NBER WORKING PAPER SERIES

EFFECTS OF AIR QUALITY REGULATION

Vernon Henderson

Working Paper No. 5118

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1995

Support from the National Science Foundation for this research is gratefully acknowledged. Mark Beardsell, Randy Becker, John Frost, Arindam Mitra, and Ari Kuncoro worked as able research assistants on this very data intensive project. Paul Portney and Leslie Papke provided helpful comments on earlier drafts of the paper and Robert Moffitt and Mark Pitt provided helpful methodological comments. The paper also benefited from comments by participants in seminars at Brown, Columbia, NYU, Marseille and NBER. This paper is part of NBER's research program in Public Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #5118
May 1995

EFFECTS OF AIR QUALITY REGULATION

ABSTRACT

This paper investigates the effects of local regulatory effort on ground level ozone air quality and on industrial location. Local regulatory effort varies by annual air quality attainment status and by state attitudes towards the environment. A switch from attainment to non-attainment status induces greater regulatory effort in a county, leading to an improvement in air quality. Air quality readings for ground level ozone improve by 3-8% depending on the exact air quality measure, following a switch to non-attainment status. Pro-environment states, which ceteris paribus, spend relatively more on pollution abatement also have cleaner air. A 1% increase in typical annual state pollution abatement expenditures leads to about a .04% improvement in local ozone readings. Heavily polluting industries show a tendency to move to counties with a record of clean air, where they are less likely to be hassled. A county switching to having a three-year record of attainment experiences a 7-9% growth in the number of heavily polluting establishments. This implies polluting industries are spreading out geographically moving from non-attainment (polluted) areas to attainment (initially less polluted) areas. Finally, for ozone, localities may improve the annual hourly extreme value reading used to measure officially local air quality, without improving measures (mean, medians, medians of daily maximum) of more typical ozone conditions. This occurs by spreading out economic activity over the day to dampen peaks of ozone inducing activity and subsequent daily ozone peaks.

Vernon Henderson
Department of Economics
Box B
Brown University
Providence, RI 02912
and NBER

This paper investigates effects of air quality regulation in the U.S.A. for the period 1977-1987. I examine regulation of ground level ozone [03]. Among the criterion air pollutants subject to national air quality standards [NAAQS's], it is the only one where, for this time period, both many localities of the U.S.A. remain out of compliance and many localities go in or out of compliance. In addition, current air quality regulatory efforts are primarily focused on 03.

The paper examines the following general issues. Does local implementation of air quality regulations matter, affecting air quality and the allocation of economic resources? Because application of regulations intentionally differs at the local level, do polluters tend to move to localities where the application of regulations is weaker? How does the exact specification of air quality standards influence outcomes? By looking at these aspects of recent air quality regulatory effort, we will better understand the consequences of future regulatory efforts.

How does air quality regulation work? At the national level, there are regulations governing the design and choice of machinery and equipment and choice of solvents, fuels and coolants to try to reduce air borne emissions. 03 is the product of emissions of volatile organic compounds [VOC] and nitrogen oxides [NO_x], as well as atmospheric conditions such as temperature. Nationally the strategy is to reduce VOC and NO_x emissions. But much of both the application and the implementation of regulations is at the local level, which will be the focus of this paper. In terms of implementation, federal air quality guidelines and standards are enforced generally at the local level. States may differ in their attitude and willingness to actively enforce federal environmental regulations, a notion I will investigate in the paper.

Application of federal regulations also varies explicitly and intentionally at the local level. In the formal process,[1] in the Federal Register each July, every county in the U.S.A.is designated for the subsequent year as being in attainment or not of NAAQS's for each criterion air pollutant. For ozone, the standard is unusual. Rather than being based on annual, quarterly, or even daily averages of hourly concentration readings it is based solely on one extreme value reading. A county is in attainment as long as the highest hourly reading not exceed 0.12 p.p.m. on more than one day a year in that county. That is, the standard is the single highest hourly reading over all hours and days of the year, except for the first day with the highest annual hourly reading. This reading is loosely called the second highest daily maximum. The reason for the focus on extreme value readings is the original belief that it is just high spikes of 03 which pose a health hazard for those with impaired respiratory systems, rather than prolonged lower level exposure. This form of regulation which ignores mean or median performance has very specific consequences for what has happened to ozone air quality.

For local application of regulations, if a county is not in attainment for a particular pollutant, its state is required to submit plans, indicating how the county will be brought into attainment by a particular date. Federal funding in various categories (such as Department of Transportation) may be at risk if reasonable progress is not made. For counties not in attainment, new manufacturing firms to the county may be subject to more stringent federal regulations governing equipment specifications. Existing firms in non-attainment areas face requirements to reduce source emissions and new firms may be required to purchase offsets (emission rights) from existing firms (Atkinson and Tietenberg 1987, Roumasset and Smith 1990). All firms in non-attainment counties are more likely to be closely monitored and subject to greater enforcement efforts (see Deily and Gray 1991 and Russel 1990, as well as Crandall 1985, Tietenberg 1985, and Portney 1990b). For ozone, in addition, auto related regulations may be tougher, requiring the state to set up auto emission inspection stations in various parts of the state. In summary, being out of attainment introduces a set of overall regulatory activities designed to reduce emissions, which counties in attainment do not face to the same extent.

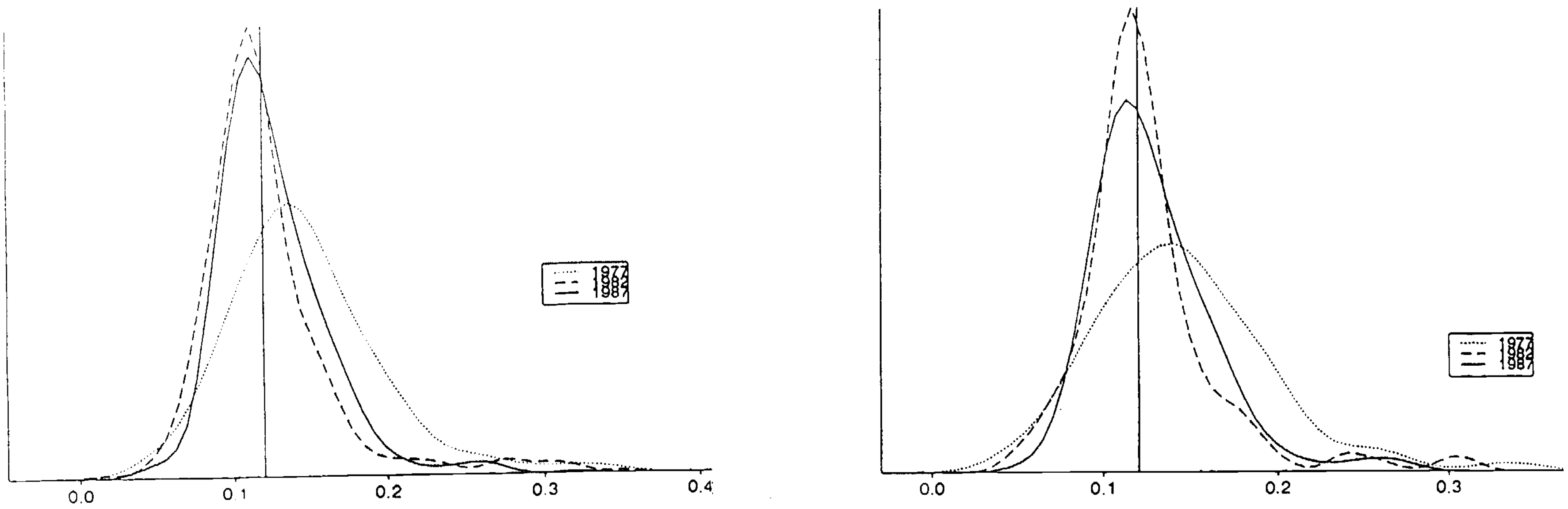
Overall, 03 air quality regulation during 1977-1987 moved the nation substantially towards achieving NAAQS's in different localities. The aggregate picture is of interest and reveals some of the regulatory issues. Ozone is monitored hourly at a variety of permanent and temporary stations around the country. I obtained data tapes for these hourly readings for 1977, 1982, 1985, and 1987, and constructed a sample of 643 monitoring stations as described in the Appendix, with the primary requirement that stations report at least twice in the four years. This sample includes more stations than the EPA bases its national trend statistics on. Coverage, or number of stations increases with time.

In Figure 1a, I picture the distribution of the second highest daily maximum readings for all stations reporting in 1977, 1982, and 1987.[2] Figure 1a gives a clear overview of the effect of regulation. Starting with the Clean Air Amendments of 1977, there is a distribution with a high standard deviation (.049) and with a high median (.140) and mean (.145), both far above the standard of 0.12 p.p.m., with the peak of the distribution also well beyond 0.12. By 1982 and then 1987, the peaks of the distributions shift left to more than meet the standard. Relative to 1977, the 1987 standard deviation, median and mean all fall significantly to .024, .119, and .125 respectively.[3] The visual image is telling. The impact of regulation is to force the peak of the distribution left to meet the standard and to narrow it in both tails. The narrowed distribution concentrates about 1/3 standard deviation below the air quality standard, so these counties around the peak achieve attainment. Air quality improves for the majority of stations but also declines for some stations initially well below the critical reading. To ensure that I have not created an illusion in Figure 1a by adding 1982 or 1987 clean stations to a 1977 set of dirty stations, in Figure 1b I report the distributions of air quality measures for the smaller sample of stations that report in all three years. The pattern is almost identical to that in Figure 1a. Means and variances of readings in 1982 and 1987 are not significantly different for the samples of stations which reported in 1977 versus did not.

In viewing Figure 1, we don't know exactly what would have happened to the distributions in the absence of the Clean Air Act. But Figure 1 suggests that localities have altered their activities to try to meet the specified standard. However, because the standard

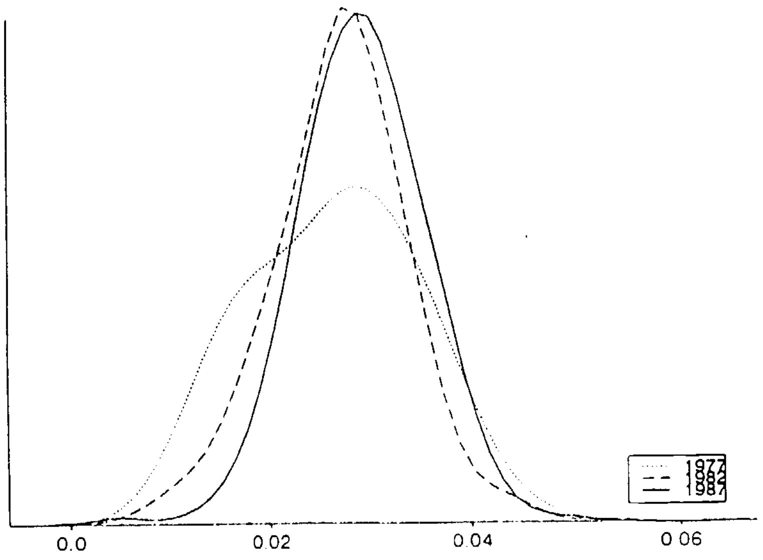
READINGS

FIGURE 1 DISTRIBUTION OF SECOND HIGHEST DAILY MAXIMUM READINGS


(A) ALL STATIONS

(B) STATIONS WHICH REPORTED IN 1977

FIGURE 2 DISTRIBUTION OF ANNUAL MEAN OF HOURLY READINGS



is a peculiar one which focuses on extreme value air quality readings, the overall picture for typical air quality readings may not be so rosy.

Figure 2 reports on the distribution across stations of the mean annual hourly reading (distributions for the median of all daily maximum readings closely mirror Figure 2 in terms of shape and positioning of the distributions). While the distributions narrow significantly after 1977, the peak does not change and the mean and median of the hour averages actually increase significantly from 1977 (.026 and .027) to 1987 (.029 and .029) (see fn. 3 for test details). Note any considerations such as the increase in the number of days monitored at different stations over time only strengthen the comparison. Some stations just monitor for the ozone "season," which is the warmer parts of the year. Lengthening the season incorporates lower reading days, which would work to lower the means over time.

Given this background information on air quality regulation and air quality changes over time, I examine three hypotheses in this paper. First, I hypothesize that local application and implementation of air quality regulations matter, per se. Ceteris paribus, the designation of county attainment status and state "attitudes" affect local air quality measures, so that non-attainment counties and pro-environment states clean-up, relative to attainment counties and other states.

Second, I examine the effect of local regulation on the location of industrial activity. Since regulatory effort differs across space, I expect a tendency for polluting industries (which are subject to greater scrutiny) to move from non-attainment ("dirtier, more regulated") counties to attainment ("cleaner, less regulated") counties. This may not be bad, but it is against the intent of the Clean Air Act. Relocation of polluting activity could be in part responsible for the narrowing over time of the distributions in Figures 1 and 2, as dirtier activities move into formerly cleaner areas.

Finally, I will show one way in which it is possible to see improvements nationally in Figure 1 in achieving NAAQS's, but not in Figure 2 in improving more typical ozone conditions. Since the air quality standard only concerns extreme value peaks, if peaks can be dampened without changing overall VOC and NO_x emissions, that presents a "loophole" in local application. I say loophole, since the intent of the Clean Air Act is to improve air quality through reduction of total emissions.

The ideas that regulation matters and specific provisions have specific consequences should not be surprising. Yet, the literature on air quality regulation often suggests otherwise. Studies suggest that local regulation has little or even perverse effects on air quality (MacAvoy, 1987) and on firm location decisions (McConnell and Schwab, 1990 and Bartik, 1988). I believe the reason why many researchers haven't found effects involves their reliance on cross-sectional data and/or estimation methods. The location of polluting activity, high concentration readings, and the designation of non-attainment status are of course all strongly positively correlated, cross-sectionally; so air quality indeed is worse in non-attainment areas, and polluting firms are predominantly found there. By use of panel data, I can control for this basic association and show, for example, how a change in status (from attainment to non-attainment, or vice versa) affects air quality and industrial location. Two recent studies which use panel data, Duffy-Deno (1992) and Gray and Shadbegian (1993), also find specific anticipated effects of regulation, although neither paper specifically examines my hypotheses.

Local Regulation Affects Air Quality

In this section I show that the local application and implementation of air quality regulations affect local air quality. Specifically, first, I hypothesize that, ceteris paribus, counties in non-attainment will have cleaner air, than those in attainment. The extra regulatory effort induced by the designation itself of non-attainment status leads counties to clean-up. In showing this, the problem is that the ceteris paribus requirement is difficult to implement since in general non-attainment counties have dirtier conditions than attainment counties. However with panel data we can implement the ceteris paribus condition by following the same counties over time. In particular, I show that a county switching from attainment to non-attainment status experiences an improvement in air quality, while the opposite switch induces a decline. Second, I hypothesize that, ceteris paribus, air quality is better in "pro-environment" states.

The Model

Annual air quality readings at a particular station are affected by the level of economic activity, creating VOC and NO_x emissions in the county, annual weather conditions, the precise location of the station relative to the location of economic activities in the county, and so on. A general specification is that

$$03_{it} = C + \beta X_{it} + \gamma Z_i + f_i + \epsilon_{it} \tag{1}$$

03_{it} is the summary ozone reading at site i in year t. X_{it} are time variant characteristics of polluting and regulatory activity, and Z_i are the time invariant characteristics of the area affecting ozone readings. f_i is a fixed/random effect specific to the monitoring station reflecting the geography of its location and ϵ_{it} is the contemporaneous error term.

The X_{it} include time dummies, overall employment levels (to control for commuting and other general economic activity), presence of specific polluting industries, annual hot weather conditions and regulatory activity. The extent of local regulatory activity in X_{it} is measured primarily by whether the county is in attainment or not of NAAQS's for that year, or lagged one period (i.e., in attainment or not the year before). Lagging may be necessary because it may take a year or so for regulatory activity to come into effect or unwind as a county changes attainment status. The Z_i are time invariant measures within the sample period relating to geographic characteristics such as coastal location and land area of the county, and attitudes in the state towards air quality regulation.

I have a four-year data panel, with each station appearing at least twice but 80% of the observations being from stations with three or more years of data. Equation (1) suggests a fixed effects estimation procedure. Fixed effects will be correlated with some of the X_{it}'s, in particular, attainment status, where the f_i affect whether or not a county is likely to be in attainment or not. Indeed, in all cases (by Hausman tests), estimation procedures reject a random effects formulation in favor of a fixed effects formulation.

Fixed effects estimation removes all time invariant variables from initial consideration. To recover these, for each monitoring station, I use the estimates of β to calculate

$$\overline{RES}_i \equiv \overline{03}_i - \hat{\beta}\overline{X}_i = C + \gamma Z_i + f_i + \left(\sum_T e_{it}/T\right). \tag{2}$$

In equation (2), $\overline{03}_i$ and $\overline{X}_i$ are respectively for each station the time average of the pollution measures and the time average of the X_{it}. $\sum e_{it}/T + f_i$ can be treated as a composite error term. Because the panel is unbalanced the errors will be heterogeneous but there are other sources of heterogeneity (from the calculation of $\overline{RES}_i$) as well. The standard errors will be White-corrected for heterogeneity.

General data sources are discussed in the Appendix and noted as we go along. For air quality summary annual measures, I present results for four measures. Besides the second highest daily maximum which defines annual (non) attainment, I use the annual mean of the hourly readings for each station. Both annual measures suffer from a critical problem. Monitoring stations record typically for only part of the year, when temperatures are higher. In evaluating hourly readings it is desirable to control for the number of days a station records. Unfortunately, the number of days recorded at a site varies annually with the ozone readings, increasing in deteriorating situations, making it difficult to treat this as an exogenous measure. While I examine this issue thoroughly below, it is difficult to effectively deal with it. To compensate, I also report summary measures for July with its high temperatures and almost universal hourly ozone readings. For many localities, July is also the worst ozone month. I will report results for the median of the highest daily readings in July and for the mean reading in July. The median measure gives a more typical peak daily reading and the mean tells us "average conditions" over the July ozone season.

The results in this section are broken into two parts. In the first I discuss the results of fixed effects estimation of equation (1) and in the second I discuss implementation and results for equation (2).

Time Variant Determinants of Air Quality Readings

I estimated equation (1) by fixed effects methods for the four measures of air quality. Results are given in Table 1. All level variables are in logs. In the top panel, I report on the critical geo-economic control variables influencing air quality over time in a county and in the bottom panel on regulatory variables. In viewing the results it is important to note that fixed effects estimation is demanding of the data, since estimates are based on time variation in the data for each monitoring station and not on cross-sectional variation in the data across stations. I start with regulatory variables.

Regulatory Variables. Included in the bottom panel of Table 1 are time dummy variables which compare 1977, 1985, and 1987 with 1982. These can be interpreted as representing the effect over time of national regulatory policy on air quality. For the second highest daily maximum and the July measures indeed there is a significant improvement in air quality from 1977 to 1982 (12-18% reductions), ceteris paribus. However, there is no change in the annual mean readings. Comparing 1982 and 1987, the picture is mixed and effects are small.

Local regulatory policy is more directly represented. First, I represent whether the county is in attainment or not of the ozone standard. With fixed effects estimation, the variable represents the effect of a change in attainment status on air quality. Some experimentation suggested (non-critically) lagging this variable one period[4] to allow for a year's lag in the implementation or relaxation of tougher enforcement and monitoring activities. In the data on ozone attainment, there is also a status of being in partial attainment, applied to a few counties. Most of these are Los Angeles and its contiguous counties which have the worst air quality in the nation. I treat them as non-attainment counties.

Earlier, I hypothesized that being designated as non-attainment would result in stricter

Table 1

Time Variant Determinants of Air Quality

	ln (2nd highest daily max.)		ln (mean annual reading)		ln (med. of daily max. July)		ln (mean July reading)	
ln (avg. daily max. temp. July)	1.09*		.213		1.85*		1.20*	
	(.179)		(.164)		(.190)		(.187)	
ln (MSA employ)	.080		.090		.137**		.212*	
	(.067)		(.062)		(.073)		(.072)	
presence of plastics	.109*		.032		-.070		-.055	
	(.056)		(.051)		(.060)		(.058)	
presence of organic ind. chemicals	.046*		.058*		.063*		.092*	
	(.022)		(.020)		(.023)		(.023)	
presence of petro. refining	.019		.014		-.0061		-.0076	
	(.018)		(.017)		(.019)		(.019)	
year 1977	.171*		.014		.133*		.115*	
	(.020)		(.018)		(.022)		(.022)	
year 1985	-.013		.041*		-.0037		.027**	
	(.013)		(.012)		(.014)		(.014)	
year 1987	.012		.058*		-.036*		-.00014	
	(.016)		(.015)		(.017)		(.017)	
county in 03 non-attainment	-.026	-.017	-.029	-.020	-.078*	-.081*	-.043**	-.040**
	(.024)	(.24)	(.022)	(.021)	(.026)	(.026)	(.025)	(.025)
index: non-attainment for other A.Q.		-.079*		-.079*		.029		-.029
		(.025)		(.022)		(.027)		(.027)
N	1864		1864		1752		1752	
Adj R^2	.66		.65		.73		.73	

Standard errors in parentheses: * significant at 5% level ** significant at 10% level monitoring and enforcement of polluting industries and stricter regulation of non-stationary sources as well, leading to improved air quality. In fixed effects estimation this is inferred by examining the effects for the same county of a change in attainment status. 18% of counties change status in the sample. A problem is that counties typically move from non-attainment to attainment status in the sample, and officials may be reluctant to wind down regulatory activity in a county achieving attainment, out of a fear of a relapse and redesignation of non-attainment.

Nevertheless, for all air quality measures, being in non-attainment has the hypothesized negative coefficient. While all t-statistics have values over one, only for the July median of the daily maximums is the coefficient actually significant at the .05 level. There a change to non-attainment status leads to an 8% improvement in air quality. Why might some measures show a less significant outcome? In general, there is the fixed effects problem of inferring only from time and not cross-sectional variation in data. For the second highest daily maximum specifically, we might expect limited explanatory power in estimation in general, for a measure of an annual ozone spike, as opposed to typical conditions. Moreover, since it is the measure keying regulatory activity, officials would be less likely to relax regulatory activity keyed to that measure.

Further investigation indicates other assets to the regulatory situation. For the other relevant air quality dimensions of regulation (sulfur oxides, carbon monoxide, nitrogen oxides and TSP (total suspended particulates)) I created an index. Relative to ozone, attainment status changes much less for any other air quality dimension, so it is difficult to treat them separately. Moreover, an index gives a summary measure of the focus of air quality regulation facing a county. The idea is that counties in non-attainment in several dimensions are going to be the focus of greater regulatory attention. Also, in terms of emissions, if, for example, ferrous-metals plants clean up sulfur dioxides emissions they may fortuitously also clean up VOC emissions. For each year, each dimension takes the value 0 if the county is in attainment in that dimension, 1/2 if the county is either in partial attainment or in violation of secondary standards, and 1 if the county is in non-attainment. I then summed these values over the four air quality dimensions for each year to get an index of the degree of non-attainment in other dimensions.

The results for this index are reported in Table 1 in the second column for each air quality measure. In two of the four cases this variable has a significant negative sign, with little effect on the magnitude of other variables. To be consistent with ozone attainment, the index is lagged one period. (For July measures the effects are stronger if the index is current.[5]) For annual measures, the effects are strong. A one standard deviation (.68) in the index (with mean .79) leads to a 7% decrease in the second highest daily maximum reading and a 6% decrease in the mean hourly reading.

In summary, either directly through ozone or through overall air quality regulation, annual ozone conditions tend to be affected significantly by county air quality attainment status. Being in non-attainment leads to greater effective regulation and improved air quality, ceteris paribus. A change to attainment status causes local regulatory effort to ease and air quality to decline.

Non-Regulatory Variables. In the top panel in Table 1 are the control variables, which I comment on briefly. The average July maximum temperature raises ozone readings with an elasticity over 1 for three of the columns. Other July temperature measures yield similar results. For the level of general economic activity including commuting, I could not really distinguish between own county versus contiguous county total employment, nor between overall manufacturing and total employment. In the end, I settled on total MSA employment since ozone is a regional problem. In all cases an increase in MSA employment increases ozone readings, but the variable is less significant than expected.

To control for industrial composition, I looked at the major national three-digit manufacturing VOC emitters. I experimented with level measures, dummies for the presence (i.e., exit or entry over the time horizon of the panel) of particular activities, and both dummies and level measures. Level measures alone are problematic, because the particular three-digit industries are only present 50% of the time. Use of both level and dummy variables seemed to demand too much of the data or fixed effects estimation, with level measures being insignificant. So Table 1 reports on the dummy variable measures. The presence of organic industrial chemicals is the only industry producing a significant impact on ozone readings, for all measures. Plastics and petroleum refining have expected positive signs for annual measures, but not for July measures. Organic industrial chemicals are relatively more polluting than plastics and probably less regulated than petroleum refining (with its larger average plant size and limited number of plants). For other industries, not reported in Table 1, the presence of non-ferrous primary metals tends to significantly raise ozone readings but not so for ferrous primary metals.

Various Econometric Issues

The estimation of equation (1), the data set, and the results in Table 1 raise a variety of econometric issues. Here I review briefly the issues I have explicitly dealt with.

As noted earlier, a major issue in analyzing annual air quality measures concerns differences in the number of days for which ozone readings are recorded. For any station, an increase in season length is likely to lower annual means (adding on tail-end days with typically lower readings) and in theory increase the expected value of the extreme value draw (by increasing the sample size alone). In attempts to control for days recorded in equation (1), it is apparent that, for a station, changes in days recorded is positively correlated with deterioration in air quality readings that year. That is, annual days recorded is a good proxy for annual pollution readings! Specifically with days recorded as the dependent variable, in fixed effects estimation, for any station, a 1% increase in the second highest daily maximum leads to a 1.17% increase in days recorded.[6] In equation (1), I experimented with instrumenting for days recorded (where then the coefficient on days recorded drops to zero), but lacked good instruments.[7] With little choice, annual measures in Table 1 do not control for days recorded. For July measures there is no issue, since coverage is universal for July across stations. That was a primary reason for focusing on July, as well as annual measures.

The remaining issues are four basic econometric concerns. I summarize the issues and outcomes of testing, footnoting details. First, in fixed effects estimation, particularly for short panels, right-hand side variables should be exogenous to contemporaneous errors in all years in the panel, not just the current year. For the second highest daily maximum reading, in particular, a bad reading in one year could lead to a change in ozone attainment status for the next year or so. For the years 1977, 1982, and 1987, this is not a problem since they are spaced far enough apart. For 1985 it is a problem, since, for example, a bad 1985 reading could lead to a change in 1986, 03 attainment status which is a 1987 right-hand side variable. To check on this issue, I re-estimated the model dropping 1985. Results are similar, with no sign switches. By a Hausman test, I could not reject the hypothesis that the coefficients are the same with and without 1985 in the sample, so 1985 is included for efficiency reasons.[8]

Second, it has been suggested that stations in "outlier" counties which experienced big air quality improvements but still remained in non-attainment may be driving my results (where, by comparison, counties switching status to attainment may have had only modest air quality improvements). To test for this I re-estimated the model dropping outlier stations. Generally coefficients on attainment status are little changed; however, the coefficient for the second highest daily maximum becomes effectively zero.[9]

Third, concerns the nature of my panel. I have a panel of monitoring stations. There are on average two stations per county in the panel. Most counties have one station but some (e.g., Los Angeles, Orange, and Cook counties) have 10 or so stations. Counties with greater representation tend not to change attainment status in the sample period, so there is a heavy weight to monitoring stations in counties which don't change status. If, however, I construct an (unbalanced) panel of counties (332 in total, for 1,063 observations) I change the weights so each county appears just once in a year.[10] Rerunning the equations in Table 1, the results are almost the same, although t-statistics for county non-attainment status rise to near significance at the 5% level for the annual and July mean reading equations.

Finally, there could be an issue of selectivity bias, concerning what counties are monitored. Of the 742 urban counties in the U.S.A., our data only cover 332. For the basic panel of monitoring stations, the county is not the unit of observation per se. In fixed effects estimation the remaining error term is the deviation about the time average for a station of the contemporaneous error term. I don't believe selectivity bias is an issue here, because these error terms relate to considerations such as annual weather conditions. Nevertheless, I looked at this issue for the panel of counties from the previous paragraph. For two sets of pairs of years (1977-82, 1982-87) I estimated a discrete-continuous choice model by maximum likelihood. The discrete choice concerns whether the county was monitored in 1977 (or 1982 for the second sample). Conditional on being monitored, the continuous choice examines the change in air quality (a first differenced equation (1)) between either 1977 and 1982 or 1982 and 1987. With four air quality measures we have eight models. For only one (the July median of daily maximum readings for 1982-1987) was the ρ coefficient significant (or near significance). Similarly Heckman two-step tests for selectivity rejected selectivity in the same seven of eight cases. I would also note that for the larger continuous sample for 1982-1987, the first differenced equation had stronger regulatory effects than in Table 1.[11]

Time Invariant Variables: State Attitude

To recover the effect of time invariant variables, I estimated equation (2) by OLS. Besides measures of geographic conditions, the key time invariant variable is a measure of the attitude of state officials towards air quality regulation. The literature suggests enforcement varies significantly across states, based on case studies of the magnitude of fines facing violators, the extent of plant site visits, citations of violators, and the like (Deily and Gray 1991 and Russel 1990). Unfortunately, there isn't a comprehensive data set covering all states over a period of years to get a real sense of how direct regulatory activity varies across states. There are indexes of "green" constructed by non-profits in evaluating state attitudes to the environment, but they all seem pretty subjective. Here I use an indirect, objective way of measuring regulatory activity.

The Bureau of the Census collects state level data on expenditures on anti-pollution activity for air borne emissions in manufacturing, for firms with 20 or more employees.[12] The data give operating costs for pollution abatement [GAC] (GAC includes depreciation, labor, materials and supplies, and services and leasing (including payments to government agencies for removal of pollutants)). There are also data on the (smaller) expenditures on new capital equipment, but I don't use it here because of significant measurement issues (expenditures are hypothetical not actual), as well as issues of timing of investments.

Abatement expenditures in a state should be determined by the levels and growth of polluting activity (total population and employment and employment in the five major two-digit polluting industries), time effects and the state record of attainment of pollution standards in different air quality dimensions. In Henderson (1994b) I estimate a fixed effects panel data model with pollution abatement expenditure as the dependent variable for 1978-1986 for 51 states (including DC) with these other measures as explanatory variables. The model has high explanatory power results. That model yields a fixed effect term for each state. I use that term as a measure of state attitudes towards air quality regulation. The fixed effect measures the extent to which firms in the state "under" or "over" spend on pollution abatement activity, controlling for the levels and growth of polluting activity and extent of non-attainment of air quality standards. Overspenders are viewed as pro-environment states, where regulations are strongly enforced.

In Table 2, I present the results of equation (2).[13] Standard errors are White corrected. For the key variable, states with a positive attitude towards environmental regulation have significantly better air quality (a significant negative coefficient). Only for the second highest daily maximum is this not the case. One is tempted to argue that negative attitude states focus their more limited total spending on this key regulatory variable — the second higher daily reading of the year — so there is no significant difference between them and other

Table 2

The Impact of Time Invariant Variables

	Second highest daily reading		**Annual mean**	**Median of July daily maxima's**		**July Mean**
constant	-1.462* (.018)	-1.11* (.216)	1.14* (.019)	-5.70* (.024)	-5.27* (.293)	-4.43* (.027)
dummy county on coast	.058* (.019)	.070* (.018)	-.166* (.022)	-.149* (.028)	-.148* (.028)	-.283* (.033)
dummy LA coast	.449* (.043)	.404* (.047)	.082 (.050)	.410* (.059)	.406* (.063)	.227* (.055)
state attitude (GAC fixed effect)	.017 (.014)	.008 (.013)	-.050* (.014)	-.043* (.019)	-.043* (.019)	-.077* (.020)
ln (county land area)		-.133* (.067)			-.129 (.088)	
ln (county land area) squared		.012* (.005)			.009 (.007)	
R^2	.23	.25	.10	.10	.11	.14
N	643	643	643	616	616	616

---

* Significant at 5% level.

states, ceteris paribus in terms of this air quality measure.

In terms of the control variables, in three out of the four cases, a general coastal dummy is negative (wind patterns and off-coast ozone absorption). However, there is an obvious problem in controlling for coastal location. Controls for metro area employment may not fully represent the effects on 03 readings of coastal corridor population concentrations in the U.S.A. near Los Angeles, New York, Boston and Houston. I did try to represent the obvious candidate by including a dummy for stations in the Los Angeles corridor including Los Angeles, San Diego, Orange, Ventura, and Santa Barbara counties. Finally, for land area, while there is no effect on mean readings. For the second highest daily reading and for the median of July daily maxima increases in land area (for smaller counties) improve air quality. Greater land area means less general density of economic activity, ceteris paribus.

Air Quality Regulation and Industrial Location

Local air quality regulation may improve air quality measures because of reduction in emissions at the various sources. However, local regulation may also lead to improved air quality simply because polluters move. In this section I investigate whether firm location decisions are affected by local environment regulation, in particular by the designation of whether a county is in attainment of air quality standards or not. As noted earlier the literature generally concludes no. Using panel data I will suggest otherwise for heavy polluters. To do so I examine the effect of attainment status on the location decisions in 742 urban counties of the five three-digit industries which are major VOC emitters in total tons, between 1978 and 1987. These are industrial organic chemicals (SIC 286), petroleum refining (SIC 291), miscellaneous plastics (SIC 307), plastic materials and synthetics (SIC 282), and blast furnace and primary steel (SIC 331). The hypothesis is that overtime plant will move into attainment areas and out of non-attainment areas.

The raw data support the hypothesis. The time period I look at covers 1978 to 1987 and the spatial coverage is the 742 urban counties in the U.S.A. which capture most industrial activity. In Table 3a I show the initial locational pattern of plants in each industry where in 1978 most plants are located in non-attainment areas. In Table 3b, I show the growth rate in the number of plants in "clean" counties which are in attainment in both 1978 and 1987 (column 1) and compare it with the growth rate in the number of plants in "dirty" counties which are in non-attainment in both 1978 and 1987. Average employment growth rates for all industries in the two sets of counties are the same for the ten years. Yet for the five high polluters, the growth rate of number of plants is much higher in the attainment counties. In column 3, we also show the growth rates for counties which changed status — going from non-attainment to attainment — which accounts for 80% of status changes in our sample for ozone. In four out of five cases, improved counties also have higher growth rates than non-attainment countries. This evidence is very suggestive of the relocation hypothesis. An econometric specification will allow us to control for other factors and to quantify effects.

The Model.

There is a substantial empirical literature on firm location decisions (Herzog and Schlottman 1991 for a review). To answer the question particular to this paper, I adapt a specific econometric model for data with two cross-sections presented in Henderson et al. (1995) to the current context which uses panel data. This work is related closely to standard work by Carlton (1983), and others. Firms in county j in a particular industry at time t have a profit function $\Pi(Y_{jt}, s_{jt}, u_{jt})$ where Y_{jt} is a vector of arguments depicting current economic and regulatory conditions. Current local scale of this industry which is measured by the number of plants, s_{jt}, represents the positive effects of own industry economies of scale and the negative effects of trying to sell more and more of the product in a limited regional market area, which may negatively impact the (unobserved) local output price. The current local supply of entrepreneurs to the industry is given by $\tilde{\Pi}(s_{jt}, d_{jt} \ldots)$ where, as local scale, s_{jt} rises, per firm profits must rise to attract more entrepreneurs. The scale of local operations is determined by the intersection of the $\Pi(\cdot)$ and $\tilde{\Pi}(\cdot)$ functions. Equating

Table 3

(a) 1978 Stock of Plants

	SIC	Percent of All Firms Located in Non-Attainment Counties
Plastic Materials & Synthetics	282	91
Industrial Organic Chemicals	286	89
Miscellaneous Plastics	307	87
Steel	331	92
Petroleum Refining	291	92

% of all urban counties with 1978 non-attainment status = 60%

(b) Total Percentage Change in Number of Plants 1978-1987 in Counties by Ozone Attainment Status

SIC	Attainment in both in '78 and '87	Non-Attainment in both '78 and '87	Non-Attainment in '78 Attainment in '87
Plast. Mat'l./ Synthetics (282)	67	14	21
Ind. Organic Chemicals (286)	19	8	36
Misc. Plastics (307)	69	20	38
Primary Steel (331)	28	2	6
Petroleum Refining (291)	15	6	-5
employ growth in counties	32	34	

and solving for s_{jt} we get a reduced-form equation of the form

$$s_j = s(Y_{jt}, e_{jt}). \tag{3}$$

Equation (3) forms the basic estimating equation. e_{jt} is decomposed into a fixed effect and contemporaneous error term. Application of the implicit function theorem, imposing "stability" in the market for establishments in an industry in a city ensures that $\partial s/\partial Y$ and $\partial \Pi/\partial Y$ have the same sign, which will be important in interpreting coefficients.

Equation (3) describes local scale for counties with the industry. Some counties in some years don't have the industry or $s = 0$ in equation (3). This means $\Pi(Y_{jt}, \overline{s}_{jt}, u_{jt}) < \tilde{\Pi}(\overline{s}_{jt}, d_{jt})$ for a critical value of $\overline{s}_{jt}$. Specifically, if I linearize (in logs) $\Pi(\cdot)$ and $\tilde{\Pi}(\cdot)$, for a continuous event $s_{jt} = \beta Y_{jt} + e_{jt}$, while for a discrete event $\beta Y_{jt} + e_{jt} < 0$, for $\overline{s}_{jt}(\text{in logs}) = 0$. I estimate equation (3) as a (fixed effects) Tobit, for the sample of urban counties where the industry appears in at least one year of the panel.

Results.

I start by examining the key regulatory variables. Initially, I measured the extent of local regulatory activity by the dummy variable for county non-attainment status used earlier in the paper. It generally had the hypothesized negative coefficient but it was only plausibly significant for two industries.[14] Experimentation revealed that a plant's response to regulatory changes may be lagged, but I didn't have sufficient information to tease out a lag structure in a short panel. More critically, a firm may be looking for a county to show a sustained record of attainment before (re)locating or staying there, especially since some counties go back and forth in status. Accordingly for this data set where I have continuous years in estimation (unlike the first section of the paper), I construct a dummy variable, which takes a value 1 if the county has been in attainment for the last three years (including the current). Thus, the action in estimation comes from counties which switch in or out of having a record (three years) of clean air.

Results are in Table 4. In column (i) for each industry, I report the coefficient for the dummy variable for being clean for three years. Except for steel which is a smaller VOC and NO_x polluter than the other industries (especially 282, 286 and 291), all coefficients are positive and significant. Counties switching to having a three-year record of attainment have 7-10% more establishments in these industries than counties with recent episodes of non-attainment. In column (ii) I allow for an additional impact of the immediate switch in and out of attainment. Only for miscellaneous plastics is the immediate impact significant. Finally in columns (iii), for plastic materials and miscellaneous plastics, I add in the index of non-attainment in other air quality dimensions. Only for these industries is the coefficient significant and robust. Here the effects of an increase in the index by 1 are large — a 11% reduction in establishments.

In summary, a county being in non-attainment for heavy VOC and NO_x emitters discourages location there. A switch to a clean ozone record increases the number of plants in the county by about 8%. In addition, for plastics, non-attainment in other dimensions is also discouraging.

In Table 4 in columns (i), I also report results on one control variable, MSA employment in other industries. Here the elasticity varies from .04 up to 1.0. Other non-reported controls are dummy variables for each county and a time dummy for the 1983 business cycle peak. I also experimented (with no effect on coefficients of regulatory variables) with price variables for annual county wages (in all other industries) and the annual national producer price index for the specific three-digit commodity. In general, these variables did not have consistent effects across industries.[15]

Estimation Issues.

Generally for models incorporating a discrete choice, fixed effects estimates are biased. I originally estimated the model with a Chamberlain (1980) conditional logit to avoid this problem. There a county is only in the estimating sample if it experiences years of both zero

Table 4 -- Industrial Location

(LN (number of establishments in county j in year t))

	Ind. Org. Chem.		Petro. Refining		Plastic Materials			Misc. Plastics			Steel	
	(286)		(291)		(282)			(307)			(331)	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)
DUMMY: CLEAN FOR LAST 3 YEARS	**.091*** **(.029)**	.077* (.032)	**.065*** **(.038)**	.051 (.041)	**.072*** **(.031)**	.083* (.034)	.056** (.033)	**.081*** **(.019)**	.045* (.021)	.065* (.019)	**-.003** **(.026)**	.011 (.029)
NON-ATTAIN. STATUS DUMMY		-.031 (.031)		-.039 (.045)		.026 (.034)			-.088* (.021)			.031 (.029)
INDEX: NON-ATTAIN. IN OTHER AQ's							-.105* (.038)			-.115* (.027)		
LN (MSA EMPLOY., ALL OTHER INDUSTRIES)	.046 (.072)		1.02* (.079)		.159** (.082)			.326* (.092)			.112 (.071)	
N [NON-ZERO, (ZERO)]	1991 (609)		1179 (725)		1869 (796)			4903 (521)			2462 (802)	

and non-zero establishments in the sample periods. An event is a sequence of 0 and 1's where a zero marks zero establishments for a year and a 1 positive establishments. The Chamberlain logit showed strong effects for SIC 286 and 307, for the non-attainment dummy. However, using the Chamberlain logit involves an enormous loss of information since it omits counties which always have the industry and whose numbers of establishments grow or decline in response to regulation.

Accordingly I switched to a fixed effects Tobit. In estimation, with the construction of the dummy for being "clean," there remain eight years (1980-87) in estimation. With that length of panel for a Tobit (and even for a probit Heckman (1981)), the extent of bias in estimation is judged to generally be very small. I did consider using a conditional Poisson estimator; but with the numbers of establishments typically ranging up to 45-75 (and ten times that for miscellaneous plastics), I felt comfortable not imposing a count model.[16]

Another issue is that the data are noisy. In particular, a single "plant" with 0-2 employees will appear in a county for a year or two and disappear. In Chamberlain logits where this feature presents the greatest problem, cleaning the data to deal with this problem did not alter coefficients (although standard errors fall).

Finally, there is the issue that, in equation (3), the Y_{jt} are assumed to be strictly exogenous. The basic problem concerns a county-industry shock in t causing local industrial growth, which could move the county into non-attainment in $t+1$. For the panel of counties always having plants, for all five industries, I conducted Hausman tests based on GMM estimates (see Henderson, 1994a, for a complete specification in another context) of continuous equations of the number of plants to test whether the Y_{jt} (including the non-attainment status dummy) are strictly exogenous versus merely predetermined (using as instruments for each year all predetermined variables). I could not reject strict exogeneity.

Nevertheless, since I was still concerned about exogeneity of the key regulatory variables I tried other experiments, focused on the two biggest (by a large margin) VOC polluters in my data — industrial organic chemicals and petroleum refining. I tried some two-stage

Tobit estimates. Two-stage results produced implausibly high regulatory effects,[17] probably due to poor instruments (now, just strictly exogenous variables, rather than predetermined ones).[18] Lacking good instruments for two-stage estimation, I tried another inefficient approach. I estimated a maximum likelihood model based on just two years, 1982 and 1987, which are far enough apart that 1982 shocks won't affect 1987.[19] The model looks at the growth rate of plants in counties between 1982 and 1987, so $\Delta s_{jt} = \beta \Delta Y_{jt} + \Delta \epsilon_{jt}$. It also incorporates the counties with zero 1982 employment and positive 1987 employment, or vice versa, in a maximum likelihood specification.[20] Dropping all other years in estimation greatly reduces efficiency and variables are generally insignificant. But the coefficients on the dummy variable for clean for three years have coefficients of .073 and .12 for industrial organic chemicals and petroleum refining respectively. Compared to Table 4, one coefficient falls modestly and the other almost doubles. Although these are imprecise point estimates, they support the general hypothesis.

Over the Day

In an urban area there is a strong daily cyclical pattern to ozone, which as we will see is related to the daily cyclical pattern of economic activity. Ozone readings peak at times around 1-2 p.m. and these peak readings are 4-5 times the trough readings at 5 a.m. Since ozone regulation is based solely on extreme value readings, if a locality can reschedule economic activity away from peak hours to off-peak hours, it can dampen the daily peaks for the same overall level of daily economic activities which lead to VOC and NO_x emissions. I hypothesize that economic activity was rescheduled in the time period 1982-87 in higher ozone areas, leading to dampening of the daily ozone peaks. In particular, we will see significant rescheduling in non-attainment areas generally and especially in California, the worst ozone state in the country. Such rescheduling will not be observed on average in attainment areas.

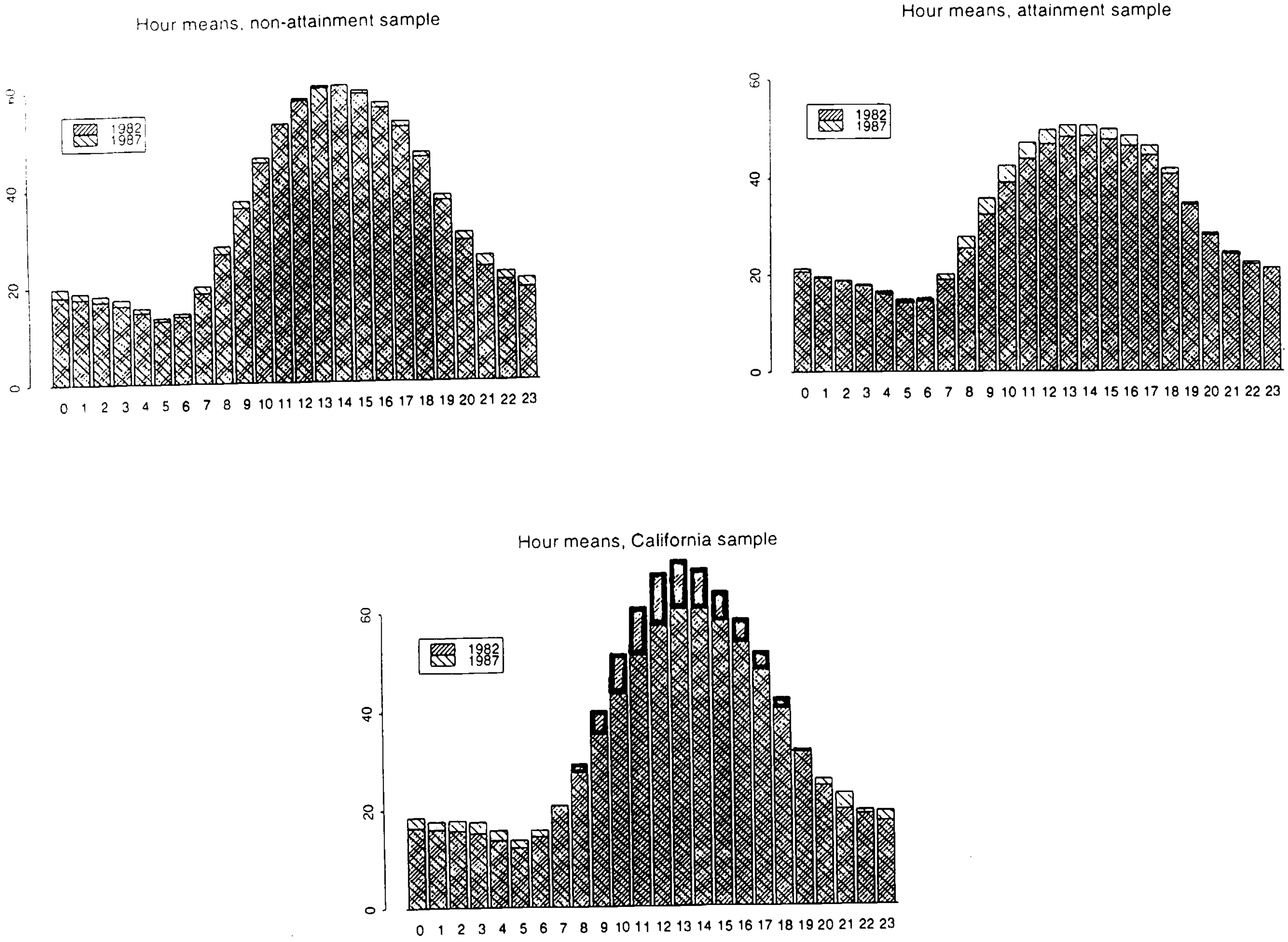
The underlying reasons for these differentials in rescheduling are less clear. In urban areas, high ozone levels are strongly correlated with high auto congestion levels. Rescheduling of economic activity in the 1980's may have been mostly connected with congestion mitigation programs, as well as the natural evolution of flex time and staggered work hour programs at the work place in congested metro areas. Today, federal funding explicitly recognizes the interconnection between congestion and ozone in the major Congestion Mitigation Air Quality (CMAQ) funding program.

The raw data support my hypothesis. I compare the changes in the daily ozone cycle between 1982 and 1987 for the month of July. I chose July because it has the most complete set of hourly readings for stations and because typically it is the worst (highest temperature) ozone month. Data are described in Appendix B. In Figure 3, I plot average hourly readings in p.p.b. over the day for July, dividing the sample of monitoring stations into 66 stations in 1982 attainment areas, 340 stations in 1982 non-attainment, and a subsample of 78 stations in California, where congestion and ozone mitigation programs have been extremely active.

In each graph, average readings for a particular hour for July are indexed, with 0 being midnight and 23 being 11 p.m. The graphs compare the 1982 and 1987 raw hourly averages. Portions of the graph which are heavily outlined or marked indicate hours where 1982 readings exceed 1987 readings — hours with improvement. In attainment areas, on average, the heights of the daily peak readings (from noon-3 p.m.) increased, so peak air quality declined between 1982 and 1987. In non-attainment areas, on average, there was little or no change in the height of the daily peak but significant increases in off-peak hours. That is, daily average ozone readings increased in these areas, but peak readings did not, in contrast to attainment areas. In very active program areas such as California, we get dramatic results — large decreases in peak ozone reading with increases in the off-peak hours.

How do congestion mitigation-air quality activities work? The daily ozone cycle does not correspond to the daily cycle of economic activity. To see this, I model hourly ozone

Figure 3 -- Hourly Average Ozone Readings for July



readings over the day for a typical station. 03_{st} is the ozone reading in day "s" at hour "t" in July for a station.

$$03_{st} = A + \sum_{\ell=1}^{m} \alpha_\ell 03_{t-\ell} + \sum_{j=1}^{23} d_{jt}\beta_j + \epsilon_{st}$$

$$d_{jt} = \left\{ \begin{matrix} 1 \\ 0 \end{matrix} \right\} \text{ if } \left\{ \begin{matrix} j = t \\ \text{otherwise} \end{matrix} \right\}. \tag{4}$$

In equation (4) $\sum_{i=\ell}^{m} \alpha_\ell 03_{t-\ell}$ represents the lag structure to ozone. Last hour's ozone reading affects this hour's since ozone persists and dissipates non-instantaneously. Typically in estimation the α_ℓ coefficients for $\ell \geq 2$ are small and become insignificant by $m = 5$. With $m = 1$, $\alpha_\ell = 0.8 - 0.9$ in estimation, with 03's measured in logs.

The d_{jt} are dummies variables where $d_{jt} = 1$ if $j = t$ and $d_{jt} = 0$ otherwise. The β_j then measure the inferred daily cycle of ozone related activities that create changes in ozone levels, over and above undissipated lagged levels. The constant term, A, reflects the midnight infusion of ozone activity. The β_j hourly ozone activities (relative to midnight) start at 1 a.m. and run through to 11 p.m. While the β_j are intended to capture the daily cycle of socio-economic activity in the region of the monitoring station, absent hourly temperature readings their magnitudes may also partially reflect daily temperature cycles. Equation (4) represents a daily cycle. From estimated coefficients, predicted hourly values may be calculated recursively, starting at 1 a.m. and returning to midnight. For $m = 1$ (only one lag), the system is simple.[21] For $m = 4$, the equations to solve the recursive model take pages. In application, we simply simulated the process, in obtaining predicted values. For a particular $\overline{A}$, we try initial values of $03_{24}, 03_{23}, 03_{22}$, and 03_{21} and iterate recalculating all predicted hourly values until our initial $03_{24} - 03_{21}$ values converge to their predicted values.

In actual estimation I pool stations within each of the three samples (attainment, non-attainment and California) for 1982 and 1987 and estimate a pooled equation where

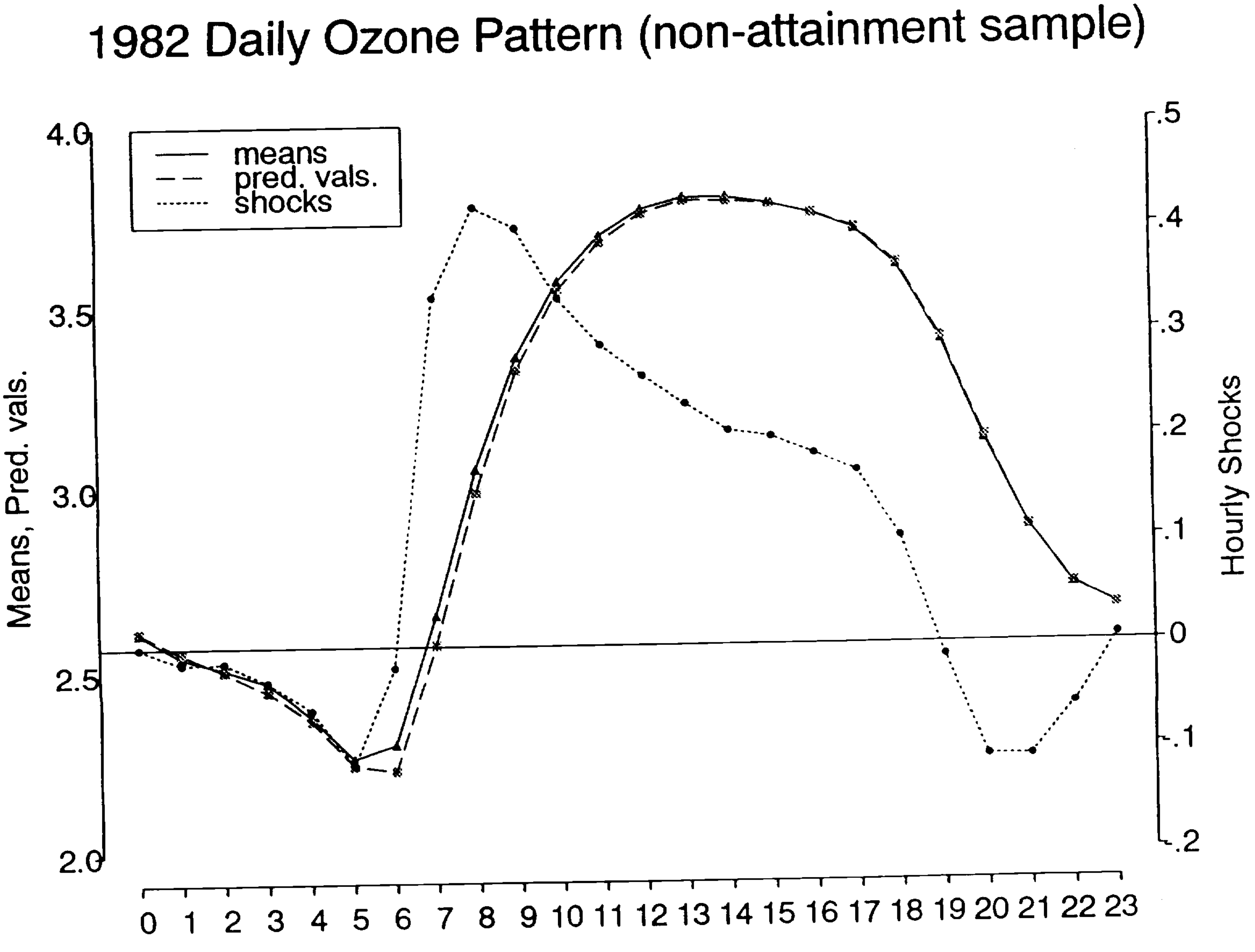
$$03_{st} = A + \sum_{i=1}^{m} \alpha_{\ell} 03_{t-\ell} + \sum_{j=1}^{23} d_{jt}\beta_j + A_{87}D_{87} + \sum_{j=1}^{23} D_{jt}\gamma_j + \ldots + \epsilon_{st} \tag{5}$$

$D_{jt} = 1$ if the year is 1987 and $j = t$; $D_{jt} = 0$ otherwise. γ_j measure any differences in hourly ozone related activity in comparing 1987 with 1982. $D_{87} = 1$ if the year is 1987 and 0 otherwise; so A_{87} represents any change in the constant term. In actual estimation, as explained in Appendix B, as additional control variables, there are dummy variables for each monitoring station and additional variables describing employment levels and July temperature conditions for each station for 1982 and 1987. Finally in estimation I also distinguish 2 daily cycles — one for Monday to Friday and one for the weekends. In the text I focus on Monday to Friday. Below, I will comment on specific estimation issues such as serial correlation (which I reject) and on some other results in Appendix B. For now I focus on the basic point.

The model is estimated by OLS for the three samples of stations from Figure 3. Sample sizes in terms of hourly readings range from 81,143 to 429,765. Level variables including 03 readings are in logs. My focus is on the daily pattern of hourly economic activity versus daily pattern of ozone readings.

Figure 4 illustrates the basic phenomenon. The right-hand vertical axis plots hourly shocks. Relative to midnight, these decline to 5 a.m., rise sharply with the morning commuting rush hour peaking at 8 a.m., and then gradually decline over the course of the day to 5 p.m. After 5 p.m. they plummet bottoming out around 8 p.m., and then rising modestly to midnight ("grave yard shift"?). There is no second peak for the late afternoon rush hour corresponding to the 8 a.m. morning rush hour peak. For example, for commuting, return journeys are spread out — part-time worker return mid-day to early afternoon, school chil-

FIGURE 4

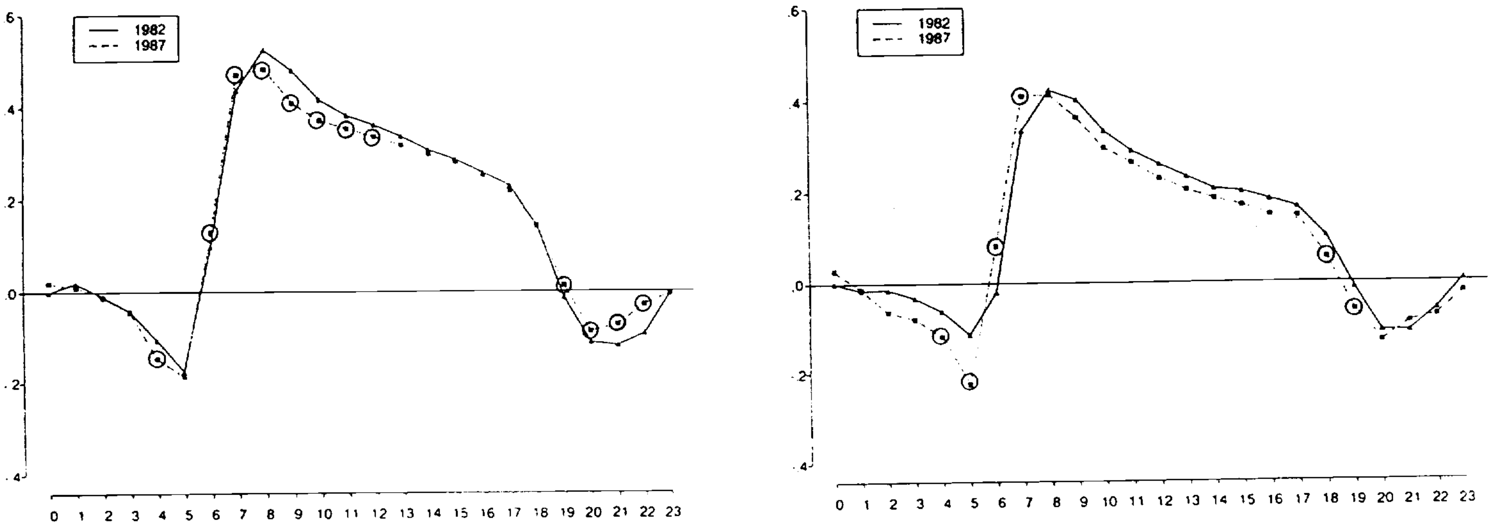

1982 Daily Ozone Pattern (non-attainment sample)
means
pred. vals.
shocks
Means, Pred. vals.
4.0
3.5
3.0
2.5
2.0
Hourly Shocks
.5
.4
.3
.2
.1
0
-.1
-.2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

dren around 3 p.m., and full-time workers over a variety of times depending on daily work and social demands.

In Figure 4, the pattern is clear. Daily economic activity peaks at 8 a.m., well before the ozone peak. The left-hand vertical axis represents actual and predicted values of 03. Because of lagged ozone accumulations from the hourly shocks, ozone peaks at 1-2 p.m., or 5-6 hours after ozone creating activity peaks at 8 a.m. The key to reducing ozone peaks is to dampen the critical 7-10 a.m. economic activity which induces later ozone accumulations.

Figure 5 plots the cycle of economic activity for 1982 versus 1987, for the three samples, based on the coefficients in Table B2 of Appendix B. 1987 coefficients are the sum of 1982 coefficients plus the 1987 differential hourly dummy. 1987 data points based on statistically significant 1987 differential dummies are circled. These 1987 differential hourly dummies which are statistically significant are also listed in Table 5 for the relevant critical hours. The discussion focuses on Table 5 for specific magnitudes and Figure 5 for a visual impression. The California case is very clear. Economic activity rescheduled away from the 7 a.m. to 10 a.m. time frame, and more towards the evening time frame. Moreover in California, hourly economic activity for the same total employment (the control variables) falls over much of the day between 1982 and 1987. In non-attainment areas the 8-10 a.m. shocks decline substantially relative to 1982 (by a total of 16 percent points). This occurs with some modest rescheduling towards the earlier morning hours (6 and 7 a.m.), as well as substantial rises in the evening. While the key hourly shocks declined, population growth between 1982 and 1987 in these areas kept overall ozone levels high. Thus, back in Figure 3, overall ozone readings rise in non-attainment areas, but the decline in hourly shocks keep the peak ozone readings the same from 1982 to 1987. For attainment areas, the key issue concerns the big rise in the (6 a.m. and) 7 a.m. shock which feeds into mid-day ozone accumulations, with no corresponding significant fall in shocks from 8 a.m.-11 a.m. This accounts for the increased peak readings in attainment areas in Figure 3.

FIGURE 5 -- DAILY CYCLE OF ECONOMIC ACTIVITY


Hourly Shocks to Ozone (non-attainment sample)

Hourly Shocks to Ozone (attainment sample)

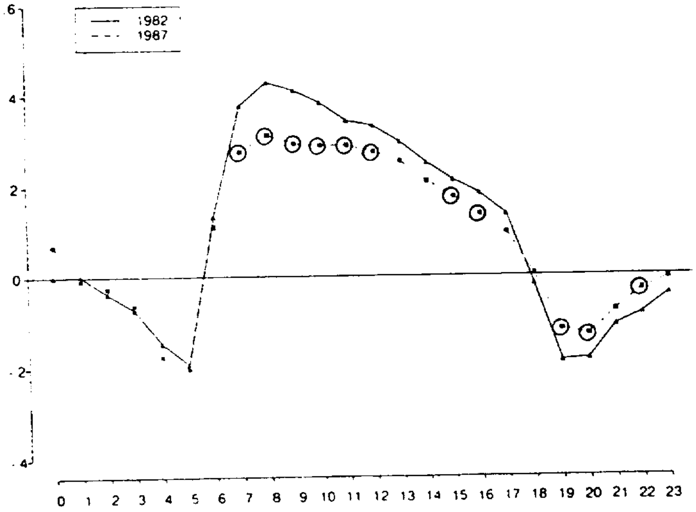

Hourly Shocks to Ozone (California sample)

Table 5

1987 Significant Hourly Shock Differentials

	Non-Attainment	Attainment	California
6 a.m.	.030	.102	-.023
7 a.m.	.034	.075	-.102
8 a.m.	-.042	-	-.115
9 a.m.	-.070	-	-.117
10 a.m.	-.044	-	-.094
11 a.m.	-.028	-	-.054
noon	-.025	-	-.058
7 p.m.	.027	-.049	.068
8 p.m.	.024	-	.053
9 p.m.	.047	-	-
10 p.m.	.065	-	.052

Other Issues. OLS estimation of equation (4) raises issues of serial correlation. For individual stations, I tested for serial correlation for the small sample of stations with almost complete sets of July readings.[22] Based on Breusch-Godfrey tests (χ^2 tests), for $m = 4$, I could reject serial correlation of order 1 for 73% of the 37 stations. Given this lack of strong evidence of serial correlation, the model is estimated by OLS. This makes it feasible to pool stations in equation (5) where, on average, the monthly time series for a station is only 85% complete.

In terms of other results, the weekend daily cycles differs from the weekday cycle (compare Tables B1 and B2). Shocks for 8 a.m.-11 a.m. are significantly smaller, while evening shocks are higher. Between 1982 and 1987 there is little change in weekend hourly patterns, except to reduce some evening shocks (a VCR effect?).

Conclusions

Local air quality regulation matters. A switch in county attainment status to non-attainment induces greater regulatory effort and results in cleaner air. Pro-environment states which spend relatively more on pollution abatement have cleaner air. Polluting industries tend to relocate over time to areas with a record of staying in attainment, so as to avoid regulatory scrutiny and effort. Finally, since the ozone air quality standard involves a single peak hour reading, localities can achieve attainment by reducing the hourly ozone peaks during a summer day, without reducing mean or median readings.

Footnotes

1. Code of Federal Regulations, title 40, part 81, subsection C.

2. The pictured distributions are based on a Kernel estimator. The Kernel is Gaussian and the window width (to which results are here insensitive) is the range covered by the middle 50% of observations.

3. All the readings in 1982 and 1987 above 0.18 are from L.A. and neighboring counties. By a t-test, the 1977 mean exceeds the 1987 mean; by an F-test the 1977 variance exceeds the 1987 variance; and by a Wilcoxon-Mann-Whitney test the 1977 median exceeds the 1987 median.

4. However, there is no 1976 or 1977 data on attainment status, so the base period value is for 1978).

5. For the median of July maxima and the mean, the coefficients and standard errors are -.016 (.027) and -.051* (.025).

6. The fixed effects model was estimated for the 643 ozone stations, with days recorded as the dependent variable. RHS variables include all those in Table 1, including the index of non-attainment for other air quality dimensions, plus MSA manufacturing employment and the second highest daily reading. Time dummies are significant (increasing with time) and so is the index of non-attainment for other air quality dimensions (with a positive expected sign). But the powerful explanatory variable is the second highest daily reading, with an elasticity for days recorded of 1.17.

7. The first stage OLS equation only has an R^2 of .14; to raise the explanatory power of such an equation (of .3 to .35) I would need to insert the (endogenous to equation (1)) annual pollution readings.

8. For nine degrees of freedom, the χ^2 value is 13.4, leading to acceptance of equal coefficients under the two estimators (with and without 1985), given the critical value is 15.9.

9. I drop 65 observations where air quality improved by 1 1/2 standard deviations (1977 measure) between 1977 and 1987 but the county remained in non-attainment. All attainment status coefficients remain negative. For the July and annual measures their absolute values fall by 9–30%.

10. For dependent variables I average mean air quality measures across stations in the same county in a year and picked the median for the air quality measures based on annual maxima or July medians of maxima. I then reran the regressions in Table 1 for the new fixed effect specification (the fixed effects applying to counties, not monitoring stations).

11. Explanatory variables in stage one included county land area, population, and percent urban in 1980, MSA total employment and employment in polluting industries in 1977 (or 1982), and a coastal location dummy. Coefficients on non-attainment status for the four dependent variables in Table 1 were respectively -.017, -.051, -.097, and -.106. The latter two are significant at the 5% level.

12. Current Industrial Reports, U.S. Department of Commerce, *Pollution Abatement Costs and Expenditures.* Annually 1977-1986.

13. $\overline{RES}_i$ in equation (2) for this table is based on an original formulation in Table 1, where the A.Q. index is entered as a current, rather than a lagged variable. Since that case is almost indistinguishable from Table 1 results, there should be no qualitative impact on Table 2.

14. These are industrial organic chemicals and miscellaneous plastics, with coefficients of

about .08.

15. The wage variable was only negative and significant for one industry and the price variable was only positive and significant for two industries. Using time dummies to control for annual price and demand effects tended to swamp equations, with insignificant time dummies (except for 1983, and/or miscellaneous plastics).

16. For the industries in Table 5, in order listed, the average number of plants for counties having the industry and the maximum number are respectively (3.5, 44), (5.2, 31), (3.5, 48), (16.2, 730), and (4.1, 73).

17. Consider for example, industrial organic chemicals. With two-stage Tobits, a switch to non-attainment status (as an immediate effect) reduces the number of plants by 22% versus 6% in ordinary Tobits. Or in a Chamberlain logit, two-stage estimates suggest a switch to attainment status increases the immediate probability of having the industry in your county from .25 to .71, whereas ordinary Chamberlain logits have the probability rising to about .35.

18. Exogenous variables include time invariant geography features and supposedly exogenous time variant features (e.g., employment) of counties other than the own in the metro area. (Predetermined variables in contrast would include own county variables in all years before the own year.)

19. A 1982 (March employment) shock could affect 1982, 1983 and 1984 attainment status, leaving "clean" defined on 1985-87 unaffected by the 1982 shock.

20. In the base year size $s_{j0} = \alpha_0 + \beta Y_{j0} + f_j + \epsilon_{j0}$ and in the final year $s_{jT} = \alpha_T + \beta Y_{jT} + f_j + \epsilon_{jT}$. If $s_{j0}, s_{jT} > 0$, the event in the LLF is $\Delta s_{jt} = \alpha + \beta \Delta Y_{jt} + \Delta \epsilon_{jt}$. If $s_{j0} = 0$ and $s_{jT} > 0$, the event is $\alpha + \beta \Delta Y_{jt} + \Delta \epsilon_{jt} > s_{jT}$. If $s_{j0} > 0$ and $s_{jT} = 0$ the event is $\alpha + \beta \Delta Y_{jt} + \Delta \epsilon_{jt} < -s_{j0}$.

21. In particular $03_\ell = \overline{A} + \alpha_1 03_{24} + \beta_1; 03_2 = \overline{A} + \alpha_1 03_1 + \beta_2 (= A + \alpha_1(03_{24} + \beta_1) + \beta_2)$, and so on. Then $03_{24} = \overline{A} + \alpha_1 03_{23} = [\overline{A}(1 + \alpha_1 + \alpha_1^2 + \alpha_1^3 + \ldots + \alpha_1^{23})] + [\alpha_1^{23}\beta_1 + \alpha_1^{22}\beta_2 + \ldots + \alpha_1\beta_{23}] + 03_{24}\alpha_1^{24}$. Equating 03_{24} and 03_{24} for a given value of $\overline{A}$ we get the initial 03_{24} for a given $\overline{A}$, or $03_{24} = \overline{A}\frac{1}{1-\alpha_1} + [\alpha_1^{23}\beta_2 + \ldots + \alpha_1\beta_{23}](1 - \alpha_1^{24})^{-1}$.

22. The sample is restricted to monitors with three or fewer missing readings out of 744 for the month (only 1 monitor has all 744 readings for a month). For missing readings I use the monthly average for that hour of the day. In terms of higher order serial correlation, Breusch-Godfrey tests rejected joint serial correlation of orders 1, 1-2, 1-3, and 1-4 for over 50% of the stations.

References

Atkinson, S. and T. Tietenberg (1987). "Economic Implications of Emissions Trading Rules for Local and Regional Pollutants," *Canadian Journal of Economics,* 20, 370-386.

Bartik, T. J. (1988). "The Effects of Environmental Regulation on Business Location in U. S.," *Growth and Change,* 19, 22-44.

Carlton, D. (1983). "The Location and Employment of New Firms," *Review of Economics and Statistics,* 65, 446-449.

Chamberlain, G. (1980). "Analysis of Covariance with Qualitative Data," *Review of Economic Studies,* 47, 225-238.

Crandall, R. W. (1985). *Controlling Industrial Pollution,* Brookings Institute: Washington.

Deily, M. E. and W. B. Gray (1991). "Enforcement of Regulations in a Declining Industry," *Journal of Environment Economics and Management,* 21, 260-274.

Duffy-Deno, K. T. (1992). "Pollution Abatement Expenditures and Regional Manufacturing Activity," *Journal of Regional Science,* 32, 419-436.

Gray, W. and R. Shadbegian (1993). "Environmental Regulation and Manufacturing Productivity at the Plant Level," NBER Working Paper No. 4321.

Heckman, J. J. (1981). "The Incidental Parameters Problem" in *Structural Analysis of Discrete Data with Econometric Applications,* C. F. Manski and D. McFadden (eds.), pp. 179-95, MIT Press.

Henderson, J. V. (1994a). "Externalities and Industrial Development," NBER Working Paper No. 4730.

———.(1994b). "State Attitudes Towards Air Quality Regulation," Brown University mimeo.

———.(1995). "Industrial Development in Cities," *Journal of Political Economy,* forthcoming.

Herzog, H. W., Jr. and Alan M. Schlottmann (1991). *Industrial Location and Public Policy,* University of Tennessee Press.

MacAvoy, P. (1987). "The Record of the EPA," in R. L. Gordon, et al. (eds.), *Energy Markets and Regulation,* Ballinger.

McConnell, V. and R. Schwab (1990). "Industry Location Decisions," *Land Economics,* 66, 57-81.

Portney, P. R. (1990a). "Policy Watch: Economics and the Clean Air Act," *Journal of Economic Perspectives,* 4, 173-182.

———.(1990b). "Air Pollution Policy," in P. R. Portney, et al. (ed.), *Public Policy for Environmental Protection,* RFF: Washington.

Roumasset, J. A. and K. R. Smith (1990). "Exposure Trading," *Journal of Environmental Economics and Management,* 18, 276-291.

Russel, C. S. (1990). "Monitoring and Enforcement," in P. R. Portney, et al. (ed), *Public Policy for Environmental Protection,* RFF: Washington.

Tietenberg, T. H. (1985). *Emissions Trading,* RFF: Washington.

Appendix A. Data Sources

Air Quality Measures. The EPA provides the hourly readings at all monitoring stations around the country for each year. I obtained these data tapes for 1977, 1982, 1985, and 1987; and on those tapes drew a sample which required (i) that the station appear at least in two years and (ii) the station be in an urban county. These restrictions were imposed so I could construct (unbalanced) panels and could relate air quality measures to county employment activity. The urban county restriction is almost costless. Almost all stations are in urban counties, and, in fact, even within urban counties only the more heavily polluted are regularly monitored. My data covers just 332 of the 742 urban counties. The requirement that stations operate at least twice loses half of reporting stations because the EPA samples all over the country with temporary stations. So a cleaner county may be monitored one year out of ten. Nevertheless I feel my coverage is good. The basic sample covers 643 monitoring stations (i.e., 643 panels), which is more monitoring stations than the EPA bases its trend statistics on in its annual reports. Coverage increases over time. There are only 187 stations in 1977, which jump to 533 in 1982 and 544 in 1987. 124 stations appear in all four years, 330 in three years (mostly starting in 1982) and 189 in two years.

Employment Data. Employment data are for 1990 urban counties from County Business Patterns for the years 1977-87. Data was kindly given to me by Bill Miracky and originally came from the Center for Governmental Study at Northern Illinois State University. These data were used to construct total and individual industrial employment measures, covering all civilian employment.

Other Data. Temperature measures are from the monthly weather station tables, published by the U.S. National Oceanic and Atmospheric Administration. Land, population, and related measures are from various City and County Data Books (U.S. Department of Commerce). Coastal location is from Rand McNally maps. Pollution abatement expenditures are from the annual **Pollution Abatement Costs and Expenditures** published by the Bureau of the Census Current Industrial Reports. County attainment status are from microfiche of the Federal Register.

Except for City and County Data Book information on tape, all data are from hard copy sources and needed to be entered by hand.

Appendix B. Hourly Ozone Reading

There are data potentially for 744 hourly readings for each station for the month of July each year. Due to missing hourly readings (for both the dependent and lagged independent variables), each station averages 630 readings in actual estimation. Readings at different stations are based on different calibrations and scales which require conversions. Instruments are calibrated so, if recorded readings are the set of non-negative integers a reading of 60 is a reading between 59.5 and 60.49. A reading of zero is possible but less than 4% of the readings are zeros. Consistent with the calibration of all other non-negative readings, we give zero readings a value of 0.25 of the minimum non-zero reading (for the set of integers, zero takes a value of 0.25). Tobit estimates were essentially identical to OLS estimates.

In Table B1 we report the coefficients for the full set of variables for the largest sample, other than M-F hourly dummies for the non-attainment sample, coefficients for lagged dependent variables, and monitor dummies. Besides employment and temperature data and their interactive terms, it is possible to interact the hourly dummies and lagged dependent variables with temperature data (e.g., July temperature differentials) to show crudely that all parts of ozone process are temperature sensitive. Our basic results on hourly shocks still stand. In Tables B2 and B3 for three samples, I report the results on hourly dummies (plotted in Figure 5) and lagged dependent variable coefficients.

Table B1

constant	-3.8576 (2.215)
1987 year dummy	.0200* (.0090)
ln (avg. July Temp.)	.8866** (.4942)
July temp. diff.[a]	-5.5077** (3.165)
ln (MSA employ)	.3153* (.1551)
ln (MSA employ) * ln (July avg. temp.)	-.0661** (.0340)
ln (July avg. temp.) * July temp. Diff.	1.6602* (.6770)
ln (MSA employ) * July temp. diff.	-.1233* (.0609)
Sat Sun dummy	.0076 (.0111)
Sat Sun Differential Hourly Dummies	
1	-.0181 (.0159)
2	.0335* (.0158)
3	.0311* (.0158)
4	.0938* (.0157)
5	.1822* (.0156)
6	.1580* (.0154)
7	-.0042 (.0154)
8	-.0780* (.0154)
9	-.0593* (.0154)
10	-.0245* (.0155)
11	-.0076 (.0155)
12	-.0096 (.0155)
13	-.0055 (.0155)
14	.0017 (.0155)
15	.0168 (.0154)
16	.0199 (.0154)
17	.0115 (.0154)

---

[a] July temperature differential is (July average daily maximum temperature - July average temperature) / July average temperature.

Sat Sun Differential Hourly Dummies (Continued)

18	.0178 (.0154)
19	.0364* (.0153)
20	.0422* (.0153)
21	.0497* (.0154)
22	.0431* (.0155)
23	.0040 (.0155)
1987 S.S. dummy	.0108 (.0162)

1987 S.S. differential dummies (relative to basic S.S. diff. hourly dumies)

1	.0239 (.0232)
2	-.0248 (.0232)
3	.0087 (.0229)
4	.0143 (.0028)
5	.0274 (.0227)
6	-.0130 (.0224)
7	-.0154 (.0224)
8	-.0073 (.0224)
9	.0243 (.0224)
10	.0065 (.0224)
11	-.0101 (.0224)
12	-.0110 (.0224)
13	-.0176 (.0225)
14	-.0217 (.0224)
15	-.0270 (.0223)
16	-.0254 (.0223)
17	-.0176 (.0223)
18	-.0160 (.0222)
19	-.0155 (.0222)
20	-.0133 (.0221)
21	-.0440* (.0223)

1987 S.S. differential dummies (to basic S.S. diff. hourly dumies) - (Continued)

22	-.0730* (.0226)
23	-.0203 (.0227)

Table B2

	Non-Attainment		Attainment Areas		California	
	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential
H1	0.0198*	-0.0092	-0.0154	0.0026	0.0024	-0.0095
	(0.0087)	(0.0124)	(0.0186)	(0.0258)	(0.0150)	(0.0224)
H2	-0.0103	-0.0022	-0.0141	-0.0500*	-0.0377*	0.0122
	(0.0087)	(0.0124)	(0.0187)	(0.0259)	(0.0147)	(0.0222)
H3	-0.0422*	-0.0020	-0.0327**	-0.0477**	-0.0729*	0.0092
	(0.0087)	(0.0122)	(0.0188)	(0.0262)	(0.0145)	(0.0213)
H4	-0.1071*	-0.0402*	-0.0621*	-0.0564*	-0.1465*	-0.0288
	(0.0086)	(0.0121)	(0.0184)	(0.0259)	(0.0143)	(0.0211)
H5	-0.1750*	-0.0111	-0.1154*	-0.1087*	-0.1920*	-0.0098
	(0.0085)	(0.0120)	(0.0177)	(0.0247)	(0.0143)	(0.0212)
H6	0.0993*	0.0301*	-0.0201	0.1024*	0.1321*	-0.0234
	(0.0085)	(0.0119)	(0.0175)	(0.0244)	(0.0142)	(0.0209)
H7	0.4364*	0.0344*	0.3349*	0.0746*	0.3774*	-0.1015*
	(0.0085)	(0.0119)	(0.0174)	(0.0243)	(0.0143)	(0.0209)
H8	0.5241*	-0.0415*	0.4221*	0.0099	0.4277*	-0.1148*
	(0.0086)	(0.0120)	(0.0175)	(0.0243)	(0.0144)	(0.0209)

Table B2 - (Continued)

	Non-Attainment		Attainment Areas		California	
	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential
H9	0.4791*	-0.0702*	0.4016*	-0.0391	0.4099*	-0.1169*
	(0.0087)	(0.0121)	(0.0178)	(0.0245)	(0.0146)	(0.0210)
H10	0.4156*	-0.0440*	0.3336*	-0.0385	0.3830*	-0.0937*
	(0.0088)	(0.0121)	(0.0180)	(0.0246)	(0.0147)	(0.0210)
H11	0.3816*	-0.0284*	0.2888*	-0.0257	0.3434*	-0.0543*
	(0.0089)	(0.0122)	(0.0179)	(0.0247)	(0.0148)	(0.0210)
H12	0.3615*	-0.0254*	0.2588*	-0.0305	0.3323*	-0.0577*
	(0.0088)	(0.0122)	(0.0179)	(0.0247)	(0.0147)	(0.0210)
H13	0.3366*	-0.0198	0.2317*	-0.0293	0.2965*	-0.0425*
	(0.0088)	(0.0122)	(0.0179)	(0.0247)	(0.0147)	(0.0209)
H14	0.3061*	-0.0088	0.2045*	-0.0211	0.2503*	-0.0404*
	(0.0088)	(0.0121)	(0.0178)	(0.0246)	(0.0146)	(0.0207)
H15	0.2847*	-0.0057	0.1985*	-0.0291	0.2122*	-0.0358**
	(0.0087)	(0.0120)	(0.0178)	(0.0245)	(0.0145)	(0.0206)
H16	0.2561*	-0.0054	0.1818*	-0.0341	0.1833*	-0.0465*
	(0.0086)	(0.0119)	(0.0176)	(0.0243)	(0.0145)	(0.0206)

Table B2 - (Continued)

	Non-Attainment		Attainment Areas		California	
	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential	Hourly Shocks	1987 Differential
H17	0.2281*	-0.0115	0.1650*	-0.0197	0.1362*	-0.0408*
	(0.0086)	(0.0119)	(0.0175)	(0.0242)	(0.0144)	(0.0205)
H18	0.1426*	0.0004	0.1017*	-0.0479*	-0.0182	0.0228
	(0.0085)	(0.0118)	(0.0175)	(0.0241)	(0.0142)	(0.0204)
H19	-0.0140**	0.0270*	-0.0121	-0.0494*	-0.1864*	0.0683*
	(0.0084)	(0.0117)	(0.0173)	(0.0240)	(0.0142)	(0.0204)
H20	-0.1140*	0.0244*	-0.1107*	-0.0217	-0.1822*	0.0530*
	(0.0084)	(0.0117)	(0.0173)	(0.0239)	(0.0141)	(0.0203)
H21	-0.1208*	0.0470*	-0.1110*	0.0211	-0.1090*	0.0321
	(0.0084)	(0.0118)	(0.0172)	(0.0239)	(0.0144)	(0.0207)
H22	-0.0942*	0.0645*	-0.0602*	-0.0148	-0.0833*	0.0522*
	(0.0084)	(0.0119)	(0.0171)	(0.0239)	(0.0146)	(0.0215)
H23	-0.0047	-0.0012	0.0059	-0.0284	-0.0414*	0.0347**
	(0.0084)	(0.0119)	(0.0172)	(0.0240)	(0.0146)	(0.0215)

Table B3

lagged dependent variables	Non-Attainment Areas	Attainment Areas	California
α_1	.9467*	.9323*	.8301*
	(.0015)	(.0035)	(.0032)
α_2	-.0979*	-.0804*	.0413*
	(.0021)	(.0048)	(.0042)
α_3	.0044*	.0074	-.0147*
	(.0021)	(.0048)	(.0042)
α_4	-.0280*	-.0030	-.0613*
	(.0015)	(.0036)	(.0032)