Development Discussion Papers

Unveiling the Income-Environment Relationship: An Exploration into the Determinants of Environmental Quality

Nazrul Islam, Jeffrey Vincent,
and Theodore Panayotou

Development Discussion Paper No. 701
May 1999

Harvard Institute for
International Development

HARVARD UNIVERSITY

HIID Development Discussion Paper no. 701

Unveiling the Income-Environment Relationship: An Exploration into the Determinants of Environmental Quality

Nazrul Islam, Jeffrey Vincent, and Theodore Panayotou*

Abstract

This paper decomposes the reduced form income-environment relationship into its structural sources. It identifies a level effect, a composition effect, and an abatement effect. This decomposition is implemented using global data on suspended particulate matter in the air. The level and composition effects are found to conform to their a priori expected linear and quadratic forms, respectively. The abatement effect is found to be generally downward sloping, though not exactly conforming to the expected inverted-J shape.

Keywords: environmental Kuznets curve, air pollution

JEL codes: O0, Q0, Z0

Nazrul Islam is an Assistant Professor in the Department of Economics at Emory University.

Jeffrey R. Vincent is a Fellow of the Institute at the Harvard Institute for International Development (HIID). His research interests include forest economics, national accounts and the environment, and environmental policy issues in transition economies.

Theodore Panayotou is a Fellow of the Institute at the Harvard Institute for International Development (HIID), Director of HIID's International Environment Program; Project Director for Central and Eastern Europe, Environmental Economics and Policy Project; and Lecturer on Economics, Harvard University. His research interests include natural resource economics (forestry, land and water, fisheries, energy, minerals and environment), agricultural economics, development economics, capital theory and growth, optimal control and duality theory, social welfare economics and policy analysis.

---

*This paper was presented earlier at seminars organized by HIID, HIID/ADB, and Department of Economics, Emory University. We would like to thank the participants of these seminars for their comments. However, we are alone responsible for remaining errors. Research assistance by Xiang Yu is gratefully acknowledged. The views expressed in this paper are authors' own and are not to be ascribed to the organizations they represent.

Unveiling the Income-Environment Relationship: An Exploration into the Determinants of Environmental Quality

Nazrul Islam, Jeffrey Vincent, and Theodore Panayotou

1. Introduction

In recent years, the relationship between economic growth and environmental quality has drawn considerable attention from researchers. This has been mainly driven by the concern whether the world as a whole will be able to sustain rapid economic growth without either running into natural resource constraints or irreparably damaging the environment. Since the most prevalent indicator of economic growth is *income level,* the question has been further concretized into one of the relationship between environmental quality and the level of income. A whole body of literature has now emerged addressing this narrowed down question. Shafik and Bandyopadhay (1992), Hettige, Lucas and Wheeler (1992), Selden and Song (1994), Grossman and Krueger (1995), Panayotou (1995), Vincent (1997), Kauffmann, Davidsdotter, and Garnham (1995) are a few examples. In these works, various measures of environmental quality have been, in many different ways, regressed on income in order to ascertain the nature of this relationship. Broadly, we refer to these as studies of the income-environment relationship (IER). Both cross-sectional and panel data have been used in this research. For some environmental indicators, the obtained relationship displays an inverted-U shaped pattern, thereby giving rise to the hypothesis of Environmental Kuznets Curve (EKC).

So far, the IER has been considered in its reduced form. As Grossman and Krueger explained, the reduced form IER gives the 'net effect' of income on the environment, and documentation of this relationship is 'an important first step' in the study of environment-income relationship. The primary reason for the reduced form approach is its easy implementability. This, in turn, has two aspects. *First,* at any point in time, environmental quality is the outcome of the interplay between pollution-generating activities, on the one hand, and pollution abatement, on the other. Yet, despite some attempts, formulation of a theoretical framework for estimating the structural relationships in these two opposing directions was not available. *Second,* even if some *ad hoc* specifications of these relationships were postulated, data for identifying them separately were difficult to obtain. It is quite understandable, therefore, why the reduced form approach has appealed to so many researchers. However, as a result of their work, we now have some understanding of the empirical nature of the income-environment relationship.

An important limitation of the reduced form approach is that it leaves us in the dark as to why the observed relationship exists. In this reduced form relationship, income serves as an omnibus representative of many factors. Hence, the obtained relationship can not throw light on their separate effects. As a corollary, it leaves us, by and large, without any handle on the relationship, any way to influence it, and, to put it more directly, any indication of the scope for policy intervention. The pertinent question therefore is whether, despite the data constraints just mentioned, it is possible to go beyond the pure reduced form income-environment relationship to distinguish and capture the underlying influences for which the income level works only as a proxy. Success in this regard will reveal more clearly the determinants of environmental quality and also indicate the possible role of conscious action to protect the environment.

In this paper, we take a first step in that direction. We begin by noting that, in the reduced form IER, income actually proxies for a number of structural forces that influence environment. We identify three such forces and name their respective effects on the environment as: (a) the Level effect, (b) the Composition effect, and (c) the Abatement effect. We derive a multiplicative relationship among these effects and, thereby, generate an equation with sufficient structural moorings. This makes it possible to recover the structural relationships from the results of the estimated equation.

The framework is implemented using a panel data set on suspended particulate matter (SPM) in the urban air around the world. We focus on SPM for three reasons. *First,* it has been included in almost all previous studies of the IER. *Second,* more data are available for it than any other pollutant. *Third,* among air pollutants, it is increasingly recognized as the one that has the greatest impact on human health (see Wilson and Spengler, 1996).

The outcome generally confirms the *a-priori* hypotheses regarding the structural relationships. The level effect is found to be monotonically increasing with output level per unit of area. The composition effect displays an inverted-U shape with respect to the share of industry in the GDP. The abatement effect proves to be generally declining with per capita lagged income level.

The paper is organized as follows. Decomposition of the reduced form IER into its structural sources is introduced in section-2. Section-3 presents the theoretical framework for the exercise. It shows how an integrated equation embodying the structural relationships can be derived, and how these structural relationships can then be recovered from the estimation results of this equation. Section-4 discusses the data and econometric issues of implementation of this framework using a global SPM data set. The results are presented and discussed in section-5. Some concluding observations are made in section-6.

2. Towards Identifying the Underlying Determinants of Environmental Quality

2.1 Decomposition of Ambient Pollution Level

The ambient pollution level in an area is influenced both by the pollution that is emitted in that area and by other natural factors such as location, topography, weather patterns etc. We term these later factors as *ancillary determinants* of environmental quality. In this study, we shall control for these ancillary determinants by incorporating site/station specific *individual effects* in our estimation procedure. Focusing on the pollution emission part, we can see that the ambient pollution level in an area can be decomposed as follows:[1]

$$\begin{aligned} Ambient\ Pollution\ Level &= \frac{Pollution\ Emission}{Area} = \frac{GDP}{Area} \times \frac{Pollution\ Emission}{GDP} \\ &= \left(GDP\ per\ unit\ of\ Area\right) \times \left(Pollution\ Intensity\ of\ GDP\right). \end{aligned} \tag{1}$$

Pollution intensity of GDP, in turn, depends on two sets of forces. On the one hand are the forces that work for *generation* of pollution. On the other hand are the forces that work for

[1] Previous studies have offered similar decomposition of pollution into its various sources. For one such example. see Griffith (1994). However, to our knowledge, no previous study has estimated a structural model based on such decomposition.

abatement and *clean up*. The actual emission, and hence the pollution intensity of GDP, is the resultant of these two opposing sets of forces. Algebraically, we can, therefore, write:

$$(2) \quad \textit{Pollution Intensity of GDP} = \frac{\textit{Pollution Emission}}{\textit{GDP}} = \frac{\textit{Pollution Generation}}{\textit{GDP}} \times \frac{\textit{Pollution Emission}}{\textit{Pollution Generation}}.$$

The *pollution generation to GDP ratio* depends on the *composition* of the GDP, while the *emission to generation* ratio depends on the *abatement* efforts. These two terms, therefore, can be referred to as the *Composition Effect (C)* and *Abatement Effect (A)*. It is clear that *GDP per unit area* represent the *Level Effect (L)*. Combining, therefore, we get the decomposition:

$$(3) \quad \textit{Ambient Pollution Level} = \frac{\textit{GDP}}{\textit{Area}} \times \frac{\textit{Composition}}{\textit{of GDP}} \times \frac{\textit{Abatement}}{\textit{Efforts}} = L * C * A.$$

We base our analysis on equation (3).[2] The relationship of GDP with the first term (*L*) is explicit. This is not so for the other two terms. Before proceeding to specification of *L*, *C*, and *A*, it is therefore worthwhile to consider the different ways in which income affects all these terms.

Level Effect

The level effect refers to income as an indicator of level of economic activity. All such activity entails interaction with environment and is generally pollution-generating and resource-depleting, to a greater or lesser extent. Hence, the higher the level of economic activity per capita or per unit of area, the higher is likely to be the level of pollution, measured in commensurate

---

[2] Note that this decomposition can be further carried forward to bring *population density* explicitly into the picture. Since GDP per unit area is the product of GDP per capita and density of population, equation (3) can also be written as

$$(4) \quad \textit{Ambient Pollution Level} = \frac{\textit{GDP}}{\textit{Population}} \times \frac{\textit{Population}}{\textit{Area}} \times \frac{\textit{Composition}}{\textit{of GDP}} \times \frac{\textit{Abatement}}{\textit{Efforts}}.$$

We use equation (3) primarily for parsimony. Also, the impact of population density can be gauged from the impact of GDP per unit of area. Another subtlety regarding this decomposition is the following. Ambient pollution level may be the result of not only the *current* emissions but also the *past* emissions. Introduction of this distinction will require further modifications of the equations above. However, certainly, the current emission level is likely to be the most important determining factor of *current* ambient level. Also, the current emission level will also be a good indicator of past accumulated level of pollution Hence, for practical purpose, equation (3) or (4) should suffice.

terms. This implies a monotonically increasing relationship between income and of pollution (Figure-1a).

Composition Effect

The composition effect depends on two other assumed relationships. Of these, the *first* is between the income level and the structure of the economy. Starting with Kuznets's pioneering work, Chenery and others have shown that the structure of the economy, as captured by the sectoral composition of output or employment, evolves in a predictable manner with the rise of income.[3] A country's rise along the income scale is associated with transformation of the structure of its economy. This transformation, in turn, reflects the underlying process of industrialization. This is generally captured by the industry's share in the output of the country. This share first increases and then declines as a country graduates from pre-industrial to industrial and then to post-industrial phases of development. The *second* assumption is that, as a producing sector, industry is more polluting and resource depleting than either agriculture or services. When combined, these two assumptions yield the inverted U-shaped relationship between pollution and income level shown in Figure-1b.

Abatement Effect

The abatement effect reflects both demand side and supply side influences. The abatement demand effect derives from an analog of Engel's Law regarding the relationship between demand for food and level of income. At low levels of income, people are more concerned with meeting their more urgent material needs, and they have less scope to worry about the quality of the environment. However, as level of income rises, they become less encumbered by pressing material needs and can better appreciate the value of environmental quality and, hence, start demanding it. This effect gives a relationship between pollution and income that slopes downward after income has reached a certain stage, yielding the shape of an inverted-*J*.

[3] This relationship between income level and the structure of the economy, in turn, has its own underlying determinants, with Engel's law regarding falling share of expenditure on food with rise of income being one.

There is a corresponding supply side to the 'abatement demand effect' story. At low levels of income, societies are less able to devote resources for abatement and pollution control measures, even if there is a perceived need and demand for better environmental quality. Higher income levels make this possible. The economy can then spend more on either importing the necessary technologies for pollution abatement and clean-up, and/or on developing these on their own by financing appropriate R&D activities. This would then yield an inverted-*J* relationship between environment and income similar in form to that for the demand effect. Hence, the overall abatement effect should also have an inverted-*J* relationship, as shown in Figure-1c.

Obviously, the precise nature of the curves depicted in Figure-1 is a matter open to debate. The descriptions and the graphs presented here are merely illustrative and depend on many assumptions that may not always hold. Thus, for example, the two assumptions needed for a composition effect of the above type may both go wrong. With inordinate use of machinery, chemical fertilizers, and pesticides, modernized agriculture may prove to be more detrimental to the environment than some types of manufacturing. Similarly, many lines of service industry, with their tendency to produce large amounts of solid and, sometimes, hazardous waste, may do more damage to the environment than many industrial sectors. Similar is the situation with the level effect. With regard to the abatement effect, it may not be the case that both the demand for better environmental quality and the supply of the pollution control and abatement measures are delayed until the attainment of a high income level. The analogy with the demographic transition can hardly be missed here. For long, it was postulated that the demographic transition can be achieved only at higher levels of income. However, recent experience has shown that this is not necessarily the case. Provided certain other conditions are fulfilled, fewer children and smaller families can become the norm even at relatively low levels of income. The might be true with environmental quality.

The purpose of the above decomposition of pollution emission into different terms and the identification of the ways in which income affects them, is not to impose certain pre-conceived patterns on the income-environment relationship. Rather, the goal is to explore how we can get at the underlying determinants of environmental quality, for which income acts as a surrogate. If these determinants can be identified, then we can also pose the question whether we can control them to affect, in a desired manner, the progression of environmental quality. These are matters we cannot investigate with a pure reduced form IER regression.

2.2 Correspondence between Various Effects and Possible Explanatory Variables

Referring to the discussion above, we ask the question, what are the variables that best represent the underlying determinants we seek to isolate? Table-1 gives a rough correspondence between the effects and the possible sets of proxy variables.

Obviously, this correspondence is approximate and not a formal one. As we have mentioned before, theoretical models formalizing the processes represented by the three effects are simply not available. Most of the theoretical work on the IER has been limited to reproducing either an inverted-U relationship between income level and pollution, or a *J*-type relationship between income level and abatement efforts (see, for example, Selden and Song (1993) and Song (1994)). These, therefore, basically provide justification for the prevalent reduced form IER regressions and do not carry us much further. Some approximation in this regard is, therefore, simply unavoidable. In the following section, we further discuss the particular issues of the correspondence between the effects and the variables.

3. Specification

3.1 General Framework

Equation (3) gives the general decomposition of the pollution level into its three structural sources. The next step in fleshing out this relationship is to determine the specification of the terms representing each of these sources. We postulate the following specifications for *L*, *C*, and *A*, and explain below the rationale for each:

(5) $$L = a_0 + a_1 Y + a_2 Y^2 ,$$

(6) $$C = b_1 Q + b_2 Q^2 + b_3 Q^3 , \text{ and}$$

(7) $$A = g_0 + g_1 I + g_2 I^2 + g_3 I^3 ,$$

where,

Y = Output (GDP) per unit of area,

Q = Indicator of composition of output,

I = lagged per capita income.

Specification of the Level Effect

With regard to the rationale of the level effect's specification, first note that most indices of environmental quality are measures of ambient levels, and they are generally measured per unit of volume of either air or water in the locales of the monitoring stations. Hence, output *per unit of area*, *Y*, instead of *per capita*, is the right variable to represent the level effect. It is only conventional to measure output by the gross domestic product (GDP). In our discussion in section 2, and in Figure-1a, we have suggested that the level effect is likely to be monotonically increasing. However, a priori, we do not know whether the increase will be linear or not. In order to allow the data to express themselves freely in this regard, we therefore allow equation (5) to be a *quadratic function* of *Y*.

Specification of the Composition Effect

For the level effect, the corresponding variable is more or less unambiguously defined and unique. However, this is not the case for the other effects. For each of the latter, there is a multiplicity of variables that can qualify as the respective corresponding variable (see Table-1). Therefore, in specifying these effects, along with the issue of the functional form, there is also the issue of which variable(s) to include. In case of the composition effect, for example, the precise way in which *Q* is to be measured can be a matter of opinion. The *first* and simplest approach, perhaps, is to take *Q* equal to the share of industry in the total output. However, one may contend that the share of industry does not contain the entire information regarding composition of output that is relevant for the relationship being investigated. Therefore, the *second* approach may be to include information regarding shares of some other sectors, like agriculture, services etc. Also, information regarding shares may be complemented by information on some other attributes of the sectors. However, such inclusion will make the overall equation extremely large. In a *third* approach, *Q* may be taken as a composite index of the sectoral shares and attributes, and be constructed and analyzed in a way similar to factor analysis. Once the variable(s) has/have been agreed upon, the next question is of appropriate form of inclusion. In our discussion in section 2, we hypothesized the composition effect to be quadratic in nature. However, this is an empirical issue. To ensure sufficient flexibility, we allow a cubic form.[4] To keep the analysis tractable, we measure Q by the first approach: we set it equal to the share of industry in total output (GDP).

Specification of the Abatement Effect

Given our discussion above, it is clear that the issue of variable choice applies equally to the specification of the abatement effect. In view of the multiplicity of variables corresponding to this effect (see Table 1), all three of the approaches discussed in the context of composition effect may be relevant. Obviously, *per capita* income level is the most important variable for the demand effect, and as discussed earlier, it is expected to influence abatement supply as well, as a proxy for R&D and environmental expenditure. However, theoretically, I in equation (4) can also be a composite index of all or a subset of the variables corresponding to the demand and supply effects. It is also possible to think of other variables, besides income per capita, entering *directly* into this equation. However, such direct entry would make the system of equations as a whole unwieldy. Moreover, reliable data across countries on environmental policies and environmental expenditure simply do not exist. Hence, we set I equal to per capita income. We use *lagged* income in order to allow some time lag between the rise in income and its effect to be transmitted to the pollution level. Although the abatement demand effect is usually hypothesized to have an inverted J-shape, we allow a cubic form in order to ensure more generality in the relationship. We use lagged income in order to allow some time lag between the rise in income and its effect to be transmitted to the level of pollution.

3.2 The Overall Equation

We now use equations (5)-(7) to substitute for L, C, and A in equation (3) to get

$$\begin{aligned} E &= L * C * A, \\ &= (a_0 + a_1 Y + a_2 Y^2) * (b_1 Q + b_2 Q^2 + b_3 Q^3) * (g_0 + g_1 I + g_2 I^2 + g_3 I^3). \end{aligned} \tag{8}$$

[4] Note that the composition effect can be operative only through levels of output. Hence its specification may not require an intercept once such a term has already been included in the specification of the level effect.

Multiplying this out we get the following expanded equation:

$$
\begin{aligned}
(9)\qquad E = {} & a_0 b_1 g_0 Q + a_0 b_2 g_0 Q^2 + a_0 b_3 g_0 YQ + a_1 b_2 g_0 YQ^2 + a_1 b_3 g_0 YQ^3 \\
& + a_2 b_1 g_0 Y^2 Q + a_2 b_2 g_0 Y^2 Q^2 + a_2 b_3 g_0 Y^2 Q^3 + a_0 b_1 g_1 QI + a_0 b_2 g_1 Q^2 I \\
& + a_0 b_3 g_1 Q^3 I + a_1 b_1 g_1 YQI + a_1 b_2 g_1 YQ^2 I + a_1 b_3 g_1 YQ^3 I + a_2 b_1 g_1 Y^2 QI \\
& + a_2 b_2 g_1 Y^2 Q^2 I + a_2 b_3 g_1 Y^2 Q^3 I + a_0 b_1 g_2 QI^2 + a_0 b_2 g_2 Q^2 I^2 + a_0 b_3 g_2 Q^3 I^2 \\
& + a_1 b_1 g_2 YQI^2 + a_1 b_2 g_2 YQ^2 I^2 + a_1 b_3 g_2 YQ^3 I^2 + a_2 b_1 g_2 Y^2 QI^2 + a_2 b_2 g_2 Y^2 Q^2 I^2 \\
& + a_2 b_3 g_2 Y^2 Q^3 I^2 + a_0 b_1 g_3 QI^3 + a_0 b_2 g_3 Q^2 I^3 + a_0 b_3 g_3 Q^3 I^3 + a_1 b_1 g_3 YQI^3 \\
& + a_1 b_2 g_3 YQ^2 I^3 + a_1 b_3 g_3 YQ^3 I^3 + a_2 b_1 g_3 Y^2 QI^3 + a_2 b_2 g_3 Y^2 Q^2 I^3 + a_2 b_3 g_3 Y^2 Q^3 I^3 .
\end{aligned}
$$

Using notations for reduced form coefficients, we can write the above as follows:

$$
\begin{aligned}
(10)\qquad E = {} & p_1 Q + p_2 Q^2 + p_3 Q^3 + p_4 YQ + p_5 YQ^2 + p_6 YQ^3 + p_7 Y^2 Q + p_8 Y^2 Q^2 + p_9 Y^2 Q^3 \\
& + p_{10} QI + p_{11} Q^2 I + p_{12} Q^3 I + p_{13} YQI + p_{14} YQ^2 I + p_{15} YQ^3 I + p_{16} Y^2 QI + p_{17} Y^2 Q^2 I \\
& + p_{18} Y^2 Q^3 I + p_{19} QI^2 + p_{20} Q^2 I^2 + p_{21} Q^3 I^2 + p_{22} YQI^2 + p_{23} YQ^2 I^2 \\
& + p_{24} Y^2 Q^3 I^2 + p_{25} Y^2 QI^2 + p_{26} Y^2 Q^2 I^2 + p_{27} Y^2 Q^3 I^2 + p_{28} QI^3 + p_{29} Q^2 I^3 \\
& + p_{30} Q^3 I^3 + p_{31} YQI^3 + p_{32} YQ^2 I^3 + p_{33} YQ^3 I^3 + p_{34} Y^2 QI^3 + p_{35} Y^2 Q^2 I^3 + p_{36} Y^2 Q^3 I^3 .
\end{aligned}
$$

The correspondence between the p ’’s and the underlying structural coefficients is obvious from the formulations above.

3.3 *Recovery of the Structural Relationships*

The immediate question that arises is whether we can recover the structural parameters from the estimated reduced form coefficients, the p ’s. As formulated above, there are 9 structural parameters and 24 reduced form coefficients. It is also obvious from the correspondence that there are restrictions on the p ’s. One way of formulating these restrictions is as follows:

$$(11)\qquad \frac{p_1}{p_{10}} = \frac{p_2}{p_{11}} = \frac{p_3}{p_{12}} = \frac{p_4}{p_{13}} = \frac{p_5}{p_{14}} = \frac{p_6}{p_{15}} = \frac{p_7}{p_{16}} = \frac{p_8}{p_{17}} = \frac{p_9}{p_{18}} = \frac{g_0}{g_1},$$

$$(12)\qquad \frac{p_{10}}{p_{19}} = \frac{p_{11}}{p_{20}} = \frac{p_{12}}{p_{21}} = \frac{p_{13}}{p_{22}} = \frac{p_{14}}{p_{23}} = \frac{p_{15}}{p_{24}} = \frac{p_{16}}{p_{25}} = \frac{p_{17}}{p_{26}} = \frac{p_{18}}{p_{27}} = \frac{g_1}{g_2}, \text{ and}$$

(13) $$\frac{p_{19}}{p_{28}} = \frac{p_{20}}{p_{29}} = \frac{p_{21}}{p_{30}} = \frac{p_{22}}{p_{31}} = \frac{p_{23}}{p_{32}} = \frac{p_{24}}{p_{33}} = \frac{p_{25}}{p_{34}} = \frac{p_{26}}{p_{35}} = \frac{p_{27}}{p_{36}} = \frac{g_2}{g_3}.$$

Numerical recovery of the structural parameters will require imposition of these constraints and checking whether we can find a *unique* correspondence between the reduced form coefficients and the structural parameters. Imposition of the restrictions will, however, make the problem one of non-linear estimation, and the intended recovery is then not guaranteed. The important thing to realize here is that, for many purposes, it is not necessary to recover all the structural parameters *separately*. This is because recovery of the structural *relations* is not the same as recovery of structural *parameters.* For the former, estimated values of some of the relevant combinations of the structural parameters may be sufficient. This is particularly true when the relationships considered are of the *polynomial* form. For these, numerical values of individual parameters are of limited value, because the effect of the explanatory variables on the dependent variable is given by a function of *several* parameters and given values of other explanatory variables. Hence, even if we fail to recover all the structural parameters separately, the above set up may prove sufficient for illuminating the underlying structural relationships to a satisfactory degree. Below, we discuss how this can be done on the basis of estimation of equation (7) above.

Recovery of the Level Effect

Based on the estimation results of the overall equation, the level effect can be traced out using the following equation:

(14) $$\begin{aligned}\hat{E} = {} & (\hat{p}_1\overline{Q} + \hat{p}_2\overline{Q}^2 + \hat{p}_3\overline{Q}^3 + \hat{p}_{10}\overline{Q}\overline{I} + \hat{p}_{11}\overline{Q}^2\overline{I} + \hat{p}_{12}\overline{Q}^3\overline{I} \\ & + \hat{p}_{19}\overline{Q}\overline{I}^2 + \hat{p}_{20}\overline{Q}^2\overline{I}^2 + \hat{p}_{21}\overline{Q}^3\overline{I}^2 + \hat{p}_{28}\overline{Q}\overline{I}^3 + \hat{p}_{29}\overline{Q}^2\overline{I}^3 + \hat{p}_{30}\overline{Q}^3\overline{I}^3) \\ & + (\hat{p}_4\overline{Q} + \hat{p}_5\overline{Q}^2 + \hat{p}_6\overline{Q}^3 + \hat{p}_{13}\overline{Q}\overline{I} + \hat{p}_{14}\overline{Q}^2\overline{I} + \hat{p}_{15}\overline{Q}^3\overline{I} \\ & + \hat{p}_{22}\overline{Q}\overline{I}^2 + \hat{p}_{23}\overline{Q}^2\overline{I}^2 + \hat{p}_{24}\overline{Q}^3\overline{I}^2 + \hat{p}_{31}\overline{Q}\overline{I}^3 + \hat{p}_{32}\overline{Q}^2\overline{I}^3 + \hat{p}_{33}\overline{Q}^3\overline{I}^3)Y \\ & + (\hat{p}_7\overline{Q} + \hat{p}_8\overline{Q}^2 + \hat{p}_9\overline{Q}^3 + \hat{p}_{16}\overline{Q}\overline{I} + \hat{p}_{17}\overline{Q}^2\overline{I} + \hat{p}_{18}\overline{Q}^3\overline{I} \\ & + \hat{p}_{25}\overline{Q}\overline{I}^2 + \hat{p}_{26}\overline{Q}^2\overline{I}^2 + \hat{p}_{27}\overline{Q}^3\overline{I}^2 + \hat{p}_{34}\overline{Q}\overline{I}^3 + \hat{p}_{35}\overline{Q}^2\overline{I}^3 + \hat{p}_{36}\overline{Q}^3\overline{I}^3)Y.\end{aligned}$$

As this equation shows, the relationship is traced out by fixing the values of Q and I at their sample means and computing the predicted values of E for different values of Y. The expressions in the three sets of parentheses give, respectively, the intercept and the coefficients of the linear and quadratic terms of this relationship.

Recovery of the Composition Effect

The recovery of the composition effect can proceed using the following equation:

$$
\begin{aligned}
\hat{E} = {} & (\hat{p}_1 + \hat{p}_4\bar{Y} + \hat{p}_7\bar{Y}^2 + \hat{p}_{10}\bar{I} + \hat{p}_{13}\overline{YI} + \hat{p}_{16}Y^2\bar{I} \\
& + \hat{p}_{19}\bar{I}^2 + \hat{p}_{22}\overline{YI}^2 + \hat{p}_{25}\bar{Y}^2\bar{I}^2 + \hat{p}_{28}\bar{I}^3 + \hat{p}_{31}\overline{YI}^3 + \hat{p}_{34}\bar{Y}^2\bar{I}^3)Q \\
& (\hat{p}_2 + \hat{p}_5\bar{Y} + \hat{p}_8\bar{Y}^2 + \hat{p}_{11}\bar{I} + \hat{p}_{14}\overline{YI} + \hat{p}_{17}Y^2\bar{I} \\
& + \hat{p}_{20}\bar{I}^2 + \hat{p}_{23}\overline{YI}^2 + \hat{p}_{26}\bar{Y}^2\bar{I}^2 + \hat{p}_{29}\bar{I}^3 + \hat{p}_{32}\overline{YI}^3 + \hat{p}_{35}\bar{Y}^2\bar{I}^3)Q^2 \\
& (\hat{p}_3 + \hat{p}_6\bar{Y} + \hat{p}_9\bar{Y}^2 + \hat{p}_{12}\bar{I} + \hat{p}_{15}\overline{YI} + \hat{p}_{18}Y^2\bar{I} \\
& + \hat{p}_{21}\bar{I}^2 + \hat{p}_{24}\overline{YI}^2 + \hat{p}_{27}\bar{Y}^2\bar{I}^2 + \hat{p}_{30}\bar{I}^3 + \hat{p}_{33}\overline{YI}^3 + \hat{p}_{36}\bar{Y}^2\bar{I}^3)Q^3.
\end{aligned} \tag{15}
$$

As the equation shows, the composition effect is traced out by fixing the values of *Y* and *I* at their respective sample means and by computing the values of *E* for different values of *Q*. The expressions in the parentheses correspond to the coefficients of the linear, quadratic, and cubic terms of this relationship.

Recovery of the Abatement Effect

The abatement effect can be recovered using the following equation:

$$
\begin{aligned}
\hat{E} = {} & (\hat{p}_1\bar{Q} + \hat{p}_2\bar{Q}^2 + \hat{p}_3\bar{Q}^3 + \hat{p}_4\overline{YQ} + \hat{p}_5\overline{YQ}^2 + \hat{p}_6\overline{YQ}^3 + \hat{p}_7\bar{Y}^2\bar{Q} + \hat{p}_8\bar{Y}^2\bar{Q}^2 + \hat{p}_9\bar{Y}^2\bar{Q}^3) \\
& + (\hat{p}_{10}\bar{Q} + \hat{p}_{11}\bar{Q}^2 + \hat{p}_{12}\bar{Q}^3 + \hat{p}_{13}\overline{YQ} + \hat{p}_{14}\overline{YQ}^2 + \hat{p}_{15}\overline{YQ}^3 + \hat{p}_{16}\bar{Y}^2\bar{Q} + \hat{p}_{17}\bar{Y}^2\bar{Q}^2 + \hat{p}_{18}\bar{Y}^2\bar{Q}^3)I \\
& + (\hat{p}_{19}\bar{Q} + \hat{p}_{20}\bar{Q}^2 + \hat{p}_{21}\bar{Q}^3 + \hat{p}_{22}\overline{YQ} + \hat{p}_{23}\overline{YQ}^2 + \hat{p}_{24}\overline{YQ}^3 + \hat{p}_{25}\bar{Y}^2\bar{Q} + \hat{p}_{26}\bar{Y}^2\bar{Q}^2 + \hat{p}_{27}\bar{Y}^2\bar{Q}^3)I^2 \\
& + (\hat{p}_{28}\bar{Q} + \hat{p}_{29}\bar{Q}^2 + \hat{p}_{30}\bar{Q}^3 + \hat{p}_{31}\overline{YQ} + \hat{p}_{32}\overline{YQ}^2 + \hat{p}_{33}\overline{YQ}^3 + \hat{p}_{34}\bar{Y}^2\bar{Q} + \hat{p}_{35}\bar{Y}^2\bar{Q}^2 + \hat{p}_{36}\bar{Y}^2\bar{Q}^3)I^3.
\end{aligned} \tag{16}
$$

In tracing out the income effect, the values of *Y* and *Q* are fixed at their respective sample means, and the values of *E* are computed for the corresponding values of *I* Expressions in the parentheses give the intercept and the coefficients of the linear, quadratic, and cubic terms, respectively.

The above completes the exposition of the framework for identification of the various determinants of environmental quality. In the remainder of the paper we present an implementation of the framework on the basis of data on suspended particulate matter (SPM) in the air of a good number of cities around the world. We begin by considering certain general issues of implementation.

4. Issues of Implementation

4.1 General Issues regarding Data

There are a host of problems involved with measurement of the dependent variable of interest, namely, environmental quality. *First* is the question whether it should be measured by pollution emission *rates* or ambient *levels*. While from the point of view of capturing the effects of the determinants, emission rates are more useful, data on them are not easily available, at least in the right form. This has led researchers to often use data on ambient levels. Switching to ambient levels, however, brings its own set of problems. This is because there exist other factors that influence ambient levels, but are not directly linked to human activity. For short, we may call these *incidental determinants* of environmental quality. The distance of a city or a site from the coast line, a desert, or a mountain range is one such example. Indicators of ambient environmental quality are also plagued by a whole range of measurement problems. The readings may be taken at different points of time, at different intervals, using different devices, instruments, and methods, and by persons with different levels of skill and motivation, etc. All these make capturing empirically the systematic relationship between environmental quality and its determinants even more difficult.

Environmental data gathering on the part of the international community is a rather recent effort, starting only during the 1970s. Although significant advances have been made in this regard with the construction and availability of the Global Environmental Monitoring System (GEMS) database, considerable problems remain. *First*, the coverage is not uniform with respect to either countries or years. *Second,* the extent of *intra*-country coverage is also not uniform. This non-uniformity of intra-country coverage makes any averaging at the country level, and any comparison thereof, seriously misleading. This implies that the analysis has to be on the basis of data by monitoring station. However, adoption of monitoring station as the unit of analysis leads to the *third* problem. The data on the determinants of environmental quality that we have mentioned above are mostly of macro nature. There is hardly any data available on these variables at the level of cities or river basins, much less monitoring stations within this areas. The prevalent solution in this regard has been to use country-wide data for all the sites in a particular country. An ideal solution would be to have a mix of site-specific and macro data on the determinants, with the mix being determined by the geographical bounds of impact of the environmental determinant in question.

For the exercise in this paper, we use the GEMS data on suspended particulate matter in urban air, as compiled by Grossman and Krueger (1995). This is our variable E. For the right hand side variables, we rely on, in addition to the Grossman-Krueger data, the World Tables and the Summers-Heston data set. For many of these variables, site-specific data are not available, and we are constrained to use the economy-wide value of these variables for all the sites in a particular country. For some of the variables, however, we try to arrive at site-specific values. Thus, the GDP per unit area (square mile) of a city is obtained by multiplying the GDP per capita for the corresponding country by the density of population of the city concerned. This gives the variable Y. It is not perfect because ideally we would like to multiply the GDP per capita of the *city* by its density of population. However, it is difficult to get data on per capita GDP for cities. Similarly, for Q, we would ideally like to have the share of industry in the GDP of the individual cities concerned. However, it is difficult to obtain data of that kind. With regard to the abatement effect we use lagged per capita GDP as the variable I.[5] This is again a case of substitution of the economy-wide value of the variable for all the cities in a country.

4.2 Some Descriptive Aspects of the Data

The data set contains 901 observations from 23 countries for the period 1977-88. Canada, China, and the USA are, by far, the most important sources of data on SPM. The total number of observations tends to be higher for the years of the early eighties than for the more recent period. Based on the country means, the value of E ranges from a low of 19 for New Zealand to a high of 530 for Pakistan. For GDP per square mile, Y, we find an enormous range of values, with a minimum of 0.0125 for Ghana and a maximum of 35.55 for Venezuela. For Q, the share of

[5] It is obtained by averaging the per capita GDP of the preceding three years, and is the same as the lagged income variable in Grossman and Krueger (1995). It would have been better to use GNP, instead of GDP, in constructing this variable. Unfortunately, the Summers-Heston data set does not provide information on GNP.

industry in GDP, we find that, based on country averages, its value ranges from a minimum of 0.1947 for Ghana to 0.4588 for Brazil. *I*, per capita lagged income, is the variable whose country distribution we are most familiar with. The range in country means is from a low of 0.645 for India to a high of 14.881 for the US.

The SPM data for each country generally come from more than one city and monitoring station. In total, the number of cities represented in the data set is 56, while the number of sites is 145. The number of sites was generally the highest for the USA, averaging to about 20. However, there were no data from the US for the two most recent years. China comes in second in terms of number of sites, Canada third. The rest of the countries generally have less than 5 sites, with Japan exceeding this mark in some recent years. It is also clear that there is considerable flux or, if you would like, turnover in the sites reporting SPM data. The highest number of sites for which we have data in any particular year is 82 (in 1982). This may be compared with 145, the overall number of sites represented in the data set. Additional information about the distribution properties of the variables concerned is presented in the form of histograms in Figures 2 to 5.

4.3 Econometric Issues

The income-environment relationship is basically a dynamic relationship with respect to a site or a country. However, to the extent that long time series on the relevant variables are difficult to come by for individual countries, researchers have mostly used cross-country data for estimating the relationship. Use of panel data is obviously an improvement in this regard because it brings into play both the time series and cross-sectional dimensions of the data. In this paper we use panel data for all available sites. Hence, the situation is one of an unbalanced panel.

There are several models to consider within the panel approach. This basically pertains to the treatment of the individual site (station) effect, which captures the *ancillary* and other site (station)-invariant factors that are not accounted for in the regression though inclusion of corresponding variables. It is quite possible that these site effects will be correlated with the other explanatory variables considered. This would call for the use of the *correlated effects* model and the associated minimum distance estimation method (see Chamberlain (1982, 1983)). However, there are several obstacles in this regard. *First* is the serious unbalanced nature of the panel data set that we have to use. *Second,* the correlated effects model is suitable for situations where the number of time periods, *T*, is small, and the number of individual units, *N*, is large. In our case, however, for many of the sites, *T* is quite large. *Third,* the correlated effects model is also geared to a situation where the number of right hand side variables is rather small. By contrast, as equation (9) shows, we are dealing with a regression that has a very large number of right hand side entries. All these make correlated effects model not quite suitable for our task.

This reduces the choice to the *fixed effects* and *random effects* models. Both have been used in the literature. Under fixed effects, the slope parameters are basically estimated from the 'within' deviations, because the procedure involves subtracting out the individual means from the variables. This also involves the loss of a large number of degrees of freedom, equaling the number of individual units (in this case, the number of sites). The main weakness of the random effects model is the assumption that the site effects are uncorrelated with the included variables. Typically, the Hausman test is used to decide which of these models to use. However, the Hausman test is actually a joint test of the specification of the model and the assumption that the individual effects are uncorrelated with the included regressors. This test does not reveal whether a rejection of the null hypothesis is caused by violation of the above assumption of uncorrelatedness or by other kinds of misspecification of the model.

In view of the above, we base our choice of the model on the following considerations. *First*, let the composite error term, e_{it} be specified as follows:

(17) $$e_{it} = m_i + u_{it},$$

where m_i is the individual effect term that is time invariant, and n_{it} is the transitory error term, which varies over both time and individual units. Then the Bruesch and Pagan Lagrangian Multiplier test, applied to the results from estimating equation (9) (discussed more fully below) overwhelmingly rejects the null $H_0 : \mathrm{Var}(m_i) = 0$. The sample value of the c^2 statistic is 701.44, and the associated *p*-value is less than 0.01 percent. This supports the random effects specification of m_i.

Second, more helpful in this regard are the results obtained from analysis of variance of the concerned variables. The results from such an analysis have been compiled in Table-2. As we can see from this table, the overwhelming part of variation in the data lies in the 'between' dimension, rather than 'within.' This is true not only for the dependent variable, *E*, but also for all the explanatory variables. Going by the sum of squares (SS), between variation accounted for 93 percent of total variation in *E*. The analogous percentages for *Y*, *Q*, and *I* are 97, 95, and 99 percent respectively. Between sum of squares (SSB) exceeds the within sum of squares (WSS) by factors of 13, 35, 21, and 111 for *E*, *Y*, *Q*, and *I*, respectively. Similarly, the between mean sum of squares (MSB) exceeds within mean sum of squares (MSW) by 69, 183, 106, and 709 times, respectively, for the same variables as above. Given this structure of the data, the fixed effects estimator, which throws out all the between variation of the data, cannot be a ideal choice. The between estimator, which can be an alternative, suffers from the opposite problem. It cannot use any of the within variation that exists in the data. Also, it cannot avoid the potential problem of correlation of m_i with the included regressors. In view of this, we choose the random effects estimator to estimate equation (9). This choice is further grounded in a third consideration, that Grossman and Krueger also used the random effects model. To the extent that our work takes the Grossman and Krueger (1995) paper as a point of departure, it is helpful to make similar econometric assumptions. This will make it easier to compare the reduced form approach of their paper with the 'structural' approach of the current work, as we shall presently see.

5. Results for Suspended Particulate Data

5.1 The Reduced Form Income-Environmental Relationship

Before we present the results of estimation of equation (2), and in order to set the stage for comparison, we want to briefly dwell upon the reduced form income-environment relationship that can be obtained from the data we are using. The graph of such a reduced form IER is presented in Figure-6. This is based on the following regression:

(18) $$E_{it} = d_0 + d_1 N_{it} + d_2 N_{it}^2 + d_3 N_{it}^3 + l_1 LN_{it} + l_2 LN_{it}^2 + l_3 LN_{it}^3 + e_{it}$$

where N_{it} and LN_{it} are current and lagged per capita income, respectively. Note that *LN* of this equation is the same as *I* of equations (5) to (16). We use different notations to avoid confusion between current and lagged values of income. As may be noted, this is the same specification as used by Grossman and Krueger. The numerical results can be seen in Table-A1. The graph in Figure-6 shows that leaving out of the incidental variables does not matter much, and the obtained graph is almost identical to that reported in Grossman and Krueger (1995, p. 363). It has a mild cubic nature, with a high and a low peak at $-7.745 and $22,530, respectively. Both of these are outside of the range of the income data and are, therefore, just numerical curiosities. Within the range, the curve slopes downward everywhere, although the slope becomes less steep as income attains higher values. The results of our analysis below show how widely different forces are working underneath to produce this seemingly placid relationship.

5.2 Regression Results

The overall outcome statistics of the regression of equation (9) are shown in Table-3. As is known, the R^2 from random effects GLS estimation do not have the classical properties. The reported values shown in the table are the R^2 from simple bivariate regressions of the fitted values on the respective set of original values. Going by these measures, we find that the regression did remarkably well. The overall fit, as measured by the R^2-overall, is 63 percent, which is a quite high value given the nature of the data. This is also confirmed by the *p*-value (less than 0.01 percent) of the c^2 statistic for the test of joint significance of the explanatory variables. As expected, the fit is much better along the between dimension of the data. R^2-between equals 0.60. However, even along the within dimension, the fit is quite respectable, with a R^2-within equaling 0.21. Looking at the estimated values of s_m, s_n, and s_e, we see that, as expected, the time-invariant individual component of the error prove to be dominant. Its standard deviation was about 2.2 times larger than that of the pure transitory component, n_{it}.

The basic regression results are provided in Table-4, which gives the parameter estimates, standard errors, and levels of significance. The total number of interactive regressors was 36. As many as 29 of them prove to be significant at least at the 10 percent level. Ten of these are significant at the 5 percent level, and another 5 prove to be significant even at the 1 percent level.

As all these regressors are basically various multiplicative combinations of only three underlying variables, the danger of multicollinearity loomed large. In view of this, the results on the significance of the individual right hand side terms of the regression are encouraging. The regression includes a time trend variable (*year*) in order to capture the secular global advance in technology and awareness. It also helps the n_{it} term to have zero mean. This variable proves to be significant at a 6.3 percent level. The overall intercept term also proves to be highly significant.

Although we may be curious about the sign and numerical magnitudes of the coefficients of the individual terms, examining these may not be that productive because the relationships of interest have been specified in the form of polynomials. Individual coefficients are of limited value in revealing these relationships. This brings us to the task of tracing out the polynomials.

5.3 Estimated Structural Relationships

Estimated Level Effect

We first turn to the recovery of the level effect. This is done using equation (14). The numerical results for this equation turn out to be the following:

$$\hat{E}_l = -15.298 + 6.425\, Y - 0.0343\, Y^2 \tag{19}$$

As equation (14) shows, this is on the basis of evaluation of the equation at the sample mean levels of *Q* and *I*. The year was set equal to 1982 which is the closest integer to the value of the sample mean of the variable "year."[6] The graph corresponding to this level effect equation is presented in Figure-7.[7] The circles, or, if you like, bubbles in the graph give information regarding the fit of the curve. These are obtained by superimposing the residuals of the regression to equation (19). The distance of these circles from the line of the graph indicate the

[6] Only the last two digits were used to denote the years. Hence, the actual number entering for the year variable in evaluating this equation is 82. The same is done to derive the numerical evaluation of the equations for composition and abatement effects below.

[7] In producing this graph and computing the level figures presented in Table-5, the constant term of equation (12) was augmented by a value of 16 to avoid negative ranges. This modification does not affect its slope and the changes in the SPM levels over intervals of *Y* computed on its basis.

size of the residuals, and the diameters of the circles are proportional to the number of observations in the corresponding interval of the variable on the horizontal axis, in the present case, *Y*.

We find that the equation for level effect is mildly quadratic, with a small negative coefficient on the quadratic term. This produces a curve that is monotonically increasing but with a declining slope. Table-5 presents the SMP levels predicted by this equation for some specified levels of GDP per square mile, *Y*. It also shows the numerical magnitudes of the slopes of the curve at those levels of *Y*. These give the marginal increases in the SPM level for a $1,000 increase in GDP per square mile, as predicted by the level effect alone. The aspects of increasing level and of decreasing slope of the level effect curve are both apparent from the numbers of column (2) and (3) of Table-5. Finally, the numbers in column (4) of this table show the change in the SMP level between the levels of *Y* shown in the first column.

Overall, it is indeed clear that the ambient level of SPM has a positive relationship with the level of GDP per unit of area, other things remaining constant. For the relevant range, this relationship is almost linear. To be precise, the slope decreases with increasing *Y*, but at a very slow pace. Extrapolation based on equation (19) shows the level effect to taper off (slope being zero) at *Y* equaling to $93,796. This indicates that the level effect is certain to bedevil the environment for a long time to come.

Estimated Composition Effect

The composition effect is recovered using equation (15). This equation when evaluated at sample means of *Y* and *I* gives the following result:

$$\hat{E}_c = 673.207 - 818.703\,Q + 30186.179\,Q^2 - 33595.984\,Q^3 \quad (20)$$

As can be seen, the equation for the composition effect has large numerical coefficients for all the terms of the cubic polynomial. It has a low peak at *Q* equal to 0.21, a high peak at *Q* equal to 0.39, and a point of inflection at a *Q* equal to 0.30. The graph corresponding to the equation is presented in Figure-8.[8] For the most part of the range of *Q*, the relationship is quadratic, with an inverted-*U* shape. It is only for very low values of *Q* (under 0.2) that the curve has a declining segment. However, as the bubble for this interval shows, the fit of the curve for this range is not satisfactory. The location of the bubble far below the curve reveals that the actual values of *E* for this range of *Q* were much lower than those predicted by the curve. Also, the size of the bubble shows that the number of observations belonging to this range is small. In view of this, it is plausible to do away with the cubic specification of the composition effect and to constrain it to be quadratic. We refrain from doing so in order to preserve the generality of the framework.

Table-6 shows some computed values of *E* on the basis of the composition effect equation. The level values are conditional on the normalization. The slope figures are presented in column (3). Note that the range of values of *Q* is theoretically limited to [0,1]. In actual data, it is limited to an even shorter interval. That is why, even though the slope figures are rather high, the relevant range over which *E* changes with changes in *Q* is not that large. This can be seen from column (4) of Table-6. The numbers in this column give the changes in the SPM level, as predicted by equation (20), across intervals of *Q* shown in column (1). We can also see that, barring the first entry, these bear out the largely quadratic nature of the curve. An initial phase of increase is followed by a decreasing phase.

Thus the estimated composition effect curve demonstrates, by and large, a hump-shaped relationship of the SPM level with the share of industry in GDP, *Q*. The peak is not reached until a long range of values of *Q* is passed, during which the SPM level increases almost linearly. Beyond that level, the ambient level of SPM seems to be negatively related with *Q*, when the levels of *Y* and *I* are held constant. Note that this inverted-*U* curve should not be confused with the humped curve that we presented in Figure-1b. There the pollutant level was graphed against income level. In contrast, what we have here on the horizontal axis is *Q*, and not income level. Nevertheless, we find a generally inverted-*U* shaped relationship between *E* and *Q*.

[8] In order to avoid the negative range, the equation was normalized upward through augmentation of the constant term by 115. Obviously, this does not affect the slopes and the changes in the SPM levels over intervals of *Q* computed on the basis of the equation.

Estimated Abatement Effect

Finally, we come to the abatement effect. This is recovered using equation (16). Evaluation of this equation at sample means of Y and Q yields the following result:

$$(21) \qquad \hat{E}_a = 272.35 - 54.10\, I + 4.233\, I^2 - 0.1007\, I^3$$

This equation has a low peak at I equaling \$9,856, and a high peak at I equaling \$18,168. It also has an inflection point at I equaling \$14,012. However, clearly, the linear and quadratic terms dominate this equation. The graph corresponding to this equation is presented in Figure-9.

In our discussion of section-2, we noted that the hypothesis most often put forward regarding this effect is one of an inverted-*J* shape. The curve in Figure-9 does not quite conform to that hypothesis. Rather than an inverted-*J*, it resembles more to a backward-*J*. After displaying a steep decline over lower intervals of *I*, the curve flattens out. In fact, after an income level of \$9,856, the curve displays an upward slope, albeit mild. However, since the cubic term in the equation enters negatively, this upward trend is soon arrested, and beyond an income level of \$18,168, the slope of the curve again turns negative.

The figures of Table-7 give a numerical illustration of this pattern of movement. The main way in which this estimated relationship differs from the generally held belief regarding the impact of income on environment (as depicted in Figure 1c) is as follows. The inverted-*J* hypothesis predicts that one has to wait until a certain (high) level of income is reached before the beneficial impact of income on pollution reduction becomes effective. In contrast, the graph in Figure (9) indicates that income starts to have large pollution-reducing effect starting at very low levels. From this point of view, the estimated relationship may generate some optimism, provided it is true.

Note that although this is a graph with pollution on the vertical axis and level of per capita income on the horizontal axis, it is not directly comparable to similarly plotted graphs that we are familiar with. While the latter give the net effect of income on level of pollution, the present curve is showing the effect of rise of income when other important variables, like *Y* and

Q, are held constant. So the interpretation is more complicated in this case. The advantage now is that it allows us to see the effect of per capita income on environmental quality when this relationship is not confounded by the effects of rising GDP per unit of area and changes in the composition of output.

6. Concluding Remarks

The exercise above has shown that it is possible to unveil the income-environment relationship, which has until now remained shrouded in mystery. We have argued that in the reduced form IER, income actually proxies for a number of different forces that influence environmental quality in different ways. We identified and named several of these, namely, (i) the level effect, (ii) the composition effect, and (iii) the abatement effect. We started by deriving a multiplicative relationship among these, and we allowed sufficient flexibility in their individual specifications. This resulted in a reduced form specification with structural moorings. We estimated the equation on the basis of data on suspended particulate matter in the air of different cities in a global sample of countries. The structural relationships were then recovered from the results of the reduced form regression.

The estimated curves of the level, composition, and abatement effects were found by and large to conform to our a priori theoretical expectations. The level effect is found to be monotonically increasing over the relevant range of values of GDP per unit of area, *Y*. The composition effect mostly displays an inverted-*U* shape. The abatement effect curve proves to be generally declining.

However, this does not mean that all the results were as expected. The composition effect shows a declining segment for the lower range of *Q*, which is something difficult to explain. The fit for that part of the curve is, however, poor. Similarly, it is not clear why the abatement effect should prove more forceful at lower levels of income than at higher. However, these discrepancies should not come as a surprise. *First,* we do not claim that the specifications used in this paper are, in any sense, final. *Second,* and more important, there were numerous gaps in our analysis between specification and implementation. Those gaps arose mainly because of data limitations. In this regard, we can point to the following, at least, three. *First,* theory requires data on the explanatory variables to be site-specific. In reality, in most cases, we had to be satisfied with country-level data. *Second,* theory pointed to a host of variables that correspond to the composition and abatement effects. In actual implementation, we used just one variable for each of these effects. *Third,* even with respect to the variables that we used, there were limitations. In view of all these shortcomings, it is noteworthy that, in terms of the qualitative properties, the estimated equations and curves for the different effects largely conformed to the theoretical predictions. We view this as promising support for the framework proposed in this paper. With more work, the framework can certainly be improved.

References

Arrow K., B. Bolin, R. Costanza, P. Dasgupta, C. Folke, C. Holling, B.-O. Jansson, S. Levin, K. G. Meller, C. Perrings and D. Pimentel (1995), "Economic Growth, Carrying Capacity and the Environment," *Science,* 268: 520-1.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18:5-46.

Chamberlain, G. (1983), "Panel Data," in *Handbook of Econometrics*, edited by Z. Griliches and M. Intrilligator, North Holland, Amsterdam, pp. 1247-1318.

Dutt, A. K. and J. Mohan Rao (1996), "Growth, Distribution and Environment: Sustainable Development in India," *World Development,* 24 (2): 287-305.

Gershenkron, Alexander (1952), "Economic Backwardness in Historical Perspective," in *The Progress of Undeveloped Areas,* edited by B. F. Hoselitz, Chicago, University of Chicago Press,

Griffith, Charles (1994), The Environmental Kuznets Curve: Examining Economic Growth and Environmental Degradation, *Environment Dissemination Notes*, No. 9, The World Bank.

Grossman Gene and Alan Krueger (1995), "Economic Growth and the Environment," *Quarterly Journal of Economics,* (May): 353-77.

Hettige H., R. E. B. Lucas and D. Wheeler (1992), "The Toxic Intensity of Industrial Production: Global Patterns, Trends, and Trade Policy," *American Economic Review,* 478-81.

Kaufmann R., B. Davidsdotter and S. Garnham (1995), "The Determinants of Atmospheric SO2 Concentrations: Reconsidering the Environmental Kuznets Curve," Center for Energy and Environmental Studies, Boston University.

Lopez Ramon (1994), "The Environment as a Factor of Production: The Effects of Economic Growth and Trade Liberalization," *Journal of Environmental Economics and Management,* 27: 163-84

Panayotou T. (1995), "Environmental Degradation at Different Stages of Economic Development," in *Beyond Rio (The Environmental Crisis and Sustainable Livelihoods in the Third World),* edited by Iftikhar Ahmed and Jacobus A. Doeleman, International Labor Organization, London, Macmillan Press Ltd.

Shafiq Nemat and Sushenjit Bandyopadhyay (1992), "Economic Growth and Environmental Quality: Time Series and Cross-Country Evidence," World Bank Policy Research Working Paper, WPS 904, Washington D. C.

Song Daqing (1993), “A Simple Static Model for Pollution Kuznets Curve,” Dept. of conomics, Syracuse University.

Selden Thomas M., and Daqing Song (1994), “Environmental Quality and Development: Is here a Kuznets Curve for Air Pollution Emissions?” *Journal of Environmental Economics and Management*, XXVII:147-62.

Selden Thomas M. and Daqing Song (1993), “Neoclassical Growth, the J-curve for Abatement, and the Inverted U-curve for Pollution,” Dept. of Economics, Syracuse University.

Vincent Jeffery (1997), “Resource Depletion and Economic Sustainability in Malaysia,” *Environment and Development Economics*, 2(1): 19-37

World Bank (1995), *World Tables*, Washington D.C.

Table-1

Correspondence Between the Effects and Candidate Variables

Type of Effect	**Variables**
Level Effect	*GDP per unit of area*
Composition Effect	*i) Sectoral shares in GDP,* *ii) Technological features of the sectors, etc.*
Abatement Effect (demand side)	*i) Per capita income,* *ii) Level of education and awareness,* *iii) Political rights and leadership,* *iv) Civil liberties,* *v) Inequality of income distribution, etc.*
Abatement Effect (supply side)	*i) Level of R&D expenditure,* *ii) Degree of openness of the economy,* *iii) Direct expenditure on pollution control and abatement, etc.*

<u>Table-2</u>

Results from Analysis of Variance of the Variables

Item	*SMP Level (E)*	*GDP per sq. mile (Y)*	*Share of industry in GDP (Q)*	*Lagged GDP per capita (I)*
Sum of squares, Total (SST)	15,239,065	73,216	4.1732	32,097
Degrees of freedom for SST	900	900	900	900
Sum of squares, Between	14,144,495	71,176	3.979	31,861
Degrees of freedom for SSB	144	144	144	144
Sum of squares, Within (SSW)	1,094,570	2,040	0.194	286
Degrees of freedom for SSW	756	756	756	756
Mean sum of squares, Total (MST)	16,932	81	0.0046	36
Mean sum of squares, Between (MSB)	98,226	494	0.0276	221
Mean sum of squares, Within (MSW)	1,448	3	0.0003	0.312
SSB as ratio of SST	0.93	0.97	0.95	0.99
SSW as ratio of SST	0.07	0.03	0.05	0.01
SSB as ratio of SSW	12.92	34.89	20.51	111.40
MSB as ratio of MST	5.80	6.06	6	6.20
MSW as ratio of MST	0.09	0.03	0.06	0.01
MSB as a ratio of MSW	67.84	183.13	106.15	709.16
Test for equality of Within and Between Variances				
Value of F-statistic	67.84	183.15	107.72	708.73
p-value of F	0.0000	0.0000	0.0000	0.0000
Bartlett's Test for equality of within group variances				
Value of Chi-square statistic	1150.63	1810.64	522.08	956.19
p-value of the Chi-square	0.000	0.000	0.000	0.000

Table-3

Panel Regression Using Particulate Data
(Overall Statistics)

Item	*Value*
Number of Observations	901
Number of countries	23
Number of cities	56
Number of sites	145
R-square, Within	0.2088
R-square, Between	0.6003
R-square, Overall	0.6341
Sample value of c^2 *for overall fit*	439.36
p-value of the c^2	0.0000
Estimated value of s_{m_i}	71.2670
Estimated value of $s_{v_{it}}$	33.7140
Estimated value of s_e	78.8392

Table-4

Detailed Results of the Panel Regression

(Dependent Variable: Median SPM level in air)

RHS Variable	*Coeffcient*	*Standard Error*	*Z-value*	*p-value of z*
Constant	797.94	290.03	2.751	0.006
Year	-1.52	0.89	-1.856	0.063
Q	-5859.07	4821.78	-1.215	0.224
Q^2	30952.03	22071.12	1.402	0.161
Q^3	-45756.06	28913.42	-1.583	0.114
YQ	753.47	441.76	1.706	0.088
YQ^2	-4760.31	2600.88	-1.830	0.067
YQ^3	6880.05	3749.30	1.835	0.067
Y^2Q	-47.64	16.86	-2.826	0.005
Y^2Q^2	262.49	98.12	2.675	0.007
Y^2Q^3	-342.41	140.42	-2.439	0.015
QI	1349.17	2235.84	0.603	0.546
Q^2I	-16692.65	12685.82	-1.316	0.188
Q^3I	32914.63	17849.66	1.844	0.065
YQI	-605.21	309.47	-1.956	0.051
YQ^2I	3944.03	1786.98	2.207	0.027
YQ^3I	-6020.66	2556.56	-2.355	0.019
Y^2QI	31.13	10.65	2.922	0.003
Y^2Q^2I	-176.65	61.92	-2.853	0.004
Y^2Q^3I	243.02	89.49	2.716	0.007
QI^2	-488.31	396.91	-1.230	0.219
Q^2I^2	3907.22	2362.68	1.654	0.098
Q^3I^2	-6759.88	3515.11	-1.923	0.054
YQI^2	104.46	53.17	1.965	0.049
YQ^2I^2	-659.35	313.74	-2.102	0.036
YQ^3I^2	992.52	461.34	2.151	0.031
Y^2QI^2	-4.43	1.81	-2.443	0.015
$Y^2Q^2I^2$	25.22	10.63	2.373	0.018
$Y^2Q^3I^2$	-35.04	15.58	-2.249	0.025

Table-4 (continued)

Detailed Results of the Panel Regression

(Dependent Variable: Median SPM level in air)

RHS Variable	*Coeffcient*	*Standard Error*	*Z-value*	*p-value of z*
QI^3	28.46	19.74	1.442	0.149
Q^2I^3	-202.36	121.73	-1.662	0.096
Q^3I^3	332.58	188.19	1.767	0.077
YQI^3	-4.71	2.58	-1.826	0.068
YQ^2I^3	29.19	15.60	1.871	0.061
YQ^3I^3	-43.61	23.60	-1.848	0.065
Y^2QI^3	0.18	0.09	2.033	0.042
$Y^2Q^2I^3$	-1.01	0.52	-1.952	0.051
$Y^2Q^3I^3$	1.41	0.77	1.830	0.067

Table-5

Level Effect

(Partial effect of GDP per square mile on the level of SPM in air)

GDP per square mile (Y)	*SPM Level in the air (E)*	*Slope of the Abatement Curve (¶E/¶Y)*	*Change in E over the interval (DE)*
100	1.344	6.418	
500	3.906	6.391	2.562
1,000	7.093	6.357	3.187
5,000	31.971	6.083	24.878
10,000	61.527	5.740	29.556
15,000	89.370	5.398	27.844
20,000	115.501	5.055	26.131
25,000	139.920	4.713	24.418
30,000	162.625	4.370	22.706
35,000	183.618	4.028	20.993
40,000	202.899	3.685	19.281

Notes:

1. The SPM level figures in the table are conditional on the normalization chosen.
2. *DE* represent change in E associated with the change in *Y* over the previous figure on this variable in the table. Thus, for example, SPM level change by 2.562 as a result of a partial effect of increase in *Y* from \$100 to \$500. These change figures do not depend on the normalization.
3. The slope figures are also invariant to normalization.

Table-6

Composition Effect

(Partial effect of the share of industry in GDP on the SMP level in air)

Share of industry in the GDP (Q)	*Level of SPM in air (E)*	*Slope of the Composition Effect curve*	*Change in E over the interval (DE)*
0.15	126.754	-1393.58	
0.20	90.546	-138.75	-36.208
0.25	104.480	612.14	13.934
0.30	143.361	859.09	38.880
0.35	181.990	602.10	38.629
0.40	195.172	-158.83	13.181
0.45	157.708	-1423.70	-37.464
0.50	44.403	-3192.51	-113.306

Notes:

1. The SPM level figures in the table are conditional on the normalization chosen.
2. *DE* represent change in E associated with the change in *S* over the previous figure on this variable in the table. Thus, for example, SPM level change by 13.934 as a result of a partial effect of increase in *Q* from 0.20 to 0.25. These change figures do not depend on the normalization.
3. The slope figures have not been presented because the value of *Q* is numerically bounded between (0,1) and hardly reaches 1.

Table 7

Abatement Effect

(Partial Effect of Lagged Per Capita GDP on SPM Level in Air)

Lagged Per Capita GDP (I)	*Level of SMP in air (E)*	*Slope of the Abatement Curve (¶E/¶I)*	*Change in the SMP level over the interval (DE)*
500	246.353	-49.938	
1,000	222.392	-45.931	-23.961
2,000	180.291	-38.371	-42.101
3,000	145.448	-31.416	-34.843
4,000	117.258	-25.065	-28.190
5,000	95.118	-19.318	-22.141
6,000	78.422	-14.175	-16.696
7,000	66.567	-9.636	-11.855
8,000	58.949	-5.701	-7.618
9,000	54.963	-2.371	-3.986
10,000	54.005	0.355	-0.958
11,000	55.471	2.477	1.466
12,000	58.757	3.995	3.286
13,000	63.259	4.908	4.502
14,000	68.372	5.217	5.113
15,000	73.493	4.923	5.120
16,000	78.016	4.023	4.523
17,000	81.338	2.520	3.322
18,000	82.855	0.413	1.517
19,000	81.962	-2.299	-0.893
20,000	78.055	-5.615	-3.907

Notes:

1. The SPM level figures in the table are conditional on the normalization chosen.
2. *DE* represent change in E associated with the change in *I* over the previous figure on this variable in the table. Thus, for example, SPM level change by -23.961 as a result of a partial effect of increase in *I* from $500 to $1,00. These change figures do not depend on the normalization.
3. The slope figures are also invariant to normalization.

Table-A1

Regression Results for the Reduced Form Income-Environment Relationship

Overall Statistics of the Regression:

Number of observations =901
R-square, Within = 0.0004
R-square, Between =0.4208
R-square, Overall = 0.4857
Chi-square = 96.40
p-value of Chi-square = 0.000

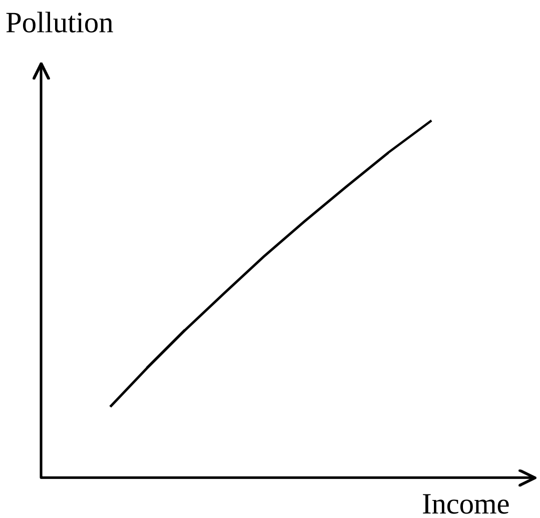
Dependent variable = Median SPM level (microgram per cubic meter)

RHS Variable	*Coefficient*	*Standard Error*	*z-value*	*p-value of z*
Constant	235.11	18.27	12.869	0.000
N	37.27	22.38	1.664	0.096
N^2	-5.06	2.96	-1.711	0.087
N^3	0.16	0.10	1.616	0.106
LN	-47.08	24.56	-1.917	-.055
LN^2	4.65	3.46	1.345	0.179
LN^3	-0.15	0.13	-1.152	0.249

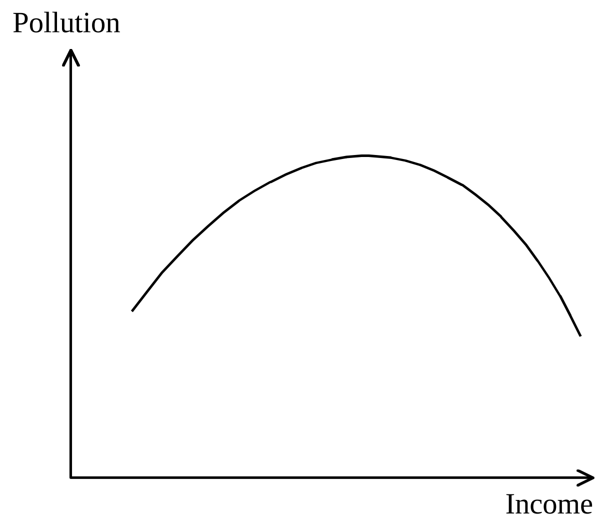
Notes:

1. N denotes per capita GDP
2. LN denotes lagged per capital GDP defined as average per capita GDP over the previous three years.

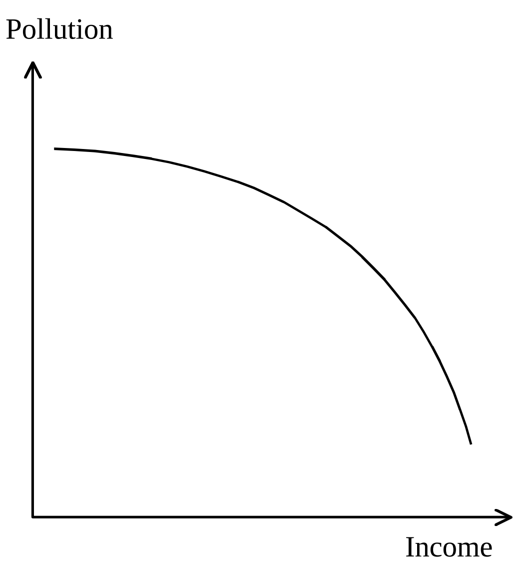
Figure 1. Different Effects of Income on Environment



a) Level Effect



b) Composition Effect



c) Abatement Effect

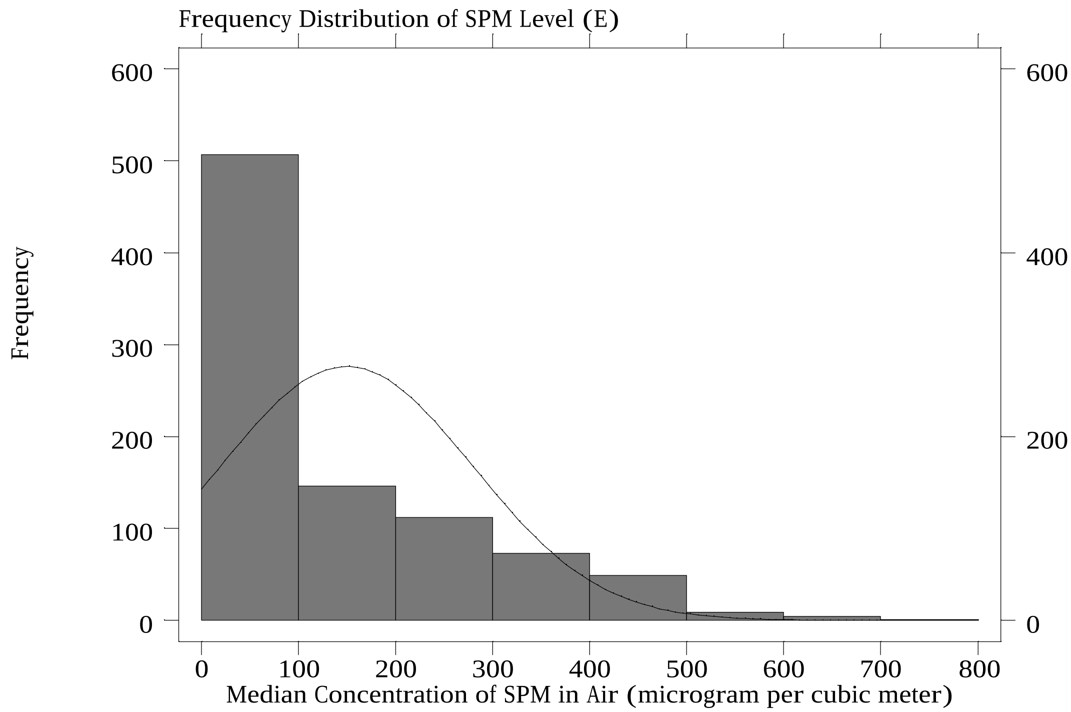


Figure-2

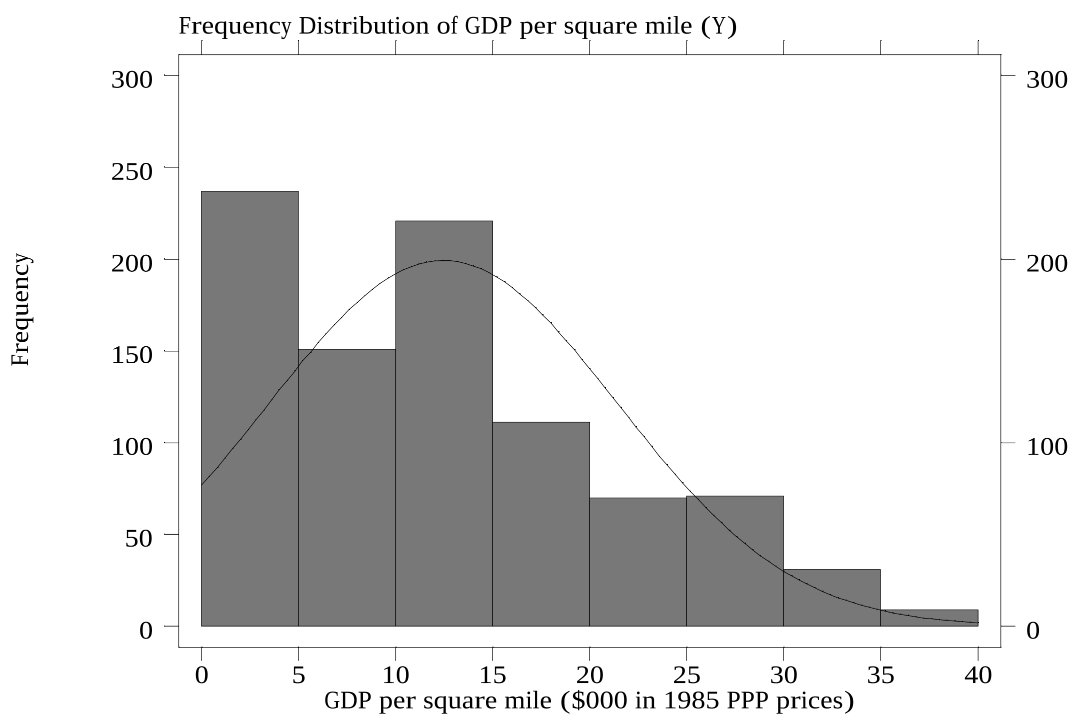


Figure-3

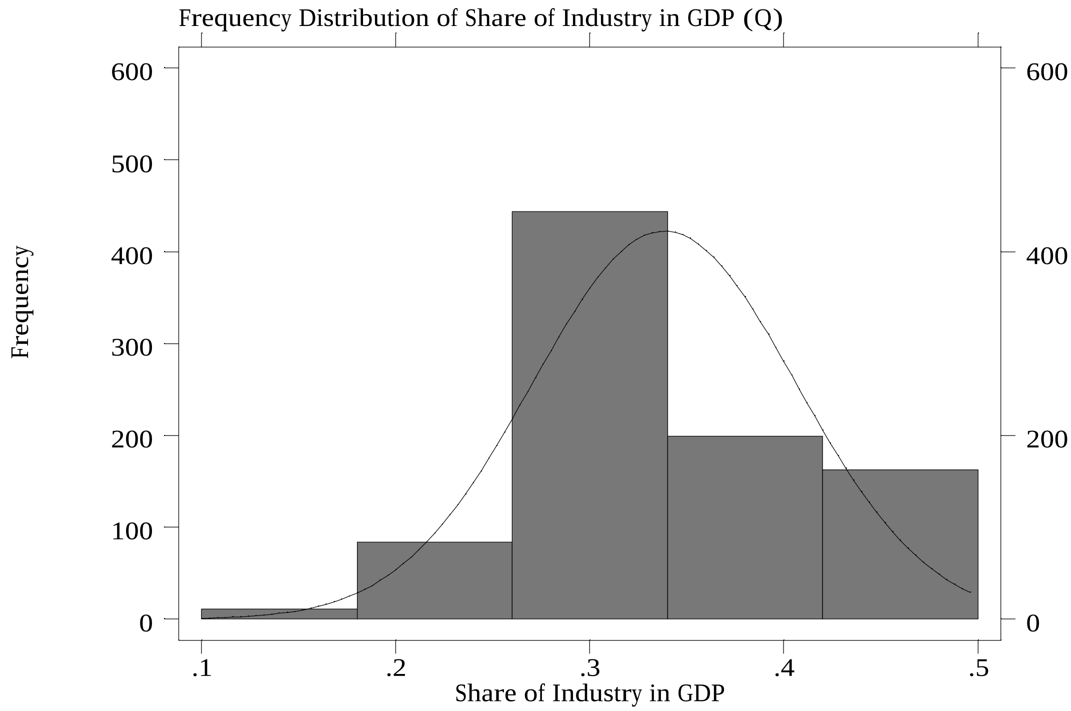


Figure-4

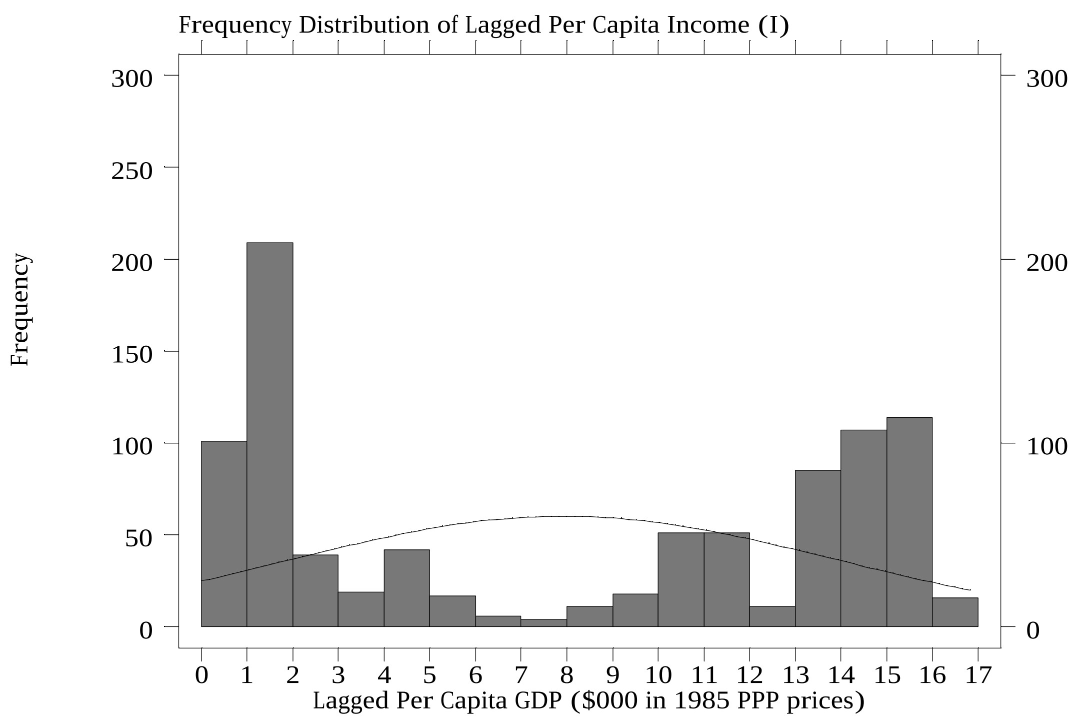


Figure-5

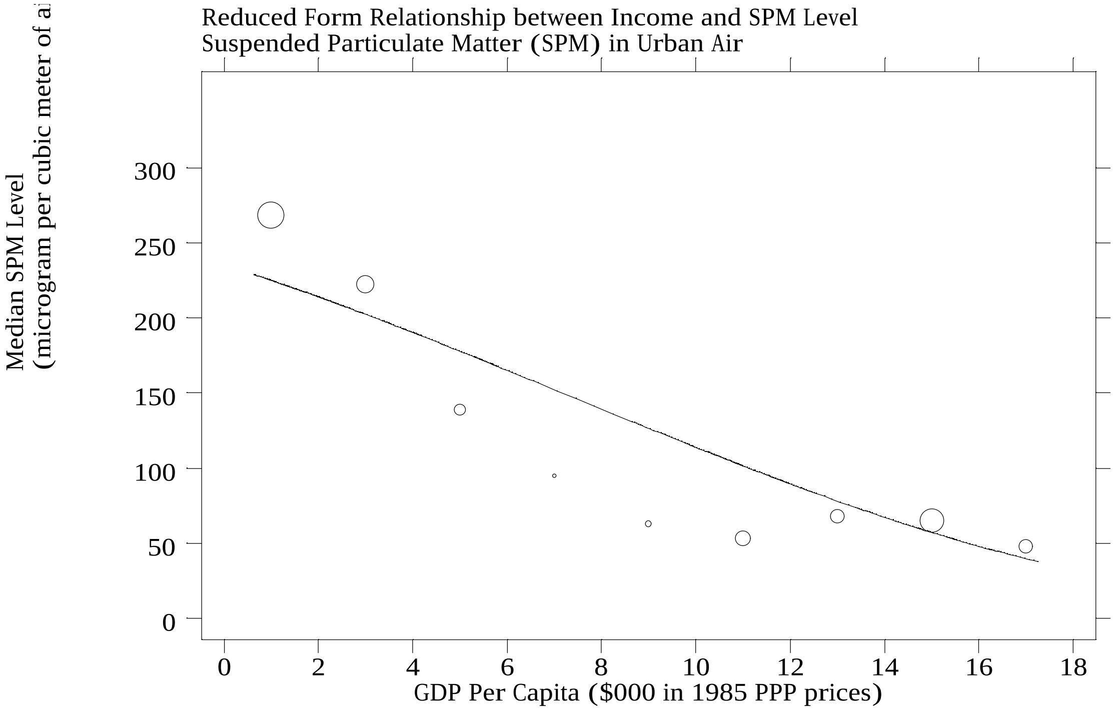


Figure-6

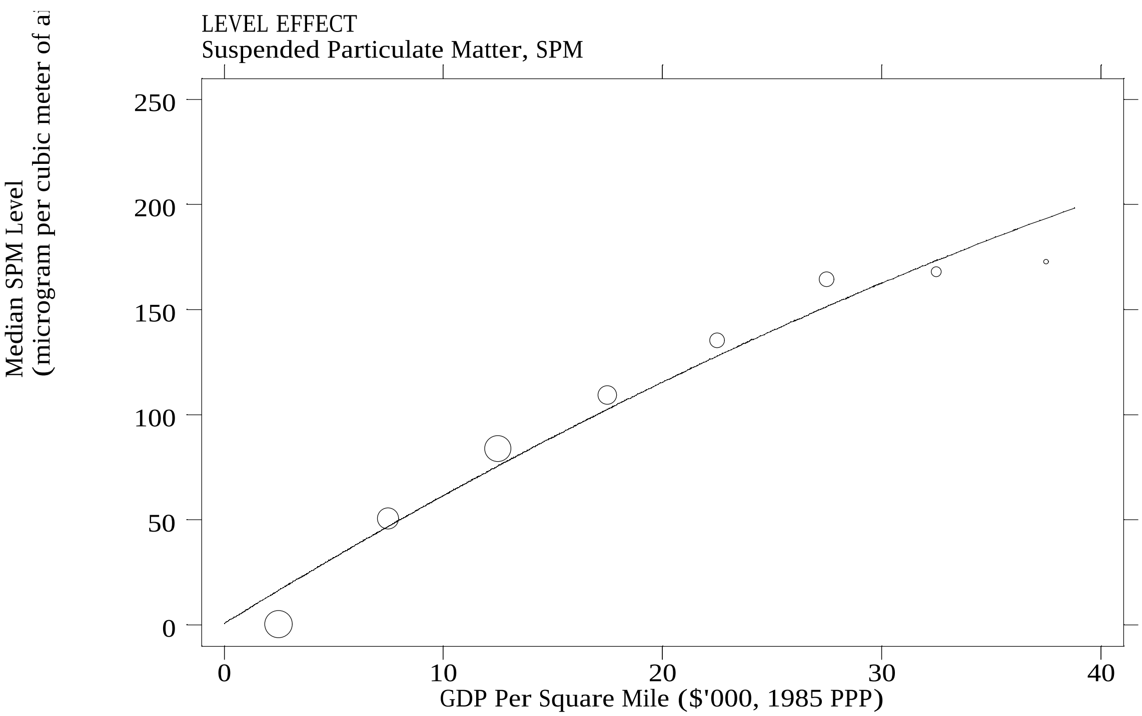


Figure-7

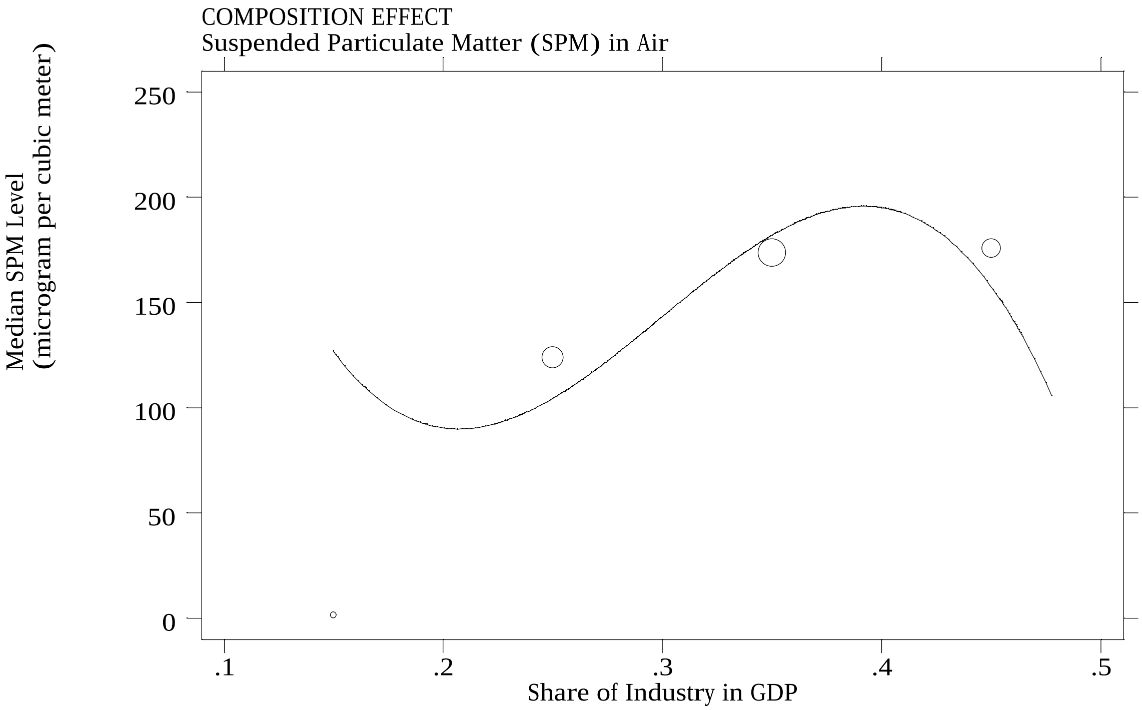


Figure-8

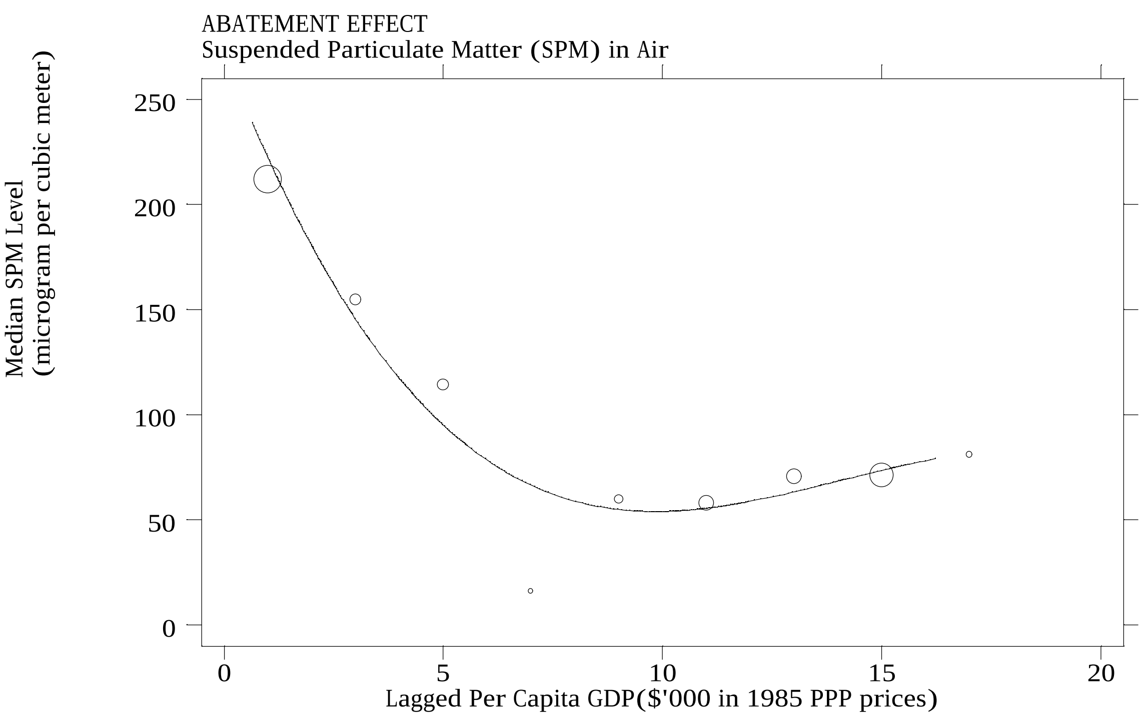


Figure-9