Seventh draft. Please do not cite or quote without permission.

MODELLING DISTRIBUTED LAG EFFECTS IN EPIDEMIOLOGICAL TIME SERIES STUDIES

by

David Maddison

CSERGE Working Paper

MODELLING DISTRIBUTED LAG EFFECTS IN EPIDEMIOLOGICAL TIME SERIES STUDIES

by

David Maddison

Centre for Social and Economic Research
on the Global Environment
University College London
and
University of East Anglia

Acknowledgements

The Centre for Social and Economic Research on the Global Environment (CSERGE) is a designated research centre of the UK Economic and Social Research Council (ESRC). The author would like to acknowledge helpful comments made on an earlier version of this paper by Luis Cifuentes, Alison Loader, Robert Maynard, Trevor Morris, David Pearce and the contributors to a meeting organised by the World Health Organisation in Bilthoven, 20-22 November, 2000. The usual disclaimer applies.

ISSN 0967-8875

Abstract

The paper argues that much of the existing literature on air pollution and mortality deals only with the transient effects of air pollution. Policy, on the other hand, needs to know when, whether and to what extent pollution-induced increases in mortality are reversed. This involves modelling the entire distributed lag effects of air pollution.

Borrowing from econometrics this paper presents a method by which distributed lag effects can be estimated parsimoniously but plausibly estimated. The paper presents a time series study into the relationship between ambient levels of air pollution and daily mortality counts for Manchester employing this technique.

Black Smoke is shown to have a highly significant effect on mortality counts in the short term. Nevertheless we find that 80 percent of the deaths attributable to BS would have occurred anyway within one week whereas the remaining 20 percent of individuals would otherwise have enjoyed a normal life expectancy.

Keywords: Air Pollution and Mortality; Time-Series Analysis; and Distributed Lags.

1. Introduction

A vast number of epidemiological studies have identified particulate matter and, less frequently, other air pollutants as being statistically related to daily mortality counts. Examples include Schwartz and Dockery (1992) for Philadelphia, Anderson et al (1996) for London, Touloumi et al (1996) for Athens, Cropper et al (1997) for Delhi, Saldiva et al (1995) for Sao Paolo and Ostro et al (1996) for Santiago to name but a few. Despite the fact that these studies have been undertaken in very different locations the methodology followed by these studies is generally same. The procedure is to use Least Squares or Poisson regression analysis to control for seasonal variations in daily mortality counts along with variations in meteorological conditions, day-of-the-week effects and one or two pollution variables.

Although these studies have alerted policy makers to the potential harm from ambient pollution concentrations the results provided by time-series studies into the mortality effects of air pollution are nonetheless turning out to be of limited value from the policy perspective. One problem relates to the current emphasis on single and dual pollutant models in the epidemiological literature that makes it difficult to attribute mortality effects to particular pollutants. The other problem, which is the main focus of this paper, involves the way in which air pollution impacts are entered into the model. Typically air pollution is included either as a contemporaneous variable or with one or two lags. Although such a methodology may succeed in demonstrating that air pollution and premature mortality are causally linked sensible policy responses cannot be formulated only on the basis of knowledge of the transient impact of air pollution on mortality. Policy needs to

know when, whether and to what extent pollution-induced increases in mortality are reversed.

Being aware of this limitation to their work contributors to the epidemiological literature are typically very careful to specify that the empirical evidence, as it stands, does not say anything about the extent to which life has been foreshortened as a consequence of poor air quality[1,2]. Indeed the epidemiological literature states in a number of places that it is impossible to measure the extent of life lost using time series studies (see for example Anderson et al, 1996, or McMichael et al, 1998).

The paper introduces a simple modelling technique in which the entire infinite lagged response of daily mortality to increases in air pollution is modelled in a plausible yet parsimonious fashion. In so doing the technique nests the kind of models that have so far been used to explore the links between air pollution and mortality as a special case. It argues that such methods provide a far superior description of variations in daily mortality rates and yield insights of greater relevance to policy. In particular, if one is able approximate the infinite distributed lagged impact then one can observe the rate at which excess mortality counts attributed to air pollution are reversed. Finally this study provides an

[1] An exception is the time series analysis of the link between particulate matter and daily mortality counts in Delhi by Cropper et al (1997). Separate regressions are run for daily mortality counts in different age groups and the assumption is made that those individuals who succumb to the effects of air pollution would otherwise have enjoyed a normal life expectancy.

[2] Obviously the extent of life lost due to the chronic effects of air pollution cannot be inferred from time series studies. These effects require a completely different approach (see for example Pope et al, 1995).

illustration of the technique in the context of a study of the links between air pollution and mortality in Manchester[3].

The following section offers a discussion and critique of current practice in modelling the distributed lag effects of air pollution on mortality. An alternative method of modelling the distributed lags is introduced and the relative advantages of the method are explained. The remainder of the paper describes the empirical implementation of the technique. Section three discusses the data used to implement the model along with the econometric modelling techniques employed. Section four discusses the implications of the results and the final section concludes.

[3] Recently, using very different techniques to those proposed here, Zeger et al (1999) claim to have have produced 'harvesting-resistant' estimates of the effects of air pollution on mortality. These authors also recognise the potential policy relevance of whether the victims of air pollution are primarily those who are already frail and whose life expectancy is already quite short. Their estimates of the health effects of air pollution are larger than those produced by conventional modelling techniques.

2. Modelling Lags in Time Series Air Pollution-Mortality Studies

Almost without exception standard practice in the statistical modelling of the relationship between daily mortality counts and ambient levels of air pollution is to include just contemporaneous, once or twice-lagged values for air pollution into a regression equation (see Gouveia and Fletcher, 2000 for a recent example). In these cases the decision about which lag to select is seldom explained in detail but often it seems that the single most significant lag is selected as for example in Katsouyanni et al (1996).

It is however unlikely that the researchers who present such models in the literature intend them to be taken too literally. For example, a researcher who seeks to explain variations in daily mortality rates by the value of a pollutant once lagged is presumably not claiming that the totality of the effect is experienced precisely one day afterwards – none before and none after. Nevertheless what such investigators actually end up estimating is the transient impact of air pollution.

An extension of this approach is to estimate the model using single lagged-values for air pollutants ever further back in time. In this way one might suppose that the lagged impacts of air pollution on mortality would emerge. Using data from Barcelona (Sunyer et al, 1996), an example of this approach is contained in Appendix 1 of the report of the UK Department of Health's Committee on the Medical Effects of Air Pollution (1998). The problem with this approach is that,

to the extent that pollutant variables are auto-correlated over time, the effects of adjacent lag terms will also be picked up[4].

Running a regression on a moving average of air pollution levels is perhaps a small improvement on including just single lags (e.g. Schwartz et al, 1996). But since it compels the lagged effects of pollutants to be exactly equal on consecutive days and then disappear it cannot be terribly realistic. In other papers researchers freely estimate the coefficients on two or more consecutive pollution-levels and present the cumulated or 'interim' impacts of air pollution (e.g. Dab et al, 1996). These constitute a further improvement but once again assume that the impact of air pollution on mortality is zero after two or three days. A more realistic model would allow for the lagged effects of pollutants gradually to decay and perhaps turn negative if the deaths of susceptible individuals were being brought forward, before vanishing.

In theory the means to explore such a possibility would be to estimate freely a model containing many lagged terms for each of the pollutants. In practice however analysts have, unsurprisingly, been reluctant to add a large number of additional regressors to their models. They claim, quite correctly, that estimation of the unrestricted regression will not be able to locate the lag structure because it will be plagued by multicollinearity between the lagged regressors.

[4] In fairness the presence of these diagrams in the report was mainly intended to show that whilst variations in daily mortality are correlated with lagged levels of air pollution, future levels of air pollution do not correlate with variations in daily mortality. Hence there is evidence of causality.

These observations on current practice prompt the following questions. First, how can a distributed lag structure be modelled parsimoniously in the context of air pollution-mortality studies (or indeed any study)? Secondly, how sensitive are the estimated relative risk ratios to seemingly arbitrary decisions regarding the period of time over which to cumulate the lagged impacts of air pollution? Thirdly, to what extent can adding a more realistic lag structure reduce the unexplained variance in a model? The first of these questions is addressed in the following paragraphs; the latter questions can be answered only by empirical research and are deferred to the second half of the paper.

A variety of techniques to approximate lag structures have been proposed in the econometrics literature and these may be useful in the context of epidemiological studies too. This is a view shared by Schwartz et al (1996) who argue that the epidemiological literature needs to pay greater attention to econometric approaches to modelling distributed lags. It is also plausible to assume that a more systematic approach to specifying lags would allow better comparison between sites.

One widely explored method of estimating lagged impacts is the polynomial approach of Almon (1965). The technique involves making the assumption that the distribution of lag coefficients can be represented by a polynomial of sufficiently high order. The coefficients of the polynomial are estimated absorbing the order of the polynomial plus one degrees of freedom.

In apparently the first epidemiological study to utilise this technique, Schwartz (2000) employs a quadratic polynomial lag with a maximum lag of five days in

an United States based analysis of the link between ambient concentrations of particulate matter and the deaths of over-65s. He finds that the use of the technique increases the measured relative risk ratios associated with particulate matter compared to those associated with a one-day lag or a two-day moving average. Schwartz argues that this method should become standard practice in the epidemiological time-series studies.

The method of polynomial lags however suffers from the defect that it is necessary to specify a finite endpoint prior to estimation. There has, in the econometrics literature, been an extensive analysis of the consequences of miss-specifying the lag length (as well as the order of the polynomial; see for example Hendry et al, 1984). Simply assuming a maximum lag length is hazardous as the Almon lags technique will genially distribute the effects over the entire lag whether this is appropriate or not[5]. Finally, the technique is acutely sensitive to missing observations and has extreme difficulty in capturing any long-tailed lag distribution of the type that might be expected in epidemiological time-series studies (see for example Maddala, 1977).

In the opinion of the author these features serve to make the polynomial lags technique quite unsuitable for use in epidemiological time-series studies. Partly because of these shortcomings the polynomial lags technique has seen relatively few recent applications in the field of applied econometrics either. Most econometricians resort to the method of 'rational lags' (Jorgenson, 1966) in situations in which the modelling of distributed lags is called for.

[5] The Schwartz study might be criticised for simply assuming a maximum lag of five days and the appropriateness of a polynomial of degree two. There are protocols for selecting the

The idea behind rational lags is that *any* infinite distributed lag function can be approximated by the ratio of two finite polynomials in the lag operator[6]. As such the rational lags technique involves nothing more than the inclusion of additional explanatory variables. Testing the significance of these extra variables is therefore very straightforward. Furthermore it is possible to retrieve the implied parameters of distributed lag function in a relatively straightforward manner enabling the analyst to observe the lagged impact of a pulse change in the independent variable (see appendix 1). The rational lag technique seems well suited to dealing with issues that arise in epidemiological time-series studies. But to the knowledge of the author this is first occasion on which its use has been proposed in such a context.

Before moving to an empirical demonstration of the use of rational lags, it is appropriate to note one further issue that was hinted at in the introduction. Although it is not the main focus of this paper, the empirical analysis that follows features an important difference that distinguishes it from much of the existing empirical literature. This is the fact that no less than four different air pollutants are simultaneously included in the model. The existing literature is by contrast

---

appropriate lag lengths and order of the polynomial but these do not appear to have been followed.

[6] The lag operator L is defined by $LX_t = X_{t-1}$. The lag operator may be applied more than once so that $L^2X_t = X_{t-2}$. It may also be handled algebraically like an ordinary variable such that $L^1L^2X_t = X_{t-3}$. Consider the following infinite distributed lag model:

$$Y_t = \alpha + \sum_{i=0}^{i=\infty} \beta_i L^i X_{t-i} + e_t$$

Rather than estimating the unrestricted model Jorgenson's Rational Lag technique involves estimating the following equation by means of non-linear least squares:

characterised by one and two-pollutant models. But given the non-zero correlations which often exist between different air pollutants single-pollutant models risk explaining what are essentially the same deaths several times over[7]. This hampers attempts to determine which out of a range of air pollutants are responsible for the empirically observed mortality impacts and prevents researchers from reaching any conclusions regarding the overall health burden imposed by pollution-generating activities. These criticisms of current practice are also reflected in the report of the UK Department of Health's Committee on the Medical Effects of Air Pollution (1998).

$$Y_t = \alpha + \frac{\gamma_0 + \gamma_1 L X_t + \gamma_2 L^2 X_t + \ldots + \gamma_j L^j X_t}{1 + \omega_1 L X_t + \omega_2 L^2 X_t + \ldots + \omega_k L^k X_t} + e_t$$

[7] Schwartz et al (1996) remark that "One occasionally sees studies that have fitted regression models using four or even more collinear pollutants in the same regression… Given the non-trivial correlation of the pollutant variables and the relatively low explanatory power of air pollution these for mortality or hospital admissions such procedures risk letting the noise in the data choose the pollutant". The author however believes that alternative procedures risk letting the researcher choose the pollutant. Matters are less clear when one recognises that ambient concentrations recorded by monitors may be a poor representation of the typical individual's exposure.

3. The Empirical Analysis

Daily data on non-accidental all-cause mortality (MORT) is taken from Manchester from the start of 1988 to the end of 1992 – a period of some 1,825 days[8]. Measures of 24-hour averages for SO_2 and black smoke both in $\mu g/m^3$ are taken from three different sites[9]. For both pollutants a single index was formed taking the geometric mean. An issue arises in the case of the SO_2 measures in that an unusual number of observations indicate zero concentrations of SO_2. These readings were thought to be indicative of alkaline contamination and in what follows these observations are treated as if they were missing[10]. Data on NO_2 is taken from Manchester Town Hall and data on 8-hour maximum O_3 is taken from the suburban site of Glazebury[11]. Both of these records are in terms of ppb and are highly fragmented. Data on daily mean temperature (TEMP) in °C and relative humidity (HUMID) as a percentage are taken from Manchester Ringway airport. The data are described in tables 1 and 2.

[8] The areas covered include Bolton, Bury, Manchester, Oldham, Rochdale, Salford and Stockport. I am grateful to Trevor Morris of the UK Department of Trade and Industry for supplying these data.

[9] These sites are Manchester 11, Manchester 15 and Manchester 21.

[10] I am grateful to Alison Loader of AEA Technology for advice on this point. Even after these observations have been discarded the SO_2 monitors continue to show only a low correlation with one another.

[11] During the period January 1988 to December 1992 there were no O_3 monitors operating in the centre of Manchester. Monitors established at a later date show a high correlation of 0.82 with the monitor in Glazebury.

Table 1: Descriptive Statistics

Variables	Mean	Std. Dev.	Minimum	Maximum
MORT	57.95	11.03	30.00	120.00
TEMP (°C)	10.41	4.92	-3.10	26.70
HUMID (%)	65.89	15.47	24.00	100.00
BS ($\mu g/m^3$)	16.10	14.75	1.00	179.85
SO_2 ($\mu g/m^3$)	42.93	18.84	8.46	241.32
NO_2 (ppb)	27.12	11.75	5.83	116.25
O_3 (ppb)	32.32	13.97	3.00	96.00

Source: See text.

Table 2: The Correlation Matrix

	MORT	TEMP	HUMID	BS	SO_2	NO_2	O_3
MORT	1.00						
TEMP	-0.47	1.00					
HUMID	0.27	-0.34	1.00				
BS	0.21	-0.36	0.21	1.00			
SO_2	0.11	0.01	-0.08	0.45	1.00		
NO_2	0.07	-0.22	0.02	0.68	0.42	1.00	
O_3	-0.23	0.46	-0.57	-0.40	-0.08	-0.08	1.00

Source: See text.

Apart from pollution and meteorological variables, a number of other variables were incorporated into the regression analysis. Six dummy variables (SUN, MON, TUE etc) were included for different days of the week. The method used to control for seasonal variations in mortality was to include eleven dummy variables for each of the different months (JAN, FEB, MAR etc)[12]. A linear time trend (TIME) was also included to capture autonomous changes in the daily mortality rate. Finally, a dummy variable (FLU) was included to test for the possibility of a structural break during the three-month influenza epidemic during the winter of 1989/90. The following equation was estimated in which L is the lag operator:

[12] Most epidemiologists would agree that controlling for seasonal effects is of paramount importance. There are many different ways of doing this and some researchers prefer to regress daily mortality on sine and cosine terms of differing frequencies. Others use the monthly dummy variable approach adopted here (see for example Schwarz, 1994) although note that this approach imposes the same seasonal pattern across each year. A yet more general analysis would allow the monthly dummies to vary across years. More recent applications have employed non-parametric smoothing techniques. The author has also used sine and cosine terms at frequencies of one, two, three, four, six and twelve months to control for seasonal effects. All the important results contained in this paper appear to be completely unaffected (further details are available upon request).

$$\log(MORT_t) = \alpha + \beta_1 SUN_t + \beta_2 MON_t + \beta_3 TUE_t + \beta_4 WED_t + \beta_5 THU_t + \beta_6 FRI_t + \beta_7 JAN_t +$$

$$\beta_8 FEB_t + \beta_9 MAR_t + \beta_{10} APR_t + \beta_{11} MAY_t + \beta_{12} JUN_t + \beta_{13} JUL_t + \beta_{14} AUG_t + \beta_{15} SEP_t +$$

$$\beta_{16} OCT_t + \beta_{17} NOV_t + \beta_{18} TIME_t + \beta_{19} FLU_t + \frac{\sum_{i=0}^{i=3} \gamma_i L^i TEMP_t}{1 + \sum_{i=1}^{i=3} \omega_i L^i TEMP_t} +$$

$$\frac{\sum_{i=0}^{i=3} \psi_i L^i TEMP_t^2}{1 + \sum_{i=1}^{i=3} \zeta_i L^i TEMP_t^2} + \frac{\sum_{i=0}^{i=3} \lambda_i L^i BS_t}{1 + \sum_{i=1}^{i=3} \mu_i L^i BS_t} + \frac{\sum_{i=0}^{i=3} \nu_i L^i SO_{2t}}{1 + \sum_{i=1}^{i=3} \pi_i L^i SO_{2t}} +$$

$$\frac{\sum_{i=0}^{i=3} \theta_i L^i NO_{2t}}{1 + \sum_{i=1}^{i=3} \sigma_i L^i NO_{2t}} + \frac{\sum_{i=0}^{i=3} \rho_i L^i O_{3t}}{1 + \sum_{i=1}^{i=3} \kappa_i L^i O_{3t}} + \frac{\sum_{i=0}^{i=3} \delta_i L^i HUMID_t}{1 + \sum_{i=1}^{i=3} \eta_i L^i HUMID_t} + e_t$$

This regression equation uses the rational lag technique to approximate an infinite distributed lag on both the weather and the pollution variables. Note that a maximum lag length of i = 3 for both the numerator and denominator of the terms in air pollution and weather is sufficient to capture quite complicated lag patterns such as that described in the preceding section. This also has the advantage of encompassing the lag lengths typically encountered in epidemiological research without any restriction being imposed (e.g. Katsouyanni et al, 1996).

Initially the error term was assumed to be normally identically and independently distributed and estimates of the parameters were obtained by using maximum

likelihood estimation techniques[13]. Examination of the residuals however pointed to the presence of autocorrelation. This phenomenon, which is by no means unusual in time series analyses of air pollution and mortality, was dealt with by quasi-differencing the data and estimating no less than four autocorrelation parameters as part of the maximum-likelihood estimation routine.

Last of all, whilst non-linear higher order effects were obvious for temperature, adding higher order terms for the time trend, humidity and the pollution variables did not result in a statistically significant increase in fit[14]. The fit of the regression was quite good (the R^2 statistic is 0.47) but given the fragmented nature of the data set, the corrections for autocorrelation and the desire to treat missing values correctly (as opposed to imputing them) only 1,047 observations were used in the analysis. Full details of the estimation results are available from the author on request.

[13] In many empirical analyses the error term is assumed to be a Poisson variable. In this analysis the daily number of deaths is typically very large and there is probably no discernible difference from modelling the error term as a Normal variable.

[14] More specifically annually averaged mean temperature was subtracted from daily temperature and then squared. The resulting variable was then added to the equation.

4. **Discussion**

The statistical analysis reveals that non-accidental mortality counts in Manchester are primarily influenced by seasonal factors. The statistical model fails to detect any autonomous change in daily mortality rates although there is evidence of elevated death rates corresponding to the influenza epidemic of 1989/90. Meteorological factors also appear to be important with terms in both daily mean temperature and its squared value highly significant. But neither humidity nor day-of-the-week effects are present.

The prime question of interest is whether the exclusion of those additional terms that allow for the possibility of infinite lagged impacts represents a statistically significant loss of fit. A likelihood-ratio test suggests that the loss of fit is highly significant, a finding that provides strong support for the use of the rational-lags technique in this context[15].

Another important question is whether the exclusion of the air pollution variables would result in a statistically significant decrease in fit. Employing a Likelihood Ratio test it can be shown that both Black Smoke and NO_2 are statistically significant at the 1 percent level of significance (see table 3). Ozone is statistically significant at the 5 percent level and SO_2 is not significant even at the 10 percent level. This indicates that, at least at conventional levels of statistical significance, three out of four pollutants have at least short-term effects on mortality.

[15] The χ^2 Statistic is 73.47 against a critical value of 38.93 at the one-percent level of significance with 21 degrees of freedom.

It is difficult to compare these results to the existing literature in anything other than purely qualitative terms. First and foremost this is because most researchers are measuring either the transient impact of air pollution at variety of lag lengths or the interim impact cumulated over an arbitrary number of days. This technique, by contrast, calculates a different relative risk ratio at each lag length. Secondly, unlike most other analyses, this study calculates the mortality effects of air pollution within the context of a multi-pollutant rather than a single-pollutant model[16].

The long run effect (i.e. the cumulated lag from time t=0 to infinity) associated with a pulse increase in each air pollutant is tested using a Wald test (see table 4)[17]. Only in the case of NO_2 is the long run impact found to differ significantly from zero and even then only at the 10 percent level of significance. The implication is that over the long term the deaths associated with air pollution are brought forward rather than caused.

In a sense however, the distinction between 'deaths brought forward' and 'deaths caused' is a false one. In the long run we are all dead and the fact that the long run impact is zero is quite different from saying that air pollution is unimportant from a public health perspective. What matters is the amount of

[16] The data was also analysed in the 'traditional' way by running single and dual-pollutant models. In these models the single most significant lagged value of each air pollutant was included as a regressor variable. The results are quite similar to those found in the literature (further details are available upon request).

[17] Following on from footnote 6 the long run impact of a unit change in variable X is given by:

life lost[18]. This can best be appreciated by calculating the implied relative risk ratio at each lag length following a pulse increase in pollution.

Table 5 indicates that the most pronounced impact of BS on mortality occurs on the same day (the relative risk ratio is 1.05 per 100 μg/m^3 with a 95 percent confidence of 1.09 to 1.02). There is then a marked reduction in the number of deaths on the fourth day and from that point onwards the cumulative impacts of BS differ insignificantly from zero. Cumulating the lags over a period of seven days the relative risk ratio is 1.01 per 100 μg/m^3 with a 95 percent confidence of 1.03 to 0.99. One therefore cannot exclude the possibility of *very* small health impacts extending over a period of up to one week. Indeed, the cumulated risk is essentially unchanged as one increases the period over which the risks are cumulated from one week to infinity. The implication is that although the majority of deaths involve short term harvesting the remainder of the victims might otherwise have enjoyed a normal life expectancy. Another way of expressing the findings is to say that for every five-individuals who die as a consequence of the acute effects of air pollution, only one would have survived for more than seven days (and probably much longer). Results similar

---

$\frac{\gamma_0 + \gamma_1 + \gamma_2 + \ldots + \gamma_j}{1 + \omega_1 + \omega_2 + \ldots + \omega_k}$

[18] The statistical significance of the long run impact of air pollution on mortality might be interpreted as a test of model specification. It would be very peculiar if the model suggested that air pollution caused deaths that would otherwise *never* have occurred. On the other hand it must be remembered that technique of rational lags provides only an approximation to the true lag function. Restricting the model in order to ensure that the long-run impact of air pollution on mortality is always zero might interfere with the ability to approximate short-run impacts.

to those for BS emerge for NO_2 and O_3, namely that we cannot reject the hypothesis that the short-run effects are quickly reversed[19].

It is worthwhile reiterating that these results are not a consequence of the modelling strategy adopted: the only restriction imposed is that the lag coefficients should move in geometric progression after lag four.

Katsouyanni et al (1997) provide estimates of the cumulative impacts of BS on mortality for 8 different cities. These impacts are cumulated over two to four days depending upon whatever yielded the 'best estimate' and presumably therefore involve only the positive impacts of air pollution. Table 5 illustrates that in the case of Manchester cumulating the impacts of BS even over the first two rather than the first four days could, inadvertently, give policy-makers and researchers from other disciplines a totally different impression of the health risks posed by the acute effects of air pollution. The results in this paper highlight danger of policy-makers over-reacting to the kind of results that have so far characterised the literature reflecting only very short-term impacts.

[19] One cannot reject the hypothesis that the mortality impacts of NO_2 and O_3 are reversed within two days.

Table 3: Air Pollution as an Influence on Daily Mortality Rates

Air Pollutant	Number of Restrictions	χ^2 Statistic
BS	7	29.28***
SO_2	7	9.87
NO_2	7	26.71***
O_3	7	17.74**

*Source: see text. Note that *** means significant at the 1 percent level of significance; ** means significant at the 5 percent level of significance; and * means significant at the 10 percent level of significance.*

Table 4: The Long Run Effects of Air Pollution

Air Pollutant	Number of Restrictions	χ^2 Statistic
BS	1	0.53
SO_2	1	1.46
NO_2	1	3.73*
O_3	1	1.50

*Source: see text. Note that *** means significant at the 1 percent level of significance; ** means significant at the 5 percent level of significance; and * means significant at the 10 percent level of significance.*

Table 5: The Cumulated Impacts of Black Smoke

Lag Length in Days	Cumulated Coefficient	T-statistic
0	5.19E-04	2.85***
1	3.83E-04	2.82***
2	4.99E-04	2.88***
3	7.93E-05	0.78
4	7.41E-05	0.72
5	7.49E-05	0.73
6	7.43E-05	0.73
7	7.42E-05	0.72

*Source: see text. Note that *** means significant at the 1 percent level of significance; ** means significant at the 5 percent level of significance; and * means significant at the 10 percent level of significance.*

5. Conclusions

This paper has noted that much of the existing epidemiological literature estimates only the very short-term transient or interim-lag impacts of air pollution. Arguably however, knowledge of these impacts is not of much policy relevance. One needs to know how soon, whether and to what extent any increase in mortality is reversed and this requires estimating the entire distributed lagged impact of air pollution on mortality

Noting the deficiencies of alternative techniques, this paper has used the method of rational lags to approximate the infinite distributed lag impact of a change in air pollution. The method is straightforward and involves including additional terms that permit one to approximate the entire distributed lag. These are shown to dramatically improve the fit of the regression equation. Using the method of rational lags we find that 80 percent of the deaths attributable to BS would have occurred anyway within one week whereas the remaining 20 percent of individuals would otherwise have enjoyed a normal life expectancy. No generality is claimed for these results which in any case have wide confidence intervals attached to them.

An obvious application of the technique of rational lags would be to the analysis of the link between air pollution and various morbidity endpoints. For example, the technique of rational lags appears able to answer the following important question: are respiratory hospital admissions caused by air pollution or merely advanced (in which case air pollution imposes no additional costs on the National Health Service)?

Because the results from any one study are too uncertain to be used as a basis for policy it would be interesting to reanalyse the data from existing studies using this technique. By computing the cumulated impact of air pollution at fixed intervals (e.g. three days, one week and one month) and combining the results it may be possible to determine the speed with which excess mortality attributed to the acute impacts of air pollution is reversed. But in order to be meaningful such comparisons must be careful to compare impacts cumulated over identical periods of time and should avoid focussing solely on the very short-term impacts.

References

Almon, S. (1965) The Distributed Lag between Capital Appropriations and Expenditures *Econometrica* Vol. 33 pp178-196.

Anderson, H., A. Ponce de Leon, J. Bland, J. Bower and D. Strachan (1996) Air pollution and daily mortality in London: 1987-92. *British Medical Journal* Vol. 312 pp665-669.

Committee on the Medical Effects of Air Pollution (1998) *Quantification of the Effects of Air Pollution on Health in the United Kingdom*. London: HMSO.

Cropper, M., N. Simon, A. Alberini, S. Arora and P. Sharma (1997) The Health Benefits of Air Pollution Control in Delhi. *American Journal of Agricultural Economics* Vol. 79 No. 5 pp1625-1629.

Dab, W., S. Medina, P. Quenel, Y. Le Moullec, A. Le Tertre, B. Thelot, C. Monteil, P. Lameloise, P. Pirard, I. Momas, R. Ferry and B. Festy (1996) Effects of Ambient Air Pollution in Paris. *Journal of Epidemiology and Community Health* Vol. 50 (Supp.) pp42-46.

Gouviea, N. and T. Fletcher (2000) Time Series Analysis of Air Pollution and Mortality: Effects by Cause, Age and Socioeconomic Status. *Journal of Epidemiology and Community Health* Vol. 54 pp750-755.

Hendry, D. Pagan A. and Sargan, J. (1984) Dynamic Specifications. In Griliches, Z. and Intriligator, I. *Handbook of Econometrics* Vol. 2, North Holland: Amsterdam.

Jorgenson, D. (1966) Rational Distributed Lag Functions. *Econometrica*, Vol. 34 pp135-149.

Katsouyanni, K., J. Schwartz, C. Spix, G. Touloumi, D. Zmirou, A. Zanobetti, B. Wojtyniak, J. Vonk, A. Tobias, A. Ponka, S. Medina, L. Bacharova and H. Anderson (1996) Short Term Effects of Air Pollution on Health: A European Approach Using Epidemiologic Time Series Data: The APHEA Protocol *Journal of Epidemiology and Community Health* Vol. 50 (Supp.) pp12-18.

Katsouyanni, K., G. Touloumi, C. Spix, J. Schwartz, F. Balducci, S. Medina, G. Rossi, B. Wojtyniak, J. Sunyer, L. Bacharova, J. Schouten, A. Ponka and H. Anderson (1997) Short Term Effects of Ambient Sulphur Dioxide and Particulate Matter on Mortality in 12 European Cities: Results from Time Series Data from the APHEA Project. *British Medical Journal* Vol. 314 pp1658-1663.

McMichael, A., H. Anderson, B. Brunekreef and A. Cohen (1998) Inappropriate Use of Daily Mortality Analyses to Estimate Longer-Term Mortality Effects of Air Pollution. *International Journal of Epidemiology* Vol. 27 pp450-453.

Maddala, G. (1977) *Econometrics* McGraw-Hill: Singapore.

Ostro, B., J. Sanchez, C. Aranda and G. Eskeland (1996) Air Pollution and Mortality: Results from a Study of Santiago, Chile. *Journal of Exposure Analysis and Environmental Epidemiology* Vol. 6, pp97-114.

Pope, C., M. Thun, M. Namboodiri, D. Dockery, J. Evans, F. Speizer and C. Heath (1995) Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of US Adults. *American Journal of Respiratory and Critical Care Medicine* Vol. 151 pp669-674.

Saldiva, P., C. Pope, J. Schwarz, D. Dockery, A. Lichtenfels, J. Salge, I. Barone, G. Bohm (1995) Air Pollution and Mortality in Elderly People: A Time Series Study in Sao Paulo, Brazil. *Archives of Environmental Health* Vol. 50 No. 2 pp159-163.

Schwarz, J. (1994) Total Suspended Particulate Matter and Daily Mortality in Cincinnati, Ohio. *Environmental Health Perspectives* Vol. 102 No. 2 pp186-189.

Schwarz, J. (2000) The Distributed Lag between Air Pollution and Daily Deaths. *Epidemiology* Vol. 11 No. 3 pp320-326.

Schwarz, J. and D. Dockery (1992) Increased Mortality in Philadelphia Associated with Daily Air Pollution Concentration. *American Review of Respiratory Disease* Vol. 145 pp600-604.

Schwarz, J., C. Spix, G. Touloumi, L. Bacharova, T. Barumamdzadeh, A. Le Tertre, T. Piekarski, A. Ponce De Leon, A. Ponka, G. Rossi, M. Saez, and J. Schouten (1996) Methodological Issues in Studies of Air Pollution and Daily Counts of Deaths or Hospital Admissions. *Journal of Epidemiology and Community Health* Vol. 50 (Supp.) pp3-11.

Sunyer, J., J. Castellsague, M. Saez, A. Tobias and J. Anto (1996) Air Pollution and Mortality in Barcelona. *Journal of Epidemiology and Community Health* Vol. 50 (Supp.) pp76-80.

Touloumi, G., E. Samoli and K. Koutsouyanni (1996) Daily Mortality and Winter Type Air Pollution in Athens, Greece – a Time Series Analysis within the APHEA Project. *Journal of Epidemiology and Community Health* Vol. 50 (Supp.) pp47-51.

Zeger, S., F. Dominici and J. Samet (1999) Harvesting-Resistant Estimates of Air Pollution Effects on Mortality. *Epidemiology* Vol. 10, No. 2, pp171-176.

Appendix 1: Calculating the Implied Lag Coefficients in the Rational Lag Model

Given the equivalence between the parameters of the distributed lag and the parameters of the rational lag function one can rewrite the equation shown in footnote 6 in the following way:

$$\left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 +^3 \ldots\right)\left(1 + \omega_1 L + \omega_2 L^2 + \omega_3 L^3\right) = \left(\gamma_0 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3\right)$$

By comparing coefficients of the various powers of L one obtains the following:

$$\beta_0 = \gamma_0$$

$$\beta_1 = \gamma_1 - \beta_0\omega_1$$

$$\beta_2 = \gamma_2 - \beta_0\omega_2 - \beta_1\omega_1$$

$$\beta_3 = \gamma_3 - \beta_0\omega_3 - \beta_1\omega_2 - \beta_2\omega_1$$

$$\beta_4 = -\beta_1\omega_3 - \beta_2\omega_2 - \beta_3\omega_1$$

Notice that after the fourth term the series follows the simple recursion:

$$\beta_k = -\beta_{k-3}\omega_3 - \beta_{k-2}\omega_2 - \beta_{k-1}\omega_1$$

These equations may now be solved recursively for each β.