multiple BEA industries correspond to a single NAICS industry, we take BEA output-weighted means to calculate a unique NAICS industry value. If a BEA industry gets split into multiple

---

[12]These are calculated as a proportion of first use energy consumption and not total energy consumption.

NAICS industries, all NAICS industries get the same value. There are several BEA industries that don't have corresponding NAICS codes—importantly, the BEA considers government utilities and private utilities separately, and only the private utility gets mapped to a NAICS utility code.

We use the indirect emissions calculated from the BEA to account for the full embedded emissions of production in two ways. One is through addition of the intermediate emissions intensities, by industry, to direct emissions intensities from CM. The second uses the CM Materials Trailer, which identifies material inputs into production by establishment. We use the BEA emissions intensity values to calculate the direct and the total emissions embedded in material inputs. The direct emissions capture the industry averages for emissions from fuel and electricity use in the production of materials. The indirect emissions use the full Leontief inverse to capture all emissions generated throughout the economy in the production of the materials, on average by industry. We add these to CM emissions intensities to calculate two versions of total emissions productivity based on material inputs. In compiling data from the CM Materials Trailer, we assign zero emissions to unspecified materials inputs (the "other industry" category). The fact that these "other industry" inputs represent a reasonable share of all inputs provides another reason why our estimates understate true dispersion in energy and CO_2 productivity.