The Effects of Exports on Facility Environmental Performance: Evidence from a Matching Approach*

Jingbo Cui

jbcui@iastate.edu

Department of Economics

Iowa State University

Hang Qian

hqi@iastate.edu

Department of Economics

Iowa State University

This version: May 22, 2013

Selected Paper prepared for presentation at the Agricultural & Applied Economics Associations's 2013 AAET & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

*We gratefully acknowledge support from GianCarlo Moschini for providing access to the National Establishment Time Series database.

Abstract

This paper employs matching techniques to investigate the effects of facility export status on environmental performance. Using facility-level criteria air emission data in the U.S. manufacturing industry, we find the industry-specific effects of export status on emission intensity, measured by emissions per value of sale. In some industries, there is consistent and robust evidence supporting the superior environmental performance of exporters relative to non-exporters in terms of emission intensity for all criteria air pollutants tracked in the paper. In other industries, we find evidence that exporters appear to have higher emission intensity than non-exporters for some pollutants but not all.

Keywords: criteria air emissions; exports; propensity score matching
JEL Classification: F18, Q56

1 Introduction

As public concerns over global warming, industrial pollution, and trade liberalization are gradually rising, economists have been long engaged in examining the environmental consequences of international trade. The empirical literature in this area, using aggregate-level (e.g., country-level) data, has provided mixed results over the past two decades. With the emergence of longitudinal micro-level data, much of the attention in the trade community has been recently directed towards understanding the firms' heterogeneity across export status. Along this line, a few studies seek to explore the firm-level relationship between export orientation and environmental performance.

In this paper, we examine the environmental effect of firms' export decisions, using evidence from polluting facilities in the U.S. manufacturing industry. To relax the widely assumed parametric assumption about the relationship between the outcome variable (e.g., environmental measure) and covariates (e.g., export status), this paper turns to a semi-parametric approach: the propensity score matching (PSM).[1] Exporting polluters are matched with similar non-exporting ones within the same industry in terms of their conditional likelihood of exporting, namely the propensity scores. To remove state-specific confounding unobservables and time trend that may have affected facility environmental behavior, we further restrict the matched pairs from the same U.S. state and the same year. The purpose of this paper is to shed light on the industrial heterogeneity of the environmental impacts of facilities' exporting decisions.

To this end, we compile a unique facility-level dataset in the U.S. manufacturing industry in years 2002, 2005, and 2008. The data include four types of facility-level criteria air emissions, i.e., Sulfur Dioxide (SO_2), Carbon Monoxide (CO), Ozone (O_3), and Total Suspended Particulates (TSPs). In addition, we have data regarding facilities' social-economic

[1] The PSM technique has been extensively used in identifying the causal effects of exports on firm size and productivity growth (Wagner, 2002; Girma, Greenaway, and Kneller, 2003, 2004; Loecker, 2007). In addition, List et al. (2003) employ this technique to identify the effects of environmental regulation on manufacturing plant birth.

characteristics and their exposure to environmental regulation. The latter is measured by pollutant-specific county nonattainment designation under the Clean Air Act Amendments (CAAA).

We obtain several interesting results. First, we find strong evidence that exporting status has statistically significant effects on emission intensity. Moreover, our empirical results show the industrial heterogeneity. In some industries, there is consistent and robust evidence supporting that exporters are superior in environmental performance relative to non-exporters for all four tracked criteria air pollutants. For example, within the industry of Chemical and Allied Products exporters have lower emissions per value of sales than their competing counterparts by roughly 42% of SO_2, 40% of CO, 28% of O_3, and 34% of TSPs. In other industries, however, there is evidence that exporters perform even worse than non-exporters for some, but not all, pollutants. For instance, in the industry of Printing and Publishing exporters pollute 52% more of CO per value of sales than non-exporters.

The paper contributes to the growing empirical literature that uses country variation panel data to explore the environmental effect of trade. Pioneering in this study, Copeland and Taylor (1994, 1995) theoretically decompose the environmental impact of trade liberalization into the scale, technique, and composition effects. The scale effect measures the increase in emissions due to the scale up of economy. The technique effect refers to lower pollution as a result of the improvement in pollution abatement technologies. The composition effect explains the mixed results of changing shares of dirty good on pollution. With this theoretical guide, a number of empirical studies document conflicting evidence on the environmental impacts of trade at country level. Specifically, Antweler, Copeland, and Taylor (2001) empirically investigate the aforementioned three decomposed effects. A potential weakness of their work is the endogeneity problem that trade may be determined simultaneously with income and environmental outcomes. To circumvent this shortcoming, Frankel and Rose (2005) employ exogenous geographic determinants of trade as instrumental variables. Using cross-country data, they find that trade appears to have beneficial effects on some measures of environmental quality, e.g., SO_2, though not all. There is little evidence that trade has detrimental effects on the environment. In line with Frankel and Rose (2005), Managi, Hibiki, and Tsurumi (2009) revisit this question with a larger and more globally representative sample including developing countries. A recent paper by McAusland and Millimet (2012) studies the environmental effects of international and intranational trade. Using trade data between U.S. states and Canadian provinces in 1997 and 2002, this study finds robust evidence that international trade intensity lowers toxic release, while intranational trade has harmful impacts on the environment. Unlike the above existing studies, we focus on understanding the consequence of exporting decisions on the environmental performance at facility level.

This paper is closely related to the literature exploring the role of exporters in environmental activity. Using plant-level data from different countries and various measures of environmental performance, some parallel studies seek to identify whether or not exporters are environmentally friendlier than non-exporters. Relative to non-exporters, exporters are found to be more likely to denote their innovation as having beneficial environmental effects in U.K. (Girma, Hanley, and Tintelnot, 2008), to have lower fuel per sale in Ireland (Batrakova and Davies, 2012), to emit less CO_2 constructed from fuel consumption data in Sweden conditional on size (Forslid, Okubo, and Ulltveit-Moe, 2011), and to release less toxic pollutants in U.S. controlling for sales (Holladay, 2010). Another recent paper by Cui, Lapan, and Moschini (2012) develops an intuitive model to explain the firm-level correlation among productivity, export decision, and environmental pollution. Productive firms are likely to select to export, while the most productive exporters are more likely to adopt environmentally friendly technology. Hence, exporters might behave better in the environmental performance than non-exporters. Using criteria air pollution data in the U.S. manufacturing industry, they find robust evidence documenting the negative correlation between the estimated total factor productivity and emission intensity, measured by pollution per sale, and the negative correlation between exporting status and emission intensity. We revisit the hypothesis of the firm-level environmental effects of export decisions with the same data but a different empirical approach, i.e., the matching method. Furthermore, we explore the industrial heterogeneity. Specifically, for each industry determined by the two-digit SIC code, we match exporters with similar non-exporters within the same industry, state and year.

The remaining paper is organized as follows. Section 2 introduces the empirical methodology. Section 3 discusses the data sources and provides summary statistics of the data. Empirical results together with robustness checks are presented in section 4. Section 5 concludes the paper.

2 Empirical Methodology

The matching technique is grounded in the potential outcomes framework developed by Rubin (1974). Let $D_i \in \{0, 1\}$ be the treatment variable of whether facility i enters the export market, and Y_{i1}, Y_{i0} be the potential logarithmic emission intensities under its exporting and non-exporting status, respectively. The observed emission intensity is given by $Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0}$. We are interested in the counterfactual question: Does an exporting facility pollute less on average than were it a non-exporter? Formally, our goal is to identify the average treatment effect on the treated (ATET), defined as $E(Y_{i1} - Y_{i0} | D_i = 1)$.

The PSM relies on the ignorablity assumption such that exposure to treatment is independent to potential outcomes conditional on a set of covariates. For our problem, we control the following variables: (i) a facility's labor productivity measured by value of sales per employment; (ii) distance to port as a proxy of trade variable cost; (iii) dummies of pollutant-specific county nonattainment designations reflecting the facility's exposure to environmental regulations; (iv) facility characteristic dummies indicating whether the facility is a subsidiary or not and whether it is public or private company; (v) year dummies controlling for time trend; and (vi) two-digit SIC capturing the industry-specific effects. The identification assumption is that conditional on these covariates, a polluter's exporting decision is independent to the potential exporter's and non-exporter's pollutant emissions.

Rosenbaum and Rubin (1983) show that the igorability assumption implies independence between the treatment and potential outcomes conditional on the propensity score, $e\left(\boldsymbol{X}_{it-1}\right) \equiv P\left(D_{it}=1 \left| \boldsymbol{X}_{it-1}\right.\right)$, where $\boldsymbol{X}_{it-1}$ is the covariate set. The true functional form of the propensity score is unknown but testable, since treatment must be independent to covariates conditional on the propensity score, which is known as the balancing condition (test). Regardless of pollutant types, we regress the binary decision of exports in the current year on the aforementioned covariates in the one-year lag while assuming that the propensity score takes a Logit form, that is, $e\left(\boldsymbol{X}_{it-1}\right) = \Lambda\left[g\left(\boldsymbol{X}_{it-1}\right)\right]$, where $\Lambda\left(\cdot\right)$ is the Logistic c.d.f. and $g\left(\boldsymbol{X}_{it-1}\right)$ is some polynomial function of $\boldsymbol{X}_{it-1}$. We first attempt $g\left(\boldsymbol{X}_{it-1}\right)$ in its linear form. In case of violation of the balancing condition, less parsimonious forms involving higher order terms are experimented until the balancing condition cannot be rejected.[2] Once an acceptable form of the propensity score is found, for each criteria air pollutant, the impact of exporting on emission intensity can be investigated by any well-developed matching estimators.

Since it is counterintuitive to match facilities from different industries, the Logit regressions and balancing tests are conducted using the subsample defined by each two-digit SIC industry. Matching also overemphasizes industry as the key covariate.[3] For each exporter, we collect a pool of non-exporters in the same industry whose propensity scores are similar to the exporter. Their difference in the emission intensity reflects the causal impact of the treatment variable. We then use three different matching estimators, i.e., the nearest neighbor matching, radius matching, and kernel matching. These alternative estimators differ in how the neighborhood of a treated unit is defined and how the weights are constructed in the averaging of the untreated pool. None of them can dominate others, while their joint consideration provides a robust assessment of the ATET estimates.

[2] See Becker and Ichino (2002) for detailed discussions.

[3] For other applications of matching based on both the propensity score and overemphasized covariates, see, among others Heckman et al. (1997), Lechner (2002), and Loecker (2007).

3 The Data

We compile a noval facility-level dataset in the U.S. manufacturing industry in 2002, 2005, and 2008. A facility is defined as a place where economic activities generate air emissions. The dataset is assembled from a variety of sources. The National Emission Inventory (NEI) database in the U.S. Environmental Protection Agency (EPA) reports facility-level criteria air pollutants for all areas of the United States.[4] The data acquired in this paper include SO_2, CO, O_3, and TSPs. The measure of O_3 is the sum of Volatile Organic Compounds (VOCs) and Oxide of Nitrogen (NO_x), as these two pollutants involve with the formation of ground-level O_3. We define TSPs as the sum of primary Particulate Matter-10 (PM-10) and primary Particulate Matter-2.5 (PM-2.5).[5]

The facility-level economic characteristics are retrieved from the National Establishment Time Series (NETS) Database.[6] The NETS database, developed through a joint venture with Dun and Bradstreet by Walls and Associates, is a unique and national wide business establishment database covering over 300 fields and 40 million unique establishments for every year since 1990. The data used in this study include the number of employees, value of sales, export indicator, subsidiary indicator, public or private firm indicator, Data Universal Number System (DUNS) number, geographic location (i.e., latitude and longitude), five-digit Federal Information Processing Standard (FIPS) county code, and two-digit SIC code. These two facility-level databases are matched through the DUNS number, which is a unique business establishment identifier assigned by Dun and Bradstreet. A detailed algorithm of matching the NEI database with the NETS data is provided in the appendix.

[4]Some major caveats of this NEI data are summarized in the appendix, for example, duplicated emission data in 2005.

[5]According to the EPA technical document, emission data for filterable and condensable components of particulate matter are incomplete through sample years, hence are not suggested to use in any aggregate level.

[6]The NETS data have been used to study issues related to job creations and destructions, business relocation, and business ownership (Kolko and Neumark, 2008, 2010; Neumark, Wall, and Zhang, 2011). Neumark, Wall, and Zhang (2011) provide a detailed description of the NETS and an assessment of the quality of the NETS database along many dimensions. One dimension related to our study is the estimated data versus actual data regarding employment size. In our study, this problem is not critical, because about 90% of employment data have indicators suggesting the actual data.

To measure polluting facilities' environmental pressure, we further augment the merged facility-level dataset with pollutant-specific county environmental regulations under the CAAA legislation. In general, polluting facilities located in nonattainment counties are subject to more stringent environmental regulations than those in attainment ones. Consequently, we adopt this county nonattainment designation as a proxy for a facility's exposure to environmental regulation. The regulatory county status information is obtained from the Green Book Nonattainment Areas for Criteria Pollutants reported by the EPA. For each of four criteria air pollutants, i.e., CO, SO_2, O_3, and TSPs, the Green Book indicates whether only the part of a county or the whole county is in nonattainment. In accordance with the Green Book, we assign a county to the nonattainment category for each pollutant if the whole or part of the county is designated as nonattainment status. For the case of O_3, a county is assigned as nonattainment if it is in nonattainment for NO_2 and/or O_3. The latter includes 1-hour and 8-hour standards. For TSPs, we classify a county as TSPs-specific nonattainment if it is nonattainment for PM-10 and/or PM-2.5.

Finally, we look for a proxy of trade cost variables as one of key factors in determining a facility's decision to export. One proxy of facility-specific trade variable cost is the geographic distance of the facility to its nearest U.S. ports.[7] This geographical distance measures the costs associated with transporting products from the manufacturing sites to the port of shipment. The World Port Source online database provides geographic location (i.e., latitude and longitude) of a total of 548 U.S. ports including harbor, river port, seaport, off-shore terminal, and pier, jetty or wharf. For each polluting facility in the merged dataset, we compute the distance to its nearest port among all 548 U.S. ports based on the "Haversine" formula, given the latitude and longitude of two points.[8]

We are interested in industries that are heavy emitters of criteria air pollutants for

[7]According to IHS Global Services, U.S. seaborne trade with the rest of the world accounts for 78.05% by volume (millions of metric tons) in 2008.

[8]The "Haversine" formula calculates the great-circle distance between two points, that is, the shortest distance over the earth's surface.

the following two main reasons.[9] First, these industries account for more than 80% of manufacturing sector-wide criteria air emissions. In the meanwhile, manufacturers in these dirty industries have been actively participating in the export market. Consequently, the environmental performance of polluters in dirty industries is likely to be sensitive to international trade. Second, each dirty industry in the merged dataset has a relatively large number of observations. Hence, for treated units (i.e., exporters), we may find out control units (i.e., non-exporters) to match with. Table 1 presents a list of dirty industries together with the number of exporters and non-exporters.

3.1 Descriptive Statistics

An unbalanced panel dataset of 29,183 facility-by-year observations is analyzed. There are 13,707 unique polluting facilities located among 1,859 U.S. counties. The value of sales is deflated by two-digit SIC industry-level Producer Price Index (PPI) provided by the Bureau of Labor Statistics (BLS).[10]

Table 2 provides summary statistics on a number of variables for the entire sample. As one may notice that, each facility emits at least one pollutant, but not all facilities have emission reports for all four criteria air pollutants. Moreover, the dataset contains some observations with extremely low emissions. As noted at the bottom of the table, these outliers only account for a small portion of total relevant observations.[11] The last two columns of Table 2 compare exporters and non-exporters. Exporters are larger than non-exporters in terms of the value of sale and number of employee. This result is consistent with the growing empirical trade literature that examines the differences between exporters and their competing counterparts. When it comes to the environmental performance, exporters

[9] As defined in Greenstone (2002), an industry is designated an dirty emitter of a pollutant if it accounts for at least 7 percent of industrial sector emissions. Please see table A2 of Annual Industrial Sector Pollutant Release by Industry in Greenstone (2002).

[10] The BLS reports PPI industry data on a SIC basis for years prior to 2003, and data on a NAICS basis for years after 2003. We convert the three-digit NAICS industry PPI to the two-digit SIC industry PPI in accordance with the conversion between 1987 SIC and 2002 NAICS.

[11] The fraction of observations with annual emissions less than 0.001 tons is as follows: 6.6 percent for SO_2, 0.46 percent for CO, 0.16 percent for O_3, and 1.17 percent for TSPs.

emit more SO_2 and TSPs but less CO and O_3 than non-exporters. In terms of pollution intensity, measured by emissions per value of sale (tons per dollar), exporters display better environmental performance than non-exporters for all criteria air pollutants that we track in the paper.

To further shed light on the industrial heterogeneity, for each pollutant and industry, Figure 1 plots scatter of mean (log) emission intensity by export status. The red dot line in this figure is the 45 degree line that implies the same mean values between exporters and non-exporters. In the case of O_3 there is a large discrepancy in the mean emission intensity across industry, while in the case of the remaining pollutants the mean emission intensities scatter along the 45 degree line. These four pollutant-specific figures clearly show that industries of SIC 26-27 and 32-33 are distinct from others. In these industries the relative emission intensities of exporters to non-exporters are either above or on the 45 degree line, suggesting that exporters appear not to perform better than non-exporters in the environmental perspective. By contrast, the petroleum and coal product industry of SIC 29 shows its relative mean value far below the 45 degree line. Within this industry exporters on average emit much less pollution per sale than non-exporters for all four pollutants tracked in the paper.

4 Empirical Results

The outcome of interest is emission intensity measured by log emissions per value of sales. For each criteria air pollutant, the ATET of exporting status on emission intensity is estimated across industry. Table 3 summarizes the key findings. Columns in this table correspond to various industries in terms of two-digit SIC, and standard errors are reported in parenthesis.

In general, we find the heterogeneous effects of facilities' exporting decisions on emission intensity, varying with industries and with pollutants. There is evidence that the environmental consequences of exporting status are mixed across industries. For many industries (i.e., SIC 24-25, SIC 28-30, SIC 34, and SIC 37), all three matching estimates consistently document the negative impacts of exports on emission intensity for each pollutant. These negative effects are statistically significant in most cases, indicating that exporters emit less pollution per value of sales than similar non-exporters within the same industry. For instance, in the industry of Furniture and Fixtures (SIC 25), exporters on average emit less criteria air pollutants per sales than similar non-exporters by 67% of SO_2, 64% of CO, 44% of O_3, and 49% of TSPs.[12]

For the remaining dirty industries (i.e., SIC 26-27 and SIC 32-33), our empirical results suggest an anomaly. The ATET results of exporting status on emission intensity vary with pollutant types. For example, relative to similar non-exporters in the Paper and Allied Products industry (SIC 26), exporters within this industry appear to have lower emissions intensity of CO but higher emission intensity of SO_2, O_3, and TSPs. None of these ATET estimates are statistically significant at conventional levels. When it comes to the Primary Metal Industry (SIC 32), there are consistently positive effects of exporting status on emission intensity across pollutant types. In the cases of CO and O_3, these effects are statistically significant for some matching estimators at the 1 percent level. This piece of evidence suggests that exporters in the metal industry perform even worse than comparable non-exporters. They on average tend to have higher emission intensity than non-exporters by approximately 33% of CO and 48% of O_3.

The magnitude of the environmental impacts of export decisions varies with industries and with pollutants. Regardless of pollutant types, among all selected dirty industries the Petroleum and Coal Products industry (SIC 29) displays the most superior environmental performance of exporters relative to non-exporters. Within this industry, exporting polluters emit remarkably less pollution per value of sales than similar non-exporting ones by roughly 91% of SO_2, 80% of CO, 51% of O_3, and 68% of TSPs. Among the SO_2 emitting industries with statistically significant ATET results, we find evidence suggesting that the Chemical

[12]The reported numbers are converted by the average of three matching estimates presented in table 3.

and Allied Product industry (SIC 28) presents the smallest environmental gains from export decisions. Within this industry exporters have 25% less of SO_2 per value of sales than similar non-exporters. For those industries contributing to CO pollution, the industry of Lumber and Wood Products (SIC 24) appears to have the smallest negative impacts of export decisions on emission intensity, while the industry of Stone, Clay, and Glass Products (SIC 32) shows the positive and statistically significant effects. In the former industry (SIC 24) exporters emit 22% less CO per value of sales than similar non-exporters, whereas in the latter (SIC 32) they tend to generate 31% more emission intensity than their competing counterparts. When it comes to industries emitting O_3, the magnitude of the ATET results ranges from -0.699 (SIC 29) to 0.369 (SIC 32). This evidence indicates that, whereas O_3 polluters in the industry of Petroleum and Coal Product gain from export decisions by roughly 50% lower emission intensity, polluters in the industry of Stone, Clay, and Glass Products lose from exports by releasing 45% more of O_3 per value of sales. Lastly, for the case of TSPs, we find evidence that the Lumber and Wood Product industry (SIC 24) tends to experience the smallest gains from export decisions in reducing emission intensity. Compared with non-exporters, exporters in this industry have approximately 20% less emission intensity of TSPs.

4.1 Robustness Checks

4.1.1 Match within State and Year

The baseline ATET results are estimated while matching exporters with non-exporters from the same two-digit SIC industry in terms of propensity scores. It is possible that one treated unit in the east coast may be accidently paired with another control unit in the west coast. These types of matched pairs may bias the estimated ATET results due to the confounding state-specific unosbervables, such as state-level environmental regulations, natural geographic advantage of shipping products abroad, etc. Another possible bias may arise from year trend. To remove these unobservables, we further restrict the matched exporters and non-exporters from the same U.S. state and the same year.[13] The price of this restricted matching procedure is the reduced sample size and matched number, hence rendering the statistical inference difficult. Table 4 summarizes the number of matched exporters and non-exporters for the radius matching estimator with and without the location-by-year matching restrictions.[14] As noted in this table, the number of matched pairs substantially declines when these restrictions apply.

Table 5 shows the ATET estimates across industries and pollutant types when matching within state and year. Compared with the results in Table 3, these additional matching restrictions do not alter the industrial heterogeneity of the estimated ATET results but change the statistical significance for a few industries. For SO_2 emitting industries, while the ATET results in the industry of Lumber and Wood Products (SIC 24) lose the significance, the results in the industry of Chemical and Allied Products (SIC 28) now become statistically significant at the 1 percent level for two matching estimators. Moreover, there is robust and consistent evidence suggesting the negative impacts of export decision on facility emission intensity in industries of Furniture and Fixtures (SIC 25), Chemical and Allied Products (SIC 28), Petroleum and Coal Products (SIC 29), Fabricated Metal Products (SIC 34), and Transportation Equipment (SIC 37). The magnitude of these negative ATETs increases substantially when the matched pairs are subject to additional location and year restrictions. For the remaining industries, there is still little evidence of the environmental effects of exporting decisions.

In the case of CO polluting industries, there are two significant changes in the estimated ATET results as compared with those in Table 3. In the industry of Printing and Publishing (SIC 27), the positive environmental effects of export decisions are now statis-

[13] We also conduct the matching approach by restricting the matched pairs only from the same state or from the same year. The results available upon request do not alter our conclusions in any significant ways.

[14] The number of matched pairs varies with the matching estimators. For nearest neighbor estimator, ideally every treated unit should find exactly one match, but if the industry/state/year does not have any control units within the common support, there is no match. For radius estimator, there might be multiple matches. For kernel estimator, one treated unit should be matched with all control units, though with different weights. Due to the limited space, we only report the number of matched pairs for the radius matching estimator.

tically significant at conventional levels. These estimated ATETs indicate that exporters on average emit 52% more of CO per value of sales than similar non-exporters within the same industry, state and year. In the industry of Stone, Clay, and Glass Products (SIC 32), however, the positive ATETs now lose the significance for the radius matching estimator when the additional matching restrictions are in place.

When it comes to O_3, in the industry of Furniture and Fixtures (SIC 25) the negative ATET estimates lose the statistical significance for all three matching estimators. Moreover, the ATET results for the Paper and Allied Product industry (SIC 26) and Primary Metal Industry (SIC 33) now become negative, indicating the potentially environmental benefits from exports. No statistically significant evidence supporting these beneficial effects is documented.

For TSPs, we now find some evidence that the positive ATETs are statistically significant in the industry of Printing and Publishing (SIC 27), This result suggests the deleterious effects of export on emission intensity at the facility level. Furthermore, in the industry of Stone, Clay and Glass Product (SIC 32), the ATETs are now mixed across matching estimators, but none of these estimates could be judged statistically significant at any conventional levels.

4.1.2 Subsample

The data include some observations with extremely low annual emissions.[15] To ensure these outliers not to drive the results, for each pollutant, we discard the top and bottom 5% of observations, then perform the matching procedure described in section 2. In addition, the state-by-year matching restriction still applies in this robustness check. Table 6 presents the estimated ATETs, while Table 4 shows the corresponding number of matched and control units. Our main findings are mainly unaffected. Industries of SIC 26-27 and SIC 32-33 still have mixed environmental effects of export decisions across pollutant types, whereas the

[15] The number of polluting facilities with emissions less than 1 kg is as follows: 1,068 for SO_2, 82 for CO, 44 for O_3, and 266 for TSPs.

remaining industries consistently display the environmental gains from exports for all four criteria air pollutants.

5 Conclusion

This paper employs matching techniques to investigate the impact of exporting status on emission intensity at facility level. We have assembled a large and unique panel dataset pertaining to the U.S. manufacturing industry. In particular, we focus on those manufacturing industries that are heavy emitters of criteria air pollutants, i.e., SO_2, CO, O_3, and TSPs. Our matching estimates suggest the heterogeneous environmental impacts of export decisions across industries and across pollutants. We find strong evidence, in some dirty industries but not all, that being an exporter has beneficial effects on emission intensity for all four criteria air pollutants. On the other hand, our empirical results present the deleterious effects of export orientation on the environmental performance for a few industries. Furthermore, there is little evidence that these deleterious effects hold for all four pollutants tracked in the paper.

The implication suggested by our empirical findings is of significance from the policy perspective. The environmental consequences of trade liberalization are likely to be complex. While lowering trade barriers may contribute to pollution reduction in some industries, it may also lead to further environmental degradation caused by other industries.

References

Antweler, W., Copeland, B., Taylor, S., 2001. Is Free Trade Good for the Environment? American Economic Review 91 (4), 877–908.

Batrakova, S., Davies, R. B., 2012. Is There an Environmental Benefit to Being an Exporter? Evidence from Firm Level Data. Review of World Economics 148 (3), 449–474.

Becker, S. O., Ichino, A., 2002. Estimation of Average Treatment Effects Based on Propensity Scores. Stata Journal 2 (4), 358–377.

Copeland, B., Taylor, S., 1994. North-South Trade and the Environment. Quarterly Journal of Economics 109 (3), 755–788.

Copeland, B., Taylor, S., 1995. Trade and Transboundary Pollution. American Economic Review 85 (4), 716–737.

Cui, J., Lapan, H., Moschini, G., 2012. Are Exporters More Envrionmentally Friendly than Non-Exporters? Theory and Evidence. Working Paper 12022, Iowa State University.

EPA, 2006. Documentation for the Final 2002 Point Source National Emissions Inventory. Technical report.
URL http://www.epa.gov/ttn/chief/net/2002inventory.html

EPA, 2008. Documentation for the 2005 Point Source National Emissions Inventory. Technical report.
URL http://www.epa.gov/ttn/chief/net/2005_nei_point.pdf

EPA, 2012. 2008 National Emissions Inventory, Version 2 Technical Support Document. Technical report.
URL http://www.epa.gov/ttn/chief/net/2008neiv2/2008_neiv2_tsd_draft.pdf

Forslid, R., Okubo, T., Ulltveit-Moe, K.-H., 2011. International Trade, CO_2 Emissions and Heterogeneous Firms. Working Paper 211429, Centre for Economic Policy Research.

Frankel, J., Rose, A., 2005. Is Trade Good or Bad for the Environment? Sorting out the Causality. Review of Economics and Statistics 87 (1), 85–91.

Girma, S., Greenaway, D., Kneller, R., 2003. Export Market Exit and Performance Dynamics: A Causality Analysis of Matched Firms. Economics Letters 80 (2), 181–187.

Girma, S., Greenaway, D., Kneller, R., 2004. Does Exporting Increase Productivity? A Microeconometric Analysis of Matched Firms. Review of International Economics 12 (5), 855–866.

Girma, S., Hanley, A., Tintelnot, F., 2008. Exporting and the Environment: A New Look with Micro-Data. Working Paper 1423, Kiel Institute for the World Economy.

Greenstone, M., 2002. The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures. Journal of Political Economy 110 (6), 1175–1219.

Heckman, J. J., Ichimura, H., Todd, P. E., 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies 64 (4), 605–54.

Holladay, S., 2010. Are Exporters Mother Nature's Best Friends. Working Paper, NYU School of Law.

Kolko, J., Neumark, D., 2008. Changes in the Location of Employment and Ownership: Evidence from California. Journal of Regional Science 48 (4), 717–744.

Kolko, J., Neumark, D., 2010. Does Local Business Wwnership Insulate Cities from Economic Shocks? Journal of Urban Economics 67 (1), 103–115.

Lechner, M., 2002. Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods. Journal of the Royal Statistical Society Series A 165 (1), 59–82.

List, J. A., Millimet, D. L., Fredriksson, P. G., McHone, W. W., 2003. Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator. The Review of Economics and Statistics 85 (4), 944–952.

Loecker, J. D., 2007. Do Exports Generate Higher Productivity? Evidence from Slovenia. Journal of International Economics 73 (1), 69–98.

Managi, S., Hibiki, A., Tsurumi, T., 2009. Does Trade Openness Improve Environmental Qaulity? Journal of Environmental Economics and Management 58 (3), 346–363.

McAusland, C., Millimet, D. L., 2012. Do National Borders Matter? Intranational Trade, International Trade, and the Environment. Working Paper.

Neumark, D., Wall, B., Zhang, J., 2011. Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series. The Review of Economics and Statistics 93 (1), 16–29.

Rosenbaum, P. R., Rubin, D. B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70 (1), 41–55.

Rubin, D., 1974. Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies. Journal of Educational Psychology 66, 688–701.

Wagner, J., 2002. The Causal Effects of Exports on Firms Size and Labor Productivity: First Evidence from a Matching Approach. Economics Letters 77, 287–292.

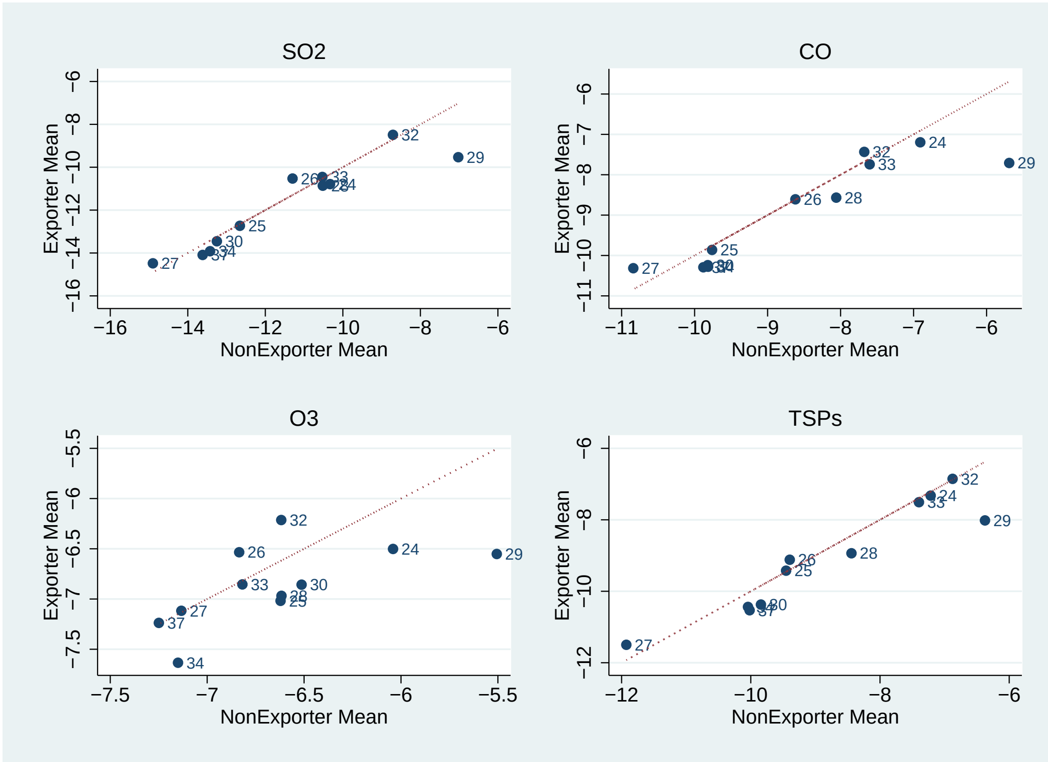

Figure 1: Scatter of Mean Log Emission Intensity (1000 US$ per ton) by Pollutant and Two-Digit SIC Industry (note: the red dot line refers to the 45 degree line)

Table 1: List of Dirty Manufacturing Industries

Industry Description	Dirty Pollutant	Number of Exporters	Number of Non-Exporters
SIC 24: Lumber and Wood Products	TSPs	459	2,045
SIC 25: Furniture and Fixtures	O_3	432	933
SIC 26: Paper and Allied Products	O_3, SO_2, CO, TSPs	453	1,227
SIC 27: Printing and Publishing	O_3	393	2,183
SIC 28: Chemicals and Allied Products	O_3, SO_2	1,142	2,521
SIC 29: Petroleum and Coal Products	O_3, SO_2, CO	90	1,703
SIC 30: Rubber and Misc. Plastics Products	O_3	833	1,981
SIC 32: Stone, Clay, and Glass Products	O_3, SO_2, TSPs	373	2,901
SIC 33: Primary Metal Industries	O_3, SO_2, CO, TSPs	665	1,664
SIC 34: Fabricated Metal Products	O_3	1,078	3,708
SIC 37: Transportation Equipment	O_3	716	1,683

Note: for the definition of dirty industry, please refer to table A2 of Annual Industrial Sector Pollutant Release by Industry in Greenstone (2002)

Table 2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	5 Percentile	95 Percentile	Exporter Mean	Non-Exporter Mean
Sale (1000 $)	29,183	27582.9	67843.6	394.1	99580.3	40003.3	23928.8
Employment	29,183	192.4	447.6	5.0	685.0	280.2	166.6
Export dummy	29,183	0.2	0.4	0.0	1.0	1.0	0.0
SO_2 (ton)	16,186	138.7	892.6	0.0006	413.2	163.4	130.7
CO (ton)	17,883	169.3	1719.6	0.02	386.8	163.5	171.2
O_3 (ton)	27,080	113.2	477.4	0.2	448.2	112.5	113.4
TSPs (ton)	22,710	44.0	197.3	0.0097	210.68	49.9	42.3
SO_2 per sale	16,186	0.049	1.393	6.06e-08	0.0242	0.012	0.061
CO per sale	17,883	0.073	3.135	2.24e-06	0.0333	0.017	0.091
O_3 per sale	27,080	0.043	0.851	3.21e-05	0.0385	0.033	0.046
TSPs per sale	22,710	0.017	0.400	9.54e-07	0.0181	0.010	0.019

Table 3: The ATET Results by Industry and Pollutant

	Estimators \ Industry	24	25	26	27	28	29	30	32	33	34	37
SO_2	Nearest Neighbor	-0.334	-1.176**	0.120	0.122	-0.143	-2.706***	-0.128	0.128	0.228	-0.382**	-0.309
		(0.288)	(0.526)	(0.486)	(0.306)	(0.265)	(0.710)	(0.274)	(0.363)	(0.257)	(0.176)	(0.279)
	Radius ($r = 0.01$)	-0.450**	-1.039**	0.295	0.089	-0.388*	-2.447***	-0.274	0.292	0.122	-0.441***	-0.510***
		(0.197)	(0.455)	(0.379)	(0.227)	(0.198)	(0.612)	(0.209)	(0.254)	(0.188)	(0.130)	(0.193)
	Kernel	-0.589***	-1.129***	0.565	0.134	-0.325*	-2.181***	-0.137	0.289	0.177	-0.497***	-0.494***
		(0.191)	(0.396)	(0.365)	(0.223)	(0.196)	(0.600)	(0.204)	(0.253)	(0.187)	(0.128)	(0.191)
CO	Nearest Neighbor	-0.129	-1.075***	-0.515*	0.201	-0.410***	-1.406***	-0.362***	0.132	-0.023	-0.479***	-0.211
		(0.252)	(0.323)	(0.294)	(0.275)	(0.165)	(0.457)	(0.177)	(0.246)	(0.206)	(0.137)	(0.226)
	Radius ($r = 0.01$)	-0.216	-0.966***	-0.326	0.273	-0.487***	-1.557***	-0.363***	0.344***	-0.094	-0.473***	-0.511***
		(0.170)	(0.276)	(0.225)	(0.194)	(0.122)	(0.369)	(0.135)	(0.169)	(0.150)	(0.102)	(0.155)
	Kernel	-0.410***	-1.050***	-0.176	0.313*	-0.467***	-1.778***	-0.374***	0.331**	-0.084	-0.477***	-0.463***
		(0.165)	(0.255)	(0.216)	(0.190)	(0.121)	(0.358)	(0.132)	(0.169)	(0.149)	(0.100)	(0.154)
O_3	Nearest Neighbor	-0.376**	-0.567**	0.044	-0.041	-0.246**	-0.489	-0.138	0.153	0.040	-0.490***	-0.011
		(0.163)	(0.281)	(0.206)	(0.147)	(0.120)	(0.420)	(0.124)	(0.206)	(0.139)	(0.101)	(0.135)
	Radius ($r = 0.01$)	-0.397***	-0.580**	0.044	-0.019	-0.287***	-0.744**	-0.236***	0.481***	0.007	-0.532***	-0.053
		(0.115)	(0.241)	(0.153)	(0.108)	(0.086)	(0.348)	(0.090)	(0.136)	(0.098)	(0.075)	(0.090)
	Kernel	-0.442***	-0.613***	0.225	0.006	-0.301***	-0.864***	-0.279***	0.472***	0.009	-0.493***	-0.039
		(0.112)	(0.228)	(0.148)	(0.108)	(0.084)	(0.338)	(0.088)	(0.135)	(0.098)	(0.074)	(0.090)
TSPs	Nearest Neighbor	-0.183	-0.720	0.072	0.106	-0.370***	-0.952**	-0.241	0.252	0.029	-0.391***	-0.407**
		(0.213)	(0.456)	(0.321)	(0.331)	(0.157)	(0.426)	(0.188)	(0.197)	(0.180)	(0.149)	(0.200)
	Radius ($r = 0.01$)	-0.228	-0.675**	0.095	0.377	-0.501***	-1.095***	-0.366***	0.071	-0.071	-0.437***	-0.572***
		(0.146)	(0.386)	(0.249)	(0.231)	(0.114)	(0.330)	(0.140)	(0.141)	(0.126)	(0.108)	(0.141)
	Kernel	-0.267*	-0.606*	0.178	0.373	-0.463***	-1.317***	-0.444***	0.065	-0.078	-0.410***	-0.552***
		(0.141)	(0.361)	(0.240)	(0.228)	(0.113)	(0.318)	(0.137)	(0.141)	(0.126)	(0.107)	(0.141)

Note: columns correspond to two-digit SIC industries. Standard errors are reported in parenthesis. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 4: The Number of Matched Treated and Control Units for Radius Matching Estimator

		Match within Country				Match within State & Year				Sub-Sample Match			
Industry	# of Matched	SO_2	CO	O_3	TSPs	SO_2	CO	O_3	TSPs	SO_2	CO	O_3	TSPs
SIC 24	exporters	316	336	427	425	239	259	302	309	234	262	307	328
	non-exporters	1,045	1,157	1,845	1,641	560	608	828	777	552	628	768	796
SIC 25	exporters	61	73	122	91	30	25	38	38	28	29	38	28
	non-exporters	78	96	186	116	34	40	83	54	29	31	53	30
SIC 26	exporters	256	274	334	272	122	120	146	116	244	246	284	235
	non-exporters	676	720	901	758	205	210	251	182	415	432	556	468
SIC 27	exporters	104	126	218	123	102	120	189	117	88	118	192	115
	non-exporters	414	557	1,522	522	349	408	1,044	427	269	359	925	316
SIC 28	exporters	755	848	1,119	995	737	822	1,082	963	652	754	945	851
	non-exporters	1,548	1,737	2,349	2,057	1,423	1,574	2,119	1,851	1175	1,306	1,695	1,484
SIC 29	exporters	67	71	80	74	53	55	62	60	56	53	60	53
	non-exporters	1,409	1,437	1,526	1,530	411	389	515	509	269	288	359	355
SIC 30	exporters	373	421	809	562	333	381	711	481	302	326	684	440
	non-exporters	808	924	1,911	1,308	650	718	1,428	980	567	607	1423	900
SIC 32	exporters	301	299	347	345	299	297	342	336	283	278	328	327
	non-exporters	1,378	1,537	1,745	2,706	1,232	1,345	1,471	2,173	1,099	1,213	1,357	1,997
SIC 33	exporters	521	523	619	615	487	500	588	594	469	470	568	566
	non-exporters	1,226	1,229	1,527	1,504	1,104	1,106	1,295	1,330	1,000	995	1,240	1,194
SIC 34	exporters	504	611	1,034	784	497	595	981	766	466	563	937	718
	non-exporters	1,686	2,012	3,439	2,678	1,534	1,789	2,876	2,315	1,310	1,527	2,682	2,078
SIC 37	exporters	339	371	712	487	309	360	666	461	306	337	653	449
	non-exporters	890	947	1,641	1,212	761	830	1,395	1,046	706	702	1357	999

For each pollutant, we discard the lowest and highest 5 percentile observations, and then match exporters with non-exporters within industry, state and year.

Table 5: The ATET Results by Industry and Pollutant: Match within State and Year

Pollutant	Estimators \ Industry	24	25	26	27	28	29	30	32	33	34	37
SO_2	Nearest Neighbor	-0.275	-1.509**	-0.251	0.265	-0.362	-3.052***	0.078	0.315	0.133	-0.295	-0.485
		(0.290)	(0.603)	(0.518)	(0.309)	(0.282)	(0.747)	(0.293)	(0.373)	(0.268)	(0.180)	(0.300)
	Radius ($r = 0.01$)	-0.251	-1.338**	0.379	0.104	-0.665***	-2.451***	-0.072	0.357	0.133	-0.433***	-0.617***
		(0.265)	(0.640)	(0.587)	(0.240)	(0.218)	(0.733)	(0.243)	(0.271)	(0.212)	(0.135)	(0.224)
	Kernel	-0.380*	-1.554***	-0.131	0.173	-0.616***	-2.892***	-0.187	0.174	0.183	-0.465***	-0.616***
		(0.205)	(0.495)	(0.412)	(0.231)	(0.204)	(0.620)	(0.217)	(0.260)	(0.204)	(0.130)	(0.218)
CO	Nearest Neighbor	0.229	-1.009***	-0.691***	0.508*	-0.430**	-2.354***	-0.118	0.302	-0.157	-0.601***	-0.499***
		(0.260)	(0.327)	(0.299)	(0.287)	(0.177)	(0.474)	(0.181)	(0.255)	(0.215)	(0.140)	(0.228)
	Radius ($r = 0.01$)	-0.091	-1.248***	-0.544	0.582***	-0.589***	-1.951***	-0.240	0.291	-0.230	-0.461***	-0.558***
		(0.239)	(0.434)	(0.360)	(0.217)	(0.134)	(0.455)	(0.154)	(0.179)	(0.175)	(0.109)	(0.175)
	Kernel	-0.154	-0.887***	-0.388	0.390*	-0.574***	-2.371***	-0.367**	0.330*	-0.183	-0.489***	-0.396**
		(0.181)	(0.299)	(0.244)	(0.200)	(0.126)	(0.379)	(0.141)	(0.172)	(0.162)	(0.102)	(0.162)
O_3	Nearest Neighbor	-0.489***	-0.173	0.002	0.114	-0.378***	-0.981**	-0.254**	0.230	-0.190	-0.456***	-0.042
		(0.170)	(0.343)	(0.212)	(0.154)	(0.127)	(0.454)	(0.126)	(0.213)	(0.144)	(0.105)	(0.139)
	Radius ($r = 0.01$)	-0.367**	-0.491	-0.023	-0.013	-0.314***	-1.243***	-0.259**	0.411***	-0.083	-0.480***	-0.083
		(0.158)	(0.412)	(0.253)	(0.133)	(0.094)	(0.431)	(0.106)	(0.147)	(0.113)	(0.085)	(0.107)
	Kernel	-0.534***	-0.374	-0.045	-0.018	-0.408***	-1.306***	-0.282***	0.378***	-0.091	-0.501***	-0.043
		(0.125)	(0.293)	(0.175)	(0.116)	(0.090)	(0.358)	(0.095)	(0.140)	(0.108)	(0.077)	(0.100)
TSPs	Nearest Neighbor	-0.242	-0.865*	-0.502	0.127	-0.378**	-1.372***	-0.440**	0.127	0.037	-0.458***	-0.448**
		(0.230)	(0.473)	(0.328)	(0.334)	(0.166)	(0.457)	(0.193)	(0.224)	(0.179)	(0.152)	(0.206)
	Radius ($r = 0.01$)	-0.199	-0.181	-0.200	0.345	-0.476***	-1.393***	-0.348**	0.020	-0.073	-0.384***	-0.639***
		(0.205)	(0.556)	(0.413)	(0.256)	(0.126)	(0.414)	(0.166)	(0.171)	(0.142)	(0.118)	(0.161)
	Kernel	-0.243	-0.719*	-0.139	0.443*	-0.460***	-1.695***	-0.478***	-0.065	-0.046	-0.341***	-0.603***
		(0.161)	(0.413)	(0.275)	(0.237)	(0.121)	(0.341)	(0.147)	(0.151)	(0.138)	(0.109)	(0.150)

Note: columns correspond to two-digit SIC industries. Standard errors are reported in parenthesis. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 6: The ATET Results by Industry and Pollutant with Subsample: Match within State and Year

	Industry / Estimators	24	25	26	27	28	29	30	32	33	34	37
SO_2	Nearest Neighbor	-0.649**	-1.368**	0.243	0.292	-0.575**	-2.551***	-0.247	0.453	0.251	-0.327*	-0.554**
		(0.283)	(0.552)	(0.394)	(0.319)	(0.259)	(0.700)	(0.283)	(0.365)	(0.253)	(0.175)	(0.277)
	Radius ($r = 0.01$)	-0.581**	-0.338	0.383	0.125	-0.631***	-1.924***	-0.175	0.494*	0.161	-0.426***	-0.459**
		(0.254)	(0.632)	(0.383)	(0.263)	(0.224)	(0.710)	(0.259)	(0.278)	(0.207)	(0.141)	(0.227)
	Kernel	-0.637***	-1.183**	0.577	0.280	-0.439***	-2.651***	-0.137	0.395	0.075	-0.472***	-0.469**
		(0.208)	(0.489)	(0.314)	(0.233)	(0.198)	(0.594)	(0.215)	(0.261)	(0.195)	(0.131)	(0.206)
CO	Nearest Neighbor	-0.543**	-0.940***	-0.003	-0.015	-0.437***	-2.275***	-0.164	0.211	-0.324	-0.451***	-0.642***
		(0.249)	(0.336)	(0.222)	(0.282)	(0.161)	(0.430)	(0.182)	(0.246)	(0.204)	(0.136)	(0.234)
	Radius ($r = 0.01$)	-0.489**	-0.697	-0.009	0.265	-0.494***	-2.050***	-0.144	0.327*	-0.096	-0.543***	-0.478**
		(0.222)	(0.419)	(0.221)	(0.222)	(0.137)	(0.441)	(0.170)	(0.184)	(0.166)	(0.113)	(0.183)
	Kernel	-0.371**	-1.048***	-0.091	0.286	-0.521***	-2.081***	-0.373**	0.285	-0.130	-0.547***	-0.543***
		(0.175)	(0.308)	(0.180)	(0.196)	(0.122)	(0.348)	(0.142)	(0.176)	(0.152)	(0.102)	(0.161)
O_3	Nearest Neighbor	-0.494***	-0.619**	0.355**	-0.100	-0.328***	-0.958***	-0.193	0.682***	-0.088	-0.482***	-0.182
		(0.154)	(0.292)	(0.171)	(0.143)	(0.112)	(0.346)	(0.120)	(0.206)	(0.135)	(0.097)	(0.131)
	Radius ($r = 0.01$)	-0.359**	-0.260	0.223	-0.085	-0.304***	-0.084	-0.212**	0.455***	0.058	-0.475***	-0.133
		(0.148)	(0.405)	(0.166)	(0.120)	(0.095)	(0.335)	(0.103)	(0.149)	(0.108)	(0.077)	(0.100)
	Kernel	-0.554***	-0.519**	0.161	-0.024	-0.300***	-0.822***	-0.268***	0.455***	0.020	-0.440***	-0.102
		(0.117)	(0.263)	(0.129)	(0.110)	(0.083)	(0.264)	(0.088)	(0.142)	(0.100)	(0.070)	(0.093)
TSPs	Nearest Neighbor	-0.309	0.265	-0.302	0.501	-0.525***	-2.349***	-0.397**	-0.111	-0.229	-0.496***	-0.486**
		(0.203)	(0.443)	(0.255)	(0.341)	(0.153)	(0.443)	(0.189)	(0.192)	(0.178)	(0.144)	(0.202)
	Radius ($r = 0.01$)	-0.349*	0.866	-0.099	0.671**	-0.388***	-1.385***	-0.238	-0.026	-0.152	-0.389***	-0.682***
		(0.187)	(0.659)	(0.255)	(0.276)	(0.131)	(0.441)	(0.172)	(0.149)	(0.140)	(0.115)	(0.159)
	Kernel	-0.372**	0.092	-0.076	0.522**	-0.508***	-1.843***	-0.454***	-0.015	-0.148	-0.403***	-0.568***
		(0.149)	(0.394)	(0.202)	(0.239)	(0.114)	(0.357)	(0.145)	(0.142)	(0.130)	(0.105)	(0.146)

Note: for each pollutant, we discard the lowest and highest 5 percentiles of observations. Columns correspond to two-digit SIC industries. Standard errors are reported in parenthesis. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

A Appendix

A.1 Caveats of the NEI Data

Some major caveats of the NEI database pertaining to point sources can be summarized as follows (EPA, 2006, 2008, 2012). First and foremost, EPA developed the 2005 NEI data based on a reduced level of effort. Part of this reduced effort involved using some 2002 NEI data in the 2005 NEI as surrogates for emission data representing 2005. Second, the 2008 NEI database was built from emission data in the Emission Inventory System (EIS). Unlike its predecessors 2002 and 2005 NEI, this 2008 database reports a different and new facility identifier, called EIS site ID, instead of previous NEI site ID. A comprehensive and updated coverage of facility identifiers may be obtained from the Emission Inventory System Gateway. This Gateway, however, is only available to EPA staff, EIS data partners responsible for submitting data to EPA, and contractors working for EPA on emissions related work. Last but not least, as noted in the EPA technical document, emission data for filterable and condensable components of particulate matter (i.e., filterable PM-10, filterable PM-2.5 and condensible PM) is not complete and is not suggested to use at any aggregate level. Users interested in PM emissions should only consider primary particulate matter, which are primary PM-10 and primary PM-2.5.

A.2 Data Matching Algorithm

Given the forgoing caveats of the NEI database, the data matching work consists of two main procedures. First, we match polluting facilities within the NEI database across years, and then retrieve DUNS numbers for these polluters from the Facility Registry System (FRS) of the EPA. Second, we match them with those appearing in the NETS database through the DUNS number.

To match polluting facilities within the NEI data across years, we first discard duplicates of 2005 data. The 2005 NEI database provides flag variables, “Start Date/End

Date" fields, to indicate which data are 2005 emissions and which data are actually taken from 2002 emissions. Around one-third of observations in the 2005 NEI have a flag variable of "Start Date" referring to year 2002. When it comes to the manufacturing industry, roughly one-quarter of observations in 2005 are duplicates of 2002 emissions. These duplicates are dropped from our study.

We then retrieve facility FRS ID from the FRS of the EPA. The FRS is a centrally-managed database that identifies facilities, sites, or places subject to environmental regulations or of environmental interests. EZ Query in the FRS provides data download options for a customized list of facilities, which are associated with NEI or EIS programs. All observations in 2002 and 2005 NEI databases have both records and FRS ID reported in the FRS, hence can be matched between these two years. However, one-eighth of 2008 NEI database is missing from the FRS, and roughly 7 percent of facilities in the manufacturing industry in this database do not have any records in the FRS. These missing manufactures are discarded in our study. With the FRS ID, facility DUNS numbers are retrieved separately through the Facility Registry System Query in the FRS. In the end, the facility-level emission dataset we compiled contains criteria air emissions, facility name, FIPS county code, zip code, facility FRS ID, and DUNS number.

In the next step, we match polluting facilities in the NEI database with those that appear in the NETS Database through the DUNS number. The EPA does not provide further information about how DUNS numbers are reported for polluting facilities and why some of them have missing DUNS numbers in the dataset. Due to an incomplete report on DUNS numbers in the FRS, approximately 80 percent of polluting facilities in the manufacturing industry collected in the NEI database have associated DUNS numbers. To circumvent this shortcoming, a pair of facilities from each source is considered as a match if the following series of criteria are satisfied. They share the same DUNS number and are located in the same area in terms of five-digit FIPS county code. More importantly, for each pair, we compare their facility names from each source to ensure the match.