Online Appendix to: Regulating Mismeasured Pollution: Implications of Firm Heterogeneity for Environmental Policy

Eva Lyubich, Joseph S. Shapiro, and Reed Walker

Appendix A 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,[1] 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

[1] 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.[2] 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).

[2]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.[3]

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

[3]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.