

NBER WORKING PAPER SERIES

AIR POLLUTION AND LOST WORK

Jerry A. Hausman

Bart D. Ostro

David A. Wise

Working Paper No. 1263

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1984

The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Air Pollution and Lost Work

ABSTRACT

A Poisson specification of the relationship between atmospheric pollution and lost work days is estimated. An important feature of the procedure is control for city-specific effects. A major source of ambiguity in interpreting the results of observational data on pollution versus health status or death rates is that pollution in a city may be correlated with other characteristics of that city that affect these outcomes but are not controlled for in the analysis. Or, individual attributes of residents may be correlated with pollution levels but not accounted for in the analysis. Our results suggest a statistically significant and quantitatively important effect of total suspended particulates on work days lost. A standard deviation increase in total suspended particulates is associated with approximately a ten percent increase in work days lost. As a concomitant of our analysis, we also find a substantial relationship between smoking by others in the individual's household and work days lost by non-smokers.

Jerry A. Hausman
Department of Economics
Massachusetts Institute
of Technology
Cambridge, MA 02139

Bart D. Ostro
Environmental Protection
Agency
401 M Street S.W.
Washington, D.C. 20460

David A. Wise
John F. Kennedy School of
Government
Harvard University
Cambridge, MA 02138

November 1983

AIR POLLUTION AND LOST WORK

by

Jerry A. Hausman, MIT

Bart D. Ostro, EPA

David A. Wise, Harvard¹

Analysis of the effects of air pollution on health must rely in large part on observational survey data. While there is growing evidence on the relationship between air pollution and health, the observational studies performed leave considerable doubt about the causal effects of pollution. The lingering and important question is whether attributes of individuals or their surroundings other than pollution, that could have caused the observed relationship, have been adequately controlled. What is the effect of left out variable bias? Two advantages of our study we hope will reduce this range of uncertainty. First, we use microdata that contain information on more individual attributes than were available to other investigators, helping to reduce excluded variable bias. Second, we use statistical techniques that allow us to control for unmeasured site effects. Though the potential effects of air pollution are manifold, we analyze work days lost to

¹We are grateful to Reed Shuldiner for expert research assistance and to very helpful comments from Jeff Harris and Will Manning.

focus attention on a direct economic cost of pollution. We make no attempt to compare the total costs and benefits of pollution reduction, however.

Several previous studies have related pollution to mortality rates. Possibly the most widely cited work is by Lave and Seskin [1971], who base their analysis on aggregate cross-section data for standard metropolitan statistical areas. The major reservation about their work has been the potential effect of unmeasured city characteristics, a motivation for several subsequent studies that addressed particular issues. Mendelsohn and Orcutt [1979] paid particular attention to migration, Crocker et al., [1979] and Gregor [1977] to medical care, and Lipfert [1980] to other socioeconomic characteristics. Nonetheless, the effect of unmeasured city characteristics has remained a major source of uncertainty. The relationship of pollution to morbidity, which is closer to our work days lost measure, was addressed by Crocker et al. [1979] using Michigan Survey Research Center interview data, by Graves and Krumm [1981] who analyzed Cook County, Illinois emergency room data, and by Seskin [1979] who studied visits to health clinics in Washington, D.C. Other researchers addressing the morbidity effects of pollution have monitored individuals over relatively long periods of time. Lunn et al. [1967] found a significant relationship between respiratory illness and air pollution among children in Sheffield, England and Ferris [1973] and

Bouhuys et al. [1978] found pollution related to respiratory disease symptoms among adults. The closest to ours is that of Ostro [1983], who analyzed the relationship between air pollution and work days lost using the annual data described below, but using a more standard statistical procedure that did not allow control for city effects.

In short, while the evidence for a relationship between pollution and health is substantial, judgments about the causal effect are clouded by uncertainty about the effect of unmeasured city-specific characteristics and the inability to control for differences among individuals in the same area. For example, more polluted areas may have more blue collar workers whose occupations are unusually dangerous, or polluted areas may also have more firms that offer employee fringe benefits like sick leave that encourage sick days. Studies that follow the same individual over time have been restricted to possibly unrepresentative samples. While the evidence from epidemiologic data must by its very nature be circumstantial, we believe that by statistical correction for city-specific effects and by better data on individual attributes, we are able to reduce the range of uncertainty.

Our analysis addresses the relationship between work days lost and atmospheric air pollution. Estimates are based on a series of national surveys of individuals conducted weekly during 1976, a time-series of cross-section surveys. In particular, survey respondents were asked how many days of work they lost due to illness or injury

during the prior two-week period. Because the number of work days lost is most often zero and always a small integer count, statistical estimation is based on the Poisson model. To allow for individual differences, the Poisson (centrality) parameter is specified as a function of individual attributes and environmental characteristics. Because we have a number of observations for the same city at several different times in the year, we are able to control for unobserved city effects, based on procedures set forth in Griliches, Hall, and Hausman [1981].

The statistical model is described in Section I. We begin with a simple Poisson "regression" model and then extend it to account for unobserved city effects. Parameter estimates are presented in Section II. Simple Poisson estimates are presented first, followed by results based on the fixed effect model. All model specifications indicate an important effect of air pollution on work days lost. As a concomitant of our analysis, and consistent with this result, we also find that cigarette smoking by others in the household also bears a strong relationship to work days lost by non-smokers in the household. Concluding discussion is in Section III.

I. The Statistical Model

The data to be described more fully below pertain to individuals j in city i in time period t . We analyze the number of work days lost, n_{itj} , in city i in period t by individual j . The period is the two weeks prior to the survey. For ease of exposition, we shall suppress the index j . Subsequent reference to summation over values of t should be taken to imply summation over j as well. A simple Poisson specification is described first and then the fixed effect version of it.

A. A Simple Poisson Model

The likelihood that n_{it} days will be lost in city i in period t (by individual j) is described by

$$(1) \quad f(n_{it}) = e^{-\lambda_{it}} \cdot \lambda_{it}^{n_{it}} / n_{it}! .$$

The expected value of n_{it} , λ_{it} , is allowed to vary among individuals according to the specification

$$(2) \quad \lambda_{it} = e^{X_{it}\beta} ,$$

with β a vector of parameters and X a vector of individual attributes and city characteristics. In this specification, the Poisson parameter λ is a deterministic function of X . The randomness in the outcome n is captured only through the Poisson specification and does not depend on

unobserved determinants of λ . The regression nature of the model is reflected in the Poisson property that $E(n_{it}) = \lambda_{it} = e^{X_{it}\beta}$. In this case the variance $V(n_{it}) = \lambda_{it}$ as well. Tests of this aspect of the model are presented in the next section.

From equation (1), the log-likelihood function is

$$(3) \quad L = \sum_i \sum_t (n_{it} X_{it}\beta - e^{X_{it}\beta} - \ln n_{it}!) .$$

The derivative with respect to β is given by

$$(4) \quad \frac{\partial L}{\partial \beta} = \sum_i \sum_t [X_{it}' (n_{it} - e^{X_{it}\beta})] .$$

In general, we obtain estimates of β using a maximum likelihood procedure. But equation (4) suggests that equivalent estimates could be obtained by an iterative non-linear weighted least squares procedure as well, with the residual given by $n_{it} - e^{X_{it}\beta}$. This observation motivates some of the specification analysis in the next section.

B. A Fixed Effect Specification

One of the major problems of drawing conclusions based on city pollution variables is that unmeasured city characteristics that are correlated with pollution may also affect health, or persons that live in a city may have common unmeasured attributes that determine their health status, or, more specifically for our purposes, the likelihood that they will miss work days. A variance components specification is a common way to address this problem in regression analysis. The fixed

effect version of this specification is often implemented by considering the deviation of individual observations from their group (e.g. city) means. The idea is to explain differences within the group, having abstracted from differences among the groups. We use an analogous procedure here, conditioning on the total number of work days lost in each city.² We thus analyze individual outcomes conditional on the total for the group, analogous to the mean for the group in the regression variance components model.

The parameter λ is now parameterized as

$$(5) \quad \ln \lambda_{it} = X_{it}\beta + u_i ,$$

where u_i is a city-specific effect with mean zero across all cities. The variance component, u_i , may be thought of as an unknown parameter, that could be correlated with X . To develop the likelihood of the individual outcomes for a city conditional on their sum, recall that if the n_{it} are independently distributed Poisson variables, then $\sum_t n_{it}$ is also distributed Poisson with parameter $\lambda_i = \sum_t \lambda_{it}$. The conditional likelihood of the n_{it} in city i is given by

²See Andersen [1970, 1972].

$$\begin{aligned}
 (6) \quad L_i &= f(n_{i1}, n_{i2}, \dots, n_{iT} \mid \sum_t n_{it}) = \Pr(n_{i1}, n_{i2}, \dots, n_{iT}) + \Pr(\sum_t n_{it}) \\
 &= \frac{e^{-\sum \lambda_{it}} \prod \lambda_{it}^{n_{it}}}{\prod n_{it}!} = \frac{(\sum n_{it})!}{\prod n_{it}!} \cdot \prod \left(\frac{\lambda_{it}}{\sum \lambda_{it}} \right)^{n_{it}}, \\
 &\quad (\sum n_{it})!
 \end{aligned}$$

with

$$(7) \quad \frac{\lambda_{it}}{\sum \lambda_{it}} = \frac{e^{X_{it}\beta} e^{u_i}}{\sum e^{X_{it}\beta} e^{u_i}} = \frac{e^{X_{it}\beta}}{\sum e^{X_{it}\beta}} = P_{it} .$$

The u_i terms cancel out.³ Therefore possible correlation between the X_{it} and the unobserved city effects, which is the major shortcoming of previous work, is eliminated.⁴

It can also be seen that this is a multinomial distribution with parameters P_{it} , since their sum is 1. And our specification of the λ_{it} yields a logit functional form for the probabilities P_{it} .⁵ Given

³Because of the incidental parameter problem, unconditional maximum likelihood estimates of individual u_i parameters need not be consistent. See, for example, Neymann and Scott [1948], Andersen [1973], and Haberman [1977].

⁴See Mundlak [1978], Hausman [1978], Hausman-Taylor [1981], and Chamberlain [1982] for discussion of similar situations.

⁵See McFadden [1974].

that there is a work day lost in city i , P_{it} can be interpreted as the probability that it is incurred by individual j in time period t . Then L_i is the likelihood of the observed proportions of work days lost in each time period, given the total number of days lost in city i . It explains differences among people and time periods in the same city. The observations for the cities are pooled in estimation of course through the assumption that the functional relationship between attributes X and work days lost is the same in each city; the vector β is the same.

In this case, the log-likelihood function is of the form

$$\begin{aligned} L &= \sum_i \ln L_i \\ (8) \quad &= \sum_i \left\{ \sum_t \Gamma(n_{it} + 1) - \sum_t n_{it} \ln \left(\sum_s e^{(X_{is} - X_{it})\beta} \right) \right\} \end{aligned}$$

Again, it is maximized with respect to the vector β .

II. Empirical Results

We begin with a discussion of the data and then present estimates based on the simple Poisson model. The first estimates are based on annual average pollution levels. Presumably these averages are the best indication of typical pollution in a city. Estimates based on average pollution levels over two-week periods are also presented. The two-week readings allow comparison of the relationships between work days lost and alternative past levels of pollution, and as we explain below provide a base for comparison with the fixed effect estimates.⁶ Before presenting the fixed effect estimates, however, we test the sensitivity of our results to the Poisson restriction that given X the variance of n is equal to its expected value. To do this we present non-linear least squares estimates that are consistent, although not efficient, under rather general assumptions on the variance structure. Having concluded that our results are not unduly affected by the Poisson variance restriction, we next present the fixed effect results. Because these estimates address a major concern inherent in all analysis based on observational data of this sort, we believe them to be the most reliable, with the limitation however, that only two-week pollution levels can be used. Thus we also present fixed effect estimates using several lagged values of pollution, under the assumption that the sum of the coefficients on past pollution levels provides a reasonable

⁶Whether average or more recent pollution is more important is likely to be specific to particular effects of pollution and is likely to vary among individuals depending on health status. We have not attempted to pursue the details of potential mechanisms.

indication of the effect of average annual pollution. In addition to the effect of pollution on work days lost, the estimated relationship between work days lost and other individual attributes is also of interest. We discuss these relationships within the context of the fixed effect specification.

Estimation is based on individual data collected through the 1976 Health Interview Survey (HIS) conducted by the National Center for Health Statistics. These data were merged with 1976 Environmental Protection Agency (EPA) data on ambient pollution and, with National Oceanic and Atmospheric Administration (NOAA) data on weather conditions, and with Census Bureau data on other urban characteristics.

The HIS is a stratified random survey of 50,000 households comprising about 120,000 individuals. Information is obtained on economic and demographic attributes of respondents, acute and chronic illnesses (identified by diagnosis), days of medical disability, work days missed because of illness, cigarette consumption, family income, occupation, education, and other individual attributes. All respondents living in 85 medium sized cities were selected for this analysis. The cities, most with populations between 100 and 600 thousand, provide geographic distribution and pollution levels that span the entire range among U.S. cities, although they do not represent a random sample of pollution levels. Our analysis also pertains to men aged 18 to 65 who are employed, and who do not smoke. About 5500 observations are included in the analysis.

The EPA pollution data (SAROAD) provide measures on ambient levels of several pollutants in these cities, including total suspended particulates (TSP), sulfur dioxide (SO₂), and sulfates (SO₄). We use the particulates measure. Pollutant measurements are based on recordings from 24-hour hi-vol samplers intended to measure outdoor pollution in residential areas. Many cities have only one "population-oriented" monitor. Others have more. Where there is more than one, an average of monitor observations is used. In addition, central city residents were matched with central city monitors and those in suburban locations were matched with monitors in those areas.

In addition to pollution, we suppose that work days lost may be affected by personal attributes such as age, health status, family income, race, occupation and by environmental characteristics such as temperature. The variables used are defined as follows:

Pollution Annual: Average annual ambient TSP level over the year in the city of residence, in ug per cubic meter.

Pollution Lag n: Average ambient TSP level over the two-week period ending n weeks before the time of the interview (e.g. Pollution Lag 4 pertains to the 5th and 6th weeks before the interview), in ug per cubic meter.

Age: In years.

Family Income: Annual income in 1000's.

Married: One if married (and living with a spouse), zero otherwise.

Spouse Works: One if man's wife works, zero otherwise.

Blue Collar Job: One if the person has a blue collar job, zero otherwise.

Industry Sick Leave: The proportion of workers nationwide in the industry in which the individual works who are covered by sick leave provisions.

Non-White: One if non-white, zero otherwise.

Chronic Conditions: One if the person has a chronic health condition, zero otherwise.

Household Cigarettes: Total number of cigarettes smoked per day by others living in the household in which the individual lives.

Population Density: Population (1977) per square mile in the individual's city of residence, in thousands.

Average Temperature: Average annual temperature in the city of residence.

Minimum Temperature: Minimum temperature in the city of residence during the two-week period prior to the interview.

January-March: One if the two-week period is in January, February, or March, zero otherwise.

April-June: Analogous to above.

July-September: Analogous to above.

The means and standard deviations of the variables are shown in Table 1.⁷ Recall that our sample includes non-smokers only, so that household cigarettes are those smoked by other people in the house in which the individual lives.

B. Poisson Parameter Estimates

Estimates based on the simple Poisson model are shown in Table 2. Average annual pollution is used in the first two specifications. The second includes seasonal variables, while the first does not. As shall be explained below, we cannot estimate (two-week) period-specific effects and still estimate a pollution effect for that period, although we shall correct for city-specific effects. The seasonal variables can be thought of as more aggregated period-specific effects. Comparison of specification 1 with 2 reveals that including the seasonal variables has little effect on the other estimates. In particular the estimated pollution coefficient does not change appreciably when the seasonal variables are added.

The estimated coefficient on annual pollution based in specification 2 is 0.0071 and is measured very precisely, with a

⁷The sample sizes vary somewhat with alternative specifications. The means reported pertain to the specifications in Table 2.

Table 1. Summary Statistics on Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Work Days Lost	0.24	1.30	0.00	14.0
Age	38.68	12.99	18.00	65.0
Education	13.90	2.28	0.00	17.0
Family Income	18.71	9.43	0.75	35.0
Married	0.78	0.42	0.00	1.0
Spouse Works	0.70	0.46	0.00	1.0
Blue Collar Job	0.54	0.50	0.00	1.0
Industry Sick Leave	0.51	0.22	0.00	0.98
Non-White	1.09	0.29	1.00	2.0
Chronic Cond.	0.10	0.31	0.00	1.0
Household Cigarettes	5.34	11.59	0.00	100.0
Population Density	6.61	3.42	0.73	14.1
Ave Temp In City	52.30	5.30	38.00	68.0
Min Temperature	33.54	17.81	-22.00	69.0
January-March	0.18	0.38	0.00	1.0
April-June	0.27	0.45	0.00	1.0
July-September	0.28	0.45	0.00	1.0
Pollution Annual	68.36	16.54	39.54	133.09
Pollution Lag 2*	72.09	30.56	19.00	292.0
Pollution Lag 4	70.62	28.98	16.58	203.5
Pollution Lag 6*	70.53	27.77	13.00	203.4

* from 4048 sample (Col. 4, Table 2); the rest are from 4234 sample (Col. 3, Table 2). The Lag 4 pollution mean for the 4048 sample is 71.28.

Table 2. Simple Poisson Model Parameter Estimates

Variable	Parameter Estimate (Standard Error)			
	(1)	(2)	(3)	(4)
Constant	-3.346 (0.192)	-3.195 (0.198)	-2.304 (0.141)	-2.512 (0.149)
Age	0.0183 (0.000987)	0.0182 (0.0010)	0.0188 (0.0010)	0.0195 (0.0010)
Education	-0.00573 (0.00578)	-0.0066 (0.0059)	-0.00867 (0.00601)	-0.0111 (0.0064)
Family Income	0.00334 (0.00147)	0.00253 (0.00151)	0.00329 (0.00152)	0.0027 (0.0016)
Married	-0.225 (0.031)	-0.208 (0.031)	-0.224 (0.031)	-0.198 (0.033)
Spouse Works	-0.282 (0.029)	-0.278 (0.030)	-0.281 (0.030)	-0.310 (0.031)
Blue Collar Job	-0.179 (0.028)	-0.168 (0.028)	-0.163 (0.028)	-0.116 (0.030)
Industry Sick Leave	0.593 (0.063)	0.596 (0.064)	0.580 (0.063)	0.589 (0.064)
Non-White	-0.336 (0.055)	-0.329 (0.055)	-0.307 (0.055)	-0.229 (0.056)
Chronic Cond.	1.0915 (0.0292)	1.0831 (0.0296)	1.068 (0.029)	1.016 (0.031)
Household Cigarettes	0.0114 (0.0010)	0.0113 (0.0011)	0.0117 (0.00105)	0.0136 (0.0011)
Population Density	0.0234 (0.0045)	0.0234 (0.0046)	0.0209 (0.0044)	0.0134 (0.0046)
Ave Temp In City	0.0148 (0.0026)	0.0138 (0.0026)		
Min Temperature			-0.00275 (0.00107)	-0.00124 (0.00116)
January-March		0.0699 (0.0399)	0.0961 (0.0399)	0.110 (0.045)
April-June		-0.245 (0.037)	-0.184 (0.043)	-0.232 (0.045)
July-September		-0.195 (0.035)	-0.129 (0.047)	-0.180 (0.050)
Pollution Annual	0.00706 (0.00082)	0.00708 (0.00082)		
Pollution Lag 2				0.00232 (0.00071)
Pollution Lag 4			0.00523 (0.00046)	0.00353 (0.00063)
Pollution Lag 6				0.00109 (0.00054)
Number of observations	4234	4234	4234	4048
Log-likelihood	-3040.35	-3032.57	-3029.12	-2909.59

standard error of about 0.0008. Since $\ln \lambda = X\beta$, this specification implies that a unit (ug/m^3) increase in annual average TSP would increase the expected number of work days lost by 0.7 percent. The mean of the annual pollution variable is 68.36, with a standard deviation of 16.1; its minimum is 39.5 and its maximum 133.1. Thus a two-standard-deviation reduction in pollution would reduce expected days lost by approximately 23 percent, almost one and a half days per year evaluated at the mean of days lost. The mean of work days lost per two-week interval is 0.24, or 6.24 per year. Thus according to these estimates, ambient TSP bears a quantitatively important relation to work days lost.

If outdoor suspended particulate pollution is associated with morbidity, one might expect a similar association with indoor air pollution to which cigarette smoke may be an important contributor.⁸ The estimated coefficient of .011 on cigarettes smoked by others in the household implies that an increase of 20 cigarettes per day, a pack, is associated with a 22 percent increase in work days lost by non-smokers. An increase of two standard deviations is associated with a 26 percent increase in days lost by non-smokers. Of course, not all cigarettes smoked by others in the household are smoked in the house when the non-smoker is present.

⁸See Repace and Lowry [1981].

The coefficients on the other variables are also of interest and we shall discuss them with reference to the estimates below that correct for city-specific effects. Although the orders of magnitude of the estimates are roughly the same as these, some of the estimates change noticeably. Since the later estimates should be preferred, we shall emphasize them.⁹

The remaining specifications in Table 2 use average pollution levels over two-week periods, with alternative time lags. These results provide a base with which to compare the fixed effects results presented below. With this comparison in mind, the population density and average temperature variables have been deleted and the two-week minimum temperature variable has been added. City variables that do not change over time cancelled out of the analysis along with the unmeasured city specific effect and thus it is impossible to estimate parameters associated with them. The alternative time lags are motivated by the possibility that pollution in the past may contribute to illness in the future. We know of no theoretical rationale for a particular lag structure, however. Thus we experimented with several. Two are reported.

While the coefficient on any individual two-week value is considerably lower than the estimated effect of annual pollution, the sum of the lagged coefficients in specification 4 is close to the annual

⁹Work days lost in a two-week period is bounded by 10 (or 14) days. Maximum likelihood estimates accounting for the truncation at this bound yield estimates not appreciably different from those in Table 2.

effect in specification 2 (.0069 versus .0071). The estimated coefficient as the single 4-week lag (entered by itself in column 3) is .0052. The estimated coefficient on a 2-week lag pollution variable in an otherwise like specification was .0043.

Although we believe that unmeasured city effects could be a potential major source of bias in these results, the Poisson model itself implies functional form restrictions that also could affect the results. In particular, the variance of work days lost, given X , is presumed to be equal to the expected value of days lost. Therefore, before we proceed we shall test the sensitivity of our results to this constraint, using specification 2 in Table 2 as a base for comparison.

C. Sensitivity to the Poisson Specification

While the Poisson model is in some ways ideally suited to these data, it also imposes restrictions on the estimates. Given X , the variance of the outcome is assumed to be equal to its mean. To test the sensitivity of our estimated pollution effect to this restriction, we present estimates based on alternative specifications that are not restricted by this assumption.

To motivate the approach, recall that equation (4) implies that estimates of the Poisson model could be obtained by maximum likelihood, as we have done, or by a non-linear weighted least squares procedure. In particular, suppose that the basic model is specified as

$$(9) \quad n_{it} = e^{X_{it}\beta} + \varepsilon_{it} \quad ,$$

where $\text{Var } \varepsilon_{it} = e^{X_{it}\beta}$. Then weighted non-linear least squares with the weight $w_{it} = \frac{1}{\sqrt{\exp(X_{it}\beta)}}$ is equivalent to our maximum likelihood estimates. It gives relatively more weight in estimation to observations with small expected values. We relax the variance assumption in two directions, relying on non-linear least squares (NLLS) estimation.

First, assume that the variance does not grow with the mean but is constant so that the weight $w_{it} = 1$. This NLLS specification gives equal weight to all observations, while the Poisson assumption gives greater weight to observations with small expected values, thus on average giving relatively greater weight to observations with zero work days lost. Non-linear least squares estimates under this assumption are shown in column 3 of Table 3.¹⁰ The estimates remain consistent under the null hypothesis that the Poisson specification holds, but they are no longer asymptotically efficient. The nonlinear least squares estimates can be compared to the Poisson estimates in column 1 of Table 3 which are estimated on an enlarged data set of 5392 observations. The NLLS estimated coefficient on average annual air pollution is 0.0096, compared to 0.0080 based on the Poisson specification. The least squares estimate however, is much less efficient than the Poisson estimate, and thus the

¹⁰In calculating the test statistic, the asymptotic standard errors for the $w_{it} = 1$ specification are calculated allowing for the presence of heteroscedasticity. The linear formulae of White [1980] are adapted for the nonlinear case.

Table 3. Non-Linear Least Squares Parameter Estimates

Variable	Parameter Estimate (Standard Error)		
	$w_{1t} = \sqrt{e^{X_{1t} + \beta}}$	$w_{1t} = \sqrt{e^{X_{1t} + \beta} + \sigma^2 e^{2X_{1t} + \beta}}$	$w_{1t} = 1$
Constant	-3.283 (.1729)	-3.887 (1.029)	-5.550 (.558)
Age	.0175 (.0009)	.0188 (.0061)	.0449 (.0095)
Education	.0004 (.0053)	-.0028 (.0289)	-.0018 (.0485)
Family Income	-.0030 (.0014)	-.0013 (.0090)	.0040 (.0147)
Married	-.2357 (.0290)	-.2117 (.1642)	-.1210 (.2819)
Spouse Works	-.2482 (.0274)	-.2415 (.1580)	-.0559 (.3064)
Blue Collar Job	-.1454 (.0265)	-.1353 (.1589)	.0129 (.2492)
Industry Sick Leave	.5412 (.0591)	.5442 (.3698)	.7625 (.5818)
Non-White	-.1548 (.0463)	-.3093 (.1606)	-1.169 (.539)
Chronic Conditions	1.035 (.0273)	1.102 (.1968)	1.434 (.2178)
Household Cigarettes	.0108 (.0009)	.0136 (.0069)	.0252 (.0053)
Population Density	.0093 (.0034)	.0105 (.0196)	.0289 (.0270)
Average Temperature	.0140 (.0024)	.0161 (.0145)	.0362 (.0233)
Pollution Annual	.0080 (.0007)	.0079 (.0042)	.0096 (.0077)
January-March	.0179 (.0343)	-.1037 (.2182)	-.7442 (.2858)
April-June	-0.2798 (.0348)	-.4203 (.2496)	-.7643 (.3811)
July-September	-.2621 (.0340)	-.3535 (.2154)	-1.226 (.3934)
σ^2	0.	.6487	----
Number of observations	5392	5392	5392

standard error is much larger. Given that the Poisson specification is correct, though, the two estimates have the same probability limit. A large sample test, based on the difference of the coefficient estimates divided by the standard deviation of the estimated difference, yields a test statistic of 0.53, which is far below standard significance levels.¹¹ However, some of the other coefficients--such as the effect of cigarettes--have changed markedly, although again the NLLS estimates are much less precise than those based on the Poisson assumption.

Thus, as another test of our specification we compare jointly all of the coefficient estimates in the Poisson and NLLS models. Under the null hypothesis the statistic

$$(10) \quad m = (\hat{\beta}_{\text{NLLS}} - \hat{\beta}_{\text{P}})' (V(\hat{\beta}_{\text{NLLS}}) - V(\hat{\beta}_{\text{P}}))^{-1} (\hat{\beta}_{\text{NLLS}} - \hat{\beta}_{\text{P}})$$

is distributed as central χ^2 with 17 degrees of freedom.¹² The test statistic is based on the idea that under the correct specification the coefficient estimates should not change much with different weights. An asymptotically equivalent statistic to m can be calculated by treating the Poisson model as a weighted NLLS specification and by then including the additional terms which arise from the NLLS specification. A joint test of the significance of the additional terms is then made. The relevant statistic is estimated to be 18.4, only slightly more than

¹¹See Hausman [1978].

¹²See Hausman [1978] and White [1981].

its expected value of 17.0 under the null hypothesis.¹³ Thus based on this test we find no evidence that our Poisson model suffers from serious misspecification.

The Poisson specification also assumes that the expected value of n_{it} is determined, given X ; there are no unmeasured determinants of λ_{it} . It is often the case that empirically observed variances are greater than this Poisson assumption would imply. This is sometimes referred to as the over-dispersion problem. We relax the Poisson assumptions and allow for greater dispersion by assuming that $\lambda_{it} = e^{X_{it}\beta} + \eta_{it} = e^{X_{it}\beta} e^{\eta_{it}}$. As long as $X_{it}\beta$ includes a constant term, we can normalize by setting $E(e^{\eta_{it}}) = 1$. Let $\text{Var}(e^{\eta_{it}}) = \sigma^2$. Within the framework of the regression specification we now have

$$(11) \quad n_{it} = e^{X_{it}\beta} e^{\eta_{it}} + \epsilon_{it} \quad , \quad \text{with}$$

$$\text{Var}(n_{it} | X_{it}) = \sigma^2 e^{2X_{it}\beta} + e^{X_{it}\beta} = e^{X_{it}\beta} (1 + \sigma^2 e^{X_{it}\beta}) \quad ,$$

maintaining the Poisson variance assumption for ϵ but adding variance due to unmeasured determinants of λ . Notice that this specification implies that the variance of n_{it} increases with its expected value, but at a faster rate than the basic Poisson variance. In this case, the

¹³The analogous F test with 17 and 5375 degrees of freedom is estimated to be 1.08.

nonlinear least squares regression weight is $w_{it} = \sqrt{e^{X_{it}\beta} + \sigma^2 e^{2X_{it}\beta}}$.

Thus our two alternative specifications "bracket" the Poisson assumption. The first gives relatively less weight than the Poisson model to observation with low expected value, while the second gives more.

If η_{it} were assumed to have a gamma distribution, then integrating over the Poisson distribution conditional on η , yields a negative binomial distribution from which a likelihood function may be formed.¹⁴ Instead of following this method, however, we adopt the pseudo maximum likelihood procedure suggested by Gourieroux, Monfort, and Trognon [1981], that yields consistent estimates of β under very general assumptions on the distribution of η . Some details of this weighted non-linear least squares procedure as well as the equal variance estimation procedure are provided in the appendix.

Estimates based on this specification are shown in column 2 of Table 3. The estimated coefficient on annual pollution in this case is 0.0079, as compared with the Poisson estimate of 0.0080. Thus the pollution estimates seem to be reasonably stable with respect to the variance assumption, although again all parameters are measured with much less precision than when the Poisson distribution is assumed. When we compare the difference of the estimated coefficients for pollution divided

¹⁴See, for example, Gilbert [1979] or Griliches, Hall, and Hausman [1981].

by the standard deviation of their difference, the statistic is .025 which is not significant as expected; the coefficient estimates are almost identical. In general the pseudo maximum likelihood estimates are quite close to the Poisson estimates, although again the Poisson estimates are considerably more precise. A test of coefficient difference, like that of equation (10), yields in this case a test statistic of 21.8 which is near its expected value (17.0) under the null hypothesis and far below conventional significance levels.¹⁵ Thus in comparing either of the NLLS specifications to the Poisson model we find that neither a test based on the estimated pollution coefficients alone nor a test based on all of the coefficient estimates leads to a rejection of the Poisson specification. We conclude that the Poisson assumption is not a serious misspecification and that the variance restriction is not unduly affecting our results. We now turn to the fixed effects estimates that address an issue that we believe is potentially much more important.

D. Fixed Effect Poisson Parameter Estimates

Estimates based on the fixed effect specification are shown in Table 4. The estimates in column 3 may be compared with those in column

¹⁵The analogous F statistic with 17 and 5375 degrees is 1.29.

Table 4: Fixed Effects Poisson Model Parameter Estimates

Variable	Parameter Estimate (Standard Error)				
	(1)	(2)	(3)	(4)	(5)
Age	0.0191 (0.0010)	0.0190 (0.0010)	0.0186 (0.0013)	0.0181 (0.0010)	0.0193 (0.0014)
Education	-0.0144 (0.0057)	-0.0124 (0.0056)	-0.0177 (0.0062)	-0.0171 (0.0063)	-0.0208 (0.0072)
Family Income	0.00337 (0.00146)	0.0034 (0.0014)	0.00371 (0.00153)	0.00340 (0.00152)	0.00436 (0.00173)
Married	-0.235 (0.037)	-0.250 (0.035)	-0.229 (0.041)	-0.228 (0.040)	-0.206 (0.042)
Spouse Works	-0.291 (0.0306)	-0.286 (0.029)	-0.280 (0.034)	-0.271 (0.032)	-0.291 (0.037)
Blue Collar Job	-0.184 (0.042)	-0.198 (0.041)	-0.181 (0.048)	-0.186 (0.047)	-0.135 (0.058)
Industry Sick Leave	0.584 (0.095)	0.603 (0.090)	0.583 (0.101)	0.594 (0.100)	0.607 (0.114)
Non-White	-0.428 (0.078)	-0.370 (0.071)	-0.414 (0.082)	-0.356 (0.073)	-0.336 (0.083)
Chronic Cond.	1.110 (0.035)	1.124 (0.034)	1.114 (0.035)	1.131 (0.035)	1.050 (0.041)
Household Cigarettes	0.0118 (0.0018)	0.0115 (0.0018)	0.0115 (0.0018)	0.0111 (0.0018)	0.0132 (0.0019)
Min Temperature	-0.0078 (0.0008)	--	-0.0027 (0.0020)	--	-0.00176 (0.00234)
January-March	--	--	0.139 (0.053)	0.0986 (0.0498)	0.124 (0.064)
April-June	--	--	-0.238 (0.072)	-0.304 (0.045)	-0.256 (0.825)
July-September	--	--	-0.137 (0.082)	-0.215 (0.050)	-0.172 (0.086)
Pollution Lag 2	--	--	--	--	0.00236 (0.00101)
Pollution Lag 4	0.0057 (0.0006)	--	0.00586 (0.00068)	--	0.00405 (0.00076)
Pollution Lag 6	--	--	--	--	0.00243 (0.00094)
Number of observations	4234	4234	4234	4234	4048
Log-Likelihood	-4447.51	-4462.78	-4442.90	-4451.62	-4170.36

3 of Table 2. The coefficient on pollution 4 weeks earlier is now .059, with a standard error of .007, as compared with .0052 in the simple Poisson model. The specification in column 5 uses three lagged values of pollution. All are statistically different from zero by standard criteria. Their sum is .0088, close to the annual pollution coefficient of .0071 in column 2 of Table 1.

Thus it seems reasonable to conclude that our estimates suggest that a standard deviation increase in pollution is associated with approximately a 10 percent increase in work days lost.

As in the simple Poisson model, the results here also suggest a substantial effect of cigarettes smoked by others in the household on work days lost. The coefficient in column 3 of Table 4 is 0.0115, with a standard error of (0.0018). This estimate indicates that 20 more cigarettes are associated with approximately a 23 percent increase in days missed. While the effect of smoking has not been the focus of our analysis here, this result seems to support other evidence that the effect of cigarette smoking is not limited to the smoker. Recent literature has suggested a significant health impact from the breathing of air polluted with tobacco smoke [Repace, 1981]. This air, called passive smoke, is believed to be chemically identical to mainstream smoke and generally more concentrated [Hoegg, 1972]. Past research has also indicated that passive smoke is associated with greater incidence of respiratory problems in children [Spiezer et al., 1980; Bonham and

Wilson, 1981; Tager et al., 1979] and greater risk of respiratory conditions and cancer among non-smoking spouses [Hirayama, 1981; and Kauffman, 1980].

The estimates in column 3 also indicate that married men miss 23 percent fewer days than the unmarried. Married men whose wives work miss 28 percent fewer days than those whose wives do not work. Blue collar workers miss 18 percent fewer days on average than those in other occupations. Not surprisingly, those with a chronic health condition miss about twice as many days as those who report no such condition. Non-whites miss about 40 percent fewer days than whites. Some of these differences, of course, may be due to differences in the types of job held by different groups. The larger the fraction of workers in the person's industry that are covered by sick leave provisions, the greater the number of work days lost. Older workers miss more days than younger ones.

E. Some Additional Results

In addition to the analysis reported in the paper, we also made preliminary comparisons using alternative samples and pollution definitions. For a sample of 763 smokers, we estimated the same specification as in Table 2, column 1. The estimated coefficient on annual pollution was 0.0125 (with a standard error of 0.0032), compared to the comparable estimate of 0.0071 for non-smokers. This result is

consistent with the hypothesis that pollution has a greater adverse affect on smokers than on non-smokers. Medical evidence suggests that smoking damages the body filters that would otherwise filter out some ambient particulate pollution. On the other hand, one might hypothesize that if an individual smokes himself, the additional effect of the cigarette smoke from others in the household may not be substantial. For smokers, we obtain an estimated coefficient on cigarettes smoked by others that is not significantly different from zero (-.004, with a standard error of .0054, versus .011, with a standard error of .0010 for non-smokers).

We also estimated the specification of Table 2, column 1 for persons who reported at least one respiratory condition. For this group, the coefficient on annual pollution was 0.010 (with a standard error of .0014), compared with 0.007 for the whole group.¹⁶ If work days lost is defined to include only days missed due to illness, and not injuries, the coefficient on annual pollution is .008 (with a standard error of .0008).

Because of the number and location of pollution monitors, our data for total particulates is much more reliable than the data for other pollutants or for components of the total. Nonetheless, we distinguished sulfates from the rest of TSP, and for central cities

¹⁶The coefficient on chronic condition also increased substantially for this group.

estimated separate coefficients for each. In our data, sulfates comprise about 13 percent of TSP. The coefficient on TSP in this specification remained large and significant (0.014 with a standard error of 0.0014), but the coefficient on sulfates was not significantly different from zero. Further work remains to be done using better data on individual pollutants.¹⁷

F. Conclusion

We have estimated a Poisson specification of the relationship between atmospheric pollution and lost work days. An important feature of our procedure is control for city specific effects. A major source of ambiguity in interpreting the results of observational data on pollution versus health status or death rates is that pollution in a city may be correlated with other characteristics of that city that affect these outcomes but are not controlled for in the analysis. Or, individual attributes of residents may be correlated with pollution levels but not accounted for in the analysis. Our data set allows us to control explicitly for many more individual attributes than were available to most previous researchers in this area. But in addition we have controlled for other unmeasured city specific effects that could be correlated with pollution levels. Thus we believe that our analysis reduces substantially the uncertainty attendant on standard analysis of the effects of pollution based on observational data.

¹⁷Monitor readings of course measure only with error the actual exposure experienced by any individual. Estimation for samples that we

Our results suggest a statistically significant and quantitatively important effect of total suspended particulates on work days lost. A standard deviation increase in TSP is associated with approximately a 10 percent increase in work days lost. According to these results if, in areas where pollution is very high, say two standard deviations above the mean, pollution were reduced to the mean, work days lost would be reduced by over 20 percent. As a concomittant of our analysis, we also found a substantial relationship between smoking by others in the individuals household and work days lost by non-smokers. Finally, the effect of atmospheric pollution on smokers appears to be greater than on non-smokers according to our results.

thought more closely matched individuals to monitors typically yielded higher estimated pollution effects, as errors in variable seasoning would suggest.

Appendix

The Poisson model can be considered as a weighted nonlinear least squares model with

$$(A1) \quad n_{it} = e^{X_{it}\beta} + \epsilon_{it}, \quad E(\epsilon_{it}) = 0, \quad \text{and} \quad V(\epsilon_{it}) = e^{X_{it}\beta}.$$

Both the functional form assumption and the variance assumption can be tested by considering different weights in estimation.¹⁸

Therefore instead of the Poisson weights $w_{it} = \sqrt{\exp(X_{it}\beta)}$ we first consider the weights $w_{it} = 1$. The nonlinear least squares estimates of equation (A1) continue to be consistent under the null hypothesis of the Poisson model. The alternative estimates can be used to form a specification test for the original Poisson specification. Note that under the Poisson null hypothesis the asymptotic covariance matrix for the non-linear least squares estimation is

$$(A2) \quad V(\hat{\beta}) = (G'G)^{-1} (G'WG) (G'G)^{-1},$$

where the elements of G are the derivatives of $e^{X_{it}\beta}$ with respect to the elements of β , and W is a diagonal matrix with the Poisson weights as the entries.

A more general specification which accounts for the "overdispersion" problem often present in Poisson models is

¹⁸White [1981] considers tests of specification for nonlinear models from this viewpoint.

$$(A3) \quad n_{it} = e^{X_{it}\beta} + \eta_{it} + \varepsilon_{it}, \quad E\varepsilon_{it} = 0, \quad E(e^{\eta_{it}}) = 0, \quad V(e^{\eta_{it}}) = \sigma^2,$$

$$\text{and} \quad V(n_{it} | X_{it}) = e^{X_{it}\beta} (1 + \sigma^2 e^{X_{it}\beta}).$$

The assumption that η_{it} is gamma distributed yields the well known negative binomial model. Recently, Gourieroux, Monfort, and Trognon [GMT(1981)] proposed an estimation method which is consistent for η a member of the exponential family. GMT prove strong consistency and asymptotic normality of their estimator. We apply their quasi-generalized pseudo ML(QML) estimator. Given the expression for the conditional variance, we have

$$(A4) \quad (n_{it} - e^{X_{it}\beta})^2 - e^{X_{it}\beta} = \sigma^2 e^{2X_{it}\beta} + u_{it} \quad Eu_{it} = 0.$$

Therefore a consistent estimator of σ^2 is the expression

$$(A5) \quad \hat{\sigma}^2 = \frac{\sum_{it} [(n_{it} - e^{X_{it}\hat{\beta}})^2 - e^{X_{it}\hat{\beta}}] e^{2X_{it}\hat{\beta}}}{\sum_{it} e^{4X_{it}\hat{\beta}}},$$

where the β is estimated from the nonlinear least squares model with $w_{it} = 1$. Then the quasi ML estimator is given by a nonlinear weighted least squares regression of equation (A1) with weights

$$(A6) \quad w_{it} = \sqrt{e^{X_{it}\hat{\beta}} + \hat{\sigma}^2 e^{2X_{it}\hat{\beta}}}.$$

Note that for $\hat{\sigma}^2 > 0$ the variance grows faster than the expected value which is a solution to the overdispersion problem. Also, the weights given in equation (A6) provide an alternative check on the specification of

equation (A1). To the extent that the estimated coefficients are similar across the three alternative weights schemes, Poisson, least squares, and QML, the possibility of functional form misspecification is decreased. We apply the Hausman (1978) specification test to evaluate the possibility of misspecification.

References

Andersen, E.B. 1970. "Asymptotic Properties of Conditional Maximum Likelihood Estimators," Journal of the Royal Statistical Society, B, 32.

Andersen, E.B. 1972. "The Numerical Solution of a Set of Conditional Estimation Equations," Journal of the Royal Statistical Society, B, 34.

Andersen, E.B. 1973. Conditional Inference and Models for Measuring. Copenhagen.

Bonham, G.S. and R.W. Wilson. 1981. "Children's Health in Families with Cigarette Smokers," American Journal of Public Health, 71, pp. 290-93.

Bouhuys, A., G.J. Beck and J.B. Schoenberg. 1978. "Do Present Levels of Air Pollution Outdoors Affect Respiratory Health," Nature, 276.

Chamberlain, G. 1982. "Panel Data," mimeo.

Crocker, T.D. et al. 1979. Methods Development for Assessing Air Pollution Control Benefits, U.S. Environmental Protection Agency.

Ferris, B.G., Jr., I.T.T. Higgins, M.W. Higgin and J.M. Peters. 1973. "Chronic Non-Specific Respiratory Disease in Berlin, New Hampshire, 1961-1967. A Follow-Up Study," Am. Rev. Resp. Dis., 107.

Gilbert, G.C. 1979. "Econometric Models for Discrete Economic Processes," given at the European Meetings of the Econometric Society, Athens.

Gourieroux, C., A. Monfort and A. Trognon. 1981. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," mimeo.

Graves, Philip E. and Ronald J. Krumm. 1981. Health and Air Quality, American Enterprise Institute for Public Policy Research.

Gregor, John J. 1977. Intra-Urban Mortality and Air Quality: An Economic Analysis of the Costs of Pollution Induced Mortality. U.S. Environmental Protection Agency, Corvallis, Oregon.

Griliches, Z., B. Hall and J. Hausman. 1981. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," National Bureau of Economic Research Technical Working Paper No. 17, forthcoming in Econometrica.

Haberman, S. 1977. "Maximum Likelihood Estimates in Experimental Response Models, Annals of Statistics, 5.

Hausman, J.A. 1978. "Specification Tests in Econometrics," Econometrica, 46.

Hausman, J.A., and W.E. Taylor. 1981. "Panel Data and Unobservable Individual Effects," Econometrica.

Hirayama, T. 1981. "Nonsmoking Wives of Heavy Smokers Have a Higher Risk of Lung Cancer: A Study from Japan," Br Med J., 282, pp. 183-185.

Hoegg, U. 1972. Environmental Health Perspectives, 2, p. 117.

Kauffman, F. 1980. "Small Airways Dysfunction in Nonsmokers," New England Journal of Medicine, pp. 303-393.

Lave, Lester and Eugene Seskin. 1977. Air Pollution and Human Health, Baltimore: Johns Hopkins University Press.

Lipfert, Frederick W. 1980. "Sulfur Oxides, Particulates and Human Mortality: Synopsis of Statistical Correlation," J. Air Pollut. Control Assoc., 31.

Lunn, J.E., J. Knowelden and A.J. Handyside. 1967. "Patterns of Respiratory Illness in Sheffield Infant Schoolchildren," British Journal Prev. Soc. Med., 21.

McFadden, D. 1974. "A Multinomial Logit Model for Qualitative Choice," in P. Zarembka (ed.), Frontiers of Econometrics, New York: Academic Press.

Mendelsohn, Robert and Guy Orcutt. 1979. "An Empirical Analysis of Air Pollution and Dose-Response Curves," Journal of Environmental Economics and Management, 6, June.

Mundlak, Y. 1978. "On the Pooling of Time Series and Cross Section Data," Econometrica, 46.

Neymann, J. and E.L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations," Econometrica, 16.

Ostro, Bart. 1983. "The Effects of Air Pollution on Work Lost and Morbidity," The Journal of Environmental Economics and Management, 10, forthcoming.

Repace, J.L. 1981. "The Problem of Passive Smoking," Bulletin of the New York Academy of Medicine, 57, p. 936.

Repace, J.L. and Lowry. 1981. Science.

Seskin, Eugene P. 1979. "An Analysis of Some Short Term Health Effects of Air Pollution in the Washington, D.C. Metropolitan Areas," Journal of Urban Economics, 63, July.

Speizer, J. et al. 1981. "Long Term Measurements of Respirable Sulfates and Particles Inside and Outside Homes," Atmos. Environ. 15, p. 23.

Tager, I.B. et al. 1979. "Effect of Parental Cigarette Smoking on the Pulmonary Function of Children," American Journal of Epidemiology, 110, pp. 15-26.

White, H. 1980. "A Heteroscedasticity Consistent Covariance Estimator and a Test for Heteroscedasticity," Econometrica, 48.

White, H. 1981. "Detection and Consequences of Misspecification in Nonlinear Models," Journal of the American Statistical Society.