A SPATIAL ECONOMETRICS APPROACH TO ANALYZING EMISSIONS SPILLOVERS[*]

SOUVIK DATTA[†]

This version: November, 2010

Abstract

This paper analyzes the presence of spillovers in pollution emission by looking at level emissions and changes in emissions; that polluting facilities around other neighbouring facilities tend to have a similar environmental performance. I use information about the location of each facility and exploit the variation in emission levels and emission changes in a large sample of manufacturing facilities in Canada by using a simple and parsimonious spatial autoregressive (SAR) model. I also use an extension of the SAR model that uses two spatial weight matrices simultaneously. The "distance" between facilities is measured by the geographical distance, which captures any demonstration effect or regulatory threat that might exist, and an industry SIC metric, which captures technological or horizontal linkages within the same sector of the industry. Spatial dependency may be the result of both these factors and the extension of the SAR model takes into account both these channels simultaneously. I find that, compared to OLS results, spatial dependency exists and is significant. My results also indicate that emission spillovers that result from being in the same industry are substantially stronger than that caused by spillovers from geographical proximity.

1 Introduction

The very nature of pollution is such that it is very rarely confined to a particular area but, depending on the substance and the medium into which it is released, spreads to neighbouring areas. This spatial aspect of pollution is an important consideration when analyzing the emissions of firms. The presence

---

[*] I am especially grateful to Werner Antweiler, Sumeet Gulati and Shinichi Sakata for valuable comments and feedback. I would also like to thank participants at the UBC Empirical Lunch Workshop and the 42nd Annual Meeting of the Canadian Economics Association for helpful suggestions. All omissions and remaining errors are my own.

[†] PhD Candidate, Department of Economics, University of British Columbia, 997–1873 East Mall, Vancouver, BC, V6T 1Z1, Canada. <souvik@interchange.ubc.ca>

of multiple polluting facilities is likely to cause the pollution of the respective plants to be spatially correlated[1], as noted by Gray and Shadbegian (2007). While the environmental performance of a particular plant could be influenced by plant-specific, firm-specific or external factors, they could also be affected by the environmental performance of neighbouring plants (Gray and Shadbegian, 2007). This would lead to spatial dependence in the emissions.

In this paper I investigate the presence of emission linkages in Canadian manufacturing plants by incorporating spatial effects in my analysis. These spatial effects are the result of spillovers that may occur when the emission of plant i affects that of plant j. I consider two channels of these spillovers and use spatial econometric methods to model this. One channel to measure spatial linkages uses geographic distance while the second channel uses the Standard Industrial Classification (SIC) to measure the similarity in a pair of polluting facilities.

This paper extends the literature in three ways. Firstly, I use the Standard Industrial Classification (SIC) code at the two-digit level to construct a supplementary measure of the closeness of two facilities. The hypothesis being that two facilities with the same SIC code will have a greater similarity in emissions than two facilities with different SIC codes. Previous research has used a decaying function of the Euclidean distance with a single parameter as a measure of closeness. The most commonly used specification is the inverse of the distance squared because of its resemblance to the gravity equation in the trade literature.

Secondly, I simultaneously use the SIC distance as well as the geographical distance in the same specification. This means that the measure of closeness involving the SIC distance captures the effect of plants in similar industries and the proximity measure using the Euclidean distance[2] captures the spatial dependence of facilities that are geographically nearer. Therefore, the two types of spatial dependencies in the dependent variable are separated out by using these two measures of closeness. Using two measures of distance or closeness simultaneously will also enable us to determine the strength and magnitude in the two types of spatial dependencies in the dependent variable.

Thirdly, the data I use is a comprehensive pollution inventory of all manufacturing facilities in Canada. Unlike other studies, this paper includes all the facilities that are required to report to the pollution inventory and is not limited to manufacturing facilities in certain regions. The advantage of

---

[1] The formal definition of spatial autocorrelation is: $\text{Cor}(y_i, y_j) = \text{E}(y_i y_j) - \text{E}(y_i)\text{E}(y_j), \quad i \neq j$, where the y's are the variable of interest and i and j refer to their respective location.

[2] I use the Haversine formula to measure the orthodromic (great circle) distance between two facilities.

using such a comprehensive database, apart from the benefits of having access to more observations, is that we can exploit the various measures of distance, especially the geographic measure by calculating it for all manufacturing facility pairs spread across the country. The standard procedure to measure closeness, as mentioned previously, has been to use a decaying function of the Euclidean distance. If facilities are distant from each other the measure of closeness will, by construction, be very low. In other words, the distance between firms i and j will be very high and, assuming a measure of closeness that is the square of the inverse of the distance, the measure will be very low.

There are primarily two questions that this paper endeavours to answer. Firstly, are there any spatial dependencies in the emissions of manufacturing facilities? There are many reasons why we would expect spatial dependence to exist. Gray and Shadbegian (2007) mention that the presence of "demonstration effects" would lead to neighbouring plants having similar environmental performance. This is caused by the pressure on a plant to improve on its environmental performance by being in the presence of other better performing plants. This would lead to a positive spatial autocorrelation in the emissions of plants. On the other hand, there could be plants that are dirtier than the surrounding ones. This may be caused by free-riders whereby some plants find it may escape attention by being in the midst of better performers as long as they are within compliance. This is likely because the regulatory authorities are concerned about the aggregate level of environmental exposure faced by the population (Antweiler, 2003). So the presence of some plants that are more polluting than others may not be a cause for concern as long as the health of the affected population is not compromised by having a maximum tolerated level of pollution.

Secondly, if there *are* any spatial dependencies, what is the channel through which they manifest themselves? Is it more pronounced through a simple geographic distance metric? Or are the similarities in the SIC codes more important? There is sufficient evidence to believe that spillovers exist and that these spillovers are greater in magnitude when firms are closer. There could be techonological spillovers so that firms within the same industry have similar equipment or similar pollution abatement technology. In that case we would expect very strong spatial dependence in the emissions of firms within the same industrial classification.

To answer these questions I use data from the National Pollution Release Inventory of Canada published by Environment Canada. This provides data on facility-level emissions. Using the coordinates of the location of the facility, I can obtain the population in the surrounding area from the Gridded

Population of the World (version 3) dataset. I also use the US Environment Protection Agency's Risk-Screening Environmental Indicators (RSEI) database (Version 2.0 b2) to calculate toxicity-weighted or health-indexed emissions.

I use various spatial econometric models to analyze the data. The spatial econometric models are based on the parsimonious spatial autoregressive regression (SAR) model. I initially analyze the data using the levels of toxic emissions. I also use the change in emission to study if there is any spatial dependence. Using differences is econometrically superior to using levels since it accounts for fixed effects. After using the two measures of closeness individually, I analyze the data using them simultaneously. I find that, compared to OLS results, spatial dependencies exist and are significant as indicated by the statistical significance of the spatial autoregressive parameters. My results also indicate that the effect of the industry SIC distance is substantially stronger than that of geographical distance.

Methodologically, this paper is closest to Gray and Shadbegian (2007) who use plant-level EPA and Census data from the US to analyze spatial effects that may affect air pollutant emissions and regulatory compliance. In this paper I use a more comprehensive set of manufacturing facilities and an SAR model with two weight matrices simultaneously to extract spatial dependencies affecting air pollutants, total pollutants and air pollutants that are causes of ground level ozone, haze and acid rain.

The rest of the paper is structured as follows. The next section describes the NPRI and population data. Section 3 discusses the spatial autoregressive model I will use to measure the pollution abatement spillovers. Spatial and non-spatial regression results and their interpretation are provided in section 4. The penultimate section discusses the empirical results and their possible causes while the last section concludes and provides directions for further research.

2 Data

This paper utilizes data from two major sources. The plant-level pollution data is the first major source and is obtained from the National Pollutant Release Inventory published annually by Environment Canada. The Gridded Population of the World (version 3) is the second major source and I obtain population data from it. I also use toxicity scores published by the US Environment Protection Agency's Risk-Screening Environmental Indicators (RSEI) database (Version 2.0 b2) to calculate toxicity-weighted or health-indexed emissions.

Table 1: NUMBER OF FACILITIES BY SIC CODES

SIC Code	SIC code Description	Frequency	Percentage
20	Food And Kindred Products	317	10.85
21	Tobacco Products	3	0.1
22	Textile Mill Products	21	0.72
23	Apparel And Other Finished Products Made From Fabrics And Similar Materials	4	0.14
24	Lumber And Wood Products, Except Furniture	359	12.28
25	Furniture And Fixtures	62	2.12
26	Paper And Allied Products	172	5.88
27	Printing, Publishing, And Allied Industries	89	3.04
28	Chemicals And Allied Products	434	14.85
29	Petroleum Refining And Related Industries	55	1.88
30	Rubber And Miscellaneous Plastics Products	238	8.14
31	Leather And Leather Products	1	0.03
32	Stone, Clay, Glass, And Concrete Products	180	6.16
33	Primary Metal Industries	198	6.77
34	Fabricated Metal Products, Except Machinery And Transportation Equipment	282	9.65
35	Industrial And Commercial Machinery And Computer Equipment	67	2.29
36	Electronic And Other Electrical Equipment And Components, Except Computer Equipment	66	2.26
37	Transportation Equipment	178	6.09
38	Measuring, Analyzing, And Controlling Instruments; Photographic, Medical And Optical Goods; Watches And Clocks	5	0.17
39	Miscellaneous Manufacturing Industries	192	6.57
Total		2923	100

2.1 National Pollutant Release Inventory

Emissions data is obtained from Environment Canada's National Pollutant Release Inventory (NPRI) that contains publicly available information on the releases and transfers of key pollutants in various communities. The NPRI was established in 1992 and legislated under the Canadian Environmental Protection Act, 1999 (CEPA 1999). Companies are required to report information on releases and transfers of pollutants on an annual basis. However, there are certain reporting criteria for reporting to the NPRI. Pollutants from mobile sources such as trucks and cars, households, facilities that release pollutants on a smaller scale and certain sector activities, such as agriculture and education and some mining activities, are not included in the NPRI but are reported under a separate program. Reporting for each NPRI substance includes an indication of whether the substance was manufactured, processed, or otherwise used and the nature of such activities and uses during a year.

The NPRI is a comprehensive database and also reports the number of employees in each facility as well as the geographic location in terms of latitude and longitude. The number of employees can be used as a proxy for the size of the facility since sales and production data are not available. Geographic coordinates are essential to perform a spatial analysis of the data. I have considered only manufacturing facilities. These are facilities with 2 or 3 as the first digit of their four-digit SIC codes. Table 1 shows the breakdown of the facilities in my sample by their two digit SIC codes. I have also considered all the provinces in Canada that have manufacturing facilities. The numbers, by each province, are reported in Table 2. As expected, the provinces of Alberta, British Columbia, Ontario and Québec make up the bulk of Canadian manufacturing facilities.

The NPRI is similar to the Toxics Release Inventory (TRI) of the United States Environmental Protection Agency (US EPA). One of the advantages, according to Harrison and Hoberg (1994), of using the NPRI over the TRI is that Canada has followed a policy of negotiations with polluters when it comes to the regulation of toxic substances rather than enforcing the regulations. This considerably lowers the presence of regulatory threat. A weak regulatory threat means that reductions could be attributed to other factors such as, for example, voluntary reductions or technological adoption. Olewiler and Dawson (1998) also report that Canadian manufacturing industries are considerably more polluting than their US counterparts which suggests that Canadian regulations are somewhat more lenient than the US. This, similar to the presence of a lower regulatory threat, is an advantage for studying the impact of voluntary pollution abatement activities since the effect of regulatory intervention, in the Canadian context, should not be very significant. Antweiler (2003) has shown that while the effect of regulatory threat in Canada is statistically significant the magnitude is very small and concludes that it is not a very effective instrument.

I use emissions data from a cross-section of facilities in all the provinces. Table 2 shows the number of manufacturing facilities that are located in the respective provinces. Most of the facilities are located in Ontario and Québec with a little more than 70% of all the manufacturing facilities in those two provinces. The time period of the data ranges from 2003 to 2006. While I will be using the average over the 2003 - 2005 period to estimate spatial regressions for emissions levels, I will consider the 2003 - 2006 period for calculating spatial regressions for differences in emissions. This is described in greater detail in the section on empirical strategy. The rationale for this time period is that there were relatively small changes in the chemicals added to the NPRI Substance List. In fact, there were no new chemicals added in 2004 and 2005 after the addition of, mainly, Volatile Organic Compounds (VOCs) in 2003. The addition of new chemicals to the list was also not very drastic in 2006 when only three Polycyclic Aromatic Hydrocarbons

Table 2: NUMBER OF FACILITIES BY PROVINCES

Province	Frequency	Percentage
Alberta	243	8.31
British Columbia	305	10.43
Manitoba	94	3.22
New Brunswick	45	1.54
New Foundland and Labrador	8	0.27
Nova Scotia	53	1.81
Ontario	1481	50.67
Prince Edward Island	7	0.24
Québec	629	21.52
Saskatchewan	58	1.98
Total	2923	100

and 15 VOCs were added. This relative lack of activity in adding chemicals to the NPRI list, compared to other years, makes this time period especially conducive to studying the emission activities. There were also no modifications to existing substances or to their reporting thresholds between 2003 and 2006 thus ensuring that there would be no compatibility issues.

There are, however, some limitations to using the NPRI data.[3] They include the fact that all emissions are self-reported, not all pollutants of interest are reported and not all sources of pollution are included. Since the emissions are self-reported there is an incentive for facilities to under-report their emissions. However, companies that meet the reporting requirements and fail to report or under-report their emissions face penalties under CEPA 1999 so a risk for improper reporting does exist.

There are very detailed reports of emissions by polluting facilities. The NPRI reports releases to the air, water and land with emissions being broken down into on-site releases and as well as transfers to off-sites and recycling. I only consider pollutants that were released on-site. Since the majority of the emissions were released into the air I use only on-site air releases as well as total on-site releases. This will facilitate the spatial analysis of any pollution abatement for the air pollutants as well as overall emissions.

A number of chemical pollutants are reported to the NPRI. Emissions of different pollutants cannot, ideally, be treated equally. For example, the health effects of being exposed to one pound of (friable) asbestos is not the same as that of one pound of silver and its compounds. The US EPA has assigned different toxicity scores to the various chemicals in its list to account for the different health impacts. For example, one pound of (friable) asbestos is 10,000 times more toxic than an equivalent amount (by mass)

---

[3] See Harrison and Antweiler (2003) for a more detailed discussion of these issues.

of silver and its compounds when ingested either orally or by inhalation. Therefore, one of the issues concerns the aggregation of various pollutants released in the production process. Instead of using just the total emissions or considering the release of individual chemicals, many authors have used a weighted sum of emissions where the weights reflect the toxicity of the chemicals (see, e.g., Hettige et al. (1992) and Horvath et al. (1995)).

Toxicity scores are obtained from the US EPA's Risk-Screening Environmental Indicators database that provides a list of chemicals as well as their toxicity based on whether they are ingested orally or inhaled. Since the NPRI provides details on the medium into which a particular pollutant was emitted I can construct a toxicity-weighted emissions variable. I use the toxicity scores for inhalation to weight the pollutants emitted into the atmosphere while the values for oral ingestion were used to weight the pollutants emitted into water bodies or the ground.

The variable of interest is the emissions from facilities. To account for the volatility of emissions I calculate the average emissions over three years. The data for emissions levels are averaged over three years from 2003 to 2005 while the emissions data used for analysing differences is averaged over the years 2003 to 2005 and also from 2004 to 2006. I consider three types of emissions to investigate if the spatial dependence varies with the classes of emissions. The first type of emission is a sum of the Criteria Air Contaminants (CACs) that consists of Total Particulate Matter (TPM), sulphur oxides (SO_x), nitrogen oxides (NO_x), VOCs, carbon monoxide (CO) and ammonia. TPM consists of PM_{10} which is Particulate Matter less than or equal to 10 microns and $PM_{2.5}$ which is Particulate Matter less than or equal to 2.5 microns. These CACs, along with some related pollutants, are the causes of air pollution such as smog and acid rain. The second emission variable is the toxicity score-weighted sum of pollutants emitted into the atmosphere while the third variable of interest is the toxicity score-weighted sum of total emissions. As noted in Antweiler and Harrison (2003), the on-site releases of chemicals in the NPRI follows a log-normal distribution. Therefore, the dependent variable in all three cases is the logarithm of the variable of interest.

As mentioned previously, there are various factors that may affect the emissions of a particular polluting facility. One factor may be the presence of a threat of government intervention to regulate emissions. To measure actual threat I use the share of regulated substances as in Harrison and Antweiler (2003). The list of regulated substances is compiled in the *Canadian Environmental Protection Act, 1999* (CEPA 1999) which is an important part of Canada's federal legislation aimed at preventing pollution and protecting the environment and human health. It has several lists that are aimed to prescribe reporting requirements for new substances. Substances that are deemed to be "toxic" under CEPA 1999 are recommended for addition to the List of Toxic Substances (Schedule 1) of the Act. As mentioned in Harrison and Antweiler (2003), chemicals in the Priority Substances List (PSL) and not in CEPA Schedule 1 can be used to measure the regulatory threat. The reason is that substances are first included in the PSL and then once a decision about the toxicity is reached[4] the substance is included in Schedule 1. However, the PSL has been reduced substantially since many substances are now included in Schedule 1.

2.2 Population Data

Population data for Canada is taken from the Gridded Population of the World (GPW) produced by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University.[5] The GPW has taken population data and transformed them into quadrilateral cells at a resolution of 2.5 arc minutes or about 5 km at the equator. The area of the cells depends on the latitude and Deichmann et al. (2001) calculate the cell size to vary from 21 km^2 at the equator to about 15 km^2 at 45°. However, Antweiler (2003) reports that since Canada's population is concentrated mostly in a narrow band between the latitudes of 42° and 53° the change in the size of the cells due to a change in latitude can be ignored.

The location of the facilities from the NPRI database is matched with the GPW data so that each firm is placed in one of the quadrilateral cells. It is very unlikely that facilities are located at the centre of a cell so I take a radius of 2 cells[6] to calculate the population around a particular manufacturing plant. Since the area around each facility is the same the population figures are essentially equivalent to the population densities. While the advantage of using the GPW data is the ability to construct radial areas to approximate the area covered by the emission, the disadvantage is that demographic and socioeconomic data from the Census cannot be matched with those quadrilateral cells. However, there is enough anecdotal evidence that suggests that concerns about "environmental justice" are not a major issue in Canada as it is in the US where research has suggested that demographics do matter (See, e.g., Arora and Cason (1999) and Hamilton (1999)).

---

[4] The process, as described by Harrison and Antweiler (2003), is not as formal in practice and regulations are introduced only if negotiations about voluntary controls fail.

[5] http://sedac.ciesin.columbia.edu/gpw/

[6] I have used various radii to calculate the population around a facility. Results, although not reported in the paper, show that the estimates are not sensitive to the choice of radius in a particular empirical specification.

Table 3: SUMMARY STATISTICS

Variable	Obs.	Mean	Std Dev	Min	Max
Log of CAC Emissions	2150	17.420	2.685	5.809	25.987
Log of HI Air Emissions	1426	18.941	3.952	1.194	29.972
Log of HI Total Emissions	1641	19.462	4.063	1.194	29.972
Fraction of CEPA-regulated CAC Emissions	2150	0.796	0.316	0	1
Fraction of CEPA-regulated HI Air Emissions	1426	0.627	0.442	0	1
Fraction of CEPA-regulated HI Total Emissions	1641	0.559	0.455	0	1
Fraction of PSL-regulated HI Air Emissions	1426	0.010	0.084	0	1
Fraction of PSL-regulated HI Total Emissions	1641	0.011	0.087	0	1
Log of Employees	2923	4.651	1.187	2.303	8.858
Log of Population	2919	10.344	1.996	2.197	13.473
Δ Log of CAC Emissions	1444	0.016	0.429	-4.615	4.344
Δ Log of HI Air Emissions	816	-0.110	1.092	-9.636	12.184
Δ Log of HI Total Emissions	907	-0.064	1.391	-12.053	16.613
Δ Fraction of CEPA-regulated CAC Emissions	1444	-0.003	0.057	-0.747	0.573
Δ Fraction of CEPA-regulated HI Air Emissions	816	-0.003	0.130	-0.997	1
Δ Fraction of CEPA-regulated HI Total Emissions	907	-0.001	0.138	-0.999	1
Δ Fraction of PSL-regulated HI Air Emissions	816	0.001	0.060	-0.956	0.982
Δ Fraction of PSL-regulated HI Total Emissions	907	-0.001	0.062	-1	0.999
Δ Log of Employees	1637	-0.016	0.212	-6.098	0.658
Δ Log of Population	1611	0.052	0.071	-0.229	0.461

I use the (natural) logarithm of the population figures for 1990 instead of the figure from contemporaneous years. Arora and Cason (1999) note that using demographic characteristics prior to the emissions release data will most likely be exogenous. We can expect that the population figures from 2005 will be affected by the emissions from the corresponding year. However, we can expect the population in 1990 to be exogenous to the emissions between 2003 and 2006. It is possible though, as explained by Arora and Cason (1999), that this assumption does not hold and that there may be some endogeneity bias if people are located in areas based on expectations of how emissions will change after 1990. I also use GPW data from 1995 to calculate the change in the population between 1990 and 1995 and use that in the difference regressions.

3 Empirical Strategy

3.1 Standard Spatial Models

The spatial dependence of pollution activities may be captured by using spatial econometric methods. The basic assumption of spatial econometrics is that observations are not independent of their location but depend on their neighbouring observations. There are two ways in which spatial dependence can be incorporated in the standard linear regression model. If we need to analyze the existence and strength of the spatial dependence then our variable of interest will have a spatially lagged dependent variable. This is referred to as a *spatial lag* model. There is also a *spatial error* model in which the spatial dependence is incorporated in the disturbance term.[7] Since my concern is the existence and strength of pollution emission spillovers I will restrict myself to the spatial lag model, also known as a mixed regressive, spatial autoregressive model.

The spatial lag model can be written as

$$y = \rho W y + X\beta + \varepsilon \tag{1}$$

where y is the emissions (level or differenced) variable, ρ is the spatial autoregressive coefficient, W is the exogenously given $n \times n$ spatial weight matrix, Wy is the spatially lagged emissions variable, X is a matrix of independent (level or differenced) variables and ε is a vector of i.i.d. error terms. The reduced form of Eq. (1) is $y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon$ where I is an $n \times n$ identity matrix. The spatial lag term Wy is, therefore, correlated with the error term. This implies that estimating the equation by OLS will be biased and inconsistent. The standard procedure is to use maximum likelihood methods for estimating the unknown parameters. The assumption for the error terms is that they follow a joint normal density function. Under this assumption, the log-likelihood function of the SAR model is:

$$\ln L(\beta, \sigma, \rho; y, X) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|I - \rho W| - \frac{1}{2\sigma^2}(y - \rho W y - X\beta)'(y - \rho W y - X\beta) \tag{2}$$

The calculation of the spatial Jacobian would complicate matters but Ord (1975) showed that it can be simplified into a function consisting of the eigenvalues ω_i of the spatial weight matrix as:

$$|I - \rho W| = \prod_{i=1}^{N}(1 - \rho\omega_i) \quad \Rightarrow \quad \ln|I - \rho W| = \sum_{i=1}^{N}(1 - \rho\omega_i)$$

Given ρ, the maximum likelihood estimators of β and σ^2 can be obtained from the first-order conditions from maximizing the log-likelihood function Eq. (2). Substituting these values in Eq. (2) will give us a concentrated log-likelihood function in ρ. Numerical optimization methods can then be used to obtain a maximum likelihood estimate of ρ. The spatial econometrics toolbox in MATLAB provided by LeSage

---

[7] See Anselin (1988) for the classic text on spatial econometrics. For more recently written introductions to this field refer to Anselin and Bera (1998), Anselin (2001) and LeSage and Pace (2009).

(1999) has been used to evaluate all these estimates.

One of the critical steps in any spatial regression estimation is the construction of the weight matrix, W. The elements of W give a notion of "distance" between each observation. This "distance" could be either geographic distance or how close one firm is to another with respect to the SIC industry code. To give an example of the latter, one could think of firms with the same industry classification code to be "closer" than firms with different industry classification codes. Geographic distance can be calculated using the Cartesian formula or the more accurate Haversine formula[8]. I use the Haversine formula to calculate the geographic distance between pairs of facilities.

The choice of the appropriate weight matrix to use, W, is also a crucial step. A common way to choose W is to obtain the maximum likelihood values of the different weight matrix specifications and choose the one with the maximum value of the likelihood function. While using the geographical distance as a measure of closeness of the facilities I have used $w_{ij} = d_{ij}^{-1}$ and $w_{ij} = d_{ij}^{-2}$ as the distance between facilities i and j. The diagonal elements, w_{ii} of W are, by convention, equal to zero. Using the inverse squared distance specification lends itself to the familiar gravity model in international trade.[9] It also assigns more weight to nearer observations. The weight matrix is also row-standardized with rows summing to one to ensure that the spatial autoregressive parameter ρ lies between –1 and +1. This also ensures that the spatial parameter ρ is comparable between models (Anselin and Bera, 1998). I have also used the SIC code to measure the "industrial distance" between plants to see if plants in closer industry classifications have similar abatement activities. Facilities that have the same 2-digit SIC codes have been assigned a distance of zero. For example, $\beta_{ij} = 0$ for two facilities i and j that have an SIC code 20 (Food and Kindred Products). Facilities that do not have the same 2-digit SIC codes but have the same 1-digit SIC code have been assigned a distance of 1 while those with different SIC codes have been assigned a distance of 2. For example, $\beta_{ij} = 1$ for a facility i that has an SIC code of, say, 20 and facility j that has an SIC code of, say, 26. If the 1-digit SIC codes are different, say one facility belongs to SIC

---

[8] Haversine formula:

$$\begin{aligned}
R &= \text{earth's radius (mean radius = 6,371km)} \\
\Delta\text{lat} &= \text{lat}_2 - \text{lat}_1 \\
\Delta\text{long} &= \text{long}_2 - \text{long}_1 \\
a &= \sin^2\left(\frac{\Delta\text{lat}}{2}\right) + \cos(\text{lat}_1).\cos(\text{lat}_2).\sin^2\left(\frac{\Delta\text{long}}{2}\right) \\
c &= 2.\text{atan2}(\sqrt{a}, \sqrt{1-a}) \\
d &= R.c
\end{aligned}$$

where the angles need to be expressed in radians.

[9] Results using the inverse distance specification are provided in Appendix A.

Table 4: LM LAG STATISTIC TESTS FOR SPATIAL DEPENDENCE

	Levels		Differences	
Dependent Variable	**W_{GEO}**	**W_{SIC}**	**W_{GEO}**	**W_{SIC}**
CAC Emissions	14.722 (0.000)	292.118 (0.000)	0.060 (0.807)	10.434 (0.001)
Health-indexed Air Emissions	0.204 (0.652)	57.820 (0.000)	0.234 (0.628)	0.267 (0.605)
Health-indexed Total Emissions	3.135 (0.077)	230.255 (0.000)	1.329 (0.249)	47.533 (0.000)

p-Values are in parentheses. Critical χ^2 values for the LM lag statistic tests are 2.71, 3.84 and 6.63 for significance levels 10%, 5% and 1% respectively. W_{GEO} and W_{SIC} are weight matrices with $w_{ij} = (\text{geographical distance}_{ij})^{-2}$ and $w_{ij} = e^{-2*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

code 2 and the other facility has an SIC code of 3 then $\beta_{ij} = 2$. The standard procedure of using the inverse of the SIC distance is not applicable in this case since $w_{ij} = \beta_{ij}^{-1} \to \infty$ and $w_{ij} = \beta_{ij}^{-2} \to \infty$ when $\beta_{ij} = 0$. Therefore, I have considered the elements of W to be $w_{ij} = e^{-\beta_{ij}}$ and $w_{ij} = e^{-2\beta_{ij}}$ where β_{ij} is the difference between the 2-digit SIC code of facilities i and j. This exponential form will ensure that when two plants have the same 2-digit SIC code, w_{ij} is defined and is equal to unity.

The w_{ij} terms should, ideally, be exogenous to the model. However, it may be argued that the location decision of a firm is endogenous. Firms that have a high pollution intensity may be located in non-urban areas while low pollution intensity facilities may be situated in a more urban setting. There may also be zoning restrictions as a form of local regulation that may affect high pollution intensity firms, as discussed in Antweiler (2003). However, using differenced variables should mitigate this problem. By differencing the emission level variable we get the rate of change of emission. This should enable us to deal with the endogeneity of the location decision.

In keeping with standard procedure, I first estimate models using OLS and use the results as a base for comparing the spatial models. With regard to dependent variables, I use the level values to analyze the spatial dependence in the environmental performance but then use the differenced values to account for any endogeneity in terms of location. Comparing the results of the spatial models with the OLS results provide an indication of how strong the spatial interactions may be. All spatial and OLS regression models are estimated using the Econometrics Toolbox for MATLAB.[10]

---

[10] The Econometrics Toolbox for MATLAB can be obtained at www.spatial-econometrics.com.

3.2 Extension of Standard Spatial Models

The standard spatial model Eq. (1) can be modified to incorporate spatial regression models with two or more weight matrices. In the previous section I have considered two channels through which pollution spillovers may work, viz. the geographical distance and the SIC code "industry" distance. They were, however, modelled separately. Modifying Eq. (1) to incorporate two spatial weight matrices can be used to analyze separate influences in the same model. The extension of the standard spatial model can then be written as

$$y = \rho_{\text{GEO}} W_{\text{GEO}} y + \rho_{\text{SIC}} W_{\text{SIC}} y + \rho_{\text{INT}} W_{\text{INT}} y + X\beta + \varepsilon \tag{3}$$

where W_{GEO} is used to capture the effect of geographic distance between neighbouring facilities and W_{SIC} captures the effect of "industry" distance. Compared to Eq. (1) the modified SAR model Eq. (3) needs a slight modification to find the estimates. The log-likelihood becomes

$$\begin{aligned} \ln L(\beta, \sigma, \rho_{\text{GEO}}, \rho_{\text{SIC}}; y, X) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|I - \rho_{\text{GEO}} W_{\text{GEO}} \rho_{\text{SIC}} W_{\text{SIC}}| \\ -\frac{1}{2\sigma^2}\left(y - \rho_{\text{GEO}} W_{\text{GEO}} y - \rho_{\text{SIC}} W_{\text{SIC}} y - X\beta\right)' &\left(y - \rho_{\text{GEO}} W_{\text{GEO}} y - \rho_{\text{SIC}} W_{\text{SIC}} y - X\beta\right) \end{aligned} \tag{4}$$

where the change in maximum likelihood estimation compared to the standard SAR model is the optimization problem involving the two spatial autoregressive parameters ρ_{GEO} and ρ_{SIC} . The MATLAB functions are obtained from the spatial econometrics toolbox.[11]

The specification with two weight matrices will be useful to test the strength and magnitude of the spatial dependence in emission between facilities that are geographically closer and in the same industry. The spatial dependency arising from the geographical distance is of a (geographically) localized nature in the sense that the emissions of a facility may be affected by the emissions of other facilities that are in its vicinity, irrespective of the industry the other facility belongs to. The other kind of spatial dependency that arises from how close or far apart the facilities are with respect to their respective industries occurs irrespective of geographical factors. While there is no *a priori* reason to rank the strength and magnitude of the two kinds of spatial dependencies we might suspect the latter effect to be stronger than the former.

---

[11] The modified MATLAB code for the SAR model with two weight matrices was provided by Donald J. Lacombe.

4 Results

4.1 Standard Spatial Models

The regression results look at both on-site releases as well as changes in the releases to see if there are any spatial dependencies in emissions. Each table considers a different emissions variable. The first table, Table 5, reports the regression results of the aggregation of all the Criteria Air Contaminants. Tables 6 and 7 contain the regression results of the health-indexed air and total emissions respectively. The results of the estimation procedure for on-site emission levels are given in columns (1), (2) and (3) of Tables 5, 6 and 7. The results for changes in emissions are reported in columns (4), (5) and (6) in each table. The base regression, to which all the other results are compared, is the OLS regression (labelled OLS in each table). Spatial regressions are indicated by $\mathrm{W}_{\mathrm{GEO}}$ and $\mathrm{W}_{\mathrm{SIC}}$ using the geographic and industry distance weight matrices, respectively. The simple SAR model with only one spatial weight matrix is considered initially. The elements of W_{GEO} are the inverse squared distance specification while W_{SIC} is the SIC code industrial distance specification described in the previous section. All OLS, $\mathrm{W}_{\mathrm{GEO}}$ and $\mathrm{W}_{\mathrm{SIC}}$ regressions have province dummies to control for province fixed effects and the OLS and $\mathrm{W}_{\mathrm{GEO}}$ regressions also have two-digit industry SIC dummies to control for industry fixed effects.

The spatial autoregressive coefficient ρ is the parameter of interest in measuring the presence and strength of the effect of neighbouring facilities' emissions and emission changes on the facility under consideration. Most of the spatial regression results show that the spatial dependence in the dependent variable is positive and significant though the strength of ρ depends on the emissions variable as well as the weight matrix considered. While the spatial dependence for CAC emissions is positive but not significant for the emissions variable when we use the geographical distance W_{GEO} as the weight matrix, the effect is positive and highly significant for the SIC distance matrix W_{SIC}. The value of ρ is also positive and very significant for CAC emission changes which suggests that polluting plants that are closer together both in terms of geographical distance as well as SIC codes tend to reduce their CAC emissions together. The magnitude of ρ is, however, lower for changes in emissions when compared to emission levels. Results for health-indexed air emissions are also similar except that ρ is surprisingly negative and significant when we use SIC distance as the spatial weight matrix in the regression for emission differences. This negative spatial dependence suggests that facilities that lower their health-indexed air emissions are surrounded by facilities that increase theirs thus creating a checkerboard-type

Table 5: OLS and Spatial Regression Models for CAC Emissions

Variable	**Emission Levels**			**Emission Differences**†		
	OLS (1)	$\mathbf{W}_{\mathbf{GEO}}$ (2)	$\mathbf{W}_{\mathbf{SIC}}$ (3)	**OLS** (4)	$\mathbf{W}_{\mathbf{GEO}}$ (5)	$\mathbf{W}_{\mathbf{SIC}}$ (6)
Intercept	15.510^{a}	13.862^{a}	2.189	-0.012	-0.012	-0.009
	(0.473)	(0.809)	(1.384)	(0.013)	(0.013)	(0.013)
Fraction of CEPA-regulated output	-1.321^{a}	-1.270^{a}	-1.509^{a}	-2.119^{a}	-2.120^{a}	-2.120^{a}
	(0.153)	(0.155)	(0.148)	(0.112)	(0.112)	(0.112)
Log of Employees	1.000^{a}	0.991^{a}	0.858^{a}	0.522^{a}	0.522^{a}	0.527^{a}
	(0.042)	(0.042)	(0.039)	(0.073)	(0.073)	(0.073)
Log of Population (1990)	-0.174^{a}	-0.163^{a}	-0.150^{a}	0.204	0.203	0.207
	(0.025)	(0.026)	(0.023)	(0.144)	(0.143)	(0.143)
ρ		0.087^{a}	0.816^{a}		0.006^{a}	0.571^{c}
		(0.024)	(0.080)		(0.002)	(0.201)
Province dummies	Yes	Yes	Yes			
SIC dummies	Yes	Yes	No			
Adjusted R^2	0.432	0.434	0.346	0.168	0.168	0.167
Observations	2150	2150	2150	2003	2003	2003
Log-Likelihood		-3798	-3891		-408	-405
Spatial Multiplier, $1/(1-\rho)$		1.095	5.435		1.006	2.331

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (CAC Air Emissions). Standard errors are in parentheses.
Specifications $\mathrm{W}_{\mathrm{GEO}}$ and $\mathrm{W}_{\mathrm{SIC}}$ are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-2}$ and $w_{ij} = e^{-2*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (4), (5) and (6) should be interpreted as being differences.

situation. However, results for health-indexed total emissions show that spatial dependence is positive and significant for both emission levels and changes in emissions. It is negative, albeit insignificant, for changes in emissions when the spatial regression includes the SIC distance weight matrix W_{SIC}.

Results from the spatial regressions show that the coefficients are not too different from the OLS regressions. However, the coefficients from OLS results are less significant than those from the spatial regressions. Including the spatially lagged emission variables in the regressions has the effect of strengthening the effect of the other explanatory variables. However, this effect is not very strong. The regressors across all the regression results tables are common, apart from the fraction of PSL-regulated output in Table 5 since none of the CAC substances are present in the PSL. Since the fraction of CEPA-regulated output and PSL-regulated output are a measure of the actual regulation and perceived threat respectively we should expect the effect of these two variables to be negative on the emission level. The higher the regulation or regulatory threat the lower should be the emissions. This prediction holds for the CAC emissions across the OLS and spatial specifications with the effect being slightly lower for the

Table 6: OLS and Spatial Regression Models for Health-Indexed Air Emissions

	Emission Levels			**Emission Differences**†		
Variable	**OLS** (1)	**W$_{GEO}$** (2)	**W$_{SIC}$** (3)	**OLS** (4)	**W$_{GEO}$** (5)	**W$_{SIC}$** (6)
Intercept	16.675^a (0.973)	16.447^a (1.027)	4.043 (2.263)	-0.013 (0.035)	-0.013 (0.035)	-0.027 (0.035)
Fraction of CEPA-regulated output	-2.251^a (0.222)	-2.248^a (0.220)	-2.470^a (0.201)	-3.948^a (0.195)	-3.949^a (0.194)	-3.951^a (0.194)
Fraction of PSL-regulated output	-4.951^a (1.072)	-4.942^a (1.06)	-5.128^a (1.066)	-5.449^a (0.530)	-5.455^a (0.529)	-5.450^a (0.529)
Log of Employees	1.077^a (0.084)	1.076^a (0.084)	0.980^a (0.073)	0.374 (0.192)	0.375 (0.191)	0.375 (0.191)
Log of Population (1990)	-0.241^a (0.052)	-0.239^a (0.052)	-0.243^a (0.048)	-0.769 (0.411)	-0.759 (0.410)	-0.769 (0.410)
ρ		0.012^a (0.002)	0.726^a (0.130)		0.011^a (0.002)	-0.193^c (0.078)
Province dummies	Yes	Yes	Yes			
SIC dummies	Yes	Yes	No			
Adjusted R^2	0.301	0.301	0.262	0.257	0.257	0.257
Observations	1426	1426	1426	1305	1305	1305
Log-Likelihood		-3218	-3250		-1285	-1285
Spatial Multiplier, $1/(1-\rho)$		1.012	3.650		1.011	0.838

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (Health-Indexed Air Emissions). Standard errors are in parentheses.
Specifications W$_{GEO}$ and W$_{SIC}$ are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-2}$ and $w_{ij} = e^{-2*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (4), (5) and (6) should be interpreted as being differences.

latter (as seen by the lower value) compared to the OLS result. The effect can also be seen in the case of the health-indexed total and air emissions variables in Tables 6 and 7.

The change in actual regulation as well as regulatory threat should also have a negative effect on the change in emissions. If the change in regulation or regulatory threat is positive we should expect the the change in emissions to be negative. This is reflected in the negative coefficient for the variables describing the fraction of CEPA-regulated and PSL-regulated outputs across the OLS and spatial regressions for all the emissions variables under consideration. The effect is negative and highly significant throughout.

The number of employees in a facility is taken as a proxy for the scale of operation so we expect the effect on the emission variables to be positive. More employees in a facility reflect a bigger operating scale and therefore, more emissions. The results show that this holds in all specifications with the elasticity being close to or greater than one in the level of emissions. Changes in the number of employees also have the same effect on changes in emission as shown by the positive coefficient in columns (4),

Table 7: OLS and Spatial Regression Models for Health-Indexed Total Emissions

	Emission Levels			**Emission Differences**†		
Variable	**OLS** (1)	$\mathbf{W_{GEO}}$ (2)	$\mathbf{W_{SIC}}$ (3)	**OLS** (4)	$\mathbf{W_{GEO}}$ (5)	$\mathbf{W_{SIC}}$ (6)
Intercept	17.586^a	16.695^a	0.951	-0.057	-0.055	-0.004
	(0.899)	(0.959)	(4.346)	(0.040)	(0.040)	(0.041)
Fraction of CEPA-regulated output	-2.625^a	-2.614^a	-2.752^a	-4.468^a	-4.468^a	-4.459^a
	(0.200)	(0.199)	(0.156)	(0.228)	(0.227)	(0.226)
Fraction of PSL-regulated output	-7.534^a	-7.515^a	-7.510^a	-6.419^a	-6.413^a	-6.367^a
	(0.938)	(0.928)	(0.913)	(0.600)	(0.599)	(0.595)
Log of Employees	0.976^a	0.973^a	0.903^a	0.340	0.344	0.345
	(0.076)	(0.075)	(0.057)	(0.223)	(0.223)	(0.221)
Log of Population (1990)	-0.209^a	-0.198^a	-0.225^a	-0.438	-0.435	-0.337
	(0.047)	(0.048)	(0.041)	(0.460)	(0.459)	(0.457)
ρ		0.044^a	0.903^a		0.026^a	0.756^a
		(0.003)	(0.222)		(0.002)	(0.131)
Province dummies	Yes	Yes	Yes			
SIC dummies	Yes	Yes	No			
Adjusted R^2	0.370	0.370	0.289	0.237	0.237	0.240
Observations	1641	1641	1641	1516	1516	1516
Log-Likelihood		-3664	-3701		-1776	-1768
Spatial Multiplier, $1/(1-\rho)$		1.046	10.309		1.027	4.098

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (Health-Indexed Total Emissions). Standard errors are in parentheses.
Specifications W_{GEO} and W_{SIC} are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-2}$ and $w_{ij} = e^{-2*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (4), (5) and (6) should be interpreted as being differences.

(5) and (6) in the regression results tables. This means that if the scale, as measured by the number of employees, increases the emission also increases. However, the effect is not statistically significant when the dependent variables are the health-indexed emission variables.

The effect of population has the expected negative effect on the emission of a facility. This effect is also significant across all the specifications and emission variables. A higher population surrounding the polluting facility will have a greater environmental exposure while we can also expect the presence of a strong consumer pressure group in more populated areas. This result is similar to findings in related literature. We would expect changes in population density to have a positive effect on pollution abatement activities or, in other words, a negative effect effect on changes in a facility's emission. My results indicate that this effect holds in the case of health-indexed air emissions but is of the opposite sign for CAC emissions and health-indexed total emissions. In almost all cases however, the effect is statistically insignificant.

Table 8: ELASTICITY OF EMISSION VARIABLES

Variable	**CAC** **W_{GEO}** (1)	 **W_{SIC}** (2)	**HI Air** **W_{GEO}** (3)	 **W_{SIC}** (4)	**HI Total** **W_{GEO}** (5)	 **W_{SIC}** (6)
Share of CEPA-regulated output	-1.107	-3.527	-1.191	-1.728	-1.869	-3.209
Share of PSL-regulated output			-0.063	-0.097	-0.100	-0.178
Log of Employees	1.085	2.125	1.556	2.559	1.196	2.303
Log of Population	-0.179	-0.273	-0.140		-0.191	-0.306

Note: $w_{ij} = (\text{geographical distance}_{ij})^{-2}$ and $w_{ij} = e^{-2*(\text{SIC distance})_{ij}}$ are elements in the weight matrices W_{GEO} and W_{SIC} respectively

After having analyzed the existence and strength of spillovers in pollution levels as well as abatement as shown by the positive and significant spatial lag parameter ρ we now turn to interpreting the spatial lag model. The effect of a change in any of the explanatory variables on the change in emissions variable at a particular facility is the sum of the direct impact as well as the induced impact and is referred to as the *spatial multiplier*. As shown in Kim et al. (2003), the spatial multiplier is expressed as $1/(1-\rho)$ if there were a unit change in every location. This means that a change in any of the explanatory variables in neighbouring facilities will have an effect on the emissions of facility i. But how much will the effect be? To find that out the elasticity of the emissions (evaluated at the mean) from a small change in the fraction of the regulated output we use

$$\varepsilon_{x_k} = \frac{\beta_k}{(1-\rho)} \bar{x}_k$$

where $\bar{x}_k$ is the mean value of x_k. The elasticity of the emissions from a small change in the neighbouring population or employees is

$$\varepsilon_{x_k} = \frac{\beta_k}{(1-\rho)}$$

For example, the elasticity of CAC emission levels from a small change in the fraction of CEPA-regulated output in column (3) of Table 5 is 3.535 while it is 0.273 and 2.125 for changes in population and employees respectively. If we compare these elasticity values with those obtained from the OLS regression we can see that they are significantly higher. We can, therefore, conclude that spatial dependence has a significant effect on the emissions of facilities. The elasticities for all the significant variables are calculated in Table 8.

4.2 Extension of Standard Spatial Models

Results for the modified SAR model with two weight matrices, Eq. (3), are presented in Table 9. There are several observations to be made from comparing results of this specification with the simple OLS and the standard SAR models. Firstly, the SIC distance parameter estimates ρ_{SIC} for all the emission variables, except for difference health-indexed air emission, is positive and significant. The geographical distance parameter estimates ρ_{GEO}, while positive throughout, are not always significant. This differs from the simple SAR model in which it was positive and significant for all the emission variables. We can, therefore, conclude that the SIC industry distance may capture the spatial dependencies in pollution better than simply the geographical distance and this effect persists even after including both those spatial dependencies in the modified SAR model. Most of the ρ_{SIC} estimates are significant at the 99% level which shows that this specification is superior to ordinary least squares. Secondly, the variation in the dependent variable is explained to almost the same extent by both the OLS models and the modified spatial models and for both emission levels and emission differences.

These results show that there is a much stronger spatial dependence that arises from similar industrial facilities as compared to the spatial dependence that may arise from how far one facility is, geographically, with respect to its neighbours. Facilities in similar industries would tend to have similar emissions due to technological similarities from using similar processes. The fact that there are positive spatial autoregressive parameters ρ_{GEO} and ρ_{SIC} suggests that changes in emissions by a particular facility is, on average, being positively influenced by its neighbours, in particular, the "SIC industry" neighbours. So, for example, there is a reduction in the emissions of facility i when facility j in a "close enough" industry code also reduces its emissions. This, however, goes the other way as well. A positive spatial dependence implies that an increase in the emissions of facility i when the emissions of facility j, in a "close enough" industry code, increase.

5 Discussion of Results

The results in the previous section indicate the presence of strong and positive spatial dependence in pollution emissions. While this holds for both emission levels as well as emission differences, the latter is a more robust result because of the econometric superiority of using differences over levels. Even after accounting for facility-specific and location-specific factors we observe a strong spatial dependence.

Table 9: SPATIAL REGRESSION MODELS WITH TWO WEIGHT MATRICES

Variable	**CAC Air** Level (1)	Diff.† (2)	**HI Air** Level (3)	Diff.† (4)	**HI Total** Level (5)	Diff.† (6)
Intercept	1.047^c (0.420)	-0.008 (0.013)	3.813^a (0.875)	-0.027 (0.035)	0.516 (0.798)	-0.001 (0.040)
Fraction of CEPA-regulated output	-1.454^a (0.150)	-2.120^a (0.112)	-2.467^a (0.205)	-3.952^a (0.194)	-2.752^a (0.180)	-4.459^a (0.225)
Fraction of PSL-regulated output			-5.122^a (1.070)	-5.456^a (0.529)	-7.516^a (0.934)	-6.361^a (0.595)
Log of Employees	0.854^a (0.039)	0.527^a (0.073)	0.980^a (0.076)	0.375^c (0.191)	0.905^a (0.067)	0.349 (0.221)
Log of Population (1990)	-0.138^a (0.023)	0.206 (0.143)	-0.236^a (0.048)	-0.758 (0.410)	-0.211^a (0.044)	-0.331 (0.456)
ρ_{GEO}	0.092^a (0.023)	0.003 (0.027)	0.026 (0.029)	0.012 (0.027)	0.054^c (0.026)	0.023 (0.027)
ρ_{SIC}	0.780^a (0.060)	0.582^c (0.209)	0.710^a (0.113)	-0.202 (0.370)	0.866^a (0.052)	0.780^a (0.130)
Province dummies	Yes	No	Yes	No	Yes	No
Adjusted R^2	0.398	0.171	0.280	0.258	0.354	0.249
Observations	2150	2003	1426	1305	1641	1516
Log-Likelihood	-9805	-5851	-6885	-4554	-7997	-5678

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (Emissions variable). Standard errors are in parentheses.
†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (2), (4) and (6) should be interpreted as being differences.

The consistent theme in the results is that spillovers are more localized when geographic distance is used. The spillovers are more global in scope when the industry distance metric is considered. We can conclude this from analyzing the magnitude of the spatial autoregressive parameter ρ which is larger for ρ_{SIC} than it is for ρ_{GEO} when we use both the simple spatial autoregressive regressive specification and the extended SAR model where both ρ_{SIC} and ρ_{GEO} are estimated simultaneously.

Griliches (1979) considered the issue of spillovers and posited various hypotheses for the reasons that spillovers might exist, specifically in R&D. They may be the result of horizontal, technological or vertical spillovers.[12] However, he did not consider the importance of geography. The importance of physical proximity has been recognized by recent researchers and is one of the cornerstones of spatial econometrics.

My results indicate that while physical proximity between facilities is important, the effect is over-

---

[12] Horizontal spillovers exist between firms in the same product market. Technological spillovers result from firms conducting similar research. Vertical spillovers exist between firms that are suppliers or retailers.

shadowed by the technological and horizontal spillovers that I capture using the industry metric. The industry metric depends on the SIC code of facilities and it captures the spillovers between firms that produce for the same market as well as firms that may be conducting similar research. Technological similarities between two firms in the same SIC code should be much higher than that of two firms that belong to different sectors within the manufacturing industry. These technological spillovers maintain their significance when I consider emission changes. This points to similar environmental performance between firms in the same sector and some form of peer effect. This effect goes beyond the borders of Canadian provinces and, hence, against some of the findings in the recent literature on spatial effects in environmental performance.

While geographical proximity cannot be discounted in my results, their importance is substantially weaker and corroborates recent findings on spatial dependence in environmental performance. There might be several reasons behind spatial dependence caused by geographical proximity. Facilities in the same region may have a similar environmental performance due to the presence of "demonstration effects" which is caused by facilities to be cleaner under peer pressure from other cleaner firms. The positive spatial autoregressive parameter ρ_{GEO} indicates that there is a possibility of the existence of "hotspots" where dirtier firms agglomerate.

6 Conclusion

In this paper I have used the spatial lag model and its extension to capture the spatial dependency of pollution emission levels and pollution emission changes between neighbouring facilities. Pollution emission changes are measured by taking the difference in emissions of Criteria Air Contaminants, health-indexed air emissions and health-indexed total emissions. I have weighted emissions with their toxicity because not all substances have the same impact on human health. The emissions data are from a comprehensive set of manufacturing plants located in Canada. Results show that spatial dependence does exist in the emissions of manufacturing plants and is positive indicating that emission activities by neighbouring plants are, on average, similar. Using differences in emissions accounts for plant-specific and firm-specific effects. It also ensures that province and industry-specific factors are accounted for. Taking differences also accounts for the endogeneity that may exist in the location of facilities. This is an improvement over the existing literature when looking for spillovers in pollution emitted by manufacturing facilities.

I have also used the estimates from the spatial regressions to construct spatial multipliers to interpret the implications of the spatial lag model. I show that the results are much stronger when we incorporate spatial effects compared to non-spatial models. However, the strength of the spatial dependence, as measured by the spatial dependence parameter ρ, is much stronger when we consider emission levels as compared to emission differences. Since studying emission differences is econometrically superior to looking at just emission levels we can conclude that spatial dependencies in pollution abatement, even though it does exist, may not be very strong when we consider geographical distance in the spatial weight matrix but appears to be much stronger when the SIC industry spatial weight matrix is used. This shows quite clearly that using only geographical distances for analyzing spillovers in a setting where individual facilities are the units of observation may be a very simplistic and insufficient way to capture linkages.

There is also much scope for further research in analyzing environmental performance by using spatial econometrics. I have used the simple SAR(1) model and its modification to study spillovers. Its highly parsimonious and fully parameterized nature is an advantage but also an Achilles heel. The limitations of the SAR model, as discussed by Pinkse and Slade (2010), include the fact that the relationship may be non-linear, the error term and independent variables may be dependent and that the entire spatial dependence structure can be represented by the the spatial lag parameter ρ. Using different specifications would also serve as a test of robustness for the SAR model. There is also a case of combining the geographic distances with industry measures of distance, like the SIC codes, to compute a more sophisticated weight matrix that captures more aspects of "distance" than just the physical distance. In this paper, I have accomplished a different version in that I have considered two separate weight matrices to indicate the two separate channels through which spatial dependecy may arise. The availability of panel data, as is the case with most toxics release inventories, is also suited for using spatial panel methods for further analysis. The disadvantage of using panel data is the change that might occur over a period of time in terms of new substances added or old substances removed from the list of reporting criteria as well as new facilities being asked to report. However, spatial panel econometrics is an area of active research and using these new methods would provide more robust and innovative ways of looking at spatial dependencies in emissions or other variables of interest.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L. (2001). Spatial Econometrics. In Baltagi, B., editor, *A Companion to Theoretical Econometrics*, pages 310–330. Blackwell Oxford.

Anselin, L. and Bera, A. K. (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. *Handbook of applied economic statistics*, pages 237–89.

Antweiler, W. (2003). How effective is green regulatory threat? *American Economic Review*, 93(2):436–441.

Antweiler, W. and Harrison, K. (2003). Toxic Release Inventories and Green Consumerism: Empirical Evidence from Canada. *Canadian Journal of Economics*, 36(2):495–520.

Arora, S. and Cason, T. N. (1999). Do community characteristics influence environmental outcomes? *Southern Economic Journal*, 65(4):691–716.

Deichmann, U., Balk, D., and Yetman, G. (2001). Transforming Population Data for Interdisciplinary Usages: From Census to Grid. *Washington DC: Center for International Earth Science Information Network*. available at http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf.

Gray, W. and Shadbegian, R. (2007). The Environmental Performance of Polluting Plants: A Spatial Analysis. *Journal of Regional Science*, 47(1):63–84.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 10(1):92–116.

Hamilton, J. (1999). Exercising property rights to pollute: Do cancer risks and politics affect plant emission reductions? *Journal of Risk and Uncertainty*, 18(2):105–124.

Harrison, K. and Antweiler, W. (2003). Incentives for Pollution Abatement: Regulation, Regulatory Threats, and Non-governmental Pressures. *Journal of Policy Analysis and Management*, 22(3):361–82.

Harrison, K. and Hoberg, G. (1994). *Risk, Science, and Politics: Regulating Toxic Substances in Canada and the United States*. McGill Queens Univ Pr.

Hettige, H., Lucas, R., and Wheeler, D. (1992). The Toxic Intensity of Industrial Production: Global Patterns, Trends, and Trade Policy. *American Economic Review*, 82(2):478–481.

Horvath, A., Hendrickson, C., Lave, L., McMichael, F., and Wu, T. (1995). Toxic emissions indices for green design and inventory. *Environmental Science and Technology*, 29(2):86.

Kim, C., Phipps, T., and Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45(1):24–39.

LeSage, J. (1999). Econometrics Toolbox. [online] http://www.spatial-econometrics.com/.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

Olewiler, N. and Dawson, K. (1998). Analysis of National Pollutant Release Inventory Data on Toxic Emissions by Industry. Technical report.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Pinkse, J. and Slade, M. E. (2010). The Future of Spatial Econometrics. *Journal of Regional Science*, 50(1):103–117.

A Appendix

A.1 Spatial Dependence Tests

Table 10: LM Lag Statistic Tests for Spatial Dependence

	Levels		**Differences**	
Dependent Variable	$\mathbf{W_{GEO}}$	$\mathbf{W_{SIC}}$	$\mathbf{W_{GEO}}$	$\mathbf{W_{SIC}}$
CAC Emissions	14.722 (0.000)	292.118 (0.000)	0.060 (0.807)	10.434 (0.001)
Health-indexed Air Emissions	0.204 (0.652)	57.82 (0.000)	0.234 (0.628)	0.267 (0.605)
Health-indexed Total Emissions	3.135 (0.077)	230.255 (0.000)	1.329 (0.249)	47.533 (0.000)

p-Values are in parentheses. Critical χ^2 values for the LM lag statistic tests are 2.71, 3.84 and 6.63 for significance levels 10%, 5% and 1% respectively. W1 and W2 are spatial weight matrices with $w_{ij} = (\text{geographical distance}_{ij})^{-1}$ and $w_{ij} = e^{-1*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

A.2 Spatial Regression Results

Table 11: SPATIAL REGRESSION MODELS FOR CAC EMISSIONS

Variable	Levels $\mathbf{W}_{\text{GEO}}$ (1)	Levels $\mathbf{W}_{\text{SIC}}$ (2)	Differences† $\mathbf{W}_{\text{GEO}}$ (3)	Differences† $\mathbf{W}_{\text{SIC}}$ (4)
Intercept	12.123^a (1.817)	-0.282 (1.362)	-0.012 (0.013)	-0.008 (0.013)
Fraction of CEPA-regulated output	-1.276^a (0.158)	-1.639^a (0.153)	-2.120^a (0.112)	-2.120^a (0.112)
Log of Employees	0.990^a (0.042)	0.889^a (0.040)	0.522^a (0.073)	0.525^a (0.073)
Log of Population (1990)	-0.161^a (0.028)	-0.177^a (0.024)	0.204 (0.143)	0.208 (0.143)
ρ	0.187^c (0.081)	0.966^a (0.071)	0.005 (0.004)	0.612^a (0.143)
Province dummies	Yes	Yes		
SIC dummies	Yes	No		
Adjusted R^2	0.4343	0.3658	0.1679	0.1675
Observations	2150	2150	2003	2003
Log-Likelihood	-3798	-3914	-408	-407
Spatial Multiplier, $1/(1-\rho)$	1.230	29.412	1.005	2.577

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (CAC Air Emissions). Standard errors are in parentheses. Specifications W1 and W2 are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-1}$ and $w_{ij} = e^{-1*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.
†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (3) and (4) should be interpreted as being differences.

Table 12: Spatial Regression Models for Health-Indexed Air Emissions

	Levels		**Differences**†	
Variable	$\mathbf{W}_{\mathbf{GEO}}$ (1)	$\mathbf{W}_{\mathbf{SIC}}$ (2)	$\mathbf{W}_{\mathbf{GEO}}$ (3)	$\mathbf{W}_{\mathbf{SIC}}$ (4)
Intercept	15.131^a (1.727)	1.390 (3.154)	-0.011 (0.035)	-0.066 (0.073)
Fraction of CEPA-regulated output	-2.238^a (0.221)	-2.617^a (0.207)	-3.950^a (0.194)	-3.951^a (0.194)
Fraction of PSL-regulated output	-4.917^a (1.06)	-5.436^a (1.078)	-5.455^a (0.529)	-5.449^a (0.529)
Log of Employees	1.071^a (0.084)	1.041^a (0.077)	0.374 (0.191)	0.375^c (0.191)
Log of Population (1990)	-0.229^a (0.054)	-0.256^a (0.049)	-0.761 (0.41)	-0.764 (0.41)
ρ	0.080 (0.065)	0.867^a (0.160)	0.029^a (0.006)	-0.739 (0.889)
Province dummies	Yes	Yes		
SIC dummies	Yes	No		
Adjusted R^2	0.3014	0.2649	0.2571	0.2569
Observations	1426	1426	1305	1305
Log-Likelihood	-3218	-3258	-1285	-1285
Spatial Multiplier, $1/(1-\rho)$	1.087	7.519	1.030	0.575

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (Health-Indexed Air Emissions). Standard errors are in parentheses. Specifications W1 and W2 are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-1}$ and $w_{ij} = e^{-1*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (3) and (4) should be interpreted as being differences.

Table 13: SPATIAL REGRESSION MODELS FOR HEALTH-INDEXED TOTAL EMISSIONS

	Levels		**Differences**†	
Variable	$\mathbf{W_{GEO}}$ (1)	$\mathbf{W_{SIC}}$ (2)	$\mathbf{W_{GEO}}$ (3)	$\mathbf{W_{SIC}}$ (4)
Intercept	14.341^a	-0.154	-0.048	0.0001
	(1.641)	(1.202)	(0.041)	(0.043)
Fraction of CEPA-regulated output	-2.601^a	-2.992^a	-4.470^a	-4.468^a
	(0.199)	(0.183)	(0.227)	(0.227)
Fraction of PSL-regulated output	-7.501^a	-8.012^a	-6.410^a	-6.402^a
	(0.927)	(0.949)	(0.599)	(0.598)
Log of Employees	0.967^a	1.003^a	0.340	0.343
	(0.075)	(0.068)	(0.222)	(0.222)
Log of Population (1990)	-0.184^a	-0.255^a	-0.435	-0.391
	(0.049)	(0.045)	(0.459)	(0.458)
ρ	0.162^c	0.966^a	0.096	0.755^a
	(0.061)	(0.045)	(0.067)	(0.179)
Province dummies	Yes	Yes		
SIC dummies	Yes	No		
Adjusted R^2	0.3710	0.3158	0.2367	0.2377
Observations	1641	1641	1516	1516
Log-Likelihood	-3663	-3723	-1776	-1774
Spatial Multiplier, $1/(1-\rho)$	1.193	29.412	1.106	4.082

Significance at the 1%, 5% and 10% levels are denoted by ^{a}, ^{b} and ^{c} respectively.
The dependent variable is Log (Health-Indexed Total Emissions). Standard errors are in parentheses. Specifications W1 and W2 are spatial regressions with $w_{ij} = (\text{geographical distance}_{ij})^{-1}$ and $w_{ij} = e^{-1*(\text{SIC distance})_{ij}}$ as elements in the weight matrices respectively.

†: For conserving space I have used the same variable names to report results from the difference specification. All regressors in columns (3) and (4) should be interpreted as being differences.