For Online Publication
Appendix for "The Size Distribution of Firms and Industrial Water Pollution: A Quantitative Analysis of China"

Ji Qi
Chinese Academy of Environmental Science

Xin Tang*
Wuhan University
International Monetary Fund

Xican Xi
Fudan University

September 19, 2019

CONTENTS

*Corresponding Author: zjutangxin@gmail.com

A National General Survey of Pollution Sources

The most important data source that we draw upon in Section I.B of the main text is the 2007 *National General Survey of Pollution Sources* (henceforth NGSPS). In this section, we provide a detailed description of the dataset, which is based on the following documents:

- *Census Program of the National General Survey of Pollution Sources* (第一次全国污染源普查方案);
- *Regulation on National General Survey of Pollution Sources [Decree of the State Council of the People's Republic of China (No.508)]* (全国污染源普查条例[中华人民共和国国务院令第508号]).
- *Technical Specifications Requirements of the First National General Survey of Pollution Sources* (第一次全国污染源普查技术规定).

For an official introduction to the NGSPS and declassified information at the aggregated level, interested readers could refer to the book *Data Collection of the National General Survey of Pollution Sources* (《污染源普查数据集》, 中国环境科学出版社, 2011), which is publicly available.

A.1 *General Introduction*

The NGSPS was a joint effort of multiple national ministries in China. The survey was organized directly by the State Council with data collection and quality control done by specially trained field staffs. According to the Regulation on National General Survey of Pollution Sources [Decree of the State Council of the People's Republic of China (No. 508)], faking data would lead to financial penalties or even lawsuits. The purpose of the NGSPS is to understand the number of pollution sources and their distribution in different industries and regions; to understand the generation, discharge and treatment of major pollutants; to establish records for key pollution sources; to build a pollution source information database and an environmental statistics platform; and to provide basis for formulating policies and plans for economic and social development and environmental protection. The term "pollution source" here refers to premises, facilities and equipment which discharge pollutants to environment in the process of production, living or other activities or have adverse impact on environment, as well as other sources that result in pollution.

The survey covers all of the pollution units from agricultural, industrial, domestic sources, and facilities for the centralized treatment of pollution within the borders of the People's Republic of China. In this study, we focus on the industrial pollution sources, which consist of businesses from 39 manufacturing industries according to the *Industrial Classification for National Economic Activities* (GB/T4547-2002).

Key and Regular Sources.—The firms in the industrial sources are divided into two groups, *key sources* and *regular sources*. A key source firm is required to file questionnaires that are more detailed than those filed by the regular source firms. A firm is categorized as a key source, if one of the following condition is satisfied:

1. All production entities that discharge pollutants that contain heavy metal, hazardous waste, and radioactive substance.
2. All production entities that belong to the 11 heavily polluting industries, which include: Paper and Paper Products; Food Processing; Raw Chemical Materials and Chemical Products; Textile; Ferrous Metal Smelting and Rolling Processing; Food Manufacturing; Production and Supply of Electric and Heating Power; Manufacturing of Leather, Fur, and Feather; Processing of Petroleum, Coking, and

Nuclear Fuel; Manufacturing of Non-metallic Mineral Goods; Ferrous Metal Smelting and Rolling Processing.

3. Firms with more than CNY 5 million in revenue that belong to the 16 key industries, which include: Beverage Manufacturing; Medicine Manufacturing; Chemical Fibers Manufacturing; Transportation Equipment Manufacturing; Coal Mining and Washing; Non-ferrous Metal Mining; Processing of Timber, and Manufacture of Wood, Bamboo, Rattan, Palm and Straw Products; Petroleum and Natural Gas Exploitation; General Purpose Machinery Manufacturing; Ferrous Metal Mining; Non-metal Mining; Apparel, Footwear and Caps Manufacturing; Water Production and Supply; Metal Products Manufacturing; Special Purpose Machinery Manufacturing; Communication Equipment, Computers and Other Electronic Equipment.

A firm is categorized as a regular source firm, if none of the above conditions holds.

Variables.—The survey collects information on the following variables for the industrial sources.

1. Firm's basic registration information, geographic latitude and longitude, the receiving water body of firm's discharged wastewater, etc.
2. The consumption of raw and intermediate inputs, including: water consumption, energy consumption (coal, petroleum, gas, electricity, etc.), the sulfur content of fuel, the consumption of hazardous intermediate input, etc.
3. The quantity of each of the products produced by the firm.
4. The type, size and number of the pollutants treatment equipment that the firms own.
5. The generation, abatement, discharge and comprehensive utilization of various types of pollutants; and the operation of various kinds of pollution prevention and control facilities.
6. The monitoring of pollutants emissions, including: the date and frequency of the monitoring; the type, quantity, and concentration of the pollutant.

Pollutants.—The pollutants included for the industrial sources are those that have general implications on pollution control. In particular, these pollutants are as follows.

1. Wastewater: Chemical Oxygen Demand (COD), Ammonian, Petrochemicals, Volatile Phenols, Mercury, Cadmium, Plumbum, Arsenium, Hexavalent Chromium, Cyanidium. For Paper and Paper Product, Food Processing, Food Manufacturing and Beverage Manufacturing Industries, Five-days Biochemical Oxygen Demand (BOD_5) is added. For Urban Sewage Treatment Plants, Total Phosphorus, Total Nitrogen, and BOD_5 are added.
2. Exhaust: Soot, Industrial Dust, Sulfur Dioxide. For Electrolytic Aluminium, Cement, Ceramic, Frosted Glass industries, Fluoride is added. For Vehicle Exhaust Census, Carbon Monoxide and Hydrocarbon are added.
3. Industrial Solid Waste: Hazardous Waste (according to the *National Catalogue of Hazardous Wastes*), Smelting Waste, Flying-Ash, Slag, Coal Refuse, Gangue, Radioactive Slag.
4. Plaster discharged by Desulfurization Facilities, Sludge generated by waste water treatment facilities, and the remaining from burning hazardous waste.

TABLE A.1—PERCENTAGE OF FIRMS WITH POSITIVE EMISSIONS BY POLLUTANTS

	Waste	COD	Petro	NH_4^+	BOD	CN	Cr^{6+}	Phenol	As	Cr
Key Sources	76.2	73.2	31.4	25.2	17.4	4.90	4.86	2.42	2.28	2.01
Regular Sources	35.2	28.3	7.91	6.49	2.56	0.13	N/A	0.04	0.07	N/A

† Source: The 2007 National General Survey of Pollution Sources (NGSPS). The acronyms are respectively referring to: Wastewater, Chemical Oxygen Demand, Petrochemicals, Ammonian, Biochemical Oxygen Demand, Cyanidium, Hexavalent Chromium, Volatile Phenols, Arsenium and Chromium.

5. Radioactive pollution sources from the utilization of Concomitant Radioactive Mineral and from Civil Nuclear Power Generation.

Table A.1 lists the percentage of key and regular firms that report positive emissions of different water pollutants, respectively.

A.2 *Methods of Measurement*

There are three different methods to measure the level of pollutant emissions, which are *Monitoring Method, Method of Emission Coefficient*, and *Material Balance Method*. In this section, we describe these three methods and explain which method is used to determine the level of pollutant emissions in practice.

Monitoring Method.—In this method, the actual quantity and concentration of pollutants are measured, which are used to infer the annual quantity of the generation and discharge of different pollutants. In practice, there are three sources from which the monitored data are obtained, ranked by the priority of adoption when data from multiple sources are available: *NGSPS monitoring, historical monitoring*, and *online monitoring*. The NGSPS monitoring data are collected by the field staff in the survey year (2007). The historical monitoring data are measured and recorded in the past three years, with the latest data having the highest priority.[1] The online monitoring data are those automatically uploaded by computerized and Internet-connected pollutants treatment equipment. Before the historical and online monitoring data are adopted, the field staff of the survey would ensure that the production condition and pollutant treatment technology have not undergone substantial changes.

A pollution source needs to report monitored data, if any of the following four criteria is met:

1. Nationally Monitored Key Pollution Sources (国控重点污染源): all firms listed in the *List of National Intensively Monitored Firms* (国家重点监控企业名单,环办函[2007]93号).

2. Facilities for Centralized Treatment of Pollution.

3. Provincially Monitored Key Pollution Sources (省控重点污染源) that in sum account for 65% of total provincial emission as is recorded in the *2005 Environmental Statistic Yearbook*. That is for each of the major pollutants, starting from the most polluting firm and adding up the quantity of emission, all firms until the summation reaches the 65% of the total provincial emission are recorded. The collection of the list of firms for all major pollutants provides the firms in this category.

4. Newly established projects since 2005 whose pollutants discharge is higher than the least polluting firms in the Provincial Monitored Key Pollution Sources.

[1]There are four different sources of historical data, ranked again by the priority of adoption when data from multiple sources are available: the historical data monitored by the local environmental authorities, the data monitored upon the completion of a newly constructed project, the data monitored by a third-party agency, and those self-reported by the firms.

These firms are later referred to as the *monitored firms* in this document.

Starting from 2007Q1, for all Nationally Monitored Key Pollution Sources, Centralized Treatment Facilities, and Provincially Monitored Key Pollution Sources, waste water sources have to be monitored at least once per quarter, and exhaust sources have to be monitored at least once per six months.

The actual monitoring practice is subject to the following guidelines:

1. The monitoring of Mercury, Cadmium, Hexavalent Chromium, Plumbum, and Arsenium must be conducted at the discharging outlets of individual factory workshops, with each outlets monitored separately.
2. Other water-based pollutants are monitored at the outlets of the factory.
3. The flow rates of wastewater and exhaust are monitored at the same time with the monitoring of pollutants discharges.
4. The monitoring of pollution sources are conducted at the representative production and polluting time periods.
5. All monitoring technical standards of wastewater pollutants follow *The Technical Specifications Requirements for the Monitoring of Surface Water and Wastewater* (地表水和污水监测技术规范, HJ/T91-2002) and *The Technical Specifications Requirements for the Monitoring of Wastewater Pollutants Emission* (水污染物排放总量监测技术规范, HJ/T92-2002).
6. For unstable wastewater sources, the samples are collected using the *Water Ratio Automatic Sampler*.
7. All monitoring technical standards of exhaust pollutants follow *The Determination of Particulates and Sampling Methods of Gaseous Pollutants from Exhaust Gas of Stationary Sources* (固定污染源排气中颗粒物测定和气态污染物采样方法, GB/T16157-1996).

Method of Emission Coefficient.—In this method, the generation and discharge of pollutants are calculated according to the *Handbook of Emission Coefficient*, edited by the *Chinese Academy of Sciences*. The benchmark coefficients are estimated based on firm's production technology, production scale, etc. The actual coefficients used are modified from the benchmark coefficients according to firm's use of intermediate inputs, its managerial practice, and the pollution treatment equipment installed. The actual pollutants generation and discharge are then calculated according to the actual production scale of the firm in year 2007.

Material Balance Method.—In this method, the quantities of the materials entering and leaving a system are measured, and therefore the material flows which might have been unknown or difficult to measure without this method are identified. The exact rule of conservation used in the analysis of the system depends on the specific context, but all rules revolve around mass conservation, i.e., that matter cannot disappear or be created spontaneously. In parctice, a firm's usage of intermediate inputs, energy consumption, water consumption, and production technology are used simultaneously, in order that the calculated pollutants generation and discharge can realistically reflect the actual production and emission of the firm in reality.

Which Method Will Be Used in Practice?—To determine which measurement method will be used in practice, the survey follows the following guidelines.

1. The data of the Key Pollution Sources are obtained mainly through monitoring and the method of emission coefficient, while the material balance method is only used when the other two methods are not feasible. More specifically, the data of all firms that are identified as the monitored firms are

obtained by monitoring. The data of other firms in the key sources are obtained by the method of emission coefficient.

2. The data of the Regular Pollution Sources are obtained mainly through the method of emission coefficient, while the material balance method is used only when the method of emission coefficient is not feasible.

3. Before being accepted, the monitored data are compared with those calculated from the method of emission coefficient. If the discrepancy is less than 20% of the monitored value, the monitored value is used. If the discrepancy is larger than 20%, the production technology and operating status of the firm are examined. If the operating status is in compliance with that specified in the technical regulations of the monitoring practice (for example, the firm has to reach at least 75% of its production capacity when the emission is monitored), the monitored data are accepted. Otherwise, the emissions are calculated using the method of emission coefficient.

B A Comparison of Firm Size Distributions for Different Samples

This section compares the firm size distributions for different samples of firms in China. In Figure B.1, we present the kernel densities of firm size estimated using different samples in NGSPS and CNEC. On the upper panels, we plot the firm size distributions for the samples that consist of the key source firms (the left panel) and all firms (the right panel) in the NGSPS, respectively. On the lower panels, we plot the firm size distributions for the *Annual Surveys of Industrial Firms (ASIF)* sample (the left panel), which consists of all state-owned firms and private firms with sales above CNY 5 million, and the sample of all firms (the right panel) in the CNEC. There are two findings that we would like to emphasize:

1. Although the key source firms are on average larger than the firms in the full sample, they include both small and large firms, and the empirical density of the log output of the key source firms resembles a bell shape. This suggests that the key source firms are reflecting the characteristics of the firms over the entire range of firm size. As a result, no systematic bias will be introduced when we use the key source firms to empirically analyze the relationship between pollution intensity/treatment technology adoption and firm size in Section I.B.

2. There are marked differences in the firm size distributions between the ASIF sample and CNEC full sample, with the ASIF sample missing most of the small firms. As a result, while the ASIF data are widely used in the studies of Chinese economy, we use the CNEC full sample in our paper, because both the small and large firms play a critical role in our empirical and quantitative analysis.

In sum, the two take away points from the four panels in Figure B.1 are that the key firms in NGSPS are representative of both large and small firms, and that we have to use the CNEC full sample in companion with the NSGPS for our purpose.

C Additional Results on Firm Size and Pollution Intensity

In this section, we present additional results on the statistical relationship between firm size and pollution intensity. In Section C.1, we show the scatter-plots and regression results for individual polluting industries and and for the whole manufacturing sector as well. In addition, we show that the negative relationship between firm size and pollution intensity holds conditional on whether the firms are using physical, chemical or

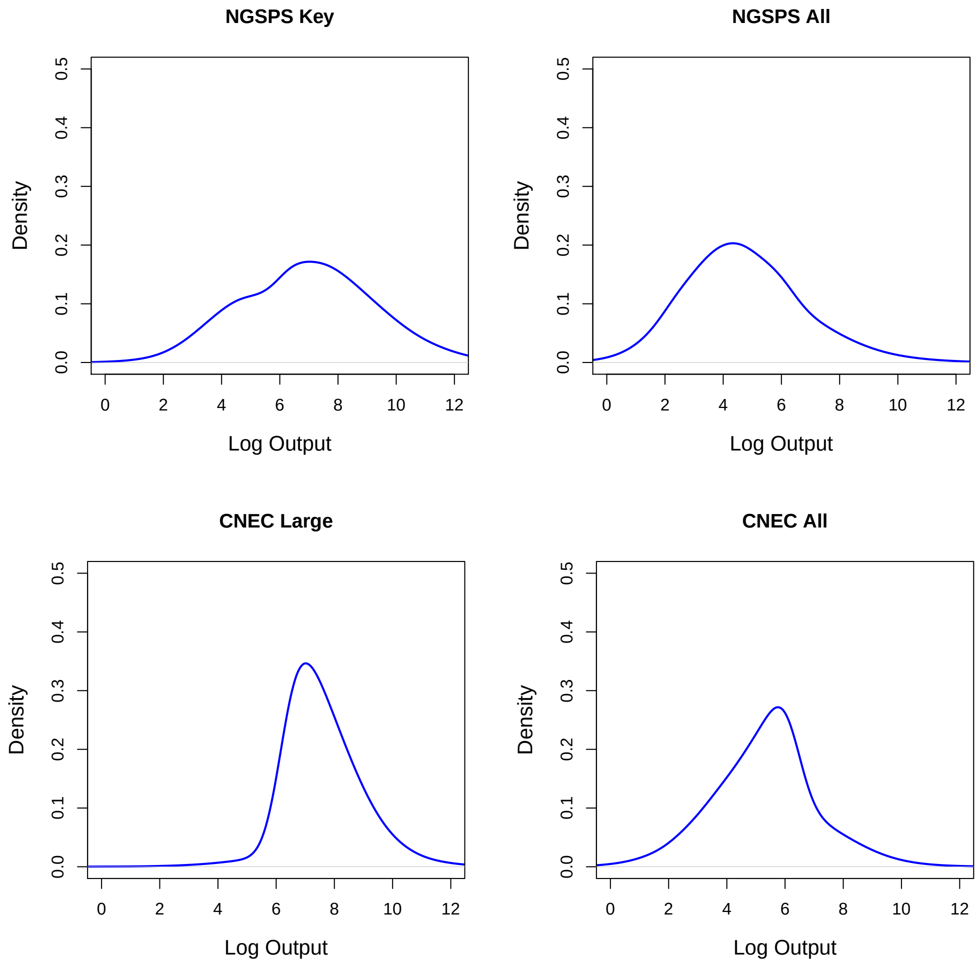

FIGURE B.1. FIRM SIZE DISTRIBUTIONS IN VARIOUS DATA SOURCES

Sources: NGSPS and CNEC. In all panels, the horizontal axes are in log-scale.

biological technologies. We also show that our results in Section I.B remain robust if we cluster the standard errors by provinces. In Section C.2, we explain our estimates on coefficients of the dummy variables, and compare our findings with those from the existing literature.

C.1 *Additional Scatter-plots and Regression Results*

For each industry among the top five polluting industries, we regress the following equation:

$$\log(\text{COD}_i) = \beta_0 + \beta_1 \log(\text{Output}_i) + \mathbf{X_p}\boldsymbol{\gamma}_1 + \mathbf{X_o}\boldsymbol{\gamma}_2 + \varepsilon_i,$$

where $\mathbf{X_p}$ is the provincial dummy, and $\mathbf{X_o}$ is the ownership dummy. The estimation results are summarized in Table C.1, with the corresponding scatterplots plotted in Figure C.1. We also include the scatterplot with the top five industries pooled together in Figure C.1 for completeness. All the estimates are significant both

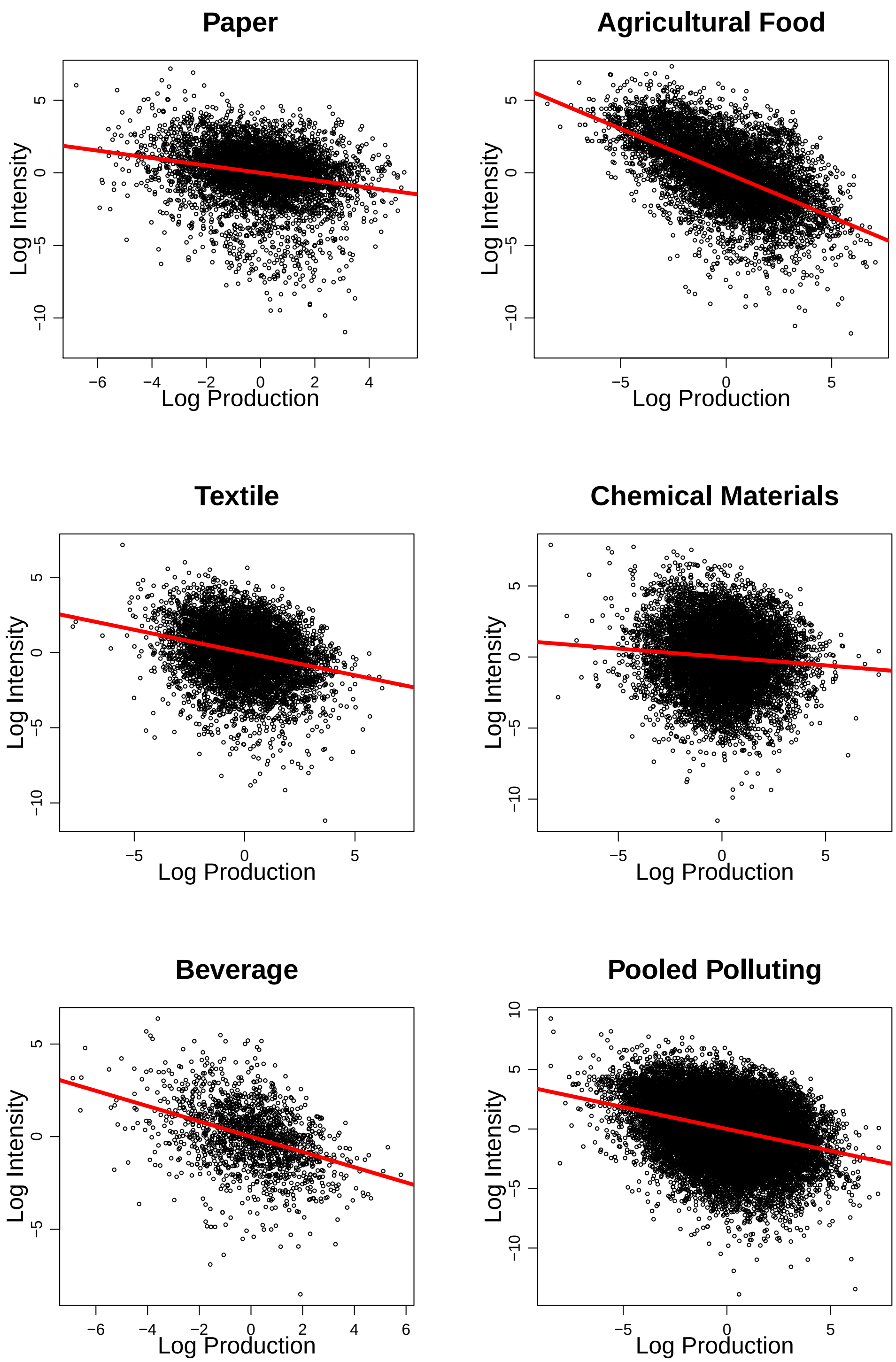

FIGURE C.1. FIRM SIZE AND POLLUTION INTENSITY

Source: NGSPS. Line: Least square fit.

TABLE C.1—FIRM SIZE AND POLLUTION INTENSITY BY INDUSTRY

Parameters	Paper	Food	Textiles	Chemical Materials	Beverages
β_0	−4.14	−2.15	−3.71	−7.12	−2.86
	(0.56)	(0.26)	(0.57)	(0.33)	(0.39)
β_1	0.74	0.39	0.70	0.88	0.59
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
R^2	0.38	0.31	0.38	0.40	0.34
N	5,632	6,893	6,284	8,686	1,558

† Data Source: NGSPS.

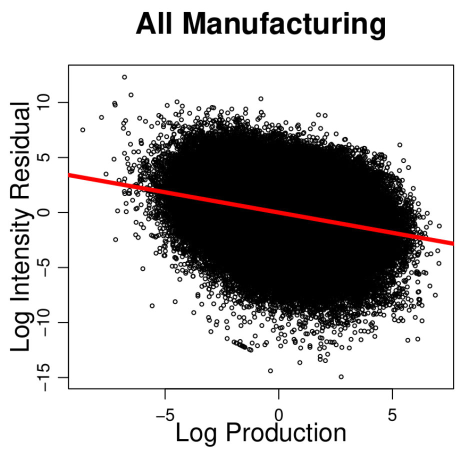

FIGURE C.2. FIRM SIZE AND POLLUTION INTENSITY: ALL MANUFACTURING INDUSTRIES

Source: NGSPS. Line: Least square fit. Residuals after controlling for the covariates are plotted.

statistically and economically with reasonable levels of R^2.

If we apply Regression (1) in Section I.B of the main text to the whole manufacturing sector, we get the following results

$$\log(\text{COD}_i) = \underset{(0.84)}{-11.05} + \underset{(0.00)}{0.63} \times \log(\text{Output}_i) + \mathbf{X}\boldsymbol{\gamma} + \varepsilon_i,$$

which implies a negative correlation between firm size and pollution intensity. However, as we mentioned in Section I.A of the main text, because industrial water pollutants are typically concentrated in a handful of industries, when we apply the above analysis to the whole manufacturing sector, we lose about two thirds of the manufacturing firms, as firms in many industries did not report positive COD emissions. Figure C.2 visualizes the above regression by plotting residuals after controlling for all the covariates. Likewise, if we apply the linear regression of technology adoption in Section I.B of the main text to the whole manufacturing sector, we get the following results:

$$y_i = \underset{(0.11)}{-0.16} + \underset{(0.000)}{0.04} \times \log(\text{Output}_i) + \mathbf{X}\boldsymbol{\gamma} + \varepsilon_i,$$

implying again a similar positive correlation between firm size and the adoption of clean technology.

For the size-intensity relationship conditional on treatment equipment, in Section I.B of the main text, we group physical and chemical equipments together with an extra dummy variable for chemical equipment.

Here, we show that the relationship holds as well when we estimate the relation for each type of equipment separately:

$$\begin{aligned}
\text{Physical:} \quad & \log(\text{COD}_i) = \underset{(0.40)}{-3.50} + \underset{(0.01)}{0.58} \times \log(\text{Output}_i) + \mathbf{X_s}\boldsymbol{\gamma}_1 + \mathbf{X_p}\boldsymbol{\gamma}_2 + \mathbf{X_o}\boldsymbol{\gamma}_3 + \varepsilon_i, \\
\text{Chemical:} \quad & \log(\text{COD}_i) = \underset{(0.49)}{-4.59} + \underset{(0.01)}{0.77} \times \log(\text{Output}_i) + \mathbf{X_s}\boldsymbol{\gamma}_1 + \mathbf{X_p}\boldsymbol{\gamma}_2 + \mathbf{X_o}\boldsymbol{\gamma}_3 + \varepsilon_i, \\
\text{Biological:} \quad & \log(\text{COD}_i) = \underset{(0.20)}{-4.37} + \underset{(0.01)}{0.67} \times \log(\text{Output}_i) + \mathbf{X_s}\boldsymbol{\gamma}_1 + \mathbf{X_p}\boldsymbol{\gamma}_2 + \mathbf{X_o}\boldsymbol{\gamma}_3 + \varepsilon_i.
\end{aligned}$$

For Regression (1) in the main text, we have also clustered standard errors at the provincial level. We confirm that the statistical significance does not change:

$$\log(\text{COD}_i) = \underset{(0.30)}{-3.75} + \underset{(0.03)}{0.63} \times \log(\text{Output}_i) + \mathbf{X_s}\boldsymbol{\gamma}_1 + \mathbf{X_p}\boldsymbol{\gamma}_2 + \mathbf{X_o}\boldsymbol{\gamma}_3 + \varepsilon_i.$$

C.2 *Locations and Ownership Rights*

The pollution intensity also exhibits rich provincial level variations. We take the regression results on the paper industry for example. Using Beijing as the benchmark, of all the remaining 29 provinces and municipalities, only the coefficient of Shanghai is negative and statistically insignificant. For the remainders, all estimates are positive and highly significant.[2] The five most polluting provinces are Ningxia, Inner Mongolia, Gansu, Shaanxi and Xinjiang, for which the rankings of their provincial GDP in 2007 are respectively 29, 16, 27, 19 and 25 out of 31 provincial administrative regions (not including Hong Kong and Macau). Geographically, these are also inland provinces. In fact, despite that the majority of the paper manufacturers are located in Zhejiang and Guangdong (ranked 1st and 4th by GDP), the pollution level is lower comparing to those inland less developed regions. These patterns are highly robust across various specifications of our regressions. The general message we get from these dummy coefficients are the following. First, Beijing and Shanghai, arguably the political and economic center of China, exert pollution control standard far more stringent than the rest provinces. Second, inland and less developed provinces in general have worse control over and are more prone to the industrial pollution problem. These two patterns are broadly consistent with the environmental Kuznetz curve and the results from previous literature. However, unlike Jiang, Lin and Lin (2014), ownership rights do not seem to have significant effects. If anything, the Foreign and Hong Kong, Macau and Taiwan based firms have only weakly better pollution controls. However, the statistical significance of the estimates never exceed 10% across all our regressions. We conjecture that this could be because Jiang, Lin and Lin (2014)'s estimates are derived using a different sample of firms (in total 1,931 firms), while our data cover more firms.

D Evidence Supporting the Modeling Assumptions

In this section, we provide some evidence in support of our modeling assumptions. In Section D.1, we present the evidence that motivates us to model the distortions as progressive taxes on output. In Section D.2, we justify our choice of modeling the adoption of advanced treatment technology as fixed costs. We

[2]One province, Hebei is significant at 10% level. Four others are significant at 1% level while all others are significant at 0.1% level. We notice here that Hebei, a province that is commonly considered as heavily affected by industrial pollution, is in fact an outlier, since the paper manufacturing is not its pillar industry.

also describe additional distributional features of the treatment technology, including the processing efficiency, designed processing capacity, and installation costs, using the Paper and Paper Product industry as an example. Section D.3 contains a brief summary of the environmental regulations in China.

D.1 *Firm-level Distortions*

Recent studies of the Chinese economy have described large distortions at both the sector and the firm levels, which negatively affect aggregate productivity and output considerably.[3] Following the seminal approach developed by Hsieh and Klenow (2009, 2014), we model and estimate firm-level distortions by using variations in the average products of capital and labor across firms.[4] More specifically, if we let τ_{z_i}, τ_{k_i}, and τ_{l_i} be the wedges that firm i faces in the product, capital, and labor markets, respectively, the profit maximization problem of firm i is:

$$\pi_i = \max_{k_i, l_i} \left\{ (1-\tau_{z_i}) z_i^{1-\gamma} (k_i^{\alpha} l_i^{1-\alpha})^{\gamma} - (1+\tau_{k_i}) R k_i - (1+\tau_{l_i}) W l_i \right\}.$$

By using the first-order conditions, the firm level average product of capital ϕ_{k_i}, labor ϕ_{l_i}, and the capital/labor ratio κ_i can be expressed as

$$\phi_{k_i} = \frac{y_i}{k_i} = \frac{(1+\tau_{k_i})R}{\alpha\gamma(1-\tau_{z_i})}, \tag{D.1}$$

$$\phi_{l_i} = \frac{y_i}{l_i} = \frac{(1+\tau_{l_i})W}{(1-\alpha)\gamma(1-\tau_{z_i})}, \tag{D.2}$$

$$\kappa_i = \frac{k_i}{l_i} = \frac{\alpha}{1-\alpha} \cdot \frac{(1+\tau_{l_i})W}{(1+\tau_{k_i})R}. \tag{D.3}$$

The above equations show that in the absence of any distortions ($\tau_{z_i} = \tau_{k_i} = \tau_{l_i} = 0$), ϕ_{k_i}, ϕ_{l_i}, and κ_i should be equalized across firms. Equations (D.1) and (D.2) state that firms that face higher distortions in the capital (labor) and/or product markets demonstrate a higher average product of capital (labor). In addition, according to Equation (D.3), the capital/labor ratio increases with the relative size of labor to the capital market wedge. By using firm-level data on total production value, the book value of capital stock, and labor compensation from the CNEC, we calculate $z_i, \phi_{k_i}, \phi_{l_i}$, and κ_i for each firm in our sample. Here we set $\gamma =$ 0.93, which is the calibrated value of returns to scale in the main text. Figure D.1 shows on log scales the scatterplots of ϕ_{k_i}, ϕ_{l_i}, and κ_i against firm-level productivity z_i for the Paper and Paper Products industry. We plot the Paper industry for demonstration purpose, qualitatively the results for all five polluting industries as well as all manufacturing industries combined are the same.

Two patterns emerge from Figure D.1. First, from the two upper panels, we see that both ϕ_{k_i} and ϕ_{l_i} are positively correlated with z_i, which suggests that more productive firms have higher average products of both capital and labor. Expressed in wedges, this means that both $(1+\tau_{k_i})/(1-\tau_{z_i})$ and $(1+\tau_{l_i})/(1-\tau_{z_i})$ are higher for more productive firms. This could be because more productive firms are subject to higher factor

[3] See Hsieh and Klenow (2009), Song, Storesletten and Zilibotti (2011), Brandt, Tombe and Zhu (2013), and Tombe and Zhu (2019), among others.

[4] Though average factor product dispersion could arise due to dynamic capital adjustment [Asker, Collard-Wexler and Loecker (2014)], for the case of China, David and Venkateswaran (2019 forthcoming) show that dynamic adjustment costs only accounts for about 10% of the variation in average factor product in the data. Naturally, we also cannot separate out variations in average factor product caused by firm-specific characteristics, for instance firm-specific markup or capital share. However, we argue that it would require empirical implausible variations of such parameters to generate the patterns of the data we observe.

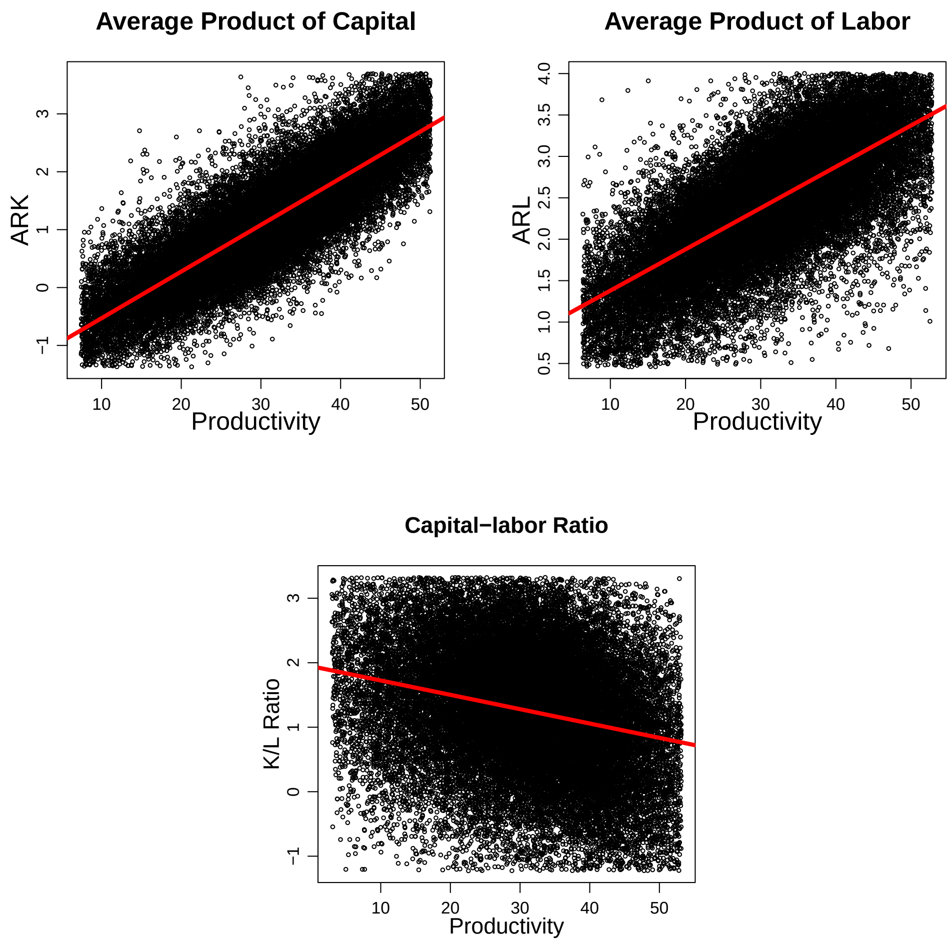

FIGURE D.1. AVERAGE FACTOR PRODUCTS AND CAPITAL-LABOR RATIO ACROSS FIRM TFP LEVELS

Source: CNEC. All panels are plot in log scale. Lines are least square fit. Data of the Paper and Paper Product Industry are plotted.

or product market distortions or both. Second, from the lower panel, we see that the capital/labor ratio is at best weakly negatively correlated with z.

Qualitatively, these findings indicate that the relative wedge firms face in the capital and labor markets does not depend strongly on the idiosyncratic productivity of firms, which in the context of our model implies $1 + \tau_{k_i} \approx 1 + \tau_{l_i}$.[5] Since we cannot separately identify the three wedges, for simplicity, we assume $\tau_{k_i} = \tau_{l_i} = 0$ and attribute all the variations in the average product of factors to the wedges in the product market τ_{z_i}. Whether we assume $\tau_{k_i} = \tau_{l_i} = 0$ or alternatively $\tau_{z_i} = 0$ does not affect our results, but the interpretations need to be changed accordingly.[6]

As a result, following Hsieh and Klenow (2014) and Bento and Restuccia (2017), we model these distor-

[5]This also explains the reason why $\phi_{k_i} \approx \phi_{l_i}, \forall i$.

[6]For example, we cannot distinguish between the data-generating process we use here and another process where τ_{k_i} and τ_{l_i} increase simultaneously, while τ_{z_i} is equal to zero.

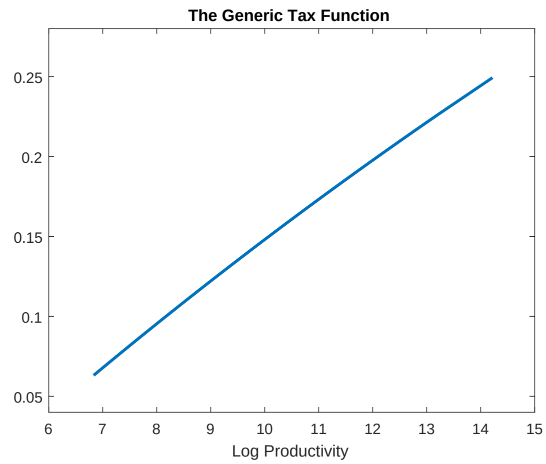

FIGURE D.2. THE OUTPUT TAX FUCTION τ_z

The actual output tax function used in the quantitative exercises, given the calibrated values for ϕ_0 and ϕ_1

tions as implicit progressive output taxes, and the tax rate function takes the form used in the main text:

$$\tau_z = \max\left\{0, 1 - \phi_0 z^{\phi_1}\right\}. \tag{D.4}$$

where the parameter ϕ_0 determines the mean level of the taxes, and the parameter ϕ_1 determines the progressiveness of the taxes. When $\phi_1 < 0$, the tax rates are increasing in firm TFP z, and the progressiveness of the taxes is decreasing in the value of ϕ_1.[7] The actual tax function used in the quantitative exercises is plotted against log productivity in Figure D.2, which is increasing and concave in z given our calibrated values for ϕ_0 and ϕ_1.

As we explained in the main text, the firm-level distortions τ_z are meant to capture a variety of policies and institutions that reallocate production factors from large productive to small unproductive firms, following the *indirect* approach in the misallocation literature [Restuccia and Rogerson (2013)]. Examples include size-dependent taxes and subsidies that favor small firms, imperfect tax enforcement that induces tax evasion among small firms, financial constraints and contractual frictions that impede the growth of productive firms [Cooley and Quadrini (2001), Clementi and Hopenhayn (2006) and Akcigit, Alp and Peters (2018)], and local protectionism and internal trade barriers that disproportionately affect large productive firms [Eberhardt, Wang and Yu (2016) and Tombe and Zhu (2019)]. While it is of great policy interest to study specific policies and institutions underlying τ_z, we leave this important task to future work.[8]

D.2 *Treatment Technologies*

In the main text, we model the cost of adopting the clean technology as a fixed cost k_E independent of the production scale of firms. In this subsection, we provide additional empirical evidence in support of this assumption using the firm-level data from the NGSPS.

[7]Notice that since ϕ_1 <0, decreasing ϕ_1 increases its absolute value.

[8]Adamopoulos and Restuccia (2014) and Wu et al. (2018) are two recent examples where observable policy distortions are investigated. Adamopoulos and Restuccia (2014) study the observable farm-level price distortions while our own work Wu et al. (2018) focuses on distortions created by migration and land distribution policies in China.

TABLE D.1—PRICES OF DIFFERENT TREATMENT EQUIPMENTS

Technologies	25%	50%	75%	Maximum	Mean
Physical	1.4	3.5	10	1000	12.4
Biological	20	50	120	3056	109.6

Source: 2007 NGSPS. All the numbers are measured in CNY 10,000.

The assumption that the cost of technology adoption is independent of firm size has two implications. First, larger firms are more likely to adopt clean technology, and as a result, we should see positive correlations between production scale with both the possibility of adoption and total expenditure in treatment technology. The linear probability model in the main text shows the first. Here we calculate the correlation between the log value of production and treatment equipment investment as well. Not surprisingly, we find that it is positive (0.64) as well.

The second and perhaps more important implication is that the ratio of treatment technology expenditure over output should decrease with the size of the firm. We find that it is also the case in the NGSPS data. For firms in the top-5 polluting industries, we calculate the average of the ratio of treatment equipment expenditure to output value within each quintile of output. The means of the ratio from the first (smallest) to the fifth (largest) quintile are 23%, 11%, 6.7%, 5.0%, and 2.7%, respectively. A linear regression model that controls for the same set of covariates as in the main text delivers the same message:

$$\log\left(\frac{\text{Investment}_i}{\text{Output}_i}\right) = \underset{(0.11)}{-0.73} - \underset{(0.004)}{0.40} \times \log(\text{Output}_i) + \mathbf{X}\boldsymbol{\gamma} + \varepsilon_i,$$

which shows a negative correlation as well. Together, the evidence supports our choice of modeling the adoption cost of clean technology as a fixed cost.

Prices of Physical versus Biological Equipment.—We assume in the main text that only the installation of biological technologies requires a fixed cost. Here we provide evidence in support of this choice. The distributions of the prices for physical and biological equipments in absolute term (in CNY 10,000) are listed in Table D.1. As is shown in the table, the prices of biological equipments are 7 to 15 times of those of physical equipments. In addition, the total installation costs of physical equipment for the top-5 polluting industries over those of biological equipment is only 0.087. These facts again point to the importance of the fixed costs associated with the adoption of the clean technology. Therefore in our model we assume that only the adoption of biological equipment is costly.[9]

Additional Technical Features of the Treatment Technologies.—We are interested in the following features of these technologies: processing efficiency, designed processing capacity and installation costs. We proxy the processing efficiency using one minus the ratio of emitted COD to generated COD. The designed processing capacity (in tons) and actual installation costs are recorded directly by the data. We then calculate the unit capacity costs by dividing the installation costs by the processing capacity. We use the unit capacity costs as an indicator of returns to scale of clean treatment equipment. In Figure D.3, in clockwise order we plot the density functions of processing capacity, installation costs, total value of industrial output and unit capacity cost by technology types. For all panels, log-scale is used in the horizontal axes.

Broadly speaking, biological technologies have the best processing efficiency, the largest processing capacity, the highest installation costs but the lowest unit capacity cost. More specifically, the mean and median

[9] In fact, what is crucial here is that the cost to install clean technology is higher than that of dirty technology.

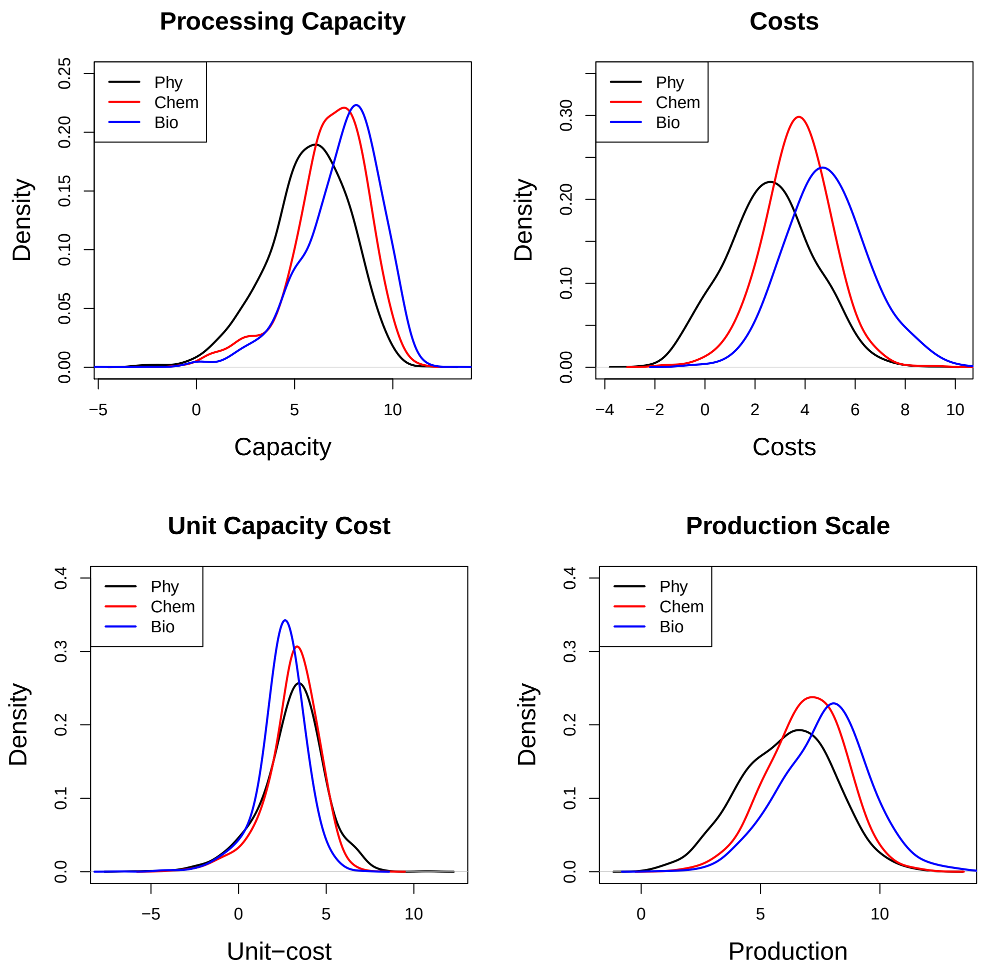

FIGURE D.3. TECHNICAL FEATURES OF DIFFERENT TREATMENT TECHNOLOGIES

Source: NGSPS. In all panels, the horizontal axes are in log-scale.

processing efficiency of biological technology are 17 and 10 percentage points higher than the physical technology. The evidence again points to a fixed costs type of mechanism behind the lower pollution intensity by large firms. That is, although biological technologies are more advanced in terms of processing capacity and efficiency, they are also more costly. Therefore, small firms lack the profit margins needed to take advantage of the returns to scale exhibited by biological technologies, and hence large firms are more likely to adopt these more advanced technologies.

D.3 *Environmental Regulations in China*

In this section, we briefly summarize the environmental regulations in China. We follow the descriptions in Lin (2013) closely here.

Environmental regulations in China are specified in *The Environmental Protection Law of the People's*

Republic of China (EPL). The EPL was launched on trial basis in 1979 and was officially enacted by the National People's Congress in 1989. It was later amended in 1993 and 2003. The Ministry of Environmental Protection (MEP, previously the State Environmental Protection Administration) is in charge of the monitoring and administration of China's environmental issues. Two major policy instruments that the MEP uses are a pollution levy system and command-and-control instruments.

The Pollution Levy System.—The pollution levy system charges firms a pollution fee based on their quantity of effluent discharge of pollutants. Before 1993, firms were required to pay for the emissions that go beyond the national standard. If a firm discharges multiple pollutants and the levels of more than one of the pollutants are above the national standard, it only has to pay for the one that exceeds the standard by the greatest amount (calculated according to the *Pollutants Equivalent Factor*). Starting from 1993, firms also have to pay for within-standard emissions. After 2003, pollution levies are calculated based on the top three above standard pollutants instead of only the top one. The 2015 revision of the EPL significantly increases the financial penalties by allowing the environmental authorities to multiply the original penalties by the number of days that firms are incompliant with the regulations specified in the law.

In practice, the MEP oversees the enforcement of the EPL and the implementation of environmental policies. Local Environmental Protection Bureaus (EPBs) are the main authorities that do the field work. In general, larger polluters are monitored by municipal EPBs while smaller ones are by district and county EPBs. All polluting firms are required to submit an annual predicted volume of emissions at the beginning of a year. Based on the actual production and abatement practice, firms are also required to revise the actual emissions during the year. The EPBs send field staff to verify the credibility of firms' reports. The inspections include consistence between emissions, output, material usage, and across historical data. Field inspectors also conduct random on-site inspections of the emissions by firms without prior warning.

Pollution levies are then calculated based on the verified reports. False reporting are subject to financial penalty with a ceiling. Serious fraudulent reports could lead to further non-financial penalties, including temporary shut down from production, administrative detention of directly responsible personnel, and even criminal charges.

The Command-and-Control Instruments.—The command-and-control instruments usually involve temporary shut down of firms incompliant with environmental emission standards, which would cost these firms a fraction of their profits. Serious violation sometimes also leads administrative detention and criminal charges similar to the case of fraudulent report of emissions. Because the pollution levy system is widely acknowledged for being ineffective in providing firms with incentives to control pollution [Dasgupta et al. (2001), Wang and Wheeler (2005) and Lin (2013)], command-and-control instruments are usually used as a deterrence for potential violation by polluting firms. The command-and-control instruments prove to be effective in many government-led campaigns against pollution. See for example, Te-Ping Chen, "China Cracks Down on Water-Polluting Industries," the Wall Street Journal, April 17th, 2015; or Trevor Nace, "China shuts down tens of thousands of factories in widespread pollution crackdown," Forbes, October 24th 2017.

E The Firm Size Distributions of Individual Industries

In the main text, we present the firm size distributions for the pooled polluting industries and the manufacturing sector as a whole. In this section, we explain in details how the firm size distributions are constructed, especially those for the U.S., and present the firm size distributions for each of the five polluting industries in China and in the U.S.

We use the 2004 CNEC and 2004 Statistics of U.S. Businesses (SUSB) data to construct the firm size

TABLE F.1—SIZE DISTRIBUTION ON POLLUTION

Methods	Paper	Agricultural Food	Textile	Chemistry	Beverage	Average
Non-parametric	39.8%	60.7%	81.6%	102.5%	103.8%	63.5%
Piecewise-linear	34.8%	69.4%	93.5%	180.2%	N/A[a]	75.4%
Parametric	43.5%	61.1%	97.5%	101.2%	89.0%	67.0%

† Note: For individual industries, the numbers reported are the pollution intensity from using the U.S. firm size distribution as percentage of that using China's distribution. We use the 1st and 3rd quartile as the ends of the output range corresponding to each employment bin in the non-parametric calculation. Column 6 (Average) calculates the weighted average of these ratios using the percentage contribution in the first row of Table 1 in the main text as weights.

[a] Since the beverage industry has fewer firms than the others, there are employment size bins with no corresponding firms in China, which invalidates the method. We set the ratio to 100% in the calculation of the last average.

distributions in China and in the U.S., respectively. The SUSB organizes data around the so-called enterprise size groups instead of firm size groups. According to the definition in SUSB, a large enterprise could consist of firms that belong to different industries. For each size bin, the SUSB reports the total number of firms, establishments and employees along with other variables summed up across all enterprises that fall in that size bin. As a result, we approximate the firm size distribution using the average firm size of a particular size group, which is calculated by dividing the total employment by the number of firms. We then assign groups of firms to different size bins according to their average size. Such imputation introduces approximation errors. To further reduce the approximation noise, we group the size bins into four main groups: 1–19, 20–99, 100–399 and 400+. The firm size distributions for each of the top polluting industries and all industries pooled together are shown in Figure E.1. As we can see, while there is substantial heterogeneity across individual industries, large firms in China account for much smaller employment shares than those in the U.S. for each industry.

F Accounting Exercises

This section provides details on the accounting exercises.

Estimation Strategies.—Ideally, we would like to have information on pollution intensity over firms' *employment* size bins. Unfortunately, the NGSPS only reports total output value and total amount of pollution at the firm level. Therefore, we need to construct pollution intensity over the number of employees. To do this, we use the CNEC to estimate a relationship between employment and output.

1. *Non-parametric*:
   - For each U.S. employment bin, we compute the 1st and 3rd quartile of output level for Chinese firms within that employment bin. The two quartiles are used as the lower and upper bounds for the output bins in NGSPS.
   - We then use the median pollution intensity of firms within the newly defined output bins as the average pollution intensity for those bins.
   - Lastly, we calculate the aggregate pollution by assigning to each bin the corresponding share of output.

2. *Piecewise Linear*:

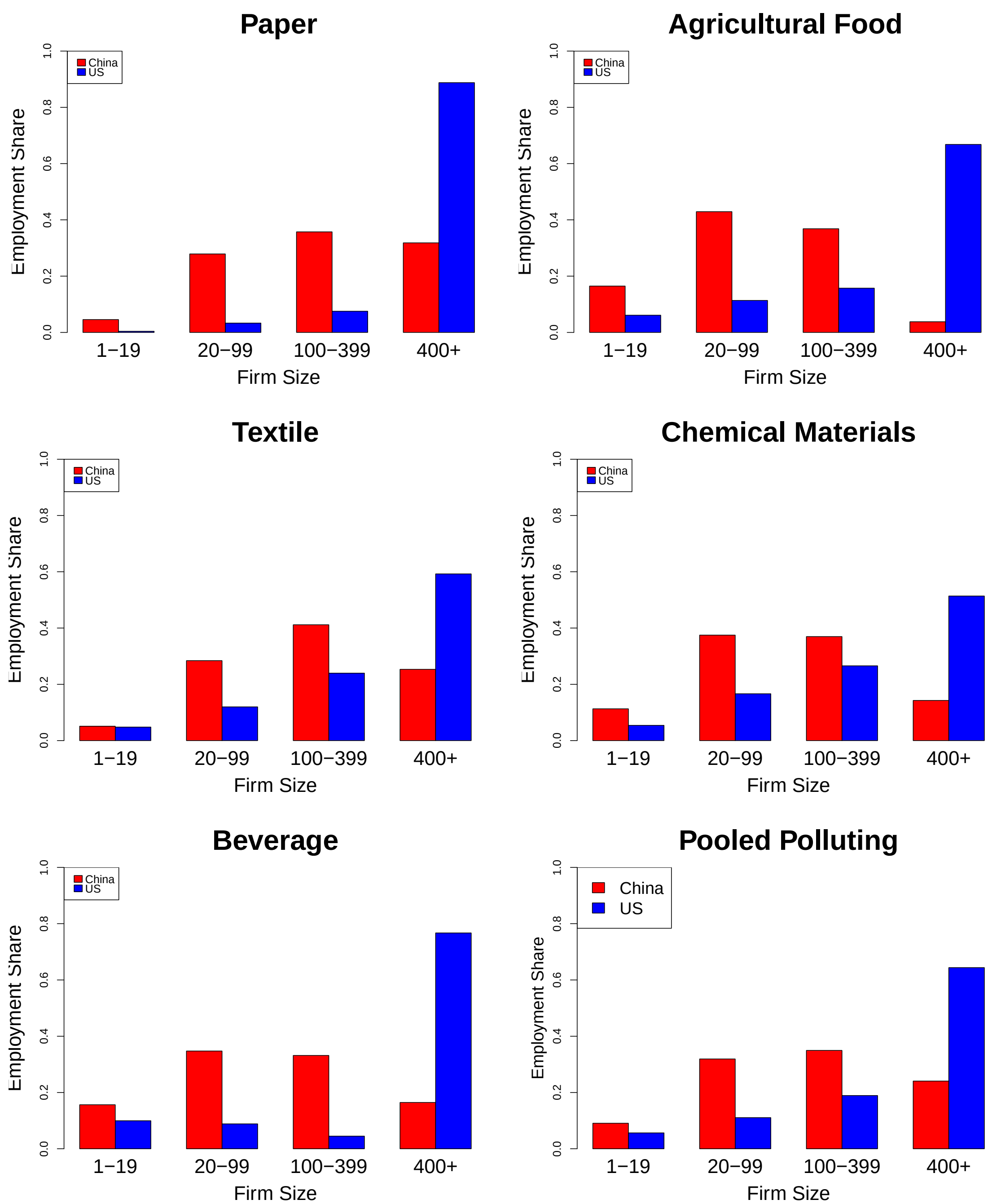

FIGURE E.1. EMPLOYMENT DISTRIBUTION

Sources: CNEC and SUSB.

- For each U.S. employment bin, we regress log-output on log-employment using the subset of Chinese firms within that employment bin. The lower and upper bounds for the output bins in this case are calculated as the predicted value of the above regression.
- We then run piecewise log-linear regression of pollution intensity on output within each new production bin. The average pollution intensity is chosen to be the predicted intensity at the midpoint of the new log-output bin.
- Lastly, the average intensity is applied to the output share distributions.

3. *Parametric*:

- Using the CNEC, we regress log-output on log-number of workers, which yields a parametric relationship between the number of workers and output.
- Using the NGSPS, we regress log-intensity on log-output, which yields a parametric relationship between intensity and output. From these two relationships, we can subsequently construct a new *parametric* relationship between intensity and number of employees. The average intensity is chosen to be the midpoint of each U.S. employment bin. Notice that in this case we have a direct functional form for employment and intensity).
- Lastly, the average intensity is applied to the output share distributions.

The estimation results are shown in Table F.1. Each of the three methods has its own advantages and disadvantages. The two non-parametric methods capture more of the variation at the local level, which could be washed out in the parametric estimation across the whole state space. However, this local nature also introduces a lot of instability on the estimates. Further, there are situations when there are gaps not covered by adjacent output bins and situations when these output bins overlap with each other. Under these conditions, some information will be lost while others are used for multiple times. Nevertheless, the results are robust across different estimation strategies. In the Section I.C of the main text, we use the results from the parametric method.

G Proofs and Additional Analytical Results

In this section, we provide formal proofs to Propositions 1 and 2 in Section II.D of the main text. The two propositions are proved respectively in Sections G.1 and G.2. We also include additional analytical results in this section. Specifically, in Section G.1, Lemma 1 and Corollary 1 highlight the key trade-off that firms face in deciding which type of treatment technology to install, when firms are and are not subject to correlated distortions respectively. Lemma 2 shows the standard results of the Lucas (1978) model where household members choose their occupations according to their comparative advantage. In Section G.2, in a simple one-sector model with a fixed number of firms and *exogenous* firm-level pollution intensity, we present the conditions under which removing the progressiveness of the distortions would reduce average pollution intensity. In Section G.3, we show the qualitative equivalence between the effects of distortions on aggregate output and pollution in the Lucas (1978) model with those in the closed-economy version of the Melitz (2003) model. Many researchers working with heterogeneous firms models are more familiar with the Melitz (2003) setup as opposed to the Lucas (1978) setting, especially in the trade and environment community. We demonstrate in this section that the differences between the two settings are immaterial to our main findings. Finally, in Section G.4, we use a simple model to study how *uncorrelated* distortions—defined as distortions that are not correlated with firm TFP—affect aggregate output and pollution.

G.1 *Correlated Distortions and Technology Adoption*

In this section, we prove Proposition 1 of the main text. To simplify notation, we set $\phi_0 = 1$ in this section.
PROOF OF LEMMA 1:

Lemma 1. *In an economy with no correlated distortions, $\pi_0(z)$ and $\pi_1(z)$ are both increasing and linear with respect to z. In addition, the slope of $\pi_1(z)$ is steeper than that of $\pi_0(z)$:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1-\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z. \tag{G.1}$$

Proof. Once the fixed cost k_E is paid, it no longer affects a firm's decision. The factor demand decisions for the two types of firms are therefore the same. The first order conditions for capital and labor are respectively

$$\frac{\partial \pi_i(z)}{\partial k}: \qquad \alpha\gamma z^{1-\gamma}k^{\alpha\gamma-1}l^{(1-\alpha)\gamma} = R, \tag{G.2}$$

$$\frac{\partial \pi_i(z)}{\partial l}: \qquad (1-\alpha)\gamma z^{1-\gamma}k^{\alpha\gamma}l^{(1-\alpha)\gamma-1} = W, \qquad i = 0, 1. \tag{G.3}$$

Dividing (G.2) with (G.3) yields constant capital to labor ratio h

$$h = \frac{k}{l} = \frac{\alpha W}{(1-\alpha)R},$$

which says more capital is demanded when technology is capital intensive (higher α) or when capital rental price R low. Notice that the system of equations (G.2) with (G.3) is log-linear and thus has closed-form solution. With some algebra, the solutions are characterized by

$$l(z) = \Phi_1 R^{\frac{\alpha\gamma}{\gamma-1}} W^{\frac{1-\alpha\gamma}{\gamma-1}} \cdot z, \tag{G.4}$$

$$k(z) = \Phi_2 R^{\frac{1+\gamma(\alpha-1)}{\gamma-1}} W^{\frac{\gamma(1-\alpha)}{\gamma-1}} \cdot z, \tag{G.5}$$

where

$$\Phi_1 = \left[\frac{(1-\alpha)^{\alpha\gamma}}{(1-\alpha)\gamma\alpha^{\alpha\gamma}}\right]^{\frac{1}{\gamma-1}} \text{ and } \Phi_2 = \frac{\alpha}{1-\alpha}\Phi_1.$$

Substitute the optimal solutions (G.4) and (G.5) back to the definition of profits functions $\pi_1(z)$ and $\pi_0(z)$, we have

$$\pi_0(z) = (1-\xi)\left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z,$$
$$\pi_1(z) = \left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z - Rk_E,$$

where

$$\Omega = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha\gamma}\Phi_1^{\gamma} \text{ and } \kappa = W^{\frac{\gamma(1-\alpha)}{\gamma-1}} R^{\frac{\alpha\gamma}{\gamma-1}}.$$

It can be seen that both functions are increasing and linear in z, and hence Equation (G.1) holds. □

PROOF OF COROLLARY 1:

Corollary 1. *Suppose that the correlated distortions are specified as* $\max\left\{0, 1 - z^{\phi_1}\right\}$ *with* $1-\gamma+\phi_1 > 0$, *then* $\pi_0(z)$ *and* $\pi_1(z)$ *are both increasing and concave with respect to* z. *In addition, the slope of* $\pi_1(z)$ *is steeper than that of* $\pi_0(z)$*:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1-\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z.$$

Proof. The proof is straightforward given Lemma 1. Substituting in the tax function, $\pi_0(z)$ and $\pi_1(z)$ now becomes

$$\pi_0(z) = (1-\xi)\left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z^{\frac{1-\gamma+\phi_1}{1-\gamma}},$$
$$\pi_1(z) = \left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z^{\frac{1-\gamma+\phi_1}{1-\gamma}} - Rk_E,$$

where Ω, Φ_1 and κ are defined as in Lemma 1.

The assumption $1-\gamma+\phi_1 > 0$ guarantees the monotonicity of the profits functions. In the terminology of Hopenhayn (2014), the assumption means that there is no rank reversal. Concavity is easily verified by taking the second order derivatives. □

PROOF OF LEMMA 2:

Lemma 2. *There exists a unique threshold* $\hat{z}$ *such that all household members with* $z \leq \hat{z}$ *choose to be workers and those with* $z \geq \hat{z}$ *become entrepreneurs. Further,* $\hat{z}$ *is pinned down by* $W = \pi(\hat{z})$

Proof. Since the overall profit function $\pi(z)$ is the upper envelope of $\pi_0(z)$ and $\pi_1(z)$, from Lemma 1 (and Corollary 1) we know that $\pi(z)$ is monotonic increasing. It is easy to verify that $\pi(0) = 0$. Therefore, as long as $0 < W < \pi(\overline{z})$, we can find a unique $\hat{z}$ such that $\pi(\hat{z}) = W$, where uniqueness follows from the monotonicity. The condition $0 < W < \pi(\overline{z})$ is guaranteed in the general equilibrium version of our model by Inada condition on the production function. □

PROOF OF PROPOSITION 1:

Proposition 1. *When* $1-\gamma+\phi_1 > 0$, *a firm chooses to install clean technology only when its productivity is higher than a threshold. Moreover, the threshold when the firm faces distortions* $\tilde{z}_f$ *is higher than that when the firm is not subject to any distortions* $\tilde{z}_n$.

Proof. Uniqueness follows from

$$\frac{\partial \pi_0(z)}{\partial z} = (1-\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z,$$

and monotonicity under both the cases with and without frictions.

We can solve for the analytical expression for $\tilde{z}_n$:

$$\tilde{z}_n = \frac{Rk_E}{\xi\left(\Omega\kappa^{\gamma} - \frac{1}{1-\alpha}\kappa\right)}.$$

Using expressions of the profits functions with correlated distortions and recall that $\phi_1 < 0$, we can show that

$$\tilde{z}_n = \frac{Rk_E}{\xi\left(\Omega\kappa^\gamma - \frac{1}{1-\alpha}\kappa\right)} < \left[\frac{Rk_E}{\xi\left(\Omega\kappa^\gamma - \frac{1}{1-\alpha}\kappa\right)}\right]^{\frac{1-\gamma}{1-\gamma+\phi_1}} = \tilde{z}_f,$$

which proves the proposition.

One caveat is that the second inequality holds only if the number in the parentheses is greater than 1. We verify this in our quantitative analysis but restrain ourselves from discussing extreme cases where the condition is not hold. □

G.2 *Correlated Distortions and Aggregate Pollution*

In this section, we present a simple one-sector model with a fixed number of firms and *exogenous* firm-level pollution intensity, which is based on Lucas (1978). We then derive the conditions under which removing the progressiveness of the distortions would reduce aggregate pollution, which are summarized in Proposition 2 of the main text. We further show that similar results hold for average pollution intensity in Corollary 2. We start with a formal description of the model setting.

The Model.—The model contains only the polluting sector. There is a representative household with a continuum of members. A measure one of the members are managers, while a measure N are workers. The occupation of household members are fixed once born. Each manager is endowed with z units of managerial talent, which is drawn from a log-normal distribution $G(z)$ with logarithmic mean and standard deviation (μ, σ). Workers supply one unit of labor inelastically in exchange for wage income, and managers run a neoclassical firms and earn profits. In the simplified economy, labor is the only production factor. We set the final good as the numeraire, and use W to represent wage rate.

The production function is given by

$$y = F(z, l) = z^{1-\gamma} l^\gamma,$$

where as usual, $0 < \gamma < 1$ is the span-of-control parameter that supports a non-degenerate distribution of firms in the equilibrium.

The production process generates pollutants e as by-products. Here we abstract from the treatment technology choice decision. Firm-level emissions now depend only on production scale y:

$$e(y) = \psi_0 y^{1+\psi_1},$$

where $-1 < \psi_1 < 1$ captures the negative correlation between output level and emission intensity as documented in Section I.B of the main text.

Firm-level distortions take the same form as in the main text:

$$\tau_z = 1 - \phi_0 z^{\phi_1},$$

where $\phi_1 < 0$ reflects the progressiveness of the distortions.

Proofs.—With the above setup, we prove the following results.

PROOF OF PROPOSITION 2:

Proposition 2. *In the simple model, there exists a threshold*

$$\hat{\phi}_1 = \frac{\psi_1(1-\gamma)}{1-\gamma(1+\psi_1)},$$

such that E is increasing in ϕ_1 when $\phi_1 < \hat{\phi}_1$, and decreasing in ϕ_1 when $\phi_1 \geq \hat{\phi}_1$.

Proof. Our goal here is to find a relationship between E and ϕ_1. With no capital, the maximization problem of the firm is

$$\pi(z) = \max_{l} \left\{(1-\tau_z) z^{1-\gamma} l^{\gamma} - Wl\right\}.$$

The first order condition is

$$l(z) = \left(\frac{\gamma\phi_0 z^{1-\gamma+\phi_1}}{W}\right)^{\frac{1}{1-\gamma}}. \tag{G.6}$$

The optimal labor demand leads immediately to output

$$y(z) = \left(\frac{\gamma\phi_0}{W}\right)^{\frac{\gamma}{1-\gamma}} z^{\frac{1-\gamma+\gamma\phi_1}{1-\gamma}}. \tag{G.7}$$

We have assumed that

$$e(y) = \psi_0 y^{1+\psi_1}.$$

Combining this equation with Equation (G.7), we get the pollution of firm-z

$$e(z) = \psi_0 \left(\frac{\gamma\phi_0}{W}\right)^{\frac{\gamma(1+\psi_1)}{1-\gamma}} z^{\left(1+\frac{\gamma\phi_1}{1-\gamma}\right)(1+\psi_1)}. \tag{G.8}$$

With Equation (G.8), all that remains is to compute the integral

$$\begin{aligned} E &= \int e(z) dG(z) \\ &= \psi_0 \left(\frac{\gamma\phi_0}{W}\right)^{\frac{\gamma(1+\psi_1)}{1-\gamma}} \int z^{\left(1+\frac{\gamma\phi_1}{1-\gamma}\right)(1+\psi_1)} dG(z). \end{aligned} \tag{G.9}$$

In Equation (G.9), we need to solve for closed-form expression for W and the integral.

Recall that if $z \sim \text{Lognormal}(\mu, \sigma^2)$, $z^a \sim \text{Lognormal}(a\mu, a^2\sigma^2)$ for $a \neq 0$. Hence

$$\int z^{\left(1+\frac{\gamma\phi_1}{1-\gamma}\right)(1+\psi_1)} dG(z) = \exp\left\{\left(1+\frac{\gamma\phi_1}{1-\gamma}\right)(1+\psi_1)\mu + \frac{1}{2}\left(1+\frac{\gamma\phi_1}{1-\gamma}\right)^2 (1+\psi_1)^2\sigma^2)\right\}. \tag{G.10}$$

We use the labor market clearing condition

$$\int l(z) dG(z) = N,$$

to solve for wage W. In particular, by substituting Equation (G.6) in the labor market clearing condition, and using the property of log-normal distribution, we get

$$N = \left(\frac{\gamma\phi_0}{W}\right)^{\frac{1}{1-\gamma}} \exp\left\{\left(1+\frac{\phi_1}{1-\gamma}\right)\mu + \frac{1}{2}\left(1+\frac{\phi_1}{1-\gamma}\right)^2 \sigma^2\right\}. \tag{G.11}$$

Substitute Equations (G.10) and (G.11) back to Equation (G.9), after some quite involved algebra, aggregate pollution can be written as

$$E = \Omega \exp\left\{ \frac{(1+\psi_1)\gamma}{1-\gamma}\frac{\sigma^2}{2}\left[2\psi_1\phi_1 + \frac{(1+\psi_1)\gamma - 1}{1-\gamma}\phi_1^2\right]\right\}, \tag{G.12}$$

where

$$\Omega = \psi_0 N^{\gamma(1+\psi_1)} \exp\left\{(1+\psi_1)\left[(1-\gamma+\psi_1)\frac{1}{2}\sigma^2 + (1-\gamma)\mu\right]\right\},$$

is a constant independent of ϕ_1. Notice that inside the bracket in Equation (G.12) is a parabola opening to the bottom. Importantly, the axis of the parabola is

$$\phi_1 = \frac{\psi_1(1-\gamma)}{1-\gamma(1+\psi_1)} \triangleq \hat{\phi}_1. \tag{G.13}$$

Proposition 2 thus follows immediately. □

Notice that by Equation (G.7),

$$Y = \int y(z)dG(z) = \int \left(\frac{\gamma\phi_0}{W}\right)^{\frac{\gamma}{1-\gamma}} z^{\frac{1-\gamma+\gamma\phi_1}{1-\gamma}} dG(z).$$

Using again the labor market clearing condition (G.11) and the property of log-normal distribution, we have

$$Y = N^{\gamma} \exp\left\{(1-\gamma)\left(\mu + \frac{\sigma^2}{2}\right)\right\} \exp\left\{-\frac{\gamma}{1-\gamma}\frac{\sigma^2}{2}\phi_1^2\right\}. \tag{G.14}$$

We see from Equation (G.14) that Y is always decreasing in ϕ_1. This echoes the finding in the literature that correlated distortions lead to losses in output by misallocation.

PROOF OF COROLLARY 2:

Corollary 2. *In the simple model, the relationship between the average pollution intensity E/Y and the progressiveness of the distortions ϕ_1 is such that*

(i) when $\psi_1 < (1-2\gamma)/\gamma$, the average pollution intensity is always increasing in ϕ_1;

(ii) when $\psi_1 \geq (1-2\gamma)/\gamma$, there exists a threshold

$$\overline{\phi}_1 = \frac{(1+\psi_1)(1-\gamma)}{1-2\gamma+\gamma\psi_1},$$

such that when $\phi_1 < \overline{\phi}_1$, E/Y is increasing in ϕ_1; and when $\phi_1 \geq \overline{\phi}_1$, E/Y is decreasing in ϕ_1;

(iii) $\overline{\phi}_1$ is increasing in both ψ_1 and γ;

(iv) $\overline{\phi}_1 < \hat{\phi}_1$.

Proof. The proof is almost the same as that of Proposition 2. By Equations (G.12) and (G.14), and again some pretty involved algebra, we can write the average pollution intensity as

$$\frac{E}{Y} = \Gamma \exp\left\{\frac{\gamma}{1-\gamma}\frac{\sigma^2}{2}\left[2\psi_1(1+\psi_1)\phi_1 + \frac{\psi_1(2\gamma-1+\psi_1\gamma)}{1-\gamma}\phi_1^2\right]\right\}, \tag{G.15}$$

where like before, Γ is a constant independent of ϕ_1.

Notice that once again, inside the bracket in Equation (G.15) is a parabola. When $\psi_1 < (1-2\gamma)/\gamma$, its opening is to the top, and hence the intensity is always increasing in distortions ϕ_1. On the other hand, when $\psi_1 \geq (1-2\gamma)/\gamma$, the parabola is opening to the bottom with the axis being

$$\phi_1 = \frac{(1+\psi_1)(1-\gamma)}{1-2\gamma+\gamma\psi_1} \triangleq \overline{\phi}_1, \tag{G.16}$$

which gives us items (ii) and (iii) of Corollary 2 immediately. Straight forward comparison of Equations (G.13) and (G.16) yield item (iv). □

Remark.—Notice that, if we assume alternatively that firm-level pollution emission is a function of its productivity z:

$$\frac{e}{y} = \psi_0 z^{\psi_1},$$

instead of as that in Equation (G.8), then Equations (G.12) and (G.15) become

$$E = \Omega \exp\left\{\frac{\gamma}{1-\gamma}\frac{\sigma^2}{2}\left(2\psi_1\phi_1 - \phi_1^2\right)\right\},$$

and

$$\frac{E}{Y} = \Gamma \exp\left\{\frac{\gamma\psi_1\sigma^2}{1-\gamma}\phi_1\right\},$$

respectively.

The two equations imply that first, the average pollution intensity is always increasing in distortions; and second, aggregate pollution is increasing in distortions, as long as the elasticity of distortions to productivity $|\phi_1|$ is larger than that of pollution $|\psi_1|$. Intuitively, this simply says that when output is more responsive to firm size, it dominates the effect of firm-level pollution intensity. Put differently, the scale effect dominates the technique effect.

G.3 *Monopolistic Competition*

In this section, we show that our results carry through to the setting of monopolistic competition, or a closed-economy version of Melitz (2003). The crucial elements of our analysis are the following. First, in the equilibrium, there needs to be a non-degenerated distribution of firms with different levels of productivity. Second, the model should feature resource allocation along the extensive margin and intensive margin, where we refer to the selection of firms into production as the extensive margin, and the distribution of factors among active firms as the intensive margin. Mathematically, what we really need is decreasing returns to scale in profits which prevents the most productive firm from taking over all the resources in the economy and profits be a function in productivity which creates selection. The curvature could come from decreasing returns to scale in production as in Lucas (1978), or from the demand and market structure as in Melitz (2003). We need curvature from one source or the other, but *not* necessarily from both.[10] To ease the exposition, we first show the equivalence using only the backbone elements with no distortions or emissions.

[10] In fact, Hopenhayn (1992) has directly assumed that the profit function is increasing in productivity [Assumption A.2.(b)] without specifying the underlying mechanism, and Bartelsman, Haltiwanger and Scarpetta (2013) have assumed diminishing returns in production and utility coexist (page 319).

We then show that a similar version of Proposition 2 holds under the monopolistic competition structure as well.

Diminishing Returns to Scale and to Utility.—The model differs from the Lucas (1978) span-of-control model in three aspects: the production function is linear, the utility function is a CES aggregation of a continuum of varieties, and the market structure is monopolistic competition. Specifically, there is still a representative household with a continuum of members. The representative household derives utility from consuming a continuum variety of goods ω aggregated according to a CES function:

$$U = \left[\int_{\omega\in\Omega} q(\omega)^{\frac{\rho-1}{\rho}} d\omega\right]^{\frac{\rho}{\rho-1}},$$

where Ω is the set of available varieties, $q(\omega)$ is the quantity consumed for each variety ω, and $\rho > 1$ is the elasticity of substitution.

Each household member is once again endowed with z units of managerial talent, where $z \sim G(z)$. Household members choose between running a firm or working as a worker. Labor is the only production factor, which the firm hires at wage W to produce a unique variety. Each variety is produced by only one firm. The production function of the firm is given as

$$y = F(z, l) = zl. \tag{G.17}$$

Because firms now have market power, they can set the price of good $p(\omega)$ as well. Since z and ω can both uniquely identify a firm in our setting, from now on, we will use z to refer to different varieties as well. Therefore, price is now written as $p(z)$, and $y = q(z) = zl$. Likewise, the set Ω is the same set of active firms, which anticipating the results later, is interval $[\hat{z}, \overline{z}]$ where as before $\hat{z}$ is the threshold of selection into production.

The profit maximization problem of the firm now becomes

$$\max_{\{p(z),q(z)\}} \{p(z)q(z) - Wl\}. \tag{G.18}$$

To solve (G.18), we need to solve for the demand function for each variety z by the household. Standard results from Dixit and Stiglitz (1977) lead to

$$q(z) = Q\left[\frac{p(z)}{P}\right]^{-\rho},$$

where

$$P = \left[\int_{\hat{z}}^{\overline{z}} p(z)^{1-\rho} dz\right]^{\frac{1}{1-\rho}},$$

and

$$Q = \left[\int_{\hat{z}}^{\overline{z}} q(z)^{\frac{\rho-1}{\rho}} dz\right]^{\frac{\rho}{\rho-1}},$$

are the price and quantity of the composite good, respectively.

If we set the composite good to be the numeraire, i.e., $P = 1$, then

$$q(z) = Q[p(z)]^{-\rho}. \tag{G.19}$$

Notice that the production function (G.17) means that $l = q(z)/z$. With Equation (G.19), firm's problem (G.18) is equivalent to

$$\max_{p(z)} Q\left[p(z)^{1-\rho} - W\frac{p(z)^{-\rho}}{z}\right],$$

where the first order condition directly leads to optimal labor demand

$$l(z) = Q\left(\frac{W\rho}{\rho-1}\right)^{-\rho} z^{\rho-1}, \tag{G.20}$$

and profit

$$\pi(z) = \frac{Q}{\rho}\left(\frac{\rho-1}{\rho}\right)^{\rho-1} z^{\rho-1}. \tag{G.21}$$

We see that both (G.20) and (G.21) are increasing in z. This means that with a non-degenerate distribution of z, l and π are also non-degenerate. Further, Equation (G.20) implies that if l is log-normally distributed, z will also be log-normally distributed, leaving us with a similar calibration strategy for $G(z)$ as in the main text. In addition, because (G.21) implies that $\pi(0) = 0$, the selection threshold $\hat{z}$ is also determined by

$$\pi(\hat{z}) = W. \tag{G.22}$$

Recall that in the Lucas (1978) model, the counterparts of (G.20) and (G.21) are respectively

$$l(z) = \left(\frac{\gamma}{W}\right)^{\frac{1}{1-\gamma}} z,$$

and

$$\pi(z) = \left[\left(\frac{\gamma}{W}\right)^{\frac{\gamma}{1-\gamma}} - W\left(\frac{\gamma}{W}\right)^{\frac{1}{1-\gamma}}\right] z,$$

and Equation (G.22) stays the same, it can be seen that the two models are equivalent, with the elasticity of substitution ρ in the monopolistic setting plays the same role as the span-of-control parameter γ. In fact, in the Lucas (1978) model, one can think of firms as earning rents over the managerial talent z.

Monopolistic Competition and Aggregate Pollution.—With the above notation, we now bring back the distortions

$$\tau_z = 1 - \phi_0 z^{\phi_1},$$

and further assume that firm level emissions are given by

$$e(z) = \psi_0[p(z)q(z)]^{1+\psi_1}.$$

We shut down the selection margin $\hat{z}$ as in Section G.2 by assuming that a measure one of the household members are born as managers and a measure N of them are born as workers. Following the same steps in the proofs of Proposition 2, we can show that the aggregate pollution in the monopolistic competition case can be shown as

$$E = \Gamma_1 \exp\left\{-\frac{\sigma^2}{2}[2\rho - \psi_1(\rho-1)^2]\phi_1^2 + \left[\mu - 2[\rho - \psi_1(\rho-1)^2]\right]\right\},$$

where Γ_1 is a constant. The expression indicates that there again exists a threshold

$$\hat{\phi}_1 = -1 + \frac{\rho - \mu/\sigma^2}{2\rho - \psi_1(\rho-1)^2},$$

such that E is decreasing in ϕ_1 when $\phi_1 > \hat{\phi}_1$, and increasing in ϕ_1 when $\phi_1 < \hat{\phi}_1$.

Similarly, the aggregate output in the monopolistic competition case is

$$Q = N \exp\left\{\mu + \frac{\sigma^2}{2}(\rho - 1) - \frac{\sigma^2}{2}\rho\phi_1^2\right\},$$

which is decreasing in the progressiveness of distortions $-\phi_1$. The impact of distortions on aggregate output in this case is equivalent to that in the Lucas span-of-control case of Section G.2 [Equation (G.14)].

G.4 *Uncorrelated Distortions*

In the main text, we focus on correlated distortions, which increase in firm TFP. However, as can be seen from Figure D.1, there are also substantial variations in average products across firms with the same TFP level.[11] In the language of Restuccia and Rogerson (2008), these variations suggest the existence of *uncorrelated* distortions, which are uncorrelated with firm TFP. In this section, we use the same model in Section G.2, except that the firm-level distortions are uncorrelated with firms' productivity z, to study the impact of uncorrelated distortions on aggregate output and pollution. We find that unlike the case of correlated distortions, uncorrelated distortions lower both aggregate output and aggregate pollution.

The uncorrelated distortions are defined as an output tax whose rate τ is randomly drawn and independent of firm productivity z. Specifically, z is drawn from a log-normal distribution $G(z)$ with log mean and standard deviation (μ_z, σ_z), and $1 - \tau$ is independently drawn from another log-normal distribution $H(\tau)$ with log mean and standard deviation (μ_τ, σ_τ).

The firm's profit maximization problem can be written as

$$\pi(z, \tau) = \max_l \left\{(1 - \tau)z^{1-\gamma}l^\gamma - Wl\right\}.$$

First-order conditions yield

$$\text{(G.23)} \qquad \begin{aligned} l(z,\tau) &= \left[\frac{\gamma(1-\tau)}{W}\right]^{\frac{1}{1-\gamma}} z, \\ y(z,\tau) &= \left[\frac{\gamma(1-\tau)}{W}\right]^{\frac{\gamma}{1-\gamma}} z, \\ e(z,\tau) &= \psi_0 \left[\frac{\gamma(1-\tau)}{W}\right]^{\frac{\gamma(1+\psi_1)}{1-\gamma}} z^{1+\psi_1}. \end{aligned}$$

The labor market clearing condition

$$\iint l(z,\tau)dG(z)dH(\tau) = N,$$

leads to

$$\text{(G.24)} \qquad N = \left(\frac{\gamma}{W}\right)^{\frac{1}{1-\gamma}} \exp\left\{\frac{\mu_\tau}{1-\gamma} + \frac{\sigma_\tau^2}{2(1-\gamma)^2}\right\} \exp\left\{\mu_z + \frac{\sigma_z^2}{2}\right\}.$$

[11] We thank a referee for pointing this out.

Combining Equations (G.23) and (G.24), we can compute the aggregate output for this economy

$$\begin{aligned} Y &= \iint y(z,\tau)dG(z)dH(\tau) \\ &= N^{\gamma} \exp\left\{(1-\gamma)\left(\mu_z + \frac{\sigma_z^2}{2}\right)\right\} \exp\left\{-\frac{\gamma\sigma_\tau^2}{2(1-\gamma)}\right\}. \end{aligned}$$

Consistent with the results in Hsieh and Klenow (2009), aggregate output Y is decreasing in the dispersion of firm-level distortions σ_τ^2.

Similarly, the aggregate emissions can be written as

$$E = \Omega(\psi_0, \psi_1, \mu_z, \sigma_z, \mu_\tau, N) \exp\left\{\frac{[\gamma(1+\psi_1)]^2 - \gamma(1+\psi_1)}{2(1-\gamma)^2}\sigma_\tau^2\right\},$$

where $\Omega(\psi_0, \psi_1, \mu_z, \sigma_z, \mu_\tau, N)$ is a constant independent of the dispersion of distortions. Since $\gamma < 1$ and $\psi_1 < 0$, E is also decreasing in the dispersion of firm-level distortions σ_τ^2.

H Additional Discussions on the Environmental Economics Literature

In this section, we explain in greater details the differences between our paper and the literature on how international trade affects the environment through the reallocation across heterogeneous firms. Specifically, our paper differs from this literature in following ways. First, we bring our general equilibrium model to firm-level data directly, and provide a quantitative assessment of counterfactual policies using an internally-consistent model disciplined by firm-level data. Second, using the firm-level data on treatment technologies, we provide an explanation for the negative correlation between firm size and pollution intensity. Our model also has different implications for the impact of environmental regulations, due to the different nature of the distortions in our paper and the trade costs emphasized by the literature.

Quantitative General Equilibrium.—Most studies in this literature focus on reduced-form analysis, in which the model is used to motivate the empirical analysis indirectly. See for example, Martin (2013), Cherniwchan (2017), Forslid, Okubo and Ulltveit-Moe (2018), and Barrows and Ollivier (2018), among others. Some studies, such as Andersen (2016), Holladay (2016) and Forslid, Okubo and Ulltveit-Moe (2018) derive qualitative predictions under strong parametric assumptions, but do not bring their models to the data directly. Instead, we provide a quantitative assessment of counterfactual policies using an internally-consistent model disciplined by firm-level data. Our quantitative analysis explicitly takes into account the general equilibrium effects of alternative policies, and it provides novel policy implications that are absent from either reduced-form or partial equilibrium analysis. The quantitative nature makes our analysis suitable for evaluating alternative policies. First, it allows us to quantify the trade-off between the technique and scale effects, that is, whether a reduction in pollution intensity couples with an increase in output would lead to an increase in aggregate pollution. This is important for our paper because the removal of distortions necessarily leads to increase in output and decrease in pollution intensity. Second, while the reduced-form analysis is able to identify how individual firm responds to changes in economic environment [for instance Cherniwchan (2017)], the aggregation of these responses would result in general equilibrium feedback that is more appropriately studied using a quantitative general equilibrium framework.

Technology Adoption.—A unique feature of our data is that we observe both the emissions and treatment technologies at the firm level. Using the data on treatment technologies, we provide an explanation for the negative correlation between firm size and pollution intensity. It allows us to not only investigate the overall effects of distortions, but also quantify an important channel through which the distortions affect emissions, namely, the adoption of clean technology. This differentiates our paper from empirical work such as Andersen (2016) and Holladay (2016). Specifically, Andersen (2016) studies the effects of credit constraints on pollution using regional data, which provides only indirect evidence on the firm behavior. While Holladay (2016) shows that on average exporters tend to be cleaner, but does not provide an explanation for this fact. One exception is Barrows and Ollivier (2018), which uses product-level data to study how trade liberalization affects pollution through the shift in the product-mix of exporting firms. Our paper differs from theirs by our focus on a completely different channel.

Implications for Environmental Regulations.—The implication of our model for environmental regulations distinguishes us from studies that emphasize the role of trade costs, most of which build their models on the Melitz (2003) model. A major difference between the distortions in our model and the trade costs in an open-economy Melitz model is that exporting is an endogenous choice of firms which brings extra revenues.[12] A productive firm can choose not to pay the trade costs if it decides to stay domestic, while the distortions in our model are imperative and are purely deadweight losses. In addition, they interact with environmental regulations in different ways. In our model, environmental regulations would worsen resource allocation along the intensive margin, but this is not necessarily the case with trade costs, because the productive exporting firms can use foreign markets as a cushion to buffer the burdens from environmental regulations after paying the trade costs. We believe that such differences are policy-relevant, because in many cases environmental authorities have to make policy decisions in the presence of distortions.

I Computational Algorithm for the Main Model

This section contains the computational algorithm we use to solve the version of the model in the main text, namely the two-sector model where products from the polluting and non-polluting sectors are perfect substitutes. The model is fairly standard to solve. In the stationary equilibrium, the Euler equation of the household implies that the net interest rate is

$$R = \frac{1}{\beta} - 1 + \delta. \tag{I.1}$$

As a result, once we find the equilibrium wage W, it is straightforward to compute the other allocations in the equilibrium. Specifically, the pseudo-code goes as follows.

1. Calculate the equilibrium interest rate by Equation (I.1).
2. Find the equilibrium wage W. This includes the following steps.
   (a) Make an initial guess of W.
   (b) Given R and W, for each firm-z in both the polluting and non-polluting sectors, we can solve for their profit functions $\pi^d(z)$ and $\pi^c(z)$. Specifically, the first order conditions of firms in the non-polluting sector are given by

$$\begin{aligned}(k):&\qquad \alpha\gamma(1-\tau_z)z^{1-\gamma}k^{\alpha\gamma-1}l^{(1-\alpha)\gamma} = R,\\ (l):&\qquad (1-\alpha)\gamma(1-\tau_z)z^{1-\gamma}k^{\alpha\gamma}l^{(1-\alpha)\gamma-1} = W.\end{aligned}$$

[12] In Section G.3, we establish the qualitative equivalence between the effects of distortions on aggregate output and pollution in Lucas (1978), and those in the *closed-economy* version of Melitz (2003).

With the optimal factor demand $k(z)$ and $l(z)$, the profit function of firms in the non-polluting sector is given by

$$\pi^c(z) = (1-\tau_z)z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - Rk(z).$$

(c) Notice that the first order conditions of firms in the polluting sector are similar, hence we can at the same time solve the potential profits of polluting firms using dirty and clean technologies $\pi_0^d(z)$ and $\pi_1^d(z)$:

$$\begin{aligned}\pi_0^d(z) &= (1-\xi)\{(1-\tau_z)z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - Rk(z)\},\\ \pi_1^d(z) &= (1-\tau_z)z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - R[k(z)+k_E].\end{aligned}$$

(d) Now $\pi_0^d(z), \pi_1^d(z)$ and $\pi^c(z)$ can be used to pin down the thresholds for occupational choices $\{\hat{z}_c, \hat{z}_d\}$ and technology adoption $\tilde{z}_d$. In particular,

$$\pi^d(z) = \max\{\pi_0^d(z), \pi_1^d(z)\},$$

and $\hat{z}^j$s are such that

$$W = \pi^j(\hat{z}_j), \qquad j = c, d.$$

When $\pi_0^d(z)$ and $\pi_1^d(z)$ satisfy single-crossing condition,

$$\tilde{z}_d = \min_z\{z \in [\hat{z}_d, \overline{z}] | \pi_0^d(z) < \pi_1^d(z)\}.$$

Notice that now for the polluting sector, $\hat{z}_d$ and $\tilde{z}_d$ define two groups of firms: those that use dirty technology $z \in [\hat{z}_d, \tilde{z}_d]$, and those that use clean technology $z \in [\tilde{z}_d, \overline{z}]$.

(e) If the labor market clearing condition

$$\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c) = \mu \int_{\hat{z}_d}^{\overline{z}} l^d(z)dG(z) + (1-\mu)\int_{\hat{z}_c}^{\overline{z}} l^c(z)dG(z),$$

does not hold, return to Step 2.(a) and update the guess for W.

3. With the equilibrium prices W and R, all the allocations of the economy can be computed following Definition 1 in the main text.

By Walras' Law, if the household budget constraint, capital, and labor market clearing conditions all hold, resource constraint is automatically satisfied. We now verify that this is the case. To simplify notation, first let us define

$$\overline{\pi}(z) = (1-\tau_z)z^{1-\gamma}(k^\alpha l^{1-\gamma})^\gamma - Wl - Rk,$$

and

$$y(z) = z^{1-\gamma}(k^\alpha l^{1-\alpha})^\gamma,$$

where the optimal factor demands are implicitly substituted in. Government tax revenue T comes from two sources here: from taxing firms using dirty technology

$$\mu\xi \int_{\hat{z}_d}^{\tilde{z}_d} \overline{\pi}(z)dG(z),$$

and from output taxes τ_z

$$\mu \int_{\hat{z}_d}^{\overline{z}} \tau_z y(z) dG(z) + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} \tau_z y(z) dG(z).$$

Recall that we have assumed that all taxes are rebated to the household as lump-sum transfers. Hence with these notations, household income I is given by

$$\begin{aligned}
I &= RK + W[\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c)] + T \\
&\quad + \mu \left[\int_{\hat{z}_d}^{\tilde{z}_d} \pi_0^d(z) dG(z) + \int_{\tilde{z}_d}^{\overline{z}} \pi_1^d(z) dG(z) \right] + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} \pi^c(z) dG(z) \\
&= RK + W[\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c)] + T + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} \overline{\pi}(z) dG(z) \\
&\quad + \mu \left[(1-\xi) \int_{\hat{z}_d}^{\tilde{z}_d} \overline{\pi}(z) dG(z) + \int_{\tilde{z}_d}^{\overline{z}} \overline{\pi}(z) dG(z) - Rk_E[G(\overline{z}) - G(\tilde{z})] \right] \\
&= RK + W[\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c)] + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} [\overline{\pi}(z) + \tau_z y(z)] dG(z) \qquad \text{(I.2)} \\
&\quad + \mu \left[\int_{\hat{z}_d}^{\overline{z}} (\overline{\pi}(z) + \tau_z y(z)) dG(z) - Rk_E[G(\overline{z}) - G(\tilde{z})] \right].
\end{aligned}$$

Because

$$K = \mu \left[\int_{\hat{z}_d}^{\overline{z}} k^d(z) dG(z) + k_E[G(\overline{z}) - G(\tilde{z}_d)] \right] + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} k^c(z) dG(z),$$

and

$$\overline{\pi}(z) = (1-\tau_z) y(z) - Wl(z) - Rk(z),$$

substitute these two equations and the labor market clearing condition into Equation (I.2), we get

$$I = \mu \int_{\hat{z}_d}^{\overline{z}} y(z) dG(z) + (1-\mu) \int_{\hat{z}_c}^{\overline{z}} y(z) dG(z).$$

Household budget constraint then leads directly to

$$C + \delta K = Y. \qquad \text{(I.3)}$$

Equation (I.3) is used to construct the aggregate consumption.

J A Two-Sector Model with CES Aggregation

In the main text, we have assumed that the goods produced by polluting and non-polluting sectors are perfect substitutes. Admittedly an extreme assumption, it is not critical to our quantitative results in the main text, because the counterfactual experiments there cause little reallocation across sectors. However, if we want to evaluate the effects of removing distortions from one sector only, this assumption is problematic because it would imply a reallocation across sectors that is too large. To address this issue, we build a two-sector model in which the final goods are produced by a representative producer, who combines the intermediate goods from the polluting and non-polluting sectors using a constant elasticity of substitution (CES) production function. We use the model to evaluate the effects of removing distortions from both sectors, and those of removing distortions from the polluting sector only. In this section, we first describe the model, and then explain the computational algorithm. In the end, we calibrate the model and report the quantitative results from the counterfactual experiments.

J.1 *The Model*

We assume that there is a single final good Y produced by a representative firm in a perfectly competitive final good market. The final good producer manufactures Y by combining the intermediate goods from the polluting sector Y_d and non-polluting sector Y_c respectively. The production function features *constant elasticity of substitution (CES)* between the two goods:

$$Y = \left[\varphi(Y_d)^{\frac{\rho-1}{\rho}} + (1-\varphi)(Y_c)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}},$$

where ρ is the elasticity of substitution, φ is the average share of polluting goods in production. We assume that the final good Y is the numeraire, and the relative prices of polluting and non-polluting goods are p_d and p_c, respectively.[13] The optimization problem of the final good producer is thus

$$\max_{Y_d, Y_c}\left\{\left[\varphi(Y_d)^{\frac{\rho-1}{\rho}} + (1-\varphi)(Y_c)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} - p_d Y_d - p_c Y_c\right\}.$$

The first order conditions yield

$$\frac{Y_d}{Y_c} = \left(\frac{\varphi}{1-\varphi}\cdot\frac{p_c}{p_d}\right)^{\rho}.$$

Since we pick the final good Y as the numeraire, the aggregate price index is

$$\left[\varphi^{\rho}(p_d)^{1-\rho} + (1-\varphi)^{\rho}(p_c)^{1-\rho}\right]^{\frac{1}{1-\rho}} = 1, \tag{J.1}$$

which implicitly defines a mapping between p_d and p_c.

Because now there are relative prices, the optimization problems for firms in the two intermediate goods sectors need to be modified accordingly. In particular, the profit maximization problem for firms in the non-polluting sector is

$$\pi^c(z) = \max_{k,l}\left\{(1-\tau_z)p_c z^{1-\gamma}(k^{\alpha}l^{1-\alpha})^{\gamma} - Wl - Rk\right\},$$

while those for firms in the polluting sector using clean and dirty technologies are given respectively as

$$\pi_1^d(z) = \max_{k,l}\left\{(1-\tau_z)p_d z^{1-\gamma}(k^{\alpha}l^{1-\alpha})^{\gamma} - Wl - R(k+k_E)\right\},$$

and

$$\pi_0^d(z) = \max_{k,l}\left\{(1-\xi)\left[(1-\tau_z)p_d z^{1-\gamma}(k^{\alpha}l^{1-\alpha})^{\gamma} - Wl - Rk\right]\right\}.$$

As before, the profit function of firms in the polluting sector is the envelope of $\pi_0^d(z)$ and $\pi_1^d(z)$:

$$\pi^d(z) = \max\left\{\pi_0^d(z), \pi_1^d(z)\right\}.$$

The rest elements of the model are identical to those in the main text.

The stationary equilibrium with CES aggregation is defined as follows.

[13] Notice that this setup is mathematically equivalent to the monopolistic competition setup in Section G.3.

Definition J.1. *A stationary equilibrium of this economy consists of prices* $\{W, R, p_d, p_c\}$*, allocations* $\{C, K, Y_d, Y_c, Y\}$*, firm's policy functions* $\{k^j(z), l^j(z), y^j(z), \pi^j(z)\}$*,* $j = c, d$*, thresholds for household members' occupational choices* $\{\hat{z}_c, \hat{z}_d\}$ *and polluting firms' technology adoption* $\tilde{z}_d$*, as well as aggregate pollution* E*, such that:*

(i) Given prices $\{W, R, p_d, p_c\}$*,* $\{C, K, \hat{z}_c, \hat{z}_d\}$ *solve the household's optimization problem;*

(ii) Given prices $\{p_d, p_c\}$*,* $\{Y_d, Y_c\}$ *solve the final good producer's optimization problem;*

(iii) Given prices $\{W, R, p_d, p_c\}$*,* $\{k^j(z), l^j(z), y^j(z), \pi^j(z)\}$*,* $j = c, d$*, and* $\tilde{z}_d$ *solve the intermediate goods producers' optimization problems;*

(iv) Prices $\{W, R, p_d, p_c\}$ *clear all markets:*

- *Labor Market:*

$$\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c) = \mu \int_{\hat{z}_d}^{\overline{z}} l^d(z)dG(z) + (1-\mu)\int_{\hat{z}_c}^{\overline{z}} l^c(z)dG(z),$$

- *Capital Market:*

$$K = \mu\left[\int_{\hat{z}_d}^{\overline{z}} k^d(z)dG(z) + k_E[G(\overline{z}) - G(\tilde{z}_d)]\right] + (1-\mu)\int_{\hat{z}_c}^{\overline{z}} k^c(z)dG(z),$$

- *Intermediate goods produced by polluting and non-polluting sectors:*

$$Y_d = \mu\int_{\hat{z}_d}^{\overline{z}} y^d(z)dG(z), \qquad Y_c = (1-\mu)\int_{\hat{z}_c}^{\overline{z}} y^c(z)dG(z),$$

- *Final Good:*

$$C + K - (1-\delta)K = Y,$$

where

$$Y = \left[\varphi(Y_d)^{\frac{\rho-1}{\rho}} + (1-\varphi)(Y_c)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}},$$

(v) Aggregate Pollution:

$$E = \mu\left[\int_{\hat{z}_d}^{\tilde{z}^d} e\left(0, y^d(z)\right)dG(z) + \int_{\tilde{z}_d}^{\overline{z}} e\left(1, y^d(z)\right)dG(z)\right].$$

J.2 *Computational Algorithm*

The stationary equilibrium can be computed in a similar way as depicted in Appendix I. The only difference here is that instead of solving for the equilibrium wage W only, we need to also find the prices for the intermediate goods, p_c and p_d. With the equilibrium prices, the allocations in Definition J.1 can be calculated accordingly. Recall that Equation (J.1) defines a 1-1 mapping between p_c and p_d, hence in practice we need to solve for W and p_d. The pseudo-code is as follows.

1. Calculate the equilibrium interest rate again by Equation (I.1).

2. Solve for the wage W and the price of the polluting sector intermediate good p_d in the equilibrium. This includes the following steps.

   (a) Make initial guesses for W and p_d, using Equation (J.1) to back out p_c:

   $$p_c = \left[\frac{1-\varphi^\rho (p_d)^{1-\rho}}{(1-\varphi)^\rho}\right]^{\frac{1}{1-\rho}}.$$

   (b) Given R, W, p_d and p_c, for each firm-z in both the polluting and non-polluting sectors, we can solve for the profit functions $\pi^c(z)$ and $\pi^d(z)$. Specifically, the first order conditions of firms in the non-polluting sector are given by:

   $$\begin{aligned} (k): &\qquad \alpha\gamma(1-\tau_z)p_c z^{1-\gamma}k^{\alpha\gamma-1}l^{(1-\alpha)\gamma} = R, \\ (l): &\qquad (1-\alpha)\gamma(1-\tau_z)p_c z^{1-\gamma}k^{\alpha\gamma}l^{(1-\alpha)\gamma-1} = W. \end{aligned}$$

   With the optimal factor demand $k(z)$ and $l(z)$, the profit function of firms in the non-polluting sector is given by

   $$\pi^c(z) = (1-\tau_z)p_c z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - Rk(z).$$

   (c) Notice that the first order conditions of firms in the polluting sector are similar, hence we can at the same time solve the potential profits of polluting firms using dirty and clean technologies $\pi_0^d(z)$ and $\pi_1^d(z)$:

   $$\begin{aligned} \pi_0^d(z) &= (1-\xi)\{(1-\tau_z)p_d z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - Rk(z)\}, \\ \pi_1^d(z) &= (1-\tau_z)p_d z^{1-\gamma}[k(z)^\alpha l(z)^{1-\alpha}]^\gamma - Wl(z) - R[k(z)+k_E]. \end{aligned}$$

   (d) Now $\pi_0^d(z), \pi_1^d(z)$ and $\pi^c(z)$ can be used to pin down the thresholds for occupational choices $\{\hat{z}_c, \hat{z}_d\}$ and technology adoption $\tilde{z}_d$. In particular,

   $$\pi^d(z) = \max\{\pi_0^d(z), \pi_1^d(z)\},$$

   and $\hat{z}^j$s are such that

   $$W = \pi^j(\hat{z}_j), \qquad j = c, d.$$

   When $\pi_0^d(z)$ and $\pi_1^d(z)$ satisfy single-crossing condition,

   $$\tilde{z}_d = \min_z\{z \in [\hat{z}_d, \overline{z}] | \pi_0^d(z) < \pi_1^d(z)\}.$$

   Notice that now for the polluting sector, $\hat{z}_d$ and $\tilde{z}_d$ define two groups of firms: those that use dirty technology $z \in [\hat{z}_d, \tilde{z}_d]$, and those that use clean technology $z \in [\tilde{z}_d, \overline{z}]$.

   (e) If the labor market clearing condition

   $$\mu G(\hat{z}_d) + (1-\mu)G(\hat{z}_c) = \mu\int_{\hat{z}_d}^{\overline{z}} l^d(z)dG(z) + (1-\mu)\int_{\hat{z}_c}^{\overline{z}} l^c(z)dG(z),$$

   and the first order condition of the final good producer

   $$\frac{\mu}{1-\mu}\cdot\frac{\int_{\hat{z}_d}^{\overline{z}} y^d(z)dG(z)}{\int_{\hat{z}_c}^{\overline{z}} y^c(z)dG(z)} = \left(\frac{\varphi}{1-\varphi}\cdot\frac{p_c}{p_d}\right)^\rho,$$

   do not hold, return to Step 2.(a) and update the guesses for W and p_d.

3. With the equilibrium prices W, R, p_d, p_c, all allocations of the economy can be computed following Definition J.1.

The resource constraint can be verified in the same way as in Appendix I, and hence is omitted here.

J.3 *Quantitative Results*

We use the model to evaluate the effects of removing distortions from one sector only, and study how cross-sector reallocation affects our results. Specifically, we compare the results from two experiments. In Experiment (i), we repeat Experiment (i) in the main text by setting $\tau_z = 0$ for both sectors. This experiment serves two purposes. First, by comparing the results with those from the main text, we show that the alternative assumption of finite and constant elasticity of substitution between the two sectors would not change our main results materially. Therefore, the assumption that the products from the two sectors are perfect substitutes is innocuous for the main results of this paper. Second, it provides a benchmark for us to compare the effects of symmetric versus asymmetric changes in distortions. In Experiment (ii), we remove the distortions from the polluting sector only, and evaluate the effects of cross-sector reallocation on our results.

The quantitative exercises in this section are for illustration purpose mainly. Therefore, we only carry out a simple calibration of the model. We keep most of the parameters intact as in Section III of the main text. There are two new parameters that we need to pin down, namely the elasticity of substitution ρ and the share of the products of the polluting sector in the production of the final good φ. Empirical estimates of ρ range from 3 to 10 for narrowly defined industries [Broda and Weinstein (2006) and Hendel and Nevo (2006)]. Hsieh and Klenow (2009) use $\rho = 3$ for 4-digit industries. Because we divide the whole economy into two broad sectors in our model, we begin with a smaller value of $\rho = 1.5$ and later consider $\rho = 3$ to gauge how the change in ρ affects our results.[14] Given ρ, we choose φ such that the revenues from the polluting sector $p_d Y_d$ is 20% of those from both sectors, $p_d Y_d + p_c Y_c$. In addition, due to the presence of relative price, we adjust the adoption cost of clean technology so that the share of firms using clean technology is again 57%.

Table J.1 presents the results. In the upper panels, we report the results for $\rho = 1.5$ and in the lower panels $\rho = 3$. For experiment (i), it is reassuring to see that for both levels of elasticity of substitution, the results are quantitatively similar to the perfect substitutes case in the main text. Following the elimination of τ_z from both sectors, aggregate output in both sectors increases by about 30%. Again, there is little reallocation across sectors, and the changes in the prices of the products in both sectors are negligible. The average pollution intensity decreases by about 42%, so again the technique effect dominates, and the aggregate pollution decreases by about 25%.

For Experiment (ii), we start with the case $\rho = 1.5$. Following the elimination of τ_z in the polluting sector, there again are a large increase in output and a large decrease in average pollution in the polluting sector. However, there is also a sizable reallocation of production factors from the non-polluting sector to the polluting sector. Specifically, since the equilibrium wage increases due to the higher labor demand from the polluting sector, firms in the non-polluting sector have to use less labor and capital. This causes a 2% decrease in the physical output of the non-polluting sector. The physical output of the polluting sector increases by 35% in this case compared to 30% in Experiment (i). This composition effect partially offsets the technique effect, and the aggregate pollution decreases by 20% compared to 25% in Experiment (i), although the decreases in average pollution intensity are similar in these two cases (41% versus 42%).

When $\rho = 3$, the cross-sector reallocation of factors would be much larger given a higher elasticity of

[14] Intuitively, the elasticity of substitution is larger across more narrowly defined industries, since it is easier to substitute apples with oranges than with smartphones.

TABLE J.1—THE EFFECTS OF REMOVING DISTORTIONS FROM POLLUTING SECTOR ONLY V.S. FROM BOTH SECTORS

	Polluting			Non-polluting		
Statistics	Benchmark	(i)	(ii)	Benchmark	(i)	(ii)
CES = 1.5						
Physical Output	100.00	129.64	134.63	100.00	129.90	98.03
Price	100.00	99.97	85.00	100.00	100.00	102.34
Revenue	100.00	129.60	114.42	100.00	129.90	100.32
# of Firms	100.00	42.63	46.44	100.00	41.88	95.38
Mean Size	64.31	152.21	176.17	51.16	123.50	49.38
Pollution	100	74.69	80.05			
Intensity	100	57.81	59.47			
Clean Share	56.18	83.80	73.10			
CES = 3.0						
Physical Output	100.00	129.83	160.51	100.00	129.89	89.45
Price	100.00	100.02	87.38	100.00	100.00	100.46
Revenue	100.00	129.85	140.25	100.00	129.88	89.86
# of Firms	100.00	41.50	48.19	100.00	42.30	89.24
Mean Size	68.36	166.29	214.24	50.16	119.84	46.09
Pollution	100	74.69	89.61			
Intensity	100	57.53	55.81			
Clean Share	56.48	84.04	74.37			

† Note: All of the values are percentages except for mean size, which is the numbers of workers.

substitution. The physical output of the polluting sector increases by 60% instead of 35% in this case, and while physical output of the non-polluting sector decreases by 11% instead of 2%. As a result, although the average pollution intensity decreases by 44%, the aggregate pollution in this case decreases by only 10% as opposed to 20% when $\rho = 1.5$. If we increase ρ further, it is possible that at some point the the cross-sector reallocation would be so strong that removing the distortions from the polluting sector would increase the aggregate pollution. However, as we mentioned above, $\rho = 3$ is likely to be upper bound of the range for empirically plausible elasticity of substitution in our model, given that we divide the whole economy into two broad sectors. Hence, our results suggest that even if we remove the distortions from the polluting sector only, the technique effect would still dominate within the empirically plausible range of ρ.

Interestingly, if we compare Experiment (ii) with (i)—recall that the only difference between these two is that we also remove distortions from the non-polluting sector in Experiment (i)—we find that removing distortions from the non-polluting sector has positive spillovers to the polluting sector, which increases the adoption rate of clean technology by about 10 percentage points. This is due to the demand effect that pulls up the price of the polluting goods, allowing many medium sized firms to earn enough profits to install clean technology. This result suggests that even the reduction of distortions in the non-polluting sector alone could potentially have positive environmental implications.

REFERENCES

Adamopoulos, Tasso, and Diego Restuccia. 2014. "The Size Distribution of Farms and International Productivity Difference." *American Economic Review*, 104(6): 1667–1697. 13

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2018. "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries." *NBER Working Paper No. 21905*. 13

Andersen, Dana C. 2016. "Credit Constraints, Technology Upgrading, and the Environment." *Journal of the Association of Environmental and Resource Economists*, 3(2): 283–319. 29, 30

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." *Journal of Political Economy*, 122(5): 1013–1063. 11

Barrows, Geoffrey, and Hélène Ollivier. 2018. "Cleaner Firms or Cleaner Products? How Product Mix Shapes Emission Intensity from Manufacturing." *Journal of Environmental Economics and Management*, 88: 134–158. 29, 30

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review*, 103(1): 305–334. 25

Bento, Pedro, and Diego Restuccia. 2017. "Misallocation, Establishment Size, and Productivity." *American Economic Journal: Macroeconomics*, 9(3): 267–303. 12

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. 2013. "Factor Market Distrotions across Time, Space and Sectors in China." *Review of Economic Dynamics*, 16(1): 39–58. 11

Broda, Christian, and David E. Weinstein. 2006. "Globalization And The Gains From Variety." *Quarterly Journal of Economics*, 121(2): 541–585. 36

Cherniwchan, Jevan. 2017. "Trade Liberalization and the Environment: Evidence from NAFTA and U.S. Manufacturing." *Journal of International Economics*, 105: 130–149. 29

Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *The Quarterly Journal of Economics*, 121(1): 229–265. 13

Cooley, Thomas F., and Vincenzo Quadrini. 2001. "Financial Markets and Firm Dynamics." *American Economic Review*, 91(5): 1286–1310. 13

Dasgupta, Susmita, Benoit Laplante, Nlandu Mamingi, and Hua Wang. 2001. "Inspections, Pollution Prices and Environmental Performance: Evidence from China." *Ecological Economics*, 36(3): 487–498. 16

David, Joel M., and Venky Venkateswaran. 2019 forthcoming. "The Sources of Capital Misallocation." *American Economic Review*. 11

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308. 26

Eberhardt, Markus, Zheng Wang, and Zhihong Yu. 2016. "From One to Many Central Plans: Drug Advertising Inspections and Intra-National Protectionism in China." *Journal of Comparative Economics*, 44(3): 608–622. 13

Forslid, Rikard, Toshihiro Okubo, and Karen Helene Ulltveit-Moe. 2018. "Why Are Firms that Export Cleaner? International Trade, Abatement, and Environmental Emissions." *Journal of Environmental Economics and Management*, 91: 166–183. 29

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637–1673. 36

Holladay, J. Scott. 2016. "Exporters and the Environment." *Canadian Journal of Economics*, 49(1): 147–172. 29, 30

Hopenhayn, Hugo A. 1992. "Entry, Exit, and firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–1150. 25

Hopenhayn, Hugo A. 2014. "On the Measure of Distortions." *NBER Working Paper No. 20404*. 21

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448. 11, 29, 36

Hsieh, Chang-Tai, and Peter J. Klenow. 2014. "The Life-cycle of Plants in India and Mexico." *The Quarterly Journal of Economics*, 129(3): 1035–1084. 11, 12

Jiang, Liangliang, Chen Lin, and Ping Lin. 2014. "The Determinants of Pollution Levels: Firm-level Evidence from Chinese Manufacturing." *Journal of Comparative Economics*, 42(1): 118–142. 10

Lin, Liguo. 2013. "Enforcement of Pollution Levies in China." *Journal of Public Economics*, 98: 32–43. 15, 16

Lucas, Robert E. Jr. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523. 19, 22, 25, 26, 27, 30

Martin, Leslie A. 2013. "Energy Efficiency Gains from Trade: Greenhouse Gas Emission and India's Manufacturing Firms." *Unpublished*. 29

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity."

Econometrica, 71(6): 1695–1725. 19, 25, 30

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics*, 11(4): 707–720. 28

Restuccia, Diego, and Richard Rogerson. 2013. "Misallocation and Productivity." *Review of Economic Dynamics*, 16(1): 1–10. 13

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review*, 101(1): 196–233. 11

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, Migration and Productivity: A Quantitative Analysis of China." *American Economic Review*, 109(5): 1843–1872. 11, 13

Wang, Hua, and David Wheeler. 2005. "Financial Incentives and Endogenous Enforcement in China's Pollution Levy System." *Journal of Environmental Economics and Management*, 49(1): 174–196. 16

Wu, Yiyun, Xican Xi, Xin Tang, Deming Luo, Baojing Gu, Shu Kee Lam, Peter Vitousek, and Deli Chen. 2018. "Policy Distortions, Farm Size, and the Overuse of Agricultural Chemicals in China." *Proceedings of the National Academy of Sciences of the United States of America*, 115(27): 7010–7015. 13