NSF award #0525251
Human and Social Dynamics Small Exploratory Grant

FINDINGS FILE

1. The construction of an index of air pollution

The structure of the problem of measuring overall air pollution has strong similarities with the structure of the problem of multidimensional deprivation. Each type of air pollution in a location can be considered to be a form of deprivation for the population in that location when it exceeds a certain level. Analytically, this is similar to the notion that when a person's education, calorie consumption, etc. falls below certain benchmarks specified for these different 'atributes', the different shortfalls represent different types of deprivation for the person concerned.

Consider the familiar structure of the problem of measuring overall deprivation of a group of individual in a multidimensional framework where we have several attributes such as education, calorie consumption, recreation, etc. Assuming that we have *n* individuals and *g* attributes, the typical conceptual structure used to measure the overall deprivation of the group of *n* individuals requires us to proceed through the following successive stages:

(i) First, we need to measure the achievement y_{ik} of each individual *i* in terms of each attribute *k*.

(ii) For each attribute *k*, we specify the benchmark level, z_k, of attribute *k*: if an individual's achievement in terms of attribute *k* falls below z_k, then he is considered to be deprived in terms of attribute *k*, the shortfall $z_k - y_k$ being the measure of his deprivation in terms of attribute *k*.

(iii) For every individual *i* and every attribute *k*, if y_k falls short of z_k, then the shortfall $z_k - y_k$ is 'normalized' in some way so that the normalized shortfall, denoted by w_{ik}, lies in the interval [0, 1]. If $z_k - y_k \leq 0$, then the normalized shortfall, $w_{ik} = 0$.

(iv) For each individual *i*, we aggregate w_{ik}, *k=1,...,g*, so as to get the overall deprivation w_i of individual *i*.

(v) Finally, we aggregate $w_1,.....w_n$ to arrive at the overall deprivation of the entire group of individuals.

To measure overall air pollution in a given location, we follow an analogous procedure. There will, however, be one significant difference. We would not have any analogue of step (v) above. The reason is as follows. The achievement, as well as the deprivation, in terms of an attribute such as calorie consumption or education can clearly vary between individuals. In contrast, we shall assume that all individuals in a given location 'consume' any particular type of air pollution to the same extent. Strictly speaking, this is not true. People who need to work mainly outdoors may ingest more of some pollutants than people who work mainly indoors. Also, some types of air pollution may affect some people more than others depending on their age and health: pollution may affect babies more adversely than it affects young adults and people who already have respiratory problems may suffer more from air pollution than people who do not have these problems. Differential impacts of any specific type of air pollution on different segments of the population in any given location cannot be completely ignored in developing a measure of overall air pollution but compiling the necessary data to address this problem is a daunting task. We shall basically treat the population in a given location as one individual for the purpose of measuring air pollution in that location. As a consequence, in developing a measure of air pollution in any given location, we would not have any counterpart of Step (v) usually involved in the measurement of overall deprivation of a group of individuals. Each of the other steps (Steps (i) through (iv)) outlined above will, however, have its counterpart for our purpose.

Let n be the number of locations and m be the number of different types of air pollutants (in our specfic empirical application, m=5). Let J denote $\{1,.....m\}$. For every period t, every location i, and and every pollutant j? J, let $x_{it}^{(j)}$ be the amount of pollutant j in the air in location i in period t. For information about $x_{it}^{(j)}$, we rely on the guidelines provided by the Environmental Protection Agency (EPA). For each of the air pollutants, EPA converts the concentration values to a scale from 0 to 500 (a higher number indicates a greater concentration of the pollutant under consideration). The range from 0 to 50 indicates the range over which one would not have any health concerns while the range from 301 to 500 indicates serious hazards. Given this, for each pollutant j, we shall choose the benchmark level of pollution, denoted by $x_{\min}^{(j)}$, as the concentration of the pollutant j corresponding to the benchmark of 50. For every location i, every period t, and every pollutant j? J, if $x_{it}^{(j)}$ does not exceed $x_{\min}^{(j)}$, we shall say that the excess of pollutant j in location i and time period t is 0; otherwise, the excess is measured by $x_{it}^{(j)} - x_{\min}^{(j)}$.

Our next step is to have a convenient normalization that will ensure that the normalized version of the excess will lie in the interval [0,1]. For every location i, every period t, and every pollutant j? J, we define the normalized excess as follows:

$$p_{it}^{(j)} = 0 \qquad if \quad x_{it}^{(j)} \le x_{\min}^{(j)}$$

$$p_{it}^{(j)} = \frac{x_{it}^{(j)} - x_{\min}^{(j)}}{x_{\max}^{(j)} - x_{\min}^{(j)}} \qquad if \quad x_{it}^{(j)} > x_{\min}^{(j)}$$

where $x_{\max}^{(j)}$ is the maximum concentration of pollutant j corresponding to the EPA benchmark of 500. Let p_{it} denote the normalized air pollution vector $(p_{it}^{1},.......p_{it}^{m})$ for location i in period t. Clearly, $p_{it} \in [0,1]^{m}$.

Finally, we have the problem of aggregating $(p_{it}^1,.......p_{it}^m)$ to get a unique scalar, I_{it} as a measure of overall air pollution in location *i* in period *t*. Let *f* be a functional rule, which, for every $p_{it} \in [0,1]^m$, specifies exactly one non-negative real number I_{it}. We write $I_{it} = f(p_{it})$.

What are the properties that one should require *f* to satisfy? Definition 1 introduces a set of properties of *f* (mathematical statements of these properties are found in our working paper).

Definition 1. Consider the function $f:[0,1]^m \rightarrow R_+$, where R_+ denotes the set of non-negative real numbers. The function *f* satisfies:

(i) **continuity**: a small change in the normalized excess of an air pollutant should not lead to a discontinuous jump in the overall measure of air pollution;

(ii) **normalization**: the overall index of air pollution is bounded between zero and one;

(iii) **monotonicity**: other things remaining the same, an increase in the normalized excess of a pollutant should lead to an increase in the overall index of air pollution;

(iv) **increasing marginal deterioration**: other things remaining the same, when the normalized excess of an air pollutant increases by a given amount, the additional damage will be greater the greater the initial level of normalized excess of that pollutant; the damage caused by an unfavorable factor tends to increase at an increasing rate as the unfavorable factor increases in volume/size/quantity.

(v) **independence**: the effect of an increase in the normalized excess of pollutant *j*, when the normalized excess of other pollutants are held fixed, does not depend on the levels at which these other normalized excesses are held fixed.

(vi) **symmetry**: different pollutants should play a similar role in the index of overall air pollution so that it would no matter whether, say, the normalized excess of pollutant *j* is 0.3 and that of pollutant *j'* is 0.2 or the other way round;

(vii) **uniform scale-invariance**: the overall index of air pollution should be preserved when there is a change of scale across all pollutants.

Continuity, monotonicity, and increasing marginal deterioration are very plausible properties of an aggregation function *f*. Normalization is a convenient convention, and, as such, it does not impose any substantive restrictions on the aggregation function *f*. Unlike continuity, monotonicity, and increasing marginal deterioration, independence, symmetry, and uniform scale invariance may not be universally acceptable. The independence property rules out the possibility of interaction between the different pollutants. This may be unrealistic in some contexts as it is possible that an increase in one pollutant may mitigate or exacerbate the effect of an increase in another pollutant. We did not, however, get any specific information about such relation of interdependence between the air pollutants with which we are concerned. We have, therefore, chosen to retain the property of independence.

The following two results are due to Chakraborty, Pattanaik, and Xu (2004) :

Proposition 2. A function $f:[0,1]^m \rightarrow R_+$ satisfies continuity, monotonicity, normalization, independence, symmetry, and uniform scale invariance if and only if, for some a >0, $f(p)=\frac{1}{m}\sum_{j\in J}(p^j)^a$, for all $p\in[0,1]^m$.

The formal structure of this class of aggregation functions is essentially the same as that of the well-known Foster-Greer-Thorbeck (FGT) class of deprivation measures. Note, however that the FGT rule aggregates the normalized deprivation levels of the individuals in a society so as to derive the measure of deprivation of the society as a whole; in contrast, in our case, the aggregation function aggregates the normalized excesses of different types of air pollutants to arrive at a measure of overall air pollution.

Proposition 3. A function $f:[0,1]^m \rightarrow R_+$ satisfies continuity, monotonicity, normalization, independence, symmetry, uniform scale invariance, and increasing marginal deterioration if and only if, for some a >1, $f(p)=\frac{1}{m}\sum_{j\in J}(p^j)^a$, for all $p\in[0,1]^m$.

The following definition focuses on some special members of the class of aggregation functions specified in Proposition 2 and also an aggregation function used by the EPA in constructing the AQI, which is not a member of that class.

Definition 4. Consider a function $f:[0,1]^m \rightarrow R_+$.We say that *f* is

(i) the average of excesses (AE) if and only if , for all $p\in[0,1]^m$, $f(p)=\frac{1}{m}\sum_{j\in J}(p^j)$;

(ii) the average of squared excesses (ASE) if and only if, for all $p\in[0,1]^m$, $f(p)=\frac{1}{m}\sum_{j\in J}(p^j)^2$; and

(iii) the maximum of excesses (ME) if and only if, for all $p\in[0,1]^m$, $f(p)=\max(p^1,.....p^m)$.

Remark 5. AE and ASE are members of the class of aggregation functions given by Proposition 2: AE is the special case when a=1 and ASE is the special case when a=2. ME is the aggregation function used by the Environmental Protection Agency in construction the Air Quality Index.

Remark 6. Since AE is a member of the class specified in Proposition 2, it satisfies continuity, monotonicity, normalization, independence, symmetry, and uniform scale invariance. It is, however, clear that AE does not satisfy the appealing property of increasing marginal deterioration since, under AE, the overall index of air pollution increases at a constant rate as the normalized excess of one pollutant increases, the normalized excesses of other pollutants remaining the same.

Remark 7. Being a member of the class of aggregation functions defined in Proposition 3, ASE satisfies continuity, monotonicity, normalization, independence, symmetry, uniform scale invariance, and increasing marginal deterioration.

Remark 8. Though ME satisfies continuity, normalization, and symmetry, it fails to satisfy the compelling property of monotonicity as well as increasing marginal deterioration and independence. The violation of monotonicity is a particularly disturbing feature of ME.

Remark 9. In order to have the same units in AE, ASE and ME, we need a transformation of ASE. A transformed measure like $\sqrt{ASE}$ will still satisfy the properties of continuity, monotonicity, normalization, symmetry, and increasing marginal deterioration, however the properties of independence and uniform scale invariance are not satisfied.

The following table provides a summary of properties of AE, ASE, $\sqrt{ASE}$, and ME.

	AE	*ASE*	$\sqrt{ASE}$	*ME* (EPA-type measure)
Continuity	yes	yes	yes	Yes
Monotonicity	yes	yes	yes	no
Increasing marginal deterioration	no	yes	yes	no
Normalization	yes	yes	yes	yes
Independence	yes	yes	no	no
Symmetry	yes	yes	yes	yes
Uniform Scale Invariance	yes	yes	no	yes

2. Exploiting the guidelines of the Environmental Protection Agency

The EPA constructs the Air Quality Index (AQI) for metropolitan areas with more than 350,000 inhabitants. EPA monitors five major criteria air pollutants: Ozone (O_3), Particulate matter (PM_{10} and $PM_{2.5}$), Carbon Monoxide (CO), Sulfur Dioxide (SO_2), and Nitrogen Dioxide (NO_2). Their concentration values are measured by monitors in several locations and converted to a scale form 0 to 500. The values in this scale are associated with health levels:

0-50	Good air quality
51-100	Moderate air quality
101-150	Unhealthy for sensitive groups
151-200	Unhealthy
201-300	Very unhealthy
301-500	Hazardous

For any day t, the index is calculated as $AQI_t = \max\{O_3, PM_{10}, PM_{2.5}, SO_2, NO_2\}$. This index is of the type *ME* provided in Definition 4.

We have calculated the indexes proposed in Definition 4 following the information provided by EPA. For a given health level, the following table establishes the equivalence among pollutants. For instance, a reading of $O_3 = 0.064\,ppm$ is equivalent (health-wise) to a reading of $CO = 4.4\,ppm$.

Original EPA cuts. Equivalence among pollutants for a given health level

EPA cut points	$O_3(ppm)$ 8 hour	PM_{10} $(\mu g/m^3)$	$PM_{2.5}$ $(\mu g/m^3)$	SO_2 (ppm)	CO (ppm)
50	0.064	54	15.4	0.034	4.4
100	0.084	154	40.4	0.144	9.4
150	0.104	254	65.4	0.224	12.4
200	0.124	354	150.4	0.304	15.4
300	0.374	424	250.4	0.604	30.4
400	0.504	504	350.4	0.804	40.4
500	0.604	604	500.4	1.004	50.4

The next step is to construct the deprivation gaps for each pollutant and for each health level. For instance, the deprivation gap for O_3 is calculated as

$$p^{(O_3)} = \frac{x^{(O_3)} - 0.064}{0.604 - 0.064}$$

where $x^{(O_3)}$ is the reading from the monitor in a given site and a given time. By calculating the deprivation gaps for all five pollutants we obtain the following table that assigns a deprivation cut-off for each health level and for each pollutant. From this table we also read the equivalence between different cut-offs. For instance a deprivation of 18.18% in PM_{10} is equivalent (health-wise) to a 3.70% deprivation in O_3.

Deprivation gaps (%) for different health levels

EPA cut points	O_3 8 hour	PM_{10}	$PM_{2.5}$	SO_2	CO
50	0.000000	0.000000	0.000000	0.000000	0.000000
100	3.703704	18.18182	5.154639	11.34021	10.86957
150	7.407407	36.36364	10.30928	19.58763	17.39130
200	11.11111	54.54545	27.83505	27.83505	23.91304
300	57.40741	67.27273	48.45361	58.76289	56.52174
400	81.48148	81.81818	69.07216	79.38144	78.26087
500	100.0000	100.0000	100.0000	100.0000	100.0000

However, to aggregate the deprivation gaps over the five pollutants we need to consider their equivalence, otherwise it would be difficult to assign a health level to the value of the aggregated index. For instance, suppose that $PM_{10} = 18.18$ (moderate health level) and $O_3 = 7.40$ (unhealthy for sensitive groups), an average of the two gaps is 12.79%. What is the health level associated with this average? If you choose the cut-off for PM_{10} the air quality is good but if you choose the cut-off for O_3 the air quality is unhealthy. Thus, our last step in the construction of the indexes proposed in Definition 4 is to convert the above deprivation gaps to a common unit. The following table shows the substitution rates among pollutants with respect to PM_{10} for each health level.

Implicit 'substitution rates' among pollutants with respect to PM10 for a given health level

EPA cut points	O_3 8 hour	PM_{10}	$PM_{2.5}$	SO_2	CO
50	NA	NA	NA	NA	NA
100	4.909091	1.0	3.527273	1.603306	1.672727
150	4.909091	1.0	3.527273	1.856459	2.090909
200	4.909091	1.0	1.959596	1.959596	2.280992
300	1.171848	1.0	1.388395	1.144817	1.190210
400	1.004132	1.0	1.184532	1.030697	1.045455
500	1.000000	1.0	1.000000	1.000000	1.000000

For instance, the substitution rate ($r^{(O_3)}$) between PM_{10} and O_3 is 4.909091 $(= \frac{18.18182}{3.703704})$, that is to say that the deprivation gap of O_3 needs to be multiplied by a factor of 4.90 to make it equivalent (health-wise) to the deprivation gap of PM_{10} corresponding to a health level of 100. Thus, the deprivation gaps of the pollutants need to be modified accordingly and it will be expressed in PM_{10} units. For instance in the case of O_3, the modified deprivation gap is

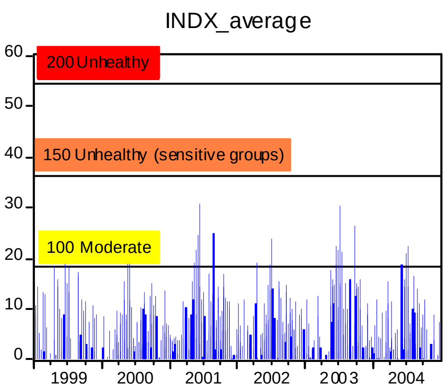
$$\tilde{p}^{(O_3)} = r^{(O_3)} \times \frac{x^{(O_3)} - 0.064}{0.604 - 0.064}$$

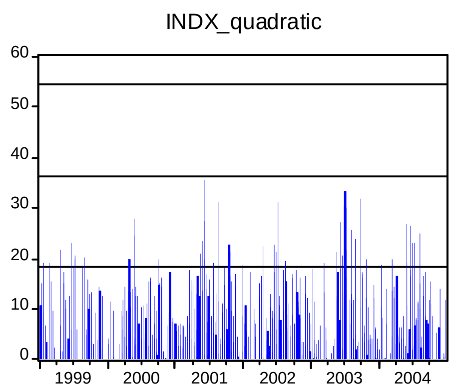
3. Some examples of pollution indexes: Riverside, San Diego, and Los Angeles

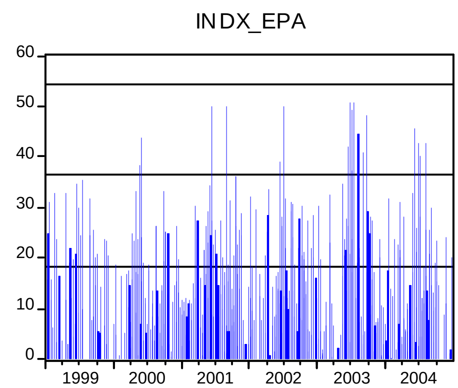
We implement the proposed air quality indexes for the 38 locations in Southern California for which we have complete readings of the five pollutants. We illustrate our results for three representative areas: Los Angeles, San Diego, and Riverside.

RIVERSIDE (monitor site 2596, Rubidoux)

AE | $\sqrt{ASE}$ | ME

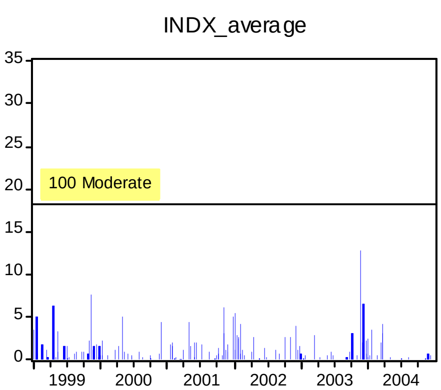


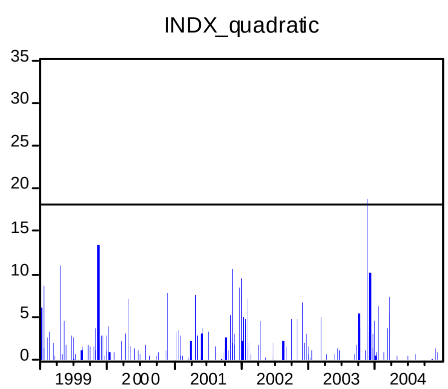


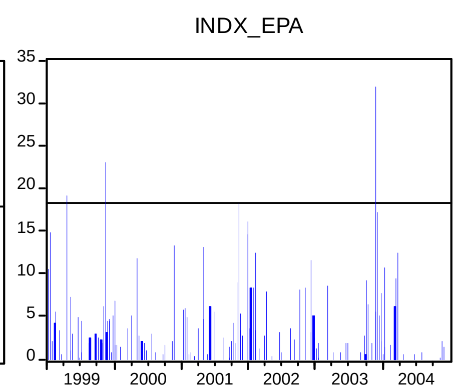


SAN DIEGO (monitor site 2327, El Cajon)

AE | $\sqrt{ASE}$ | ME

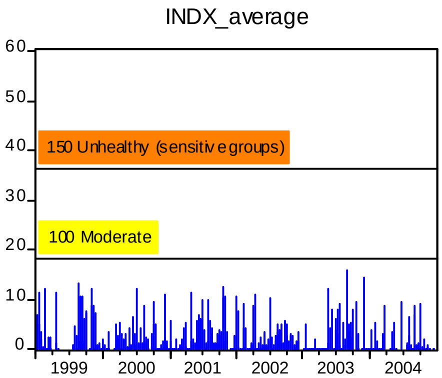


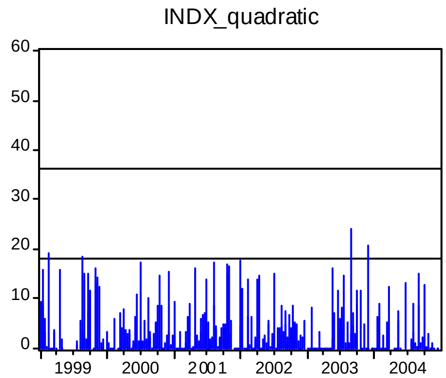


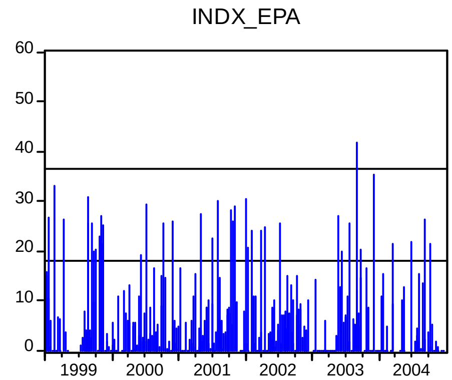


LOS ANGELES (monitor site 2484, Azusa)

AE $\sqrt{ASE}$ *ME*







In general, the EPA index by its own construction shows the largest readings across locations reaching the Unhealthy (sensitive groups) zone more often than any of the other indexes. For a given location, the three indexes are highly correlated over time but qualitatively they can be very different depending on the location. For instance, for San Diego there are not major differences among the three indexes, this location has healthy air; for Los Angeles, there are not major differences between the readings of the *AE* and the $\sqrt{ASE}$ indexes but there are differences with the *ME*; according to *ME* the air quality in this area is borderline moderate with some episodes of unhealthy air but if we were to read *AE* and $\sqrt{ASE}$ the air quality would be overall good. In Riverside we find the largest differences across indexes. The *ME* index offers a very alarmist picture with some episodes of very unhealthy air, however the $\sqrt{ASE}$ index is more moderate with worrisome episodes limited to some days in the summer months. In this sense the EPA is a very conservative index.