

ENERGY LABORATORY
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

REGULATING THE AUTOMOBILE

by

John B. Heywood
Henry D. Jacoby
Lawrence H. Linden
Howard Margolis
David Iverach
E. Allen Jacobs
Frank Lerman
Michael K. Martin

M.I.T. Energy Laboratory Report 77-007

November, 1977

Based on Research Supported by

DIVISION OF POLICY RESEARCH AND ANALYSIS
NATIONAL SCIENCE FOUNDATION
Under Grant No. OEP 76-00284

Preface and Acknowledgments

This report is the result of an effort initiated at the M.I.T. Energy Laboratory in September, 1975. It is based upon research supported by the National Science Foundation under Grant No. OEP 76-00284. The principal investigators on the effort were John B. Heywood and Henry D. Jacoby. Lawrence Linden also contributed significantly to the conception and the leadership of the project. The interactions between the leadership of the project and the other participants, and among the participants themselves, have been important to each individual effort. However, we have identified as authors the individual(s) that performed the bulk of the research and writing for each individual chapter.

Many individuals from government and industry contributed their time and ideas in discussions with the authors; we gratefully acknowledge their assistance. We have also benefitted from the sympathetic monitoring of this project at the National Science Foundation. Dr. Leonard Topper helped initiate the effort at the National Science Foundation's Office of Energy R&D Policy, and over the last year it has been monitored at the Division of Policy Research and Analysis by Dr. Ronald M. Powell; we appreciate the assistance rendered us by both.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

Preface and Acknowledgements	i
Table of Contents	ii
Introduction and Summary	iv
Chapter 1. The American System of Automotive Regulation, by Lawrence H. Linden and David Iverach	1-1
1. Introduction	1-1
2. A Brief Review of the Evolution of the Automotive Regulatory System	1-4
3. Issues of Standard-Setting and Technological Change	1-10
4. Issues of Behavior and Measurement	1-37
5. Summary and Conclusions	1-47
Chapter 2. Uncertainty in Automotive Emissions Regulations, by Michael K. Martin	2-1
1. Introduction and Overview	2-1
2. Uncertainties in Measuring and Predicting Aggregate Emissions	2-9
3. Uncertainties in Measuring and Predicting Trends in Air Quality	2-22
4. Analysis of Air Quality Trends at Kenmore Square, Boston	2-68
5. Summary and Conclusions	2-91
Chapter 3. Average Fuel Economy Standards and the Automobile Industry, by E. Allen Jacobs and Lawrence H. Linden	3-1
1. Introduction	3-1
2. The Energy Policy and Conservation Act	3-4
3. Evaluating Potential Industry Responses	3-7
4. Profit-Maximizing Strategy: Violating the Standards	3-9
5. Pricing to Meet the Standard	3-18
6. Impact of the Law Upon the Automotive Industry	3-22
7. Possible Solutions	3-26
8. Summary	3-29

Chapter 4. Standards versus Taxes: The Political Choice, by Frank Lerman	4-1
1. Methods of Control--Standards vs. Taxes	4-3
2. Alternate Theories of Societal Choice	4-6
3. Conclusion	4-29
Chapter 5. Another View of the Politics of Auto Emissions Control, by Howard Margolis	5-1
1. A Conceptual Framework	5-3
2. Reviewing the History	5-8
3. Who Got What	5-20
4. What Went Wrong	5-34

INTRODUCTION AND SUMMARY

Year after year, it seems, federal regulation of the automobile is a subject of public controversy and legislative debate. No one should be surprised that this is so. We are dealing with a sophisticated, mass-produced machine which is fueled by a complex chemical. It is operated under diverse road and climatic conditions by motorists with a wide variety of living and driving habits, and serviced by mechanics who range from the expert to the incompetent and unscrupulous. The side effects we try to control are large: Among the technical devices of modern society the automobile has emerged as the largest source of air pollution, the largest consumer of fuel, and the largest cause of accidental death. Yet the vehicle itself is not the sole cause of any of these problems. Thus legislative and administrative action must be taken in the face of imperfect knowledge of the cause-effect relationships between improvements in cars and safer travel, healthier air, and lowered national energy use.

Moreover, the industry being regulated is one of the largest and most highly concentrated in the country: GM, Ford, and Chrysler are three of the top 10 firms in the "Fortune 500." The stakes are tremendous in terms of economic cost and employment impact, and the industry response to different forms of regulation may be difficult to foresee.

The resulting problems are evident in the current regulatory framework. As late as July 1977 the auto manufacturers were preparing to turn out model year 1978 vehicles which, under current legislation, could not be sold. Naturally, the law will be changed, even if at the last

minute. But it is not clear that the regulatory framework that produces these confrontations will be fundamentally improved. Indeed, in the face of these difficult circumstances, there very likely is no fundamental restructuring of the regulatory system which would easily resolve all the problems being encountered. There simply is no solution which can resolve all the difficulties, including the political ones, at the same time.

This is not to say that dramatic solutions are not suggested. There is a large body of work on the use of selective financial instruments--such as gasoline taxes, effluent charges, and taxes and subsidies based on fuel economy--to control automotive impacts. No doubt in terms of economic efficiency and administrative simplicity, these market instruments would work pretty much as their proponents argue. But these measures have never come close to political acceptance, and it is not evident that they ever will.

Others argue that the answer lies in the establishment of a single administrator or commission, with the power to choose regulatory methods, plan and establish future standards, and coordinate the various functions now spread among the EPA, DOT, and ERDA. Such a change might lead to improved tradeoffs among conflicting objectives of regulation, but there are many problems that centralization would not solve, even if regulatory authority could be so concentrated. The role of the Congress in the details of regulation would remain, and the fundamental societal conflicts (which lie behind so many of the problems) cannot be ameliorated by any governmental reorganization.

So there is no master stroke which will lead to well-informed, efficient, and equitable regulation of this sector. What we may look for

are marginal improvements which will produce better knowledge of consequences, help avoid some of the gross waste, and yield more equitable tradeoffs among the several competing objectives. The purpose of this study is to contribute to the search for these improvements by developing a better understanding of how our current system works, how it evolved, and why.

We have approached this task by exploring selected aspects of three questions about the automotive regulatory system:

- What do we do? Our current system evolved from three independent and largely uncoordinated legislative efforts. How does one develop a coherent framework for thinking about the system, and for analyzing potential changes?
- What effects does it have? The regulations are supposed to improve air quality, safety, and fuel economy. How much do we know about the results of all this effort, and what does available analysis show?
- Why do we do it this way? Analysis of improvements to the regulatory system must take account of the political context out of which these laws and administrative rulings come. What does the legislative and bureaucratic history reveal about the American political system, and its influence on the style and content of auto regulation?

Five studies were carried out in order to shed light on these questions. The resulting descriptive material, insights, and analysis have been distilled into the five chapters of this report. The intent of the report is to form a basis for more detailed and mission-oriented studies of specific changes in the structure of automobile regulation.

What Do We Do?

Chapter 1, "The American System of Regulating the Automobile, looks across the three primary areas of regulation and seeks a common framework for describing the system's structure and its effects. Several insights result from such an exercise. The regulatory systems fall into three rough categories, as shown in the accompanying table. From the top

-
1. Equipment Standards
 2. Performance Standards
 - 2a. On the vehicle
 - 2b. On a fleet average
 3. Financial Penalties and Incentives
-

of the list down, the schemes differ in that greater and greater flexibility is allowed the automobile manufacturer. Though often proposed, financial penalties are not an important part of the current regulatory system (why they are not used is discussed in Chapter 4), and so the focus of discussion is on equipment standards (used in safety regulation), and performance controls on the vehicle (emissions) and on the fleet (fuel economy).

The choice of regulatory form at this level determines the character of the detailed aspects of standard-setting and implementation. In the case of safety, it has not proved possible to define adequate standards which apply to the vehicle as a whole (though work goes on in this area), so the regulations consist of requirements for devices or design features which the regulatory agency can demonstrate are safer. In the case of emissions and fuel economy, on the other hand, it is possible to define numerical measures of performance (grams of pollutant emitted per mile, miles per gallon of fuel consumed) which can be related directly to national goals. Thus one determinant of differences in regulatory form is essentially a technical one: it may simply be impossible to define a reasonable performance standard for a whole vehicle or for the fleet.

If performance standards are feasible, then another possibility opens up. The regulations can require performance beyond that attainable with existing technology (as opposed to an equipment standard, where regulators must prove the device works before they can mandate its use). This the Congress did in the air pollution area, and by going into such numerical detail, the Congress was able to establish a set of targets far stricter than could have been justified by an administrative agency working under general Congressional guidelines.

Thus in the area where numerical performance standards can be set, it has been possible for the Congress not only to regulate the manufacturers but to try to "force" technological change. Judged by the emissions goals set in the 1970 Amendments to the Clean Air Act, it has not worked. On

1978 models, emissions are still far above the limits set for the 1976 model year, and it is not clear that the standards set in 1970 will ever be achieved. In this circumstance the system has shown a crude feedback mechanism. Time after time automobile manufactures have proved unable to meet the legislated deadlines, within acceptable cost, and fuel economy penalties, and the statutory standards naturally have been postponed. Considerables pressure has been kept on the industry to improve emissions performance, however, and the resulting industry investments in control technology have been large. And, in fact, emissions of new vehicles are substantially lower now than they were in 1970.

In the case of fuel economy, a similar approach has been followed. An almost arbitrary set of performance goals has been set, with penalties if industry fails to meet them. In the fuel economy legislation, the controls are on fleet-average consumption, so the manufacturers have more flexibility than in the emissions case. For example, if technology improvement does not bring fuel consumption down as much as needed, the companies can take measures to shift the fleet mix toward smaller cars. Whether the legislated targets will be met remains to be seen. Clearly this approach provides all the necessary incentives to industry investment in new technology--though as noted in Chapter 4, the incentives may prove much more complex than appears on the surface.

Given that we have chosen performance standards for emission control and fuel economy, another problem arises--that of durability. Proof of performance is established by tests on factory prototype vehicles, and these may not bear a close relation to actual vehicle behavior on the road given all the variation in operating conditions. The deterioration problem may

not be too bad for fuel economy, but for emissions control it is very serious. If equipment standards were used, then this problem might be ameliorated: the federal government could mandate control devices that would not deteriorate, and could not be disconnected by the motorist. The path has not been followed in any but the safety area, and probably this is for the best. The technology is simply not now well enough established to make such specific design controls reasonable.

Of course, many of these problems--the safety issue, performance of emissions controls, and fuel economy--could be greatly reduced if the government could but influence the driving and maintenance behavior of the motorists themselves. Given that this has not proved possible--and probably won't in the future--the regulators of the automobile will continue to face these fundamental choices. One can set performance standards, but it is very difficult to enforce them on the road. To get controls that stand up on the road may require detailed design standards, but then the government must take over the role of technology development and the task of proving that designs are "necessary" and "practicable."

Thus there seems to be no sweeping change in regulatory structure that would yield vast improvement over the way we do it now. On the other hand, a review of the history indicates that there should be a continuing monitoring of the regulatory structure in search for those areas handled by performance standards that might better be tied down by a design standard or vice versa. As technology matures and new data are gained, the relative emphasis on insuring durability (as opposed to pushing technology) will shift, and the regulations should very likely change with them.

What Impact Does It Have?

No doubt the automotive regulatory system has resulted in some changes, and will produce others. New cars have catalytic converters, head restraints, and other devices which are unlikely to have appeared without legal coercion. Cars are getting smaller, major bureaucracies have been established--both in government and in industry--to ensure that regulations are developed and complied with. Some factories are closing, others opening; jobs are being created in some areas and wiped out in others.

Given that the stakes are so large, one would hope for the best possible information about the relationship between various controls and the ultimate objectives of the regulatory system--better health, fewer deaths and injuries, and energy conservation. There needs to be a constant balancing of conflicting social objectives, and to do this in any but the most crude manner, some data must be available on effects. No doubt, many claims have been made concerning the results of federal regulation--past, present, and future. But unfortunately the facts are distinctly unconvincing. The relationship of public health to ambient air pollution levels is not well understood. Historical trends in ambient air quality are difficult to detect at all, due to the pervasive influence of meteorological fluctuations. Even more difficult is the attribution of air quality trends to reductions in emissions from the motor vehicle fleet. What is known is that very small numbers of specially-prepared new vehicles meet stringent criteria when operated by trained professionals over driving cycles designed to simulate conditions in downtown Los Angeles. As discussed earlier, uncertainties in the levels of emissions are very large due to deterioration of control equipment (due to natural causes or tampering), variations in driving patterns, differences in certification and production vehicles,

uncertainties in the contribution of non-automotive sources of pollution, and a host of other phenomena. A related but distinct problem is the forecasting of the impact on future air quality of present and future emissions standards. In this case the picture is clouded by changing traffic patterns as well as by the complex relationship of emissions to ambient air pollution levels.

The analytical difficulties in determining the past and future impact of safety are equally severe. It is very difficult to disentangle the effects of the behavior of vehicle operators and the effects of changes in speed limits and road design.

Along with the primary impacts, there are secondary effects which cannot even be well enumerated. The impact of emissions standard on fuel economy is a prominently discussed technical issue. The indirect effects of the emissions standards on vehicle sales and employment may be very important, but are poorly understood. For purposes of forecasting the impact of future standards, the behavioral response of the automobile industry--in developing and introducing new technology, in changing marketing techniques and prices, and in altering the types of vehicles sold--is crucial, but not well understood.

We have chosen to address two particular questions in an attempt to illuminate the problems in determining the effects of automotive regulation. Chapter 2, "Uncertainty in Automotive Emissions Regulations," examines the technical question of how well we can detect and forecast the impact of emission regulations on ambient air quality. If we are to know the effects of regulation, two types of issues arise, even after the vehicle designs are known. First, there is a problem of measurement of the actual state of the system at any one time. At each stage, of course, there are

uncertainties in measurement; this applies from the determination of ambient pollution levels all the way back to the imputation of fleet performance from samples of prototype vehicles.

Whatever the difficulties of measurement, they limit our ability to deal with the second batch of uncertainties--those involved in predicting changes in air quality and public health. For prediction involves the possibility of unforeseen combinations of meteorological conditions, population distribution, driving habits, etc. For example, the original setting of standards in the 1970 amendments to the Clean Air Act involved a prediction of the results, in terms of air quality, of certain performance controls on new cars. Such a calculation involves (1) the industry response in terms of new technology, (2) the translation of performance data from the test stand to the field, (3) the link from aggregate emissions to air quality, and (4) the link from air quality to population exposure and human health.

One would hope to monitor (i.e., measure) developments at these various points and feed the data back into the regulatory process as evidence that standards were too lenient or too strict, or otherwise in need of change. Unfortunately, this is very difficult to do (a difficulty which is increased by the low effort devoted to data collection), and coherent statements of effects are confounded by inherent uncertainties and the long time lags before trends can be established statistically.

In an effort to explore this problem, Chapter 2 looks in detail at several of the key measurement and prediction problems. Attention is given to the measurement and prediction of automotive emissions and to the measurement and prediction of ambient air quality. The analysis includes one experiment which attempts to quantify uncertainties in air quality--

and air quality trends--using data from a particular site in Boston. The relationship of ambient air quality to public health, and the associated measurement and prediction problems, were not examined.

Several important conclusions emerge from the analysis. The most important is this: if we want to know what is in fact happening as a result of regulation, then surveillance programs must be larger than those now in place. Moreover, we must be prepared to sustain them for many years if we hope to establish resulting trends in either aggregate emissions of ambient air quality. If we fail to do this, then we may proceed with a set of standards that may be far too tight (and therefore wasteful) or too lenient (because effects are worse than early predictions show). Current standards were based on an analytical model with almost no empirical basis whatever, and with current data and data-collection systems, there is no basis for determining whether they are anywhere near correct. The time lags are too long for any kind of tight feedback loop, but it should be possible to generate some reliable guidelines for future regulation. At this point, it would seem, there is a feedback adjustment when jobs or other severe economic damage are threatened (as with the 1978 new vehicle fleet), but no such process--or the data to support it--for the standards themselves. Needless to say, the implications in terms of vehicle cost and national energy consumption are very great.

To complement the technical study on the physical impact of emissions regulation, we studied the behavioral impact on the automotive industry of the fuel economy standards. It is reported in Chapter 3, "Average Fuel Economy Standards the the Automotive Industry." The results illustrate the complex incentive effects that regulations can have on industry behavior. The fuel economy regulations of the Energy Policy and Conservation Act are particularly interesting in this regard, because the averaging

process does in fact provide manufacturers with considerable flexibility. However, the technical details of the averaging process may give automakers incentives which were quite unforeseen by those who drafted the legislation.

The study looks at the situation, which may occur in the early 1980's, where the fleet average fuel economy standards are constraining. That is, the firms must either break the standards and pay the penalties, or take some other short-run action to meet the standards. Today, the companies are in the process of making technical alternations to their vehicles--downsizing, improving engine efficiency, etc.--to meet the standards. And as a result, new vehicle fuel economies for all sizes of vehicles, are being improved. The structure of the demand for new vehicles, by fuel economy (i.e., size), with historical pricing policies, is moving slowly toward smaller cars. The question, then, is what incentives will be faced by the manufacturers if this combination of changes does not bring them within the fleet standard.

To explore this prospect we assume that an automobile manufacturer behaves as a profit-maximizing firm under some degree of competition, and thus will act to bring marginal return in line with marginal cost in each of his vehicle lines. No doubt this is a primitive description of the automobile market, but it does give some reliable insights to the complex incentives that may be offered by the current law.

We can look at two types of response by the firm. On the one hand, the firms may be willing to violate the standard and pay the legally-established penalties. Then they will add to the marginal manufacturing cost of each car type a "marginal penalty," so it will bear its full marginal cost, including fines. The marginal penalty

is the change in the total civil penalty (before tax) incurred in violating the standard, due to the production of one incremental vehicle. Such a penalty has the desired effect of raising the competitive price of vehicles with poor fuel economy and lowering that of vehicles with high fuel economy. The level of the marginal penalty that a firm would use is calculated from the harmonic averaging procedure specified in the law.

The application of this straightforward calculation can lead to some very surprising results. We examine the case of a simple two-vehicle industry. There are circumstances where a car whose fuel economy is better than the fleet average may be assigned a positive marginal penalty, i.e., the over-all fine is increased by production of a relatively efficient car! This is due to the fact that the fine is based on the difference between the actual fleet average and the standard, but the average is calculated as the (non-linear) harmonic average. In the extreme, the penalty reduction for production of a very high fuel economy vehicle--say, a "souped-up" go-cart--is so high that the firm may profit by giving these away!

The second form of response would occur if companies decided that they must meet the standard, even though it requires new pricing schemes which encourage the sale of small cars and discourage large ones. If this is the path the industry chooses, the impact on the structure of the industry could be dramatic. Essentially, each firm would attempt to cross-subsidize the sale of low-priced smaller vehicles with high profits from high-priced large cars. The actual prices necessary are a function of the demand for vehicles, by class, and the degree of deviation of the firm's unmodified structure of sales from that it desires (i.e., just

reaching the standard). In this case the firm with the largest capacity in large cars would offer the lowest prices on small cars, establishing the price in that market. Firms with less large car production would not have such cross-subsidies available, and would take major losses. Thus, under this scenario, American Motors gets quickly driven out of business as General Motors and Ford balance off revenues among car sizes. Of course, this sort of pricing behavior is termed, in other circumstances, "predatory pricing," and is patently illegal. In this case the industry giants could rightly claim that it is necessary to meet national energy goals!

While this simple analysis of possible industry actions is highly uncertain, one inescapable conclusion is that more detailed analysis is needed of the likely effects of fuel economy regulation under various foreseeable circumstances. Overall industry impacts, through pricing and fleet mix changes, could be far more disruptive than currently foreseen.

Why Do We Do It This Way?

As has already been made clear, there are substantial problems with the present system of automotive regulation in the United States. Probably the most salient of the apparent difficulties is the inability to agree upon national automotive goals and the technical measures to meet them. It is in the very nature of the process established in the emission and fuel economy regulatory systems that regular battles over the legislated standards will occur; the standards were chosen explicitly without close attention to the technology which might be available to meet them. In the safety area the seat belt interlock and air bag debates demonstrate a similar indecisiveness, in the form of a grand-scale bargaining situation

with many powerful players and large stakes. A question crucial to any proposals to change the automotive regulatory system is just how and why the present system was chosen. We address two specific questions concerning the process by which the basic structure of the American automotive regulatory system has been laid out.

First, Chapter 4 addresses the issue of "Standards versus Taxes, the Political Choice." Probably the most important single feature of the regulatory system is that it utilizes mandatory standards applied to the particular attribute of concern, rather than a system of monetary incentives. Many analysts have claimed the superiority of "emissions taxes" over fixed standards, ever since serious analysis of environmental issues was first undertaken. In a world of economic competition, emissions charges could offer the incentive to attain economically efficient levels of control, so that the trade-off between pollution costs and costs of control is optimally made. A taxation scheme would offer a continuing incentive to install pollution control devices and modify motor vehicle designs, and, over the long run, to develop new automobile technology to replace that which evolved in an age when pollution was not important. The annual confrontations, where industry and government threaten each other with the dire economic consequences of quantum changes in emission levels, could be avoided.

However, while financial incentive schemes offer enough apparent advantages, generally they do not even appear to be part of the choice set from which policy has been made. The question of why emissions charges have been so clearly rejected is the subject of this chapter. Five important areas are examined for clues to this puzzle. First is the identification of pollution as a health problem, thus invoking the traditional

responses to health issues. In general, health issues are regulated by constraining the relevant level of activity to eliminate the health hazard, with health viewed as an all-or-nothing affair. Second is the predominance of lawyers as opposed to engineers, economists and others in the policy process. A view of pollution control as a matter of rights, to be resolved by law and the courts, and a lack of knowledge of (or faith in) the operation of the market system will tend to dampen enthusiasm for market-oriented solutions. Third, the use of tough regulatory standards as symbols in the political arena is attractive compared to taxes. Fourth, the positions of the key interest groups in influencing pollution control legislation and the commitment of those groups to particular policies and, fifth, certain other features of the Congressional system, seem to point toward outcomes which are apparently "tough" and "certain," and do not confuse revenue-raising with pollution control.

In short, there are underlying forces that have mitigated against the use of taxes, and there seems no reason to expect these to go away. Thus, one should look for marginal improvements to the current system; a radical change toward the use of financial penalties is not likely to be acceptable politically.

Of course, the issue of taxes vs. standards is a very general one, and does not deal with the particular style with which we use the standard-setting process in the United States today. In automotive emissions control the process is characterized by the establishment of standards which the industry claims are not technically possible, in order to "hold the industry's feet to the fire," followed by administrative proceedings and legislative debate to determine whether or not those standards should

be postponed. So far they have always been postponed and some suitably effective but painful interim standards implemented. The differences in air quality which would result from the differences in proposed standards are so small as to appear trivial over the subsequent five to seven years, and the difference in health and welfare benefits are highly speculative even if existent. This is a distinctly American way of setting emission standards; in other countries the accomodation is less painful, and, in the view of many, less likely to involve great debates over trivial benefits.

Chapter 5, "Another View of the Politics of Auto Emissions Control," detail at just how this particular way of regulating automotive emissions came out of the legislative process. The chapter is centered around an analysis of the players in the process, other affected-but-not-playing parties, and the costs and benefits each would feel from the various legislative possibilities faced in 1970. The analysis reveals what may be an important flaw in the Congressional policy process--one that goes well beyond the bounds of automotive regulation. In the legislative process it was assumed that those who played an important role represented a complete set of the interested parties. In this case it appeared that the environmental groups represented the gainers from stricter standards, and the industry represented the losers, with the Congress having the job of balancing these interests. The difficulty comes from the fact that the interests of these players were different in important ways from the real gainers and losers--i.e., the air-breathing and car-consuming public.

Conclusion

Although it is not the purpose of this report to develop recommendations for change in automotive regulation, we are tempted to add a few concluding observations. First, if we continue to use mandatory product standards as our regulatory instrument, then there is no revision that would make things fundamentally better. The difficult choices of environment vs. energy and consumer cost are basic to this problem, and one way or another these value tradeoffs will be worked out in the political arena. At this stage in our history, the Congress is deeply involved in the details of this type of automobile regulation and this brings with it both the great power of the Congressional will, and the sluggishness and peculiar politics of the legislative process.

Surely the process could be better informed--by more analysis of the effects of policy on both the manufacturers and their customers, and by an improved system of monitoring and surveillance of actual regulatory results. To argue otherwise is to support the spending of billions while wearing a blindfold. But the task is not easy, for the facts of weather and the mass American market make the uncertainties particularly troublesome. At any rate, a program of improved monitoring, surveillance and analysis would require several years investment before it could produce meaningful results on emission and health, or on safety.

Another issue relates to the standards themselves, and how they may be revised over time. The precise numbers written into the clean air standards have taken on a life of their own; they have a symbolic value that seemingly makes them unchangeable, even in the face of errors in their definition, clear conflicts with other goals of a much heightened priority

(i.e., fuel economy) and the fact that they may never be attained. Yet the industry lives year after year with one-or-two-year postponements, with the predictable effects on investments in research and development and technology adoption. A similar fate may befall the fuel economy standards if the crunch comes and it appears that jobs are threatened.

If numbers do take on a symbolic life of their own, then there is little to be done other than what is being done now. However, it would be very useful from the standpoint of research and development and new product planning if emission standards could be set at interim levels for some reasonable length of time (say five or six years) and, perhaps some provision made for fleet averaging.

Finally, in the emissions area, regulation must address the durability question. Both clean air and fuel conservation might be served by a shift away from greater tightening of standards on prototypes and toward greater concern for durability (even at a somewhat higher level of measured emissions of prototypes). Improvements in this area may call for a reconsideration of the role of equipment or design standards in the process of automotive regulation.

Chapter 1

THE AMERICAN SYSTEM OF REGULATING
THE AUTOMOBILE

by

Lawrence H. Linden and David Iverach*

1. Introduction

The automobile has long been, in economic terms, the most important consumer product manufactured in the United States. It is the second largest purchase made by most households, and the three principal domestic firms which manufacture it have aggregate annual sales of about \$90 billion. Until about a decade ago, the technological attributes of automobiles were determined almost entirely by manufacturers. Today, the situation is very different. Mandatory federal regulations determine to a significant degree the type of vehicle that can be offered for sale.

There is no question that federal regulation of automobile safety and emissions has had significant effects on American society, and that in the future the fuel economy standards will do the same. What is unclear is whether the results have been, or will be, worth the economic costs and the political pain. More people are beginning to question whether or not the use of mandatory product regulation on the automobile companies is the best way to achieve the goals of the safety, emissions, and fuel economy programs. In some cases, this feeling stems from a belief that an alternative system, such as financial incentives, would be better. In others, the feeling derives from a belief that there simply must be a better way.

*The authors wish to thank Christopher R. Frey and Philip A. Lorang for their research assistance and critical review.

In the light of these attitudes, it is remarkable how little we understand the present system--its strong points and its weaknesses, how it developed, how it is changing, etc. People seem to agree on only one thing--that we should be able to improve it.

This chapter is an examination of the American automotive regulatory system. It is principally descriptive, but sets forth hypotheses concerning the nature of the system and how it might be improved. All three important areas of regulation -- safety, air pollutant emissions, and fuel economy--are examined simultaneously in the hope that such a cross-sectional view will stimulate perceptions and hypotheses that might not otherwise be obtained. In the present administrative structure for regulating the automobile there is no organization where emissions regulation is viewed together with safety and fuel economy. This is true in both the legislative and executive branches of the government. The only exceptions are the highest levels of leadership, which have never been strongly involved in consideration of anything but particular standards. Thus there are substantial differences in both the evolution and the structure of the three areas of regulation, as well as important similarities, and the comparative analysis approach yields a provocative set of insights into the automotive regulatory system.

This chapter is laid out as follows. First, in Section 2, a brief history of the evolution of the present system is presented. This is designed principally to provide important background material for the remainder of the chapter.

In Section 3, the issues associated with the standards used to force changes in passenger car technology are examined. The attempts to change passenger technology are the centerpieces of the automotive regulatory

system, and are the focus of this paper. First the procedures -- both institutional and analytical -- for setting automotive regulatory standards are compared. The procedures available for use by the federal government depend on the structure of the standards and this is examined next. The types of standards used in the three regulatory areas -- individual vehicle equipment standards, individual vehicle performance standards, and new car sales fleet average performance standards -- are contrasted. Since the very purpose of the standards is to change motor vehicle technology, research and development activities by the government and the industry are important parts of the system for developing and meeting the standards. The extent to which this is the case is examined in the third subsection, with a focus on the role of government-supported R&D and the incentives to industry for R&D in these areas.

In Section 4, two other issues are examined and compared across the three regulatory areas. First, the role of consumer and operator behavior in meeting the national automotive goals is addressed. In none of the three areas have government-induced changes in human behavior made an important contribution. Second, the role of monitoring and surveillance is examined. In all three areas the government faces measurement and analysis problems which are substantial, if not overwhelming. Monitoring, surveillance, and modeling efforts are crucial to the formulation of regulatory standards which have some semblance of rationality.

Finally, the last section summarizes and reviews the conclusions that the authors have reached concerning the nature of the automotive regulatory system and how it might be improved.

2. A Brief Review of the Evolution of the Automotive Regulatory System

The present structure of the automotive regulatory system is very much a function of the particular political, legislative, and administrative activities out of which it evolved. The purpose of this section is to provide this evolutionary perspective as background for the analysis of that structure which is the subject of the remainder of the chapter. This section, then, is a brief review of the history of the automotive regulatory system.

2.1 Safety¹

Automobile occupant safety had become a public issue by the end of the 1920's, but at that time and for some decades thereafter concern and regulations were directed mainly at the driver. Some states had vehicle standards requiring headlights and windshield wipers, etc., to be fitted to automobiles. With a rising road death toll in the 1960's, increasing attention was focused on the influence of vehicle design on occupant safety. After publication in 1965 of Ralph Nader's book, Unsafe at Any Speed,² which focused attention on the vehicle and not the driver, concern with vehicle design increased dramatically. On March 2, 1966, in response to the growing concern, President Johnson delivered to Congress a message on transportation and traffic safety, together with the proposed National Traffic and Motor Vehicle Safety Act of 1966. In his address, President Johnson urged that the Secretary of Commerce³ be given authority to determine and establish necessary safety performance criteria for all motor vehicles. Prior to the "Safety Act," the automobile had remained essentially free of federal regulatory controls.

The Safety Act: (1) required the unconditional imposition of mandatory standards, at the earliest possible date; (2) reduced the role of the individual states in establishing vehicle standards to a consultative one; (3) legislated a policy that the federal government would develop an independent technical capacity for performing research on accident and injury prevention, for testing, for initiating innovation in safety design to serve as a yardstick for measuring industry performance, and for developing and implementing vehicle safety standards; (4) established mandatory procedures for notification of consumers and corrective repairs for safety-related defects; and (5) gave the federal government a responsibility to provide information on new motor vehicles to assist consumers in evaluating the safety of competing cars.

The Act was approved by the President on September 9, 1966. It required that the initial Federal Motor Vehicle Safety Standards (FMVSS) be issued by the end of January 1967, and that they be applicable to vehicles produced after January 1, 1968. Because of the time constraints it was accepted that the initial FMVSS would have to be based on existing standards. In fact, most of the FMVSS were direct adaptations of the General Services Administration's standards for model year 1968 passenger cars to be purchased for the federal government. Some were also taken from voluntary industry standards, especially those of the Society of Automotive Engineers, from the previous work of the National Bureau of Standards, and from several other sources. The initial accident prevention standards addressed vehicle controls, automatic transmissions, windshield defrosters, windshield wipers, brakes, brake hoses, reflective surfaces, lamps and lighting, and rearview mirrors.

Initial crash protection standards specified lap and shoulder seat belts, energy-absorbing steering columns, padded interiors, recessed instrument knobs, a requirement for doors to remain closed upon a crash, laminated windshields, seat belt anchorage capabilities, and limited hub cap projection. The only post-crash standard issued dealt with the integrity of fuel tanks and related equipment.

Since the initial FMVSS a number of additional standards have been promulgated and a number have been strengthened. New crash protection standards have been written prescribing head restraints, specifying the security of windshield mounting, constraining the characteristics of child seating systems, and specifying side door strength and roof crush resistance. FMVSS 208, passenger restraints, has been by far the most controversial of the safety standards. The requirement for a passive restraint -- essentially the "air bag" -- has been on the regulatory docket and in and out of Congress and the courts since 1970. What was considered by many to be a more modest compromise proposal, the ignition-seatbelt interlock, was made effective for model year 1974 vehicles, but overturned by Congress in mid-1974 presumably in an expression of voter sentiment. A second post-crash standard was issued in 1970 limiting the flammability of materials used in car interiors.

Generally considered an adjunct to safety regulation, regulations to lower to cost of damage in automotive accidents are promulgated under the Motor Vehicle Information and Cost Savings Act, which became law in 1973. Damageability regulations promulgated under that Act are principally concerned with requirements that motor vehicles' bumpers be designed to withstand low-speed contacts with no damage. It required that federal bumper standards be set in order to reduce damages and economic loss resulting from automobile accidents.

2.2 Emissions⁴

As a result of the air pollution situation in Los Angeles and the identification of the part played by automobile exhaust, California, in 1964, adopted the first exhaust emission standards in the country. These standards, applicable in the 1966 model year, controlled hydrocarbon (HC) and carbon monoxide (CO) emissions. Following California's lead, Congress in 1965 passed the Motor Vehicle Air Pollution Control Act which permitted the Secretary of Health, Education, and Welfare to set nationwide emission standards for vehicles as he deemed necessary to protect the public health. In 1966, the Secretary adopted the California standards and made them applicable nationwide for the 1968 model year. Later, in 1968, the Secretary promulgated stricter HC and CO standards for the 1970 model year and an evaporative emissions standard for model year 1971.

By 1970, the "environmental" movement was in full swing. Congress, led by Senator Edmund S. Muskie and his Public Works Subcommittee on Air and Water Pollution responded with tough new regulation in the Clean Air Amendments of 1970.⁵ This law required a 90% reduction from existing controlled levels of HC and CO emissions by model year 1975 and a 90% reduction by model year 1976 of the then uncontrolled levels of nitrogen oxides (NO_x) emissions. Almost simultaneously, regulatory responsibility was transferred to the newly formed Environmental Protection Agency. Importantly, the EPA was given permission, under strict guidelines, to administratively delay each 90% reduction standard for up to one year.

Since 1970 the automobile companies, claiming technological infeasibility, have fought hard to delay and prevent the implementation of the 90% goals. The original one year delays were granted by the EPA in 1973, and Congress in 1974 legislated another one year delay of each standard. A further

delay of the HC and CO standards was granted by the EPA in 1975, so that currently the 90% goals are at this writing (June 1977) scheduled to take effect in model year 1978. A further delay and possible revision of the goals is expected of Congress shortly. At each delay, interim standards of increasing stringency have been established. Thus, while the attainment of the original goals remains elusive, substantial progress in controlling new car emissions has been made.

2.3 Fuel Economy

The legislative and administrative history of the automobile fuel economy standards is much shorter and more recent than the others. Following the 1973-4 oil embargo the federal government decided that there was a need to reduce automotive gasoline consumption. During the 1973 and 1974 Congressional sessions, numerous bills were proposed to institute fuel economy (miles per gallon) standards, vehicle weight standards, and excise taxes on large vehicles. Few of these bills were acted upon, although one did pass the Senate in late 1973, only to die in the House. In 1974 the Ford Administration arranged a voluntary agreement with the domestic manufacturers in which a goal was set calling for a 40% improvement by 1980 over 1974 levels in the industry-wide new car fleet fuel economy average.

This voluntary agreement never had much effect due to Congressional action in 1975. Both the House and Senate Commerce Committees passed bills requiring a series of fuel economy standards. The House Ways and Means Committee meanwhile passed a bill calling for increased gasoline taxes and an excise tax on high-gasoline-consumption vehicles as its solution to cutting gasoline consumption. This bill was later rejected by the whole

House and a standards bill, a title of the Energy Policy and Conservation Act, was enacted in December 1975. This Act calls for each manufacturer's new car fleet to average 18, 19, and 20 mpg in model years 1978, 1979, and 1980 respectively, and 27.5 mpg by 1985. Standards for model years 1981 - 1984 are to be set by the Secretary of Transportation.

The pattern behind these major thrusts in automotive regulation seems to hold rather well. In each case the public and its representatives came, over the period of a year or two, to reach the conclusion that a particular automotive impact had attained an intolerable level. In the safety case, it was the rising highway death toll of the early 1960's and the publication of Unsafe at Any Speed; in the emissions case, it was the rise of the environmental movement; and in the fuel economy case, it was the Arab oil embargo and the "energy crisis" which acted to bring the issue to the forefront of domestic politics. In each case exhortation and some relatively mild form of government intervention preceded the passage of stringent legislation. In the emissions case there was the 1965 legislation; in safety there was a set of standards promulgated by the General Services Administration for vehicles for government usage; and in fuel economy there was the voluntary agreement between the automobile manufacturers and the Ford Administration. In each case, in the face of growing public interest, the Congress determined these less stringent measures to be inadequate, even though they had had little time to yield significant impact. Finally, a piece of strong legislation was passed, imposing a set of mandatory measures on the automotive manufacturers -- the Safety Act of 1966, the Clean Air Amendments of 1970, and the Energy Policy and Conservation Act. This pattern is, of course, not terribly different from the way many major policy initiatives make their way into law.

There are two notable differences between the course of events in the safety area and that in emissions and fuel economy. The first is that the national move toward cleaner and more efficient vehicles was a subordinate component of major initiatives to reduce aggregate air pollutant emissions and natural petroleum consumption, respectively. In both cases the automobile was identified as a major contributor to the national problem, and was singled out for special and specific legislative treatment, but reductions in the impact of the automobile were related, at least conceptually, to national aggregates. In the safety case there was no analogous and clearly identifiable national issue of which the automotive impact was a component. The legislation on automotive safety deals with that alone.

This difference, however, is of less operational significance than another related distinction. In the emissions and fuel economy cases numerical standards and timetables were legislated -- 90% reduction in HC, CO, and NO_x emissions in four to five years, a doubling of average fuel economy in nine years. In the safety case numerical standards for vehicle safety were not available; the level of understanding for describing vehicle safety in terms of a simple set of numbers was (and remains) inadequate. Thus the Congress could not set numerical vehicle targets. This difference between safety and the other two regulatory areas is of major significance; its implications are the focus of this paper and will be explored at length below.

3. Issues of Standard-Setting and Technological Change

The principal objectives of the automotive regulatory system have been to reduce health and property losses, ambient air pollution levels, and gasoline consumption. The principal strategy for accomplishing this has

been to effect changes in the technology of the automobile by the imposition of mandatory product standards to be met by the automobile manufacturers. However, the situation is an extremely complex case, where the extent of the changes actually made depends heavily on the goals that are set, the structure of the standards and the associated incentives to change automotive technology, and the ability of the industry and the government to bring forth the desired changes. In this section we examine and contrast these features of the regulatory system for the three areas of automotive regulation.

It will be seen that the most important difference between the three areas of regulation is in the structure of the standards. In the safety area standards are generally set as numerical performance criteria or as design criteria on vehicle equipment or subsystems; i.e., no overall safety index is used, and each vehicle must meet the standards. In air pollution the three emission standards are a vector which represents an overall measure (albeit three-dimensional) of emissions performance, and, as in safety, each vehicle must meet the standards. Fuel economy standards are set as a scalar value for each manufacturer's new car fleet average in any model year. These differing standard structures have coincided with differing allocations of standard-setting authority between the Congress and the Executive Branch. In the emission and fuel economy cases the availability of simple quantitative indices has made it relatively easy for the Congress to retain the standard-setting function to itself. The results of these procedural differences, and the analytical differences which accompany them, are explored in Section 3.1 below. With a given standards structure, varying areas and degrees of freedom may be left to the manufacturers; it is therefore natural to expect differing responses from firms. This is explored in Section 3.2. Advances in technology are crucial features of what must occur if the national automom-

tive goals are to be met. The incentives for, and activities in support of, technological change in the regulated areas are explored in Section 3.3.

3.1 Standard-Setting Procedures

There are two conceptually distinct aspects to the way in which automotive regulatory standards are set. First there is the set of institutional arrangements by which data are gathered, arguments made, and some set of standards is chosen (and subsequently modified if necessary). Second, whatever the institutional form, there is the underlying conceptual structure used to define goals or balance costs and benefits, and ultimately to derive a specific set of standards. It will be seen here that, while these two aspects of the standard-setting procedures are conceptually distinct, in practice they are heavily coupled.

Most importantly, standards established by Congress (with the concurrence of the President) are subject to review only by the courts, and only for their constitutionality. Therefore the choices open to Congress are far less constrained than those of an agency. Outcomes are obviously influenced by arguments from all sides, but the arguments from those disagreeing with the outcome do not have to be refuted or even listened to. On the other hand, when an Executive Branch agency proposes a standard it has to comply with the requirements laid down in the Administrative Procedure Act which, among other things, requires the agency to give notice of proposed rule-making and to publicly answer any criticisms leveled at the proposals. Hence the extent to which the Congress chooses to establish the regulatory standards itself is a key determinant of the possible stringency of the standards.

The National Traffic and Motor Vehicle Safety Act of 1966 specifies a set of national objectives and requires that the Secretary "promulgate by order appropriate Federal motor safety standards. Each such Federal motor vehicle safety standard shall be practicable, shall meet the need for motor vehicle safety, and shall be stated in objective terms."⁶ A "motor vehicle safety standard" is defined as a "minimum standard for motor vehicle performance, or motor vehicle equipment performance."⁷ "Motor vehicle safety" is "the performance of motor vehicles or motor vehicle equipment in such a manner that the public is protected against unreasonable risk of accidents occurring as a result of the design, construction or performance of motor vehicles and is also protected against unreasonable risk of death or injury to persons in the event accidents do occur."⁸ Thus the standards must be "performance standards," they must be "appropriate" (for the vehicle or piece of equipment to which they pertain), "practicable" (within the state-of-the-art at the time of implementation), and "objective" (measurable), and they must protect the public against "unreasonable risk." The only affirmative requirement in opposition to these constraints is that the Secretary was required to promulgate an initial set of FMVSS by a certain date (January 31, 1967) and modify them one year later, with this initial set to be based principally on the General Services Administration and Society of Automotive Engineers Standards then extant. Thus, beyond mandating nationwide promulgation of these relatively inexpensive and technologically unchallenging standards, the Congress left the further determination of safety measures in the hands of the Executive Branch.

In fact, aside from the controversial passive restraint standard (effectively requiring air bags), very little in the way of major new safety requirements has actually been promulgated. Partly this has been

due to difficulties NHTSA has had in fulfilling the requirements of the Administrative Procedures Act, in that the courts, at the instigation of the regulated firms, have required the agency to in fact meet the criteria specified in the Safety Act. This reflects the substantive difficulties the agency has had in identifying new technical options and then proving that they meet the legislated constraints. The legislative procedures were also supplemented, during the Ford Administration, by Executive Branch procedures requiring cost-benefit analyses and inflation impact statements. While these were strictly procedural requirements, they indicated that aggressive pursuit of the discretionary activities under the Safety Act would not have been consistent with the ideological preferences of the Ford Administration.

The procedures adopted in setting the early standards for automobile emissions were similar to those described for the safety area. In 1965 the Congress passed legislation requiring the Secretary of Health, Education, and Welfare to prescribe emission standards after giving "appropriate consideration to technological feasibility and economic costs." The first standards were promulgated in 1966 and were the same as the industry had already met in California; that is, the industry did not need to do any new research to ensure compliance with the standards. Additional standards were promulgated in 1968 and early 1970 but again technological feasibility and costs were fully considered before promulgation.

The Clean Air Amendments of 1970 represented a dramatic strengthening of the Congress' approach. For the first time standards of performance were set by the Congress. The Administrator of the Environmental Protection Agency (EPA) was given almost no room to move -- he could delay the introduction of the standards by one year providing certain conditions were met, but he could not relax the tough performance goals of the legislation.

In arriving at these new standards Congress did not pay any attention to technological feasibility or to economic effects. This was not an oversight on the part of the Congress -- it was a deliberate adoption of a new process to stimulate the industry into meeting what were seen as its obligations. As a result, the implementing agency, the EPA, was placed in a much stronger position than it, or its predecessors in emissions (and safety) regulation had been. The industry was required to demonstrate progress -- the EPA could point to the standards and be openly critical of the industry's attempts to meet them. The Administrator could postpone the standards for one year only if (among other things) the manufacturers could show that effective control technology would not exist by the compliance date. The fact that the burden of proof was on the manufacturer (to demonstrate that the technology was not "available") and not on the regulatory agency (to demonstrate that the standard was "practicable") made a great difference. Further, the manufacturer was also required to make "good faith efforts" to meet the standards by the deadlines in order to qualify for the one-year extensions. Thus not only was the standard-setting procedure changed by the 1970 amendments in that Congress retained control over what had previously been the Executive Branch's duty, but the groundwork was laid for a much stronger dose of annual coercion in the government's dealings with the industry.

It has already been suggested that the Congress can do what an agency cannot when it comes to setting standards. A close examination of the procedures used by the Congress to arrive at the 90% control standards of the 1970 amendments supports such a thesis. It is widely accepted that the Barth report⁹ influenced the Congress -- it was virtually the only analysis in existence at the time that related automotive emissions to air quality levels. The Barth report concluded that reductions in emissions in

excess of 90% were necessary if the air quality standards were to be met. There were a number of important and controversial assumptions made in the Barth report¹⁰ -- but Congress chose not to analyze the effect of these. Congress also chose not to adopt the precise reductions calculated in the Barth report. Instead, a more symbolic 90% reduction for each of the three major pollutants was adopted. The fact that 90% was adopted for each of the three major pollutants is good evidence that symbolism was placed above science in the selection of the standard.¹¹

For all the symbolism and arbitrariness of the demand for 90% reduction there was the feeling that at least the standards were related to the objective of meeting the national goals -- in this case the ambient air quality standards. This gave the proceedings some apparent rational basis that was lacking in the safety area because of the lack of a quantifiable aggregate goal.

In 1975 the Congress set fuel economy standards. The procedure adopted was one that again an agency would have found difficult to justify. With a goal of reducing oil imports (by some unspecified amount and by some future date) the Congress set about deciding the fuel economy standards for future automobiles. What emerged can hardly be said to have been a result of the adoption of any particular analytical procedure. Standards for the near future (1980) appear to have been based on what was considered practical. Those for the longer term (1985) seem again to have been chosen principally for their symbolic value -- a doubling of the economy of existing new cars. As with the amendments to the Clean Air Act in 1970, the Energy Policy and Conservation Act gives the administering authority (now NHTSA) little chance to change the standards. Hence if the analogy with emissions standards persists NHTSA will be able to adopt far more aggressive procedures in implementing the fuel economy standards than it.

have been able to do in the safety area.

In summary, it can be seen that a different procedure was adopted in setting automobile-related standards starting in 1970. Prior to that date public concern was reflected in legislation allowing a designated agency to promulgate standards. The agency had to be in a position to defend its proposed standards at extensive and penetrating hearings, and before a Court of Appeals. After 1970 a new style was introduced in that the Congress itself dictated the standards. As a consequence the emission and fuel economy standards that resulted were almost certainly more stringent than an agency could have promulgated. The effects from the changed procedure include more resources invested in active research by the manufacturers to meet the standards (as compared to the resources invested in arguing with the agencies' proposals) and a more aggressive implementation of the standards by the agencies which are in a position to demand progress from manufacturers. The key motivating factors were the Congress' and the public's growing unhappiness with the responsiveness of the automobile industry to the national goals, and the Democratic Congress' concern that the Republican Administration would not be aggressive in pursuing those goals within discretionary regulatory authority. A necessary precondition for this change was that numerical performance standards were readily available for the Congress to use in establishing its relatively arbitrary goals in the emissions and fuel economy areas.

3.2 The Structure of the Regulatory Standards

Although the Congress is always involved in establishing the strategic goals of a national regulatory program and sometimes, as of late with emissions and fuel economy, in setting the actual standards, the implementation of the

standards is left to the administrating agency. It has already been pointed out that in each of the three automotive regulatory areas under discussion the federal government has chosen to use some sort of mandatory standard as its regulatory tool, rather than taxes or some monetary incentive scheme. Thus, product changes are demanded by various rules specifying requirements to be met by manufacturers in each model year.

Three significantly different types of automotive regulatory standards have been promulgated: equipment standards and two types of performance standards -- one which applies to individual vehicles, the other to the new car sales fleet of each manufacturer. The way the standard is implemented both by the agency (enforcement) and the industry (compliance) is very much a function of the type of standard. The automobile industry is wholeheartedly against equipment standards, arguing that they are in the best position to decide what equipment is best to meet a given goal. They insist that all that is required is that an overall goal be specified -- as a performance standard. As we shall see there are arguments for and against these types of standards, but first it is worth reviewing the standards in place at present.

Automotive air pollutant emission standards are written in terms of numerical limits on each of the three major pollutants. Each standard specifies that a certain number of grams of pollutant must not be exceeded on the average for each mile the vehicle is driven over a specified driving cycle, chosen to simulate urban driving. For example, the model year 1976 standards are: 1.5 gm./mi. hydrocarbons, 15 gm./mi. carbon monoxide, and 2.0 gm./mi. oxides of nitrogen. These exhaust standards are supplemented by a standard on evaporative hydrocarbon emissions (2 grams evaporated per test), and a crankcase blowby emission standard which specifies no emissions from the crankcase. (There also is a specification on the lead and phosphorus content of gasoline to be met by gasoline refiners and distributors, designed

to protect emission control equipment such as catalysts from deterioration.) The standards must be met by each vehicle produced.¹² No particular equipment is required by the regulations; it is up to the manufacturer to determine how best to meet the standard, but he is obligated to design all models to meet it.

Fuel economy requirements are specified as a numerical standard for the minimum average fuel economy of the fleet of vehicles each manufacturer produces in a given year (sales-weighted harmonic average).¹³ For example, the model year 1978 standard is 18.0 miles per gallon. The manufacturer's production vehicles are divided into classes, and the fuel economy of each class is determined by testing on a specific driving cycle chosen to simulate national driving characteristics. The manufacturer must meet an average which is weighted according to the sales of each vehicle class. No equipment is specified, nor is there any constraint on the fuel economy of any individual vehicle or model.

Automotive safety standards are required by the Safety Act to be standards of performance for motor vehicles or motor vehicle equipment. In fact, however, they go so far beyond the emission and fuel economy standards in specifying actual designs that we will refer to them here as equipment standards. Each and every vehicle of a given regulated class is required to meet some criterion applied to some, often quite subordinate, aspect of the vehicle. In some cases a piece of equipment is required to be mounted, sometimes with a specification of an aspect of its performance. For example, FMVSS 202 requires that head restraints meeting certain specifications be installed in passenger cars and FMVSS 108 requires that headlights, turn signals, etc., be installed and meet certain specifications. In some cases

a type of design is banned; e.g., FMVSS 211 forbids wheel projections which might injure pedestrians. In other cases specifications are prescribed without explicitly requiring new equipment; e.g., FMVSS 219 specifies side door strength and FMVSS 216 specifies roof strength.¹⁴

The most obvious difference between the three types of standards is the degree of flexibility available to the manufacturer in meeting the standard in any given year. Equipment standards leave relatively little flexibility; the manufacturer provides the specified equipment within the specified design limits. In the case of vehicle performance standards, the manufacturer may choose just how he is to meet the standard. Thus differences in design between vehicle classes and between firms are allowed; each manufacturer chooses that system which he determines will provide the optimum set of vehicle attributes (initial cost, fuel economy, drivability, etc.), while meeting the standard. If the standard stays constant in time, he can change the technology from year to year¹⁵ as experience is gained -- either by each manufacturer individually, or by one firm observing the superior features (or higher sales) of the vehicles of another firm. In the case of the fleet average standard the manufacturer has even more freedom, in that he can emphasize changes for improved fuel economy in those models where they can be made most readily. Furthermore, since there are differences in fuel economy among the vehicle classes, he can change his average by fostering sales of high fuel economy vehicles with low prices (or rebates), advertising, dealer contests, etc.¹⁶

Thus the three types of standards offer different degrees of flexibility to the manufacturer. Flexibility is important because it helps deal with problems arising from uncertainty. If the regulatory agency knew with certainty the best set of configurations well before the effective date of the standard, then a set of equipment standards could be imposed that would

give exactly the same result -- i.e., exactly the same vehicles would be produced. The uncertainty indicates the need for a dynamic decision-making process, so that new information can be incorporated as it is learned, over time. More flexibility allows year-to-year changes in equipment to take advantage of this. Often, within one or two model years, a "dominant design"¹⁷ emerges, where all the manufacturers converge on a certain technical approach to meet a given standard presumably because it is the optimal approach. Then equipment standards could be specified with little loss (and, as discussed later, possibly with some gain).

A crucial distinction between equipment standards and performance standards is that the latter can be gradually tightened, giving the manufacturer a continuing incentive to develop new technology for improving the regulated attribute. This is the present system in emissions and fuel economy regulation. It is very difficult to do this with equipment standards. In general to attain higher levels of safety, new standards must be added, but these standards must be based on design changes or equipment which are known to the regulatory agency; there is little the agency can do to force the manufacturer to develop new designs or equipment. In some cases NHTSA has attempted to increase vehicle safety by tightening performance standards on required equipment. This has proved difficult. In the case of the standard on hydraulic brakes (FMVSS 105) for example, there are approximately fifty distinct attributes which must be specified to describe the required equipment. NHTSA did successfully tighten it once.

Even if an equipment standard is promulgated, or tightened, its relation to the regulatory goal is often unclear. For example, a standard that specifies the minimum strength of a vehicle side door, thereby presumably improving safety in collisions on the side, may cause a deterioration in

the energy management characteristics for front-end collisions. There is no a priori way to ensure that overall vehicle safety is actually improved. Of course, if such a phenomenon is actually discovered to be the case, then the specification can be rewritten to protect against it, or the standard can be withdrawn if the technical options are not there. This type of sub-optimization will occur for any standard short of one which specifies some single overall criterion for the social value of an automobile. Thus, for example, even in the case of the vehicle performance standard, e.g., on emissions, meeting a standard on one pollutant may increase the emissions of another (either regulated or unregulated). However, a CO standard will very likely lower CO emissions, and a fuel economy standard similarly will almost certainly reduce gasoline consumption.¹⁸

One can think of a continuum existing from strict equipment standards to fleet performance standards with the equipment standards furthest, conceptually from the actual goal of the program. Each of the programs is designed to improve some problem: energy security or human safety or health. Naturally, the motor vehicle regulatory programs only attempt to control the motor vehicle's contribution to these problems. A fleet or vehicle performance standard attempts to deal with this contribution directly; equipment standards are indirect, allowing more room for unexpected side effects.

As discussed in the previous subsection, Congress has utilized these features of performance standards to prod the industry into developing and adopting new emissions control technology, and is hoping to do the same with fuel economy. This is possible because a reasonably good performance measure is available, thus allowing a specification to be made without knowing what technology might be used to meet it, and in confidence that the result will have a positive impact on the relevant automotive goal. A reasonable, though

unprovable, conclusion is that a specified vehicle performance standard could, if available, be similarly used in the safety area. NHTSA gives at least formal approval to an ultimate consolidation of its standards into a smaller set (four) of numerical vehicle performance standards in key areas: crashworthiness, crash avoidance, etc. The proposed passive restraint standard (FMVSS 208) has been expected to be the first step along these lines; difficulties there have apparently severely hampered the consolidation effort, and it seems to have been given a very low priority within the agency since it was first proposed in 1971.¹⁹

Fleet standards, as compared to individual vehicle standards, allow the manufacturer to phase in new technological developments to meet a given standard. This is generally much more difficult with individual vehicle standards, where each vehicle must meet the same standard. Some phasing of new equipment can be done under individual performance standards, but it is by definition impossible with individual vehicle equipment standards. The ability to phase in innovations is important in an industry with massive quantities of specialized production facilities and a consumer set whose response to innovation is not readily predictable.

On the other hand, a difficulty with fleet average standards is that, if the product is highly differentiated on the regulated attribute, the final new car fleet average is dependent on the composition of the actual fleet sold. This composition is not well controlled by the manufacturers. Furthermore, it is hard to predict the fleet composition, adding another dimension of uncertainty to the manufacturers' product development plans. Of course, this difficulty goes hand-in-hand with the advantage of fleet standards -- that, where the regulated attribute is highly differentiated among vehicles it would be inefficient to make every vehicle meet the identical standard. For example, if all vehicles had to meet a single fuel

economy standard, many models would be eliminated (the number depending on the standard) and some probably would not be constrained at all (indicating that gains could be made at low cost for these vehicles). Emissions are not so strongly differentiated technologically among the vehicle fleet as fuel economy, although NO_x and CO emissions tend to increase with the weight of the vehicle. Safety is highly differentiated with heavy cars significantly safer; without a good safety index this is difficult to quantify except from observed accident data. Measures taken to meet fleet average standards may have serious disrupting effects on the industry because they give manufacturers incentives, beyond those felt due to real manufacturing cost differences, to change prices. If there are important differences among firms in the average levels of the regulated attribute then the results could be very harmful to one or more of the manufacturers, and competition in the industry reduced. Fleet average standards have to be viewed with some concern when the regulated attribute is highly differentiated.²⁰

While the flexibility inherent in performance standards has a great deal to commend it, equipment standards may have an important place in automotive regulation. Although at the time performance standards are promulgated no one is certain of the precise nature of the equipment or design modifications that will be used to meet the standard, as the implementation date nears. and passes most manufacturers have generally elected to use similar equipment. As discussed above, a dominant design emerges that, because of its superior attributes, is widely used. Examples in the emissions area are exhaust gas recirculation systems for nitrogen oxide control, catalytic converters for hydrocarbon and carbon monoxide control and charcoal canisters for control of evaporative sources of hydrocarbons. However, and this results in many problems, the manufacturers design and build the equipment just well enough to meet the performance standard -- or, more to the point, just

well enough to avoid difficulties with the enforcement agency. There is little incentive for manufacturers to design and construct the devices so that they will be durable and reasonably tamper-proof for the life of the vehicle. In fact, any manufacturer who did this would be penalized in the marketplace because of higher costs. Hence, manufacturers inevitably select the least expensive equipment with only secondary attention being paid to durability in the hands of consumers. In the emissions case where the deterioration problem is most serious, regulators try to overcome this by including a durability requirement in the test procedure, but in fact the deterioration found when the manufacturer drives a vehicle for its 50,000 mile test does not relate closely to that observed in in-service vehicles. Manufacturers claim this is due to lack of proper maintenance, hence their support for mandatory inspection and maintenance. However, if the regulatory agency had some authority to control the construction of the devices used by manufacturers, it is likely that more durable and tamper-proof devices could be brought forth.

Automotive emission control systems are technically sophisticated and complex, and are strongly coupled to the design and operation of the entire propulsion system. There are a large number of design variables that can be traded off in order to achieve the optimal combination of vehicle attributes while meeting a given emission standard. Therefore, the role for equipment standards is a limited one. However, if the regulatory agency were given more authority to specify certain design features, the emissions system might be improved in ways that are hard to deal with through a performance standard. For example, an equipment standard might specify the minimum size of a catalytic converter, if that device is installed to meet the vehicle performance standard; this would enhance

the durability of the system. Similarly, it might be useful to specify directly what sorts of adjustments are to be allowed for maintenance of fuel and ignition systems so that tampering would be more difficult. While the ability to innovate under a performance standard is highly desirable and should certainly be maintained, equipment specifications could serve as a useful supplement where a stable dominant technology has emerged and there are significant problems not being adequately handled by the performance standard alone.

3.3 Research and Development in the Regulatory Programs

Government-supported research and development (R&D) programs have been prominent parts of the safety and emissions regulatory efforts, and are becoming more prominent in the fuel economy area. In each case, the government programs have attempted to complement the industry's programs, filling in key technical gaps and avoiding work where the industry's efforts are strong. The industry's efforts, in turn, reflect the incentives felt by the automotive firms under the regulatory system -- the standard-setting procedures and the structure of the standards themselves as discussed earlier in this section. Thus, the government's R&D programs have been related to the structure of the government's regulatory programs. In this subsection, the relationship between government and industry-supported R&D programs on regulated attributes and the regulatory structure will be explored. The emphasis will be on the emission and safety programs where data on government and industry behavior are available.

3.3.1 Safety R&D. From its inception, the U.S. government's automotive safety program was intended by the Congress to incorporate a strong R&D program. There has been a continuous series of efforts for the development of instrumentation and test procedures and, of the most long-run significance, efforts aimed explicitly at developing the technology for making motor vehicles safer. The safety R&D has sought innovations of two sorts --

incremental changes to the present passenger car, and whole new "safety" vehicles. The incremental changes studied by NHTSA have included new lighting systems, new braking systems, glare reduction techniques, etc.

The safety vehicle R&D programs have been the largest and most important -- in terms of funding and public attention -- of NHTSA's R&D efforts. The Experimental Safety Vehicle (ESV) program was initiated in 1970, with competitive contracts let to two firms. Somewhat later, one-dollar arrangements were made for the participation of Ford and General Motors. There were three phases to the program: design studies, fabrication and testing of two vehicles from each firm, and procurement of a larger number of vehicles (twelve). The principal objectives of the program were: (1) to demonstrate the feasibility of advanced automotive safety performance by designing, fabricating, and testing experimental vehicles; (2) to develop the engineering data from which to formulate new safety standards; (3) to heighten public awareness of automotive safety technology; and (4) to encourage the automotive industry to increase and accelerate its own R&D efforts.²¹ The technical focus of the program was the development of vehicles able to provide protection to occupants in a 50 mph head-on collision. This was to be accomplished by using an "integrated systems" approach to the vehicle development -- i.e., all things related to safety were to be considered. A weight specification and other criteria were to be met as well.

Of the vehicles which resulted, only one (Ford's) met the crashworthiness standard. The vehicles were from 800 to 2000 pounds heavier than comparable production vehicles, and they showed the poor fuel economies implied by this added weight. The principal technical features utilized in the effort to meet the safety standard were crushable frames and/or large hydraulic energy absorbers, both designed to decelerate carefully the vehicle and its occupants. There was little in the way of "integration" or "systems" technology generated. Nor did any of the many incremental

safety features of the safety vehicles make their way into new regulatory standards. These results led to the cancellation of the third phase of the program, in favor of the present Research Safety Vehicle (RSV) program. The RSV effort does not attempt to influence the near-term course of safety regulation, rather it explores the trade-offs that would be available for vehicles in production during the next decade.²²

3.3.2 Emissions R&D. R&D conducted in connection with the air pollutant emissions control program has included the development of test procedures and the gathering of vehicle emissions data. In the technology development area, the focus has been almost exclusively on the development and demonstration of advanced low-emissions automotive powerplants. There has been little government support for the development of new incremental technical options.

The Advanced Automotive Power Systems (AAPS) program was initiated in 1970 by President Nixon.²³ The focus of the program was the development of automotive engines which could meet the stringent emission standards then proposed for model year 1980. Since the automotive manufacturers were putting their effort into incremental changes to the conventional internal combustion engine (ICE), the designers of the AAPS program planned to have their development work completed by 1975, so that the advanced engines could be put into production by 1980, in case the ICE could not be successfully modified. In late 1970, Congress passed the Clean Air Amendments of 1970, which mandated that stringent standards very similar to the 1980 proposal of the Administration, be met by the manufacturers in model year 1975, thereby undermining the rationale for the AAPS program. However, the rationale was adjusted to reflect a longer-term view ("infor-

mation gathering"), and the program proceeded, intending to keep roughly the same timing (vehicle demonstrations in 1975).

Initially, the Environmental Protection Agency's efforts in the program were focused on two alternative automotive engines: the Rankine cycle ("steam") engine and the gas turbine. They were chosen principally because they satisfied the mandatory constraint on the selection process -- they had the clear potential to meet the statutory 1975 emissions standards. (The Stirling engine and direct injection stratified charge system also were competitors on this account.) Once selected, the systems were developed and designed to meet that constraint, while doing as well as possible with respect to the other important vehicle attributes -- fuel economy, cost, drivability, etc. Also the systems were designed to fit into conventional vehicle bodies (or nearly so).

Four different Rankine cycle systems were studied and carried into initial development; one (water with a conventional reciprocating expander) was developed finally, fabricated, and tested (on a dynamometer) in 1975. In the gas turbine case one principal contractor was chosen (Chrysler) whose previous work was extended to its next "generation", while simultaneous efforts were made with other contractors to develop components with the intention of incorporating them into the "upgraded" Chrysler engine. The program was redirected in 1973 to a lower power output (suitable for a compact car instead of an intermediate) and stretched out to 1977. The resulting "upgraded" gas turbine engine is now being tested on a dynamometer and in vehicles. In early 1975, the program was transferred to the newly created Energy Research and Development Administration (ERDA). In 1976 funding began on the development of an automotive Stirling engine.

The AAPS program was, without question, successful in developing automotive engines which could meet the statutory emission standards; the difficulties arise with the engines' other attributes. In the case of the Rankine cycle system, the fuel economy projected for vehicles powered by engines similar to that developed by ERDA would be on the order of 50% lower than that of comparable ICE-powered vehicles (at intermediate emission levels).²⁴ Furthermore, studies do not indicate that developments available in the near future would significantly alleviate this difficulty. There are other relative advantages and disadvantages of the steam system, but the fuel economy problem alone was clearly sufficient (in the eyes of the government planners) to terminate the program. In the case of the gas turbine, the engine to be completed in 1977 will have a fuel economy which is somewhat (about 6%) superior to that of the comparable 1977 production ICE. However, the problem of manufacturing cost for the special steel components of the engine remains unsolved. It now appears that the gas turbine will be competitive in cost and fuel economy with future ICE's only after major technological advances are made in the manufacturing of ceramic components. Thus, the gas turbine program is continuing, along with the Stirling engine effort, but neither holds the promise of becoming a commercially viable passenger car powerplant for at least a decade.

In the meantime, the statutory 1975 emissions standards have been postponed until 1978 and almost certainly beyond. The automotive industry has not been able to develop ICE-powered vehicles which can meet the standards with acceptable levels of other attributes - especially fuel economy and cost. Furthermore, fuel economy regulation became a major new government effort, under the Department of Transportation. This agency has begun its own set of studies, development of test procedures, tests of new devices

developed elsewhere, etc., leaving the "long-range" R&D in the area to ERDA. Thus the AAPS program is not now connected closely to any regulatory effort.

3.3.3 R&D within the Overall Regulatory Program. The role of R&D within the overall government program in the safety and emissions areas needs to be examined in the context of the incentive structure faced by the industry for R&D in each area, and the recognition that the incentive structure is dominated by the regulatory system itself. The incentives are strong functions of the structure of the standards, as discussed in Section 3.2. In particular, the areas of flexibility for each type of standard will be exploited by a firm constrained to meet the standard. The timing of the standards is crucial as well, as the industry will focus its R&D efforts on innovations which could be available for production at the time the standard becomes seriously constraining.

The types of R&D necessary to operate, and to comply with, a regulatory program can be divided between that which actually improves the technology available for installation in new vehicles, and that needed to develop and support the regulatory effort. The latter class would include the development of satisfactory testing systems and the gathering and analysis of data to monitor the impact of the programs. These are clearly the direct needs of the regulatory agency itself. While it may sometimes be in the industry's self-interest to do work in these areas (e.g., GM developed the specifications for the dummy now used for crashworthiness testing), EPA and NHTSA will generally have to fill their own requirements, and they do.

R&D to advance the actual state-of-the-art in a regulated area may be divided into three general classes: (1) at a given level of the regulated attribute, work can be done to lower the cost or improve the other nonregu-

lated attributes; (2) attempts can be made to attain levels of the regulated attribute which are higher than those imposed (at any given time) by the development of incremental changes in passenger car technology, and (3) by developing radical changes to improve the regulated attribute.

Examples of the first sort include lowering the cost or improving the fuel economy or drivability at a given level of emissions, or lowering the cost or increasing the level of comfort or ease of utilization of a safety device (e.g., the addition of inertia reels to seat belts). Once a regulation is irreversibly in place, it is clear that the industry has all the usually competitive incentives to do work of this kind. In fact, the automotive firms have done substantial work of this sort and the regulatory agencies have not supported this type of work.

Incremental changes to advance the level of the regulated attribute would include, in the emissions area, exhaust gas recirculation, carbon canisters for gasoline vapor collection and catalytic converters. In the safety area, such items as seat belts, steering wheel locks, side door beams, etc., are incremental advances. In the emissions case, the existence of long-term statutory goals, with virtually annual debates over whether they should be put off for another year or two, or not at all, has forced the industry to thoroughly explore all possible incremental advances. Thus, as long as there are statutory requirements on the books for fixed near-term deadlines, the EPA does not have to worry about developing new incremental emissions control options, and in fact they have not supported them in the past.

In the safety area, on the other hand, the regulatory system provides no incentive for the automotive industry to develop the technology to attain higher levels of safety. As discussed in Section 3.2, the equipment standards used in the safety area do not allow the agency to readily tighten safety

safety requirements unless a new technical options is well enough in hand for the development of a standard requiring the utilization of that option. Thus the burden of developing safety innovations falls on NHTSA. The agency recognizes this and has supported numerous efforts aimed at developing and testing incremental changes to improve the safety of passenger cars. In a small number of cases, these efforts have resulted in new or tightened safety standards. Examples include the strengthening of vehicle roofs and doors, decreased stopping distances and increased fuel system integrity requirements; each has required an R&D effort. Many more innovations have been studied which have not resulted in improved safety of production vehicles -- including modified vehicle frames, various alternative configurations for headlights and taillights, automotive brakes, periscopes, convex mirrors, speed-limiting devices, and drunk-driver interlocks. A number of these innovations were formally proposed as standards but were not adopted as NHTSA was unable to make an adequate case that they were both "practicable" and alleviated "unreasonable risks."

Non-incremental or radical advances in regulated attributes -- as with the crash energy management systems of the ESV's and the low-emissions alternative powerplants -- are much more difficult for the regulatory systems to deal with. A radical innovation cannot be manufactured and placed in all new vehicles in one year. This is due to the requirements for the development of new production capacity, modifications to vehicle designs and the relevant vehicle production equipment, etc. Since both the safety and emissions regulatory systems require all new vehicles to meet the identical standard,²⁵ they cannot be used to force a radical innovation into production. Thus, even though the emissions regulatory system provides some continuing incentive for advances which would lower pollutant emissions, the

incentive is weak with respect to radical innovation. The stringent standards which have been only one to three model years away at any given time have further forced the industry to concentrate on incremental modifications.²⁶ In the safety area, as discussed above, the system provides little incentive for advances -- incremental or radical -- providing a higher level of safety.

The lack of industry incentive would seem to leave the radical innovations as an appropriate area of government support. In fact, the ESV and AAPS programs were initiated in part out of a feeling that the industry was not being sufficiently aggressive in its pursuit of major changes in these areas. Based on the considerations discussed here, it might be expected that a socially desirable innovation (i.e., an innovation in which the sacrifices in unregulated attributes are acceptable to obtain the gain in the regulated attribute) of the radical sort would not be able to be successfully implemented by the regulatory system. This would seem to be the case; however, neither the ESV or AAPS programs produced a socially desirable innovation. As described above, both programs led to the development of vehicles which offered substantial improvement on the regulated attribute, but were decidedly inferior in the other attributes, and not socially acceptable overall. In both cases, the initial enthusiasm of the Congress and the agency led to technical program choices which had the consequence that the program could not contribute to the higher level goals of the regulatory program -- lowered ambient pollution levels or lives saved and injuries prevented. The initial failure of both programs is indicated by the need for the substantial redirection which each underwent. The final phase of the ESV program was never funded; it was clear that the production and testing of multiple copies of any of the ESV's would not have served any useful purpose. Rather, the program was terminated, and the RSV program begun; the goals of the RSV program do not

include the development of technology which would contribute to improved safety within the next decade or so. Similarly in the AAPS program, the Rankine cycle engine was terminated when it became obvious that a low-emissions, poor fuel economy, system was not a socially desirable innovation. The gas turbine and Stirling engine programs are now envisioned as long-range efforts, providing options to the automotive industry in the middle or late 1980's.

In the fuel economy area, the respective incentives for government- and industry-supported R&D programs are somewhat different from those in either the safety or missions areas. Because there is substantial consumer interest in fuel economy, the industry has had under way for the last several years substantial programs for the development of incremental fuel economy improvements. More efficient transmissions, electronic ignition, and new ways of using lightweight materials are examples of small technical improvements for fuel economy. The increasing stringency of the regulatory program over the next decade will add to the incentives for this type of change. Because the fuel economy standards are written as a new car sales fleet average, there is no problem phasing in innovations, as contrasted with the emissions and safety cases. In addition, the schedule of standards allowed five to eight years from the time of their promulgation until they became seriously constraining, so the innovation planning horizon under the standards is considerably longer than the one to three years under the Clean Air Act. Thus the advantages of a radical innovation in the fuel economy area are, if still somewhat cloudy, at least not so obviously missing as in the emissions case. In fact, innovations such as the diesel and direct injection stratified charge engines, which are forthcoming, reflect these circumstances. Furthermore, this mixed incentive picture has led both the industry and the government to invest R&D resources in radical

fuel economy innovations (advanced power systems and advanced transmissions).

Thus we see that in each case the role of government-supported R&D as a part of the regulatory program is a strong function of the structure of that program. In the emissions and safety areas, R&D in the development of testing instrumentation and methods is of clear importance to both EPA and NHTSA and is being actively pursued. The same is true of investigation of the impact of the standards, modeling and forecasting studies, collection of field data, etc. The role of work on incremental innovations to improve the regulated attribute is quite different. In the safety case, the industry has little incentive to pursue such innovations; NHTSA does pursue them and on occasion has been able to implement the results through regulation. The emissions regulatory system does provide ample incentives to the industry in this area; EPA does not pursue them; and the regulations are changed to reflect the industry's efforts. Finally, in both cases there is a lack of incentives to industry to develop radical innovations to improve the regulated attribute. EPA and NHTSA developed programs to fill this gap, but the initial efforts produced systems which offered major improvements in the regulated attribute at the cost of unacceptable decrements of the other attributes, and the programs have been restructured. There remain substantial questions as to how the radical innovations from these restricted programs might be incorporated by regulation into new passenger cars. In the fuel economy area, there are not as many data from which to draw insights, but it appears that the incentives for industry-supported R&D are not as deficient as those in the emissions and safety areas.

4. Issues of Behavior and Measurement

While the issues of standards and technological change are, and should be, at the heart of the automotive regulatory programs, two other issues will be briefly reviewed here. They are interesting because in each case similar problems have been encountered in the safety and emissions areas, and sometimes in fuel economy as well. This allows some insights to be gained into the nature of the automotive regulatory system.

4.1 The Impact of Human Behavior on the Attainment of Automotive Goals

In each of the three areas of automotive regulation, the vehicle buyer and operator can substantially affect attainment of the relevant national goal through his behavior. In each of the three areas, substantial efforts have been made to exploit the possibilities inherent in this, as complements to the efforts at changing the technology of the automobile. In each case it has proved easier to change the technology than to change people's behavior.

Behavioral influences on automotive impacts are numerous. First, the new car buyer has a strong influence on automotive safety and fuel economy through his choice of vehicle. Most significantly, it is well known that larger cars are safer, but that they consume more fuel in providing their transport service. The fuel economy standards explicitly allow this differentiation through their averaging procedure; in safety it results from the use of equipment standards which allow large differences in overall vehicle safety. The emissions standards are uniform across all vehicles and in fact systematic differences in emissions between the lifetime emissions of various vehicles are

probably not as large as differences between other regulated attributes. Under the Safety Act, NHTSA has operated a program to provide safety-related information to vehicle purchasers. This program has never attracted wide interest. The Energy Policy and Conservation Act includes in its provisions a mandatory vehicle fuel economy labelling program, which is complemented by the publication by the Environmental Protection Agency of the results of its fuel economy tests. Here the effect may be greater, as the published fuel economies do receive wide attention in the press, and, presumably due to the manufacturer's perception of consumer demands, are the object of competition between the manufacturers. These information programs may help people make decisions more consistent with their underlying preferences, but are not likely to affect those preferences. In any case, while it is undoubtedly impossible to sort out empirically any influence these programs might have had on consumer behavior, it is hard to imagine that their influence has been significant. At this writing, there is before the Congress an Administration proposal for taxes and rebates on new cars, differentiated by their fuel economy.²⁷ If it becomes law, this would be by far the most significant attempt yet to affect consumer behavior in the automotive sector, and the consequences could be important.

If an individual owns a car, he must still, in any given instance, decide whether or not to utilize it. When faced with a decision about taking a given trip, he may decide to forego or postpone it, to take public transit, to go as a passenger in someone else's car, etc. In most cases these alternatives will cause less emissions, consume less fuel, and result in less destruction of person and property. Transportation control plans, to assist in meeting local ambient air

quality goals have been proposed by EPA; they would include bridge tolls, parking restrictions, preferential driving lanes for busses and car pools, etc. Over the past five years these plans have been the subject of intense political controversy, especially as local politicians, reflecting local sentiments, have resisted their imposition by the EPA. The number of such provisions in effect now is far fewer than was once envisioned, and this is not due to improved air quality. The resistance is a measure of the degree to which the automobile, used in the driver-only mode, is very often the strong first choice among travel options. Car pooling has received increased governmental support as energy conservation joined emission control as a public goal; again it is hard to imagine that publicity campaigns and moral suasion have any significant incremental effect on the choice of mode in travel.

Given a decision to drive a car for a given trip, the manner in which it is driven can significantly affect the safety, emissions, and fuel consumption of the trip. Rapid starts, hard braking, and high speeds all increase the fuel consumption, emissions, and safety hazards of driving. Driving under the influence of alcohol is well known to increase safety hazards and likewise the use of seat belts can prevent death and injury. Penalties for violating speed limits and driving under the influence of alcohol are long standing efforts to affect driver behavior for the purpose of increasing road safety. Since 1974, of course, the national 55 mph speed limit has been utilized as a fuel-saving measure as well. The success of these measures has been mixed at best. Violation of the national speed limit is widespread - in 1975 most states had average speeds over 55 mph on rural interstate roads, and in 14 states over 75% of the drivers on such roads were violating the speed limit.²⁸ Despite intensive and long-standing

publicity campaigns, widespread knowledge of their effectiveness, and mandatory installation, the utilization rate of seat belts remains at only about 20%.²⁹ While in both cases substantial benefits have been obtained by the present levels of compliance, the potential is so much greater that one can only conclude that most vehicle operators value the time-saving, comfort, and convenience of high-speed driving without personal restraints more highly than fuel saved, personal risk, and legal liabilities involved.

Finally, the owner of a vehicle affects its safety, emissions, and fuel economy by the manner and degree to which he maintains it. Proper maintenance of the engine will reduce the deterioration in emissions that occur with vehicle use. The condition of a vehicle's brakes, steering, tires, and other safety-related equipment have a great influence on the safety characteristics of the vehicle. Although the owner of a vehicle rarely deliberately tampers with any of the safety-related items (this was not the case when the safety-belt buzzers and ignition interlocks were installed), tampering with pollution control equipment does occur. Clearly, such tampering reduces the effectiveness of the vehicle emission program. Maintenance does have an effect on fuel economy, though this effect is small relative to the changes being sought through the regulatory program. Mandatory inspection-maintenance programs which have been proposed as part of EPA's Transportation Control Programs have met with as much political resistance as the other proposed features.

This discussion points to the relatively simple conclusion that public policies aimed at meeting national automotive goals through behavior modification seem to have been relatively unsuccessful, at least when measured against the program's own goals. It seems that Americans are strongly

attached to their large automobiles, the freedom of using them in the driver-only mode, operating them as they see fit, and avoiding maintenance and other inconveniences. The private automobile is just that - its owner feels he can go where he likes, when he likes, in the company that he chooses (if any), and in relative comfort. The true costs, dangers, and environmental impacts that result from his using his vehicle are difficult for him to perceive and, when perceived, not valued as highly as public policy would seem to suggest is appropriate. Further, because the individual sees that he makes only an infinitesimal contribution to the total damage (or risks or costs) associated with automobile usage, he feels little incentive to alter his behavior. He feels it is too easy to look foolish and that the procedure of voluntary compliance may not be equitable. The costs - either direct or indirect, financial, or emotional - of modifying his behavior for the benefit of society as a whole do not seem worth it to him. Automobile safety, emissions, and fuel economy do not seem like real crises to most Americans.

The difficulty of changing individuals' attitudes so that the impact of the automobile is lessened means that alternative solutions are usually sought. Normally these other solutions are aimed at modifying vehicle technology. Passive restraints are proposed instead of compulsory seat belt utilization. Improved fuel consumption is mandated rather than a limit on gasoline consumed. Emission regulations are promulgated rather than limiting the number of vehicles in a region. On the basis of the record to date, the technological approach seems more likely to have an impact than any behavioral approach.

4.2 Difficulties in Measuring Regulatory Program Impact

As has been emphasized in the preceding sections, the automotive regulation attempts to influence a highly complex system. The connection between the promulgation of regulatory standards and an actual change in the impact of the automotive fleet is not as strong as often assumed. Promulgated regulations result in some change in new passenger car technology, but the regulations only guarantee that a carefully selected fleet of new vehicles meet specified requirements on special tests; whether or not the marketed vehicles are actually lower in emissions, safer, or more efficient, is less clear. Once the vehicles are in the hands of owners and drivers, reductions in emissions, bodily and property damage, and fuel consumption, depend on details of how they are maintained and utilized. In the case of air pollution, the emissions are diffused and convected as they influence ambient air quality, and, further, the changes in air quality affect in some complex way the public health and welfare. At each point in the connecting chain between the standards and the impact, there are factors external to the system whose influence on actual outcomes is important. Because the connection is not a simple one, measurements of the actual impacts are necessary to insure that the desired effects are in fact taking place. However, the lack of clarity and the noise in the system make it difficult to sort out exactly what effect each important variable - including the regulations - is having.

In the safety area, the annual toll of traffic-related deaths and injuries is known well. Generally, this is reduced to a rate of deaths (and/or injuries) per passenger (or vehicle)-mile, to correct for the gross scale effect of changes in the total amount of driving. Beyond that, correction to sort out the underlying causes in changes in the

death rate is difficult. The highway death rate declined about 3 1/2% per year from 1947 to 1960, rose very slightly through 1966, and has declined monotonically since then by several per cent per year, except for a 15% drop in 1974 as compared to 1973.³⁰ Thus, even before the initiation of national motor vehicle safety regulation with model year 1968, there were important factors lowering the highway death toll. Factors influencing the highway death rate include: the age distribution of the driving population (young drivers are more accident-prone), the spatial distribution of driving (e.g. urban v. rural), the range of distribution of the weights of vehicles on the road (a mix of small and large cars results in more dangerous collisions than a uniform population of either), the distribution of speeds (deaths are more likely in high speed collisions, and more collisions are likely when speeds are diverse), the use of alcohol, the safety of highway structures (which has been increasing due to government programs in that area), changes in the availability and quality of emergency and long-term medical care, etc.

Clearly the most important change of a safety-related variable in one year occurred when the national 55 mph speed limit was imposed in early 1974,³¹ and there is little doubt that it was an important contributor to the large reduction in the death rate that year. However, the efforts to actually estimate its impact indicate some of the difficulties with the much more complicated problem of estimating the impact of the vehicle standards. According to one recent survey,³² more than 30 studies have been carried out to explain improvements in highway safety between 1973 and 1977. Table 1 shows the partial results of three of the studies. Whether the range of disagreement is reasonable or not depends on the particular requirements at hand. From the point of view of defending a

Table 1Factors Affecting Reduction in Highway Deaths, 1973-74³³

<u>Factor</u>	<u>AASHTO</u> ^e	<u>NSC</u> ^f	<u>GM</u> ^g
Reduced and/or more uniform speeds	48	46	35
Reduction in travel	22	21	25
Reduction in vehicle occupancy	--	13	--
Reduction in night driving	--	8	24
Switch in roads used	--	4	1
Switch to weekday driving	--	--	5
Greater use of safety belts	--	4	--
Historical trend	--	--	14 ^a
Other	30 ^b	8 ^c	--
Decrease in safety due to age of drivers, use of small cars, motorcycles, and pedal cycles	<u>--</u>	<u>-4</u>	<u>-5</u>
Total	<u>100</u>	<u>100</u>	<u>100</u> ^d
Notes	<u>---</u>	<u>---</u>	<u>---</u>

^aIncludes better roads, better cars, and increased use of safety belts.

^bIncludes improved driver behavior, daylight savings time, safety belt usage, better roads, cars, and traffic programs.

^cIncludes better cars, roads, and law enforcement.

^dDoes not add due to rounding.

^eAmerican Association of State Highway and Transportation Officials, "Effects of the 55 mph Speed Limit", November, 1974

^fNational Safety Council, "Factors Contributing to the Reduction of Motor Vehicle Traffic Fatalities (January - April 1974 vs. January-April 1973)", October, 1974

^gJ.F. Carpenter, "Traffic Fatalities and the Energy Crisis--Four Month Analysis, January - April 1974", Environmental Activities Publication No. A-3176 (General Motors Technical Center, Warren, Michigan), November 20, 1974.

new safety standard in a court, the relevant criterion would be whether there had been a reduction in "unreasonable risk," that is (in this case) whether or not the impact on safety was significant. In this case, the criterion would presumably be met. Actual motor vehicle safety standard cases are generally much more ambiguous, as no single vehicle design change has had an impact of nearly this magnitude in any given year.

Of course explaining, afterwards, the causes of a change in the highway death toll cannot help in meeting a legislated criterion before the promulgation of a new regulation. This generally requires an understanding of the impact of a change in motor vehicle technology on safety of the vehicle. Such information is crucial both to selecting areas of the vehicle for the promulgation of standards, i.e. the areas causing safety problems, as well as justifying them afterwards in court. However, as discussed in Section 3.2 above, equipment standards are used by NHTSA, and the connection of the changes resulting from such standards to overall vehicle safety is generally not well understood.

In the air pollution area, many of the same problems arise. Chapter 2 of this report is a detailed analysis of the measurement and analysis problems that arise in automotive emissions regulation. In the emissions area, the problems arise both in conjunction with measuring the impact of the standards on the actual emissions output of the in-use fleet, and in understanding how those emissions affect ambient air quality.

Monitoring ambient air quality and understanding how changes in air quality arise are very difficult technical problems. Even if we assume perfect measurements of the instrument then, as discussed in Chapter 2, the large variations in local meteorology, compared to the rate of improvement expected from the regulations, imply that several years' worth of data are required to reliably measure a trend. However, during those years a

number of variables concerning the monitoring site can change - the instrument may be replaced or modified; the site may be changed; a building may be constructed or demolished nearby; an industrial plant may close, or open, or increase or decrease throughput (thus affecting emission output); the number of vehicles driving in the vicinity of the monitor may change or the way they are driven may alter. All these things, and more, will affect the quality of air the monitor is measuring.

To overcome the lack of vehicle-related results from the ambient monitoring system, EPA obtains data on vehicle emissions by measuring these emissions directly. To do this a specified test procedure is used, the same one in fact that is used for certification. However, in actual use there are many important factors that differ in a systematic fashion. For example, ambient temperature, pressure, and average driving patterns can vary from region to region and over time. In addition, the degree to which a customer maintains his car is likely to be less than factory recommendations, even discounting the possibility of deliberate tampering with emission controls. There are many other factors which can vary in actual use as discussed in Chapter 2. Hence it is uncertain how well the changes observed from year to year in the vehicle surveillance tests reflect true changes in vehicle emissions. Even if this is adequate (and it probably is) the problem remains of identifying the reasons for the high emissions of in-service vehicles.

Thus, in both the safety and emissions cases it is extremely difficult to relate regulatory standards to changes in the performance of vehicles in-use, and also to relate changes in vehicles in-use to the proximate goals of the program. In the emissions case, there is the additional problem - probably the toughest empirical problem of all - of relating changes in ambient air quality to changes in public health and

welfare. In fuel economy many of the same issues arise.

The results of this situation are reasonably clear. As discussed in Section 3.1, in the emissions and fuel economy areas, the Congress has set relatively arbitrary statutory goals for vehicle performance. In the safety area, where regulatory standards are much cruder, a set of legislative criteria are applied to standards set by NHTSA; NHTSA has had a very difficult time meeting these criteria and advances in automotive safety seem to have been much less than in emissions or what is expected in fuel economy. The gathering of data on the effect of changes in automotive regulations and the improvement of our understanding of the physical and behavioral processes underlying such changes are crucial to improvements in the performance of our regulatory system.

5. Summary and Conclusions

This analysis of the automotive regulatory system has been based on a view across the three areas of regulation. The most important similarity between the three regulatory systems is that the principal policy tool utilized is some sort of mandatory technical constraint -- i.e. a regulation -- rather than any sort of monetary incentive (although the fuel economy regulation is closer to an incentive scheme than the others). The most important difference between the systems is the type of standard utilized within the regulatory system; this has great influences on nearly every other aspect of the system. In emissions control, the regulation is a set of three quantitative emissions limits--performance standards on the vehicle as a whole -- which must be met by every vehicle produced. In safety, the standards are generally equipment standards in that they specify in a relatively constraining fashion the requirements for particular vehicle subsystems. In fuel economy, the standard is an average imposed in each manufacturer's entire sales fleet. Thus, in air pollution

and fuel economy, as compared to safety, numerical standards are written which bear directly on the relevant national goal.

This distinction between vehicle (or fleet) performance standards and equipment standards is crucial because of the way the standards are used. In both the emissions and fuel economy cases the Congress has taken it upon itself to choose very stringent, and more-or-less arbitrary (90% reduction in emissions in five years, doubling of fuel economy in ten), numerical goals, not readily attainable with the technology available at the time of promulgation. Thus, they have put tremendous pressure on the industry to develop technological fixes which could be implemented within the timing constraints of the standards. This has meant very "incremental" technical changes in the case of air pollution; more substantial changes can be accomplished under the timing and fleet-averaging of the fuel economy regulations. In safety the situation has been very different. The burden of discovering, developing, and proving the feasibility of new safety technology lies almost entirely on the regulatory agency. The agency must support technical efforts to develop safety innovations, and then defend them as necessary to prevent "unreasonable risk", and as "appropriate", "objective", and "practicable", through the regulatory procedure and often into the courts. Thus the emission and fuel economy programs have been able to be more aggressive than the safety program.

The safety agency is working to develop quantitative vehicle safety performance criteria. However, these efforts have not been a high priority item since the initial passive restraint standard, which was written as a quantitative vehicle performance standard, was thrown out in court (for not being "objective") in 1972. The analysis here would indicate that the setting of quantitative performance criteria in the legislation itself can have significant advantages for the regulatory agency, well beyond those

associated with a particular technical option such as the air bag.

Insights in a number of other areas come from laying out and comparing the three regulatory systems. In each case there have been and are substantial government-supported research and development programs designed to produce major technical advances -- e.g., the advanced automotive power systems and the "safety vehicles" -- but the government has not been able to work these innovations into the system. In fact, a major flaw in the safety and emissions regulatory systems is their inability either to stimulate, or even accept, major technological innovations. The use of fleet averaging in the fuel economy regulations offers the flexibility to handle major changes, and the incentive structure encourages such changes more than those of the emissions or safety systems.

In each case the regulatory programs have been accompanied by attempts to modify the vehicle buying or operating behavior of the public, but to little apparent avail. The difficulty of changing people's habits in this area is best illustrated by the low percentage of motorists who wear seat belts. Talk of compulsory seat belt laws has been heard for fifteen years, but most individuals still do not seem ready to accept this "behavioral fix." Similarly, programs to save fuel by driving slower and to reduce emissions by forming car pools or using public transport have not been notably successful.

In each case there are formidable difficulties in measuring the real impact of the programs beyond the visible changes in new vehicles. This leaves the benefits of the regulatory standards open to serious questions. After approximately a decade of safety and emissions regulation, it is still not possible to predict accurately what effect a proposed standard will have on death and injury rates, or on air quality. This statement reflects as much the difficulties of the problem as the lack of attention to it, although

in both the safety and emissions areas there would appear to be a significant role for improved and expanded monitoring and surveillance systems.

Aside from offering a perspective on the automotive regulatory system which is seen nowhere outside of the automotive manufacturers themselves, the comparative view of the three areas of automotive regulation offers insights into the system which might not otherwise be available. Thus, while there are significant flaws and problems in the automotive regulatory systems, some of which are consistent across the three areas of regulation, there are important points where the differences seem to offer room for useful improvements.

FOOTNOTES

1. This review is taken from D.L. Rothberg, "The U.S. Government's Automotive Safety Program; A Preliminary Examination", MIT Energy Laboratory Working Paper MIT-EL 76-029 WP (MIT, Cambridge, December, 1976).
2. R. Nader, Unsafe at Any Speed (Grossman, N.Y., 1965).
3. Automotive safety regulation is now administered by the National Highway Traffic Safety Administration (NHTSA) in the Department of Transportation.
4. A detailed description of the legislative and administrative histories of the Clean Air Act is given in C.R. Frey, "A Legislative and Administrative History of Automotive Emissions Control," unpublished MIT Energy Laboratory Internal Working Note, February, 1977.
5. See Chapter 5 of this report for a detailed analysis of the events leading to the passage of the Clean Air Amendments of 1970.
6. P.L. 89-563, 8 stat. 718, Sec. 103(a).
7. Ibid., Sec. 102(2).
8. Ibid., Sec. 102(1).
9. D.S. Barth, et al., "Federal Motor Vehicle Emission Goals for CO, HC, and NO_x Bases on Desired Air Quality Levels," presented at annual meeting of the Air Pollution Control Association (St. Louis, June, 1970), and reprinted in "Air Pollution -- 1970, Part-5," Hearings Before the U.S. Senate Public Works Committee (Government Printing Office, Washington, September, 1970).
10. See Chapter 2 of this report for a detailed discussion of the assumption underlying the rollback model used by Barth.
11. See Chapter 4 of this report for a discussion of the importance of symbolism in regulation, and Chapter 5 for the details of the role of the 90% figure in the formulation of the 1970 amendments.
12. Although the law is clear in its intent, in practice a significant number of vehicles are produced which do not meet the standards.
13. See Chapter 3 of this report for a discussion of the fuel economy averaging procedure.
14. It should be noted that the penalty structure (for violations) in the three areas is different. Most significantly the fuel economy penalty is non-prohibitive (\$50 civil fine per car for each mile-per-gallon the actual fleet average differs from the standard), while in emissions and safety it is prohibitive. Of course a manufacturer may

decide that the cost of violating the fuel economy standard is significantly more than just the fine, so it may in fact be viewed as prohibitive as well. Here we will not consider the possibility of conscious violations of the fuel economy standards.

15. Or even during a production run -- in 1975 Ford found that its competitors' models had better fuel economy and were thus selling better. As a result Ford modified the emission controls so as to obtain better mileage. These models were known as the "MPG" models.
16. Chapter 3 of this report analyzes the alternatives to meeting the fleet average fuel economy standards with technology changes.
17. See W.J. Abernathy and J.M. Utterback, "Innovation and the Evolution of Technology in the firm, unpublished Harvard Business School Paper HBS 75-18R, for a discussion of the concept of a "dominant design."
18. While this is very likely, the degree to which it is true is not well known. See Chapter 3 of this report for a discussion of the relation of the emissions performance standard to actual in-use emissions.
19. The consolidation was first proposed in "Program Plan for Motor Vehicle Safety Standards," (NHTSA, June, 1970).
20. This is the subject of Chapter 3 of this report.
21. U.S. Government Accounting Office, "Improvements Needed in Planning and Using Motor Vehicle Safety Research", (Government Printing Office, Washington, 1974), p.32.
22. See Rothberg, op. cit., for a fuller discussion of the objectives and activities of the ESV and RSV programs and the relationships between the two programs.
23. See Appendix of J.B. Heywood, H.D. Jacoby and L.H. Linden, "The Role for Federal R&D on Alternative Automotive Power Systems", MIT Energy Laboratory Report No. MIT-EL 74-013 (MIT, Cambridge, 1974) for a detailed history of the AAPS program.
24. S. Luchter and R. Renner, "An Assessment of the Technology for Rankine Engines for Automobiles," ERDA Rept. 77-54 (U.S. Energy Research and Development Administration, Washington, 1977), p. 76.
25. The emissions legislation does permit a differentiation of standards to be met by vehicles sold in California. NHTSA has on occasion allowed certain small segments of the new car sales to meet a standard less stringent than the remainder. In both cases there is no explicit legislative authority for such differentiation and the agencies would likely run into significant legal problems were they to explicitly allow a phasing in of a regulatory change.
26. See L.H. Linden, et al., "Federal Support for the Development of Alternative Automotive Power Systems; the General Issue and the Stirling, Diesel and Electric Cases", MIT Energy Laboratory Report MIT-EL 76-014 (MIT, Cambridge, 1976), Section 3.2, for a detailed

analysis of the incentives seen by the automobile industry for the development of advanced powerplants.

27. See Chapter 3 of this report for a brief discussion of the proposal tax-rebate as a supplement to the present fuel economy standards.
28. U.S. Government Accounting Office, "55 mph Speed Limit; Is It Achievable?," U.S.G.A.O. Report No. CED-77-27 (Government Printing Office, Washington, 1977). p.6.
29. W.T. Coleman, "Decision Concerning Motor Vehicle Occupant Crash Protection," (U.S. Department of Transportation, Washington, December 6, 1976), p. 29.
30. National Highway Traffic Safety Administration, 1970 Report on Activities, (U.S. Department of Transportation, Washington, 1971), p. A12, and _____, 1975 Report on Activities (U.S. Department of Transportation, Washington, 1976), p. 7.
31. The national limit effectively became law on January 4, 1975, when the Federal Aid Highway Amendments of 1974 (P.L. 96-643), were signed. It prohibits the Secretary of Transportation from approving any federal aid for highway construction in states that do not establish and enforce a 55 mph speed limit. It followed a similar but temporary law (P.L. 93-239) which had been signed on January 2, 1974, and implemented by all fifty states by March 3, 1974.
32. U.S.G.A.O., op. cit., p. 7.
33. Adapted from U.S.G.A.O., op. cit.



Chapter 2

UNCERTAINTIES IN REGULATION OF AUTOMOTIVE EMISSIONS

by

Michael K. Martin

1. Introduction and Overview

One of the most striking aspects about government regulation of automotive air pollution is the numerous uncertainties which exist in the underlying analysis linking regulation to improvements in public welfare. It is sufficient merely to examine the responses of other countries to the problem of air pollution, based on essentially the same information, to realize that reasonable people can disagree over what can and should be done about the problem. Some of this difference is of course due to political and economic considerations, but much is attributable to the fact that substantially different proposals can be supported by essentially the same scientific data.

This chapter argues that substantial uncertainties do exist in the underlying basis for emissions regulation, that they are important, and that recognition of these uncertainties leads to a particular view of how regulation should work in this area. The basic contrast is between a static and a dynamic view of regulation. The current system tends to reflect the former view--goals are set based on a combination of judgment and available data, and regulations are promulgated to meet these goals. These regulations become the focus for controversy and, once set, tend to remain fixed. The dynamic view, on the other hand, takes as central

the fact that the informational basis for regulation is uncertain, and that best judgments in key areas are apt to shift with time as knowledge improves. In this view, the focus is on the quality of current information and on the establishment of mechanisms to ensure that that information is continuously updated. This viewpoint will be set forth in more detail later; the reasoning is akin to that of a legislator advocating legislative oversight, or an engineer building in system feedback.

This chapter is an analysis of the sources and magnitudes of uncertainties in our knowledge of the automotive emissions regulatory system. Ability to predict the future results of a policy-related action--especially a change in standards--will be addressed along with our ability to know, i.e., to measure what is happening at any given time.

Before proceeding, however, it is well to consider in a bit more detail just why uncertainties matter. That is, how regulatory behavior should be different as a result of these imperfections in knowledge, as compared to the actions which would be taken if there were perfect knowledge.

- (i) The need for flexibility. As time goes on it is likely that many uncertainties will be resolved by improved modeling, surveillance data, etc. It is important that regulation not be "locked in" based upon false confidence in the certitude of predicted effects.
- (ii) The need for surveillance. The greater the degree of uncertainty, the greater the need for surveillance to assure that progress is being made in meeting goals. Surveillance often can both detect and aid in diagnosing problems; an example is automobile field surveillance. The amount of funds that can

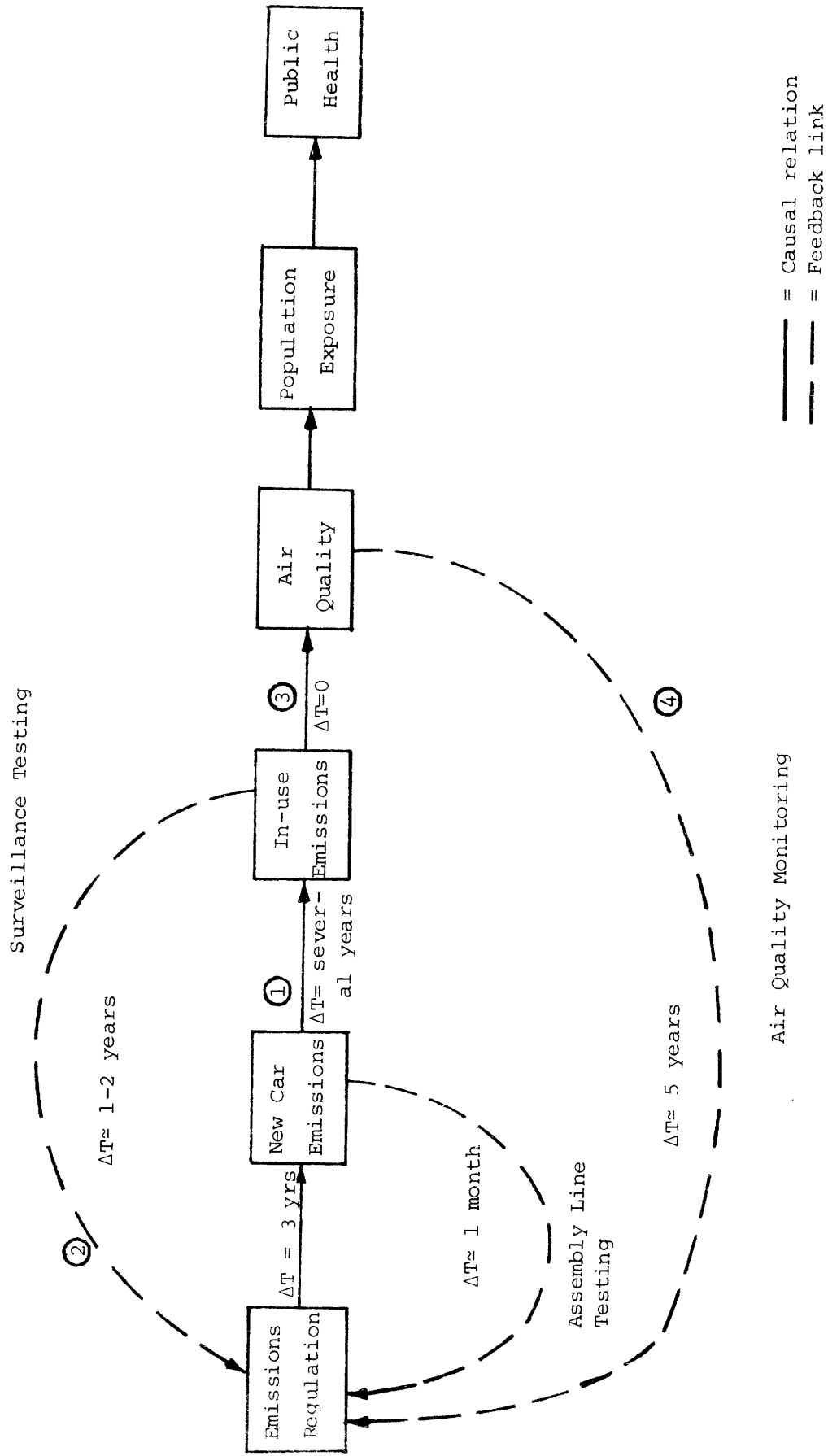
be justified for such a program is proportional to the social costs incurred in over or undershooting the emission goal.

For automobiles, this is a large sum of money.

- (iii) The need for safety factors. Society, like most persons, is risk averse--a sure thing is preferred to a gamble--and extra expenditures can be justified for improved assurance of meeting goals. In the automotive case, presumably it is desirable to aim for emissions less than those thought to be tolerable, to allow for the possibility that the actual air quality improvement will, for some unforeseen reason, be less than predicted. The "safety factor" to use depends on both the degree of uncertainty, which can be roughly estimated quantitatively, and society's degree of risk aversion, which can only be estimated judgmentally.
- (iv) The need for research. The most obvious response to uncertainty is to attack it directly. For example, it was exploratory research that discovered the possible problem of sulfate emissions from catalyst-equipped vehicles. In areas where uncertainties are very large, research may be the only practical response available.

Uncertainties are gaps in our knowledge. An analysis of the uncertainties in emissions regulation must begin with a look at the knowledge which is or could be utilized in that system. Figure 1 is a schematic showing the key variables in the system, and the crucial connections between those variables. The system is divided into the conventional units of analysis.

Figure 1: Feedback Links in Automotive Emissions Regulation



Two types of connections are shown. The set of links going from left to right indicates the chain of events which is initiated when new regulations are promulgated. Each link represents the causal relationship whereby one variable, say in-use emissions, is dependent on a previous one, say new car emissions. Each of these links combines technical and behavioral elements depending on both the underlying laws of nature and the choices of people or institutions. An understanding of these relationships is the key to the ability to predict the future impact of a change in emission regulations, and less than perfect knowledge leads to predictive uncertainties.

The set of links going from right to left are feedback links. They indicate that information on the state of the system at any given time can be used to change the emissions regulations. These are policy relationships, in that the regulatory authority can set the emissions standards. Here the interest is not in prediction, but instead in measurement; it is not possible to have useful feedback unless the actual state of the system at any given time can first be determined. Uncertainties here refer to the limits of our ability to know that state exactly. Three feedback links are shown in Figure 1.

Assembly line testing generates information concerning actual new car emissions; surveillances rating is used to measure the emission of the in-use vehicle fleet; and air quality monitoring tests the ambient impact of the vehicle emissions and can be used to make inferences about those emissions. Each can help in setting emissions standards.

Because an understanding of the causal relationships is necessary background for the remainder of this chapter, a brief discussion of the forward links of Figure 1 now follows. Starting at the left, a given

schedule of standards will typically have as its aim the introduction into general use of new emission control technology, and thus changed emissions of the vehicle fleet. However, it must be recognized that the process of forcing new technology into use is dynamic, in that the time pattern of introduction is sensitive to details of the regulation; and it is uncertain, since creation of new technology cannot always be induced by legislation or other pressure. Neither of these aspects of the problem are well understood.

One example of this is provided by the "original" model year 1976 automotive emissions standards. These standards, mandated by the Clean Air Amendments of 1970, were intended to "put industry's feet to the fire"--to force the development of new and presumably unconventional technology. However, due to the short lead times in the act, the industry concentrated on the safer course of evolutionary changes in the ICE, instead of the revolutionary advances anticipated by Congress. Attempting to force technological change is clearly risky; the mere promulgation of a standard does not ensure that it will be met.

The second forward link is that between the performance of new technology under laboratory conditions and its performance in the field, in the hands of ordinary car owners. In a standard auto emissions test, the vehicle is required to stand for 12 hours at 68-86° F; it is then pushed onto a chassis dynamometer, connected to measuring equipment, and driven through a speed-time schedule representative of Los Angeles metropolitan traffic. Twenty minutes later the vehicle is restarted and run through the first part of the driving cycle again. These "cold start" and "hot transient" emissions measurements are com-

bined in a weighted average (meant to reflect the typical ratio of hot starts to cold starts) and the resultant values are the reported emissions. The vehicles used for certification testing are carefully pre-tuned to factory specifications and follow a prescribed maintenance schedule as they accumulate mileage.

Obviously in actual use there are many important factors that differ from those experienced under standard test conditions. For example, ambient temperature, pressure, and average driving cycle can vary systematically from region to region and over time. In addition, the degree to which a customer maintains his car is likely to be less than the factory recommends, even discounting the possibility of tampering with emission controls.

The third forward link is the one which relates emissions, aggregated over some suitable time scale and area, to a measure of air quality. Though only automotive emissions have been mentioned so far, obviously a detailed knowledge of all emissions is necessary to complete this link. Presently a variety of mathematical models are available for quantifying this relationship, although none is very accurate.

Under the current regulatory structure, air pollution goals are set in terms of levels not to be exceeded more than once a year--presumably anywhere within an air quality region. With this definition of the goal of emission regulation there is no reason to carry a technical analysis any further. In a more basic sense, however, it is clear that improvement in public health (and welfare) is the real attribute of interest, and so it is worthwhile to look at how this is related to air pollution levels.

The next link of interest, then, is the one relating air pollution levels to population exposure. Little is currently known about this link, since it is difficult to characterize the movements of people in and through urban areas, and to relate air pollutant levels inside vehicles and buildings to those outside. Obviously, however, most of the population is exposed to levels less than the maximum for most of the time. Therefore, most of the population should enjoy an additional margin of safety over and above that provided by the ambient air quality standards.

The final forward link relates pollutant exposure to effects on human health and welfare. This is a subject with its own vast literature. The most recent review of the evidence on the health effects of air pollutants is that of the National Academy of Sciences¹, which concluded that although the database available for setting the standards was unsatisfactory, there did not exist substantial basis for changing the standards. On the whole, there seems to be a consensus that the health effects of low levels of pollutants are known only very approximately. In addition, there is a growing feeling that the traditional six pollutants (HC, CO, NO_x, O_x, SO_x, TSP) represent only the tip of the iceberg, and that unregulated pollutants (e.g., sulfates, PAN, nitrates) may in fact represent a greater danger.

In this chapter uncertainties in four of the links in the automotive regulatory system are analyzed; measurement uncertainties are analyzed on two feedback links, and predictive uncertainties at two forward links. In Section 2 current ability to measure and to predict in-use emissions is addressed. Sections 2.1 and 2.2 deal with link 2, focussing on EPA's

surveillance program and the extent of the resulting knowledge concerning in-use emission; and Section 2.3 deals with link 1, the relationship of in-use emissions to new car emissions. A thorough understanding of this relationship is necessary for predictions of in-use emissions, given new car emissions. Section 3 deals with uncertainties in air quality. Sections 3.1 and 3.2 address problems in the definition and measurement of air quality, respectively; together they deal with how well it is possible to know what the air quality is, so that it can be used in a feedback mode in link 4. Section 3.3 deals with the predictive side of air quality, link 3. For each of these links the goal of the analysis is to provide a quantitative estimate for the uncertainty in our knowledge of the relevant parameters, but no attempt is made to actually improve that knowledge.

In Section 4 such an attempt is made, as link 4 is explored in detail and a model developed to allow inferences to be made from monitoring data about how air quality is actually changing at a particular site in Boston. The crux of the model is the relationship between ambient air quality and meteorological conditions.

Finally, in Section 5, the results of these analyses are summarized with a crude aggregation of the uncertainties, and a set of conclusions is drawn.

2. Uncertainties in Measuring and Predicting Aggregate Emissions

In keeping with the thrust of this chapter, only uncertainties in automotive emissions will be discussed. (Uncertainties in other sources of HC, CO, and NO_x matter as well, of course). Uncertainties in

measuring aggregate emissions arise both from uncertainties in the field surveillance data used and from uncertainties in the realism of the method used to form the measure. Therefore, after a brief summary of the EPA surveillance program, these two sources of uncertainty will be examined in turn. Finally, and in light of this examination, uncertainties in predictions of aggregate emissions will be discussed.

2.1 The EPA Field Surveillance Program

Virtually every year since the late 1960's has witnessed some study to improve our characterization and evaluation of emissions from vehicles. Table 1 summarizes a number of these studies. Over the years, the Federal Test Procedure (FTP) for exhaust emissions has changed several times, and so much of the earlier data are not directly comparable to more recent data using the 1975 FTP.

Although the 1975 FTP is the basis for certification of vehicles in meeting exhaust emission standards, some effort has gone into characterizing emissions under other conditions. Thus in more recent programs, emissions over a wide range of speeds and loads have also been measured for each test vehicle. As will be discussed later, these modal emission can (with appropriate assumptions) be used to estimate emissions over driving cycles different from that of the 1975 FTP. In addition, some effort has gone into investigating the effects of vehicle loading, air conditioner usage, and ambient temperature. More recently, highway fuel economy testing (HwFET) has been added, and key mode and various short test cycles have been investigated in an attempt to find a less expensive way of measuring emissions.

Table 1*

Past and Ongoing Emission Factor Programs

Year	Program Title	No. of Vehicles	Type of Vehicles	Test Sites	Test Procedures
FY 68	Rental Vehicle Surveillance Program	705	1968 and 1969 automobiles	Los Angeles Detroit	7 X 7 FTP
FY 69	The Great Plains Surveillance Program	2029	1968 and 1969 light-duty	Houston Kansas City	7 X 7 FTP
FY 70	The National Surveillance Program-Phase I	2101	1970 light-duty	Houston Kansas City Los Angeles Detroit Denver Washington	7 X 7 FTP
FY 70	The National Surveillance Program-Phase II	369	1971 light-duty	Houston Los Angeles Detroit Denver	7 X 7 FTP
FY 70	A Surveillance Study of Smoke from heavy-duty Diesel-Powered Vehicles-Southwestern USA	64	1970 and 1971 heavy-duty	San Antonio	Chassis Dynamometer adaption of 1970 FTP
FY 70	A Surveillance Study of 145 Emissions from heavy-duty Gasoline-Powered Vehicles-Southwestern USA		1970-1972	San Antonio	Chassis Dynamometer version of 1970 FTP, Over-the road testing
FY 71	A Study of Emission from Light-Duty Vehicles in Six Cities, FY 71	1020	1957-1971 light-duty	Chicago St. Louis Denver Washington Los Angeles Houston	1972/1975 CVS FTP, Evaporative Emissions, Modal Emissions
FY 72	A Study of Emissions from Light-Duty Vehicles in Six Cities, FY 72	1020	1966-1972	Chicago St. Louis Denver Washington Los Angeles Houston	1972/1975 CVS FTP, Evaporative Emissions

* reproduced from the reference in footnote 2.

Table 1 (continued)

Year	Program Title	No. of Vehicles	Types of Vehicles	Test Sites	Test Procedures
FY 73	Acquisition of Diesel Truck Operational Parameters and Emissions	10	1970-1973	San Antonio	Chassis Dynamometer version of 1970 FTP Over-the-road testing
FY 73	A Study of Emissions from Light-Duty Vehicles in Six Cities FY 73	1080	1967-74	Detroit St. Louis Denver Newark Los Angeles Houston	1972/1975 CVS FTP, Modal Emissions, Evaporative Emissions, Loaded Vehicle Tests, Air Conditioning Tests, Aldehydes and Light HC, Sound Level Tests, I/M Correlation Tests
FY 73	Mass Emissions from Precontrolled Heavy Duty Gasoline-Powered Vehicles	10	1965-1969	San Antonio	Chassis Dynamometer version of 1970 FTP, Over-the-road testing
FY 73	Ambient temperature and Light-Duty Vehicle Emissions	26	1967-1975	Bureau of Mines	1975 FTP
FY 74	A Study of Emissions from Light-Duty Vehicles in Seven Cities	1968	1965-1975	Chicago Denver Houston Phoenix Washington St. Louis Los Angeles	1975 FTP FTP, HWFET, Modal, Key Mode, Two Short Cycles, Two Other Transient Cycles, Aldehydes and Light HC
FY 74	Heavy-Duty Emission Factor Testing Program (Gas and Diesel)	30	1965-1975	San Antonio	Steady State, Sinusoidal, Transient Cycles, FTP
FY 75	Study of Emissions from 6000-8500 pound GVW Vehicles	250	1972-1976	To be determined	1975 FTP, HWFET, Modal
FY 75	Study of Emissions from Motorcycles	200	1973-1976	To be determined	Proposed FTP, HWFET, Modal
FY 75	Study of Emissions from Light-Duty Vehicles in 2220 Seven Cities		1968-1976	Chicago St. Louis Phoenix Houston Denver Washington Los Angeles	1975 FTP, HWFET, Modal Tests, Short Cycles

This continuing program, involving only a few thousand vehicles each year, is the primary basis for monitoring progress in meeting emission goals. The following section will look at the accuracy of estimation of current emissions from light duty vehicles and the limitations inherent in the test program.

2.2 Uncertainties in Current Aggregate Emissions

EPA has set forth a straightforward and logical methodology for estimating aggregate emissions from light duty vehicles.² The methodology takes as its starting point measured emissions, using the certification test procedure, from in-use vehicles of various model years. It is explicitly recognized that emissions in the field will differ from those of the test procedure due to differences in operating patterns and ambient conditions. Thus, there are a number of "correction factors" which are used to modify the basic "emission factors." The various data used can be summarized as follows:

- 1) base exhaust emission factors, disaggregated by model year, calendar year, altitude, and location (California vs. U.S.)
- 2) base evaporative and crankcase HC emissions, by model year and location
- 3) correction factors for average driving cycle speed, by model year, altitude and location.
- 4) correction factors for average ambient temperature
- 5) correction factors for cold/hot start weighting
- 6) data on age distribution of vehicles and average mileage accumulation rate of vehicles by age of vehicle

Data are readily available to estimate the uncertainties in base exhaust emission factors. Table 2 shows the large variability of measured emissions among individual vehicles of a given age and model year. In some cases, the sample standard deviation even exceeds the mean, indicating a highly skewed distribution since emissions must be positive. In calculating aggregate emissions, however, average emission factors are used, and these will both have a smaller sample variance and be more symmetrically distributed about their respective means.

Table 2 shows the uncertainties in base emission factors (i.e., emissions over the 1975 FTP, averaged over all vehicles of the same model year) for each model year, as measured by two times the sample standard deviation s (approximate 95% confidence intervals) and expressed both in absolute terms and as a percentage of the corresponding average, \bar{e} . These differ systematically for each pollutant, are most stable when expressed as a percentage of the corresponding mean, and are strongly affected by the sample size. It is worth looking at the resultant uncertainty in aggregate emissions, setting all correction factors to 1.0 (i.e., no correction), and this is done for CO in calendar year 1975 in Table 4. The weighting factors represent the fraction of total vehicle miles travelled attributable to autos of various ages; they were excerpted from the reference given in footnote 3, and are roughly correct for this application. As can be seen, the uncertainty stemming from emission factor uncertainty is rather small--about 4% for CO. It is slightly higher for HC and lower for NO_x .

Buried in this analysis, however, is an implicit assumption that base emission factors are the same for all non-California non-high-altitude

*
 Table 2: FY 74 Emission Program
 Emission Results for All Cities except Los Angeles and Denver
 1975 FTP

Year	N	Mean Miles (K)	Hydrocarbons (GM/MI)		Carbon Monoxide (GM/MI)		NO _x (GM/MI)	
			Arithmetic Mean	SD	Arithmetic Mean	SD	Arithmetic Mean	SD
65-67	126	80.8	8.93	7.51	108.54	52.31	2.89	1.44
1968	77	69.5	6.70	6.72	82.59	44.76	3.60	1.78
1969	88	62.5	5.98	3.64	78.46	38.51	4.25	1.71
1970	99	58.9	5.34	7.67	63.88	31.83	3.66	1.33
1971	113	48.5	5.21	6.20	52.69	37.55	3.90	1.33
1972	176	41.9	4.23	4.50	51.79	48.71	4.03	1.48
1973	128	29.0	3.33	1.78	45.31	40.42	3.01	1.62
1974	193	20.2	3.58	2.37	41.77	25.69	2.89	1.40
1975	587	8.8	1.32	1.03	22.92	23.56	2.44	1.01
75LDT	50	8.2	1.48	1.55	18.82	19.03	2.45	1.02

* excerpted from the reference in footnote 3.

Table 3: Uncertainty in Base Exhaust Emission Factor in 1975
All cities except L.A. and Denver ⁴

<u>Year</u>	<u>N</u>	$2s/\bar{e}$ (%)		
		<u>HC</u>	<u>CO</u>	<u>NO_x</u>
65-67	126	15	9	9
68	77	23	12	11
69	88	13	10	9
70	99	29	10	7
71	113	22	13	7
72	176	16	14	6
73	128	9	16	10
74	193	10	9	7
75	587	8	8	3

<u>Year</u>	<u>N</u>	$2s$ (g/mi)		
		<u>HC</u>	<u>CO</u>	<u>NO_x</u>
65-67	126	1.34	9.32	.257
68	77	1.53	10.2	.406
69	88	.776	8.21	.365
70	99	1.54	6.40	.267
71	113	1.17	7.06	.250
72	176	.678	7.34	.223
73	128	.314	7.14	.286
74	193	.341	3.70	.201
75	587	.085	1.94	.083

Table 4: Sample Calculation of Uncertainty for CO in 1975
Due to Base Emission Factor Uncertainty

<u>Year</u>	<u>W</u>	<u>e</u>	<u>2 S (e)</u>	<u>W² (2S(e))²</u>
pre 68	.155	109	9.32	2.08
68	.072	82.6	10.2	.539
69	.083	78.5	8.21	.464
70	.086	63.9	6.40	.302
71	.106	52.7	7.06	.560
72	.122	51.8	7.34	.801
73	.125	45.3	7.14	.796
74	.135	41.8	3.70	.250
75	.116	22.9	1.94	.051
		<hr/>	<hr/>	
		$\bar{e} = 60.7 \text{ g/mi}$	$2S(\bar{e}) = 2.42 \text{ g/mi}$	

$$2 S(\bar{e})/\bar{e} = 4.0\%$$

cities. This assumption has not been verified; the data are insufficient to either accept or reject it as a statistical hypothesis. There are reasons to expect that "city effects" do exist. For example, air conditioner usage is probably much higher in Houston than Seattle, and vehicle maintenance is presumably better than average in places which have mandatory emissions inspection and maintenance programs. City effects could be quite large and still remain undetected, since the number of vehicles tested in any one city is relatively small. If the possibility of city effects is allowed, then uncertainty in aggregate emissions for any given city is probably two to four times the figure originally calculated.

Data for base evaporative emissions are somewhat sketchier, and so warrant a crude approach here. Reference 5 gives data for diurnal and hot soak evaporative emissions, and the respective standard deviations, for 126 Los Angeles vehicles. For this pooled data, using the method of AP-42 supplement 5³ to convert to g/mile figures, the uncertainty in the average is about .14 g/mi or 6%. This is for an average evaporative emission rate of 2.35 g/mi, and so adds a few percent to uncertainty in aggregate HC emissions for uncontrolled and partially controlled vehicles. Fuel tank running losses add uncertainties of the same order.

Crankcase HC emissions are usually estimated to be eligible for post-1967 vehicles (post-1963 for California). For uncontrolled vehicles, emissions are about 4.1 g/mi, but it is not clear how accurate this number is--it is referenced in AP-42 to an unpublished internal EPA report. It is possible only to speculate about the merits of the assumption that current crankcase HC controls are completely effective; there is evidence⁶ that many PCV systems are partially or fully clogged on in-use vehicles, but no measurements of resultant emissions could be found in the

literature. If the average effectiveness were 90% instead of 100%, this would result in HC emissions as large as the statutory exhaust standard.

In recognition of the fact that emissions in normal operation differ systematically from those measured by the 1975 FTP, several "correction factors" have been developed. These are applied as multiplicative factors to modify the base emission factor and, as currently used, attempt to compensate for deviation from test procedure values of three parameters: average driving cycle speed, ambient temperature, and proportion of starts from a fully-cooled down state.

Uncertainties in speed correction factors can arise in several ways. These factors are derived by using the modal emission model⁷ and a number of representative driving cycles at various average route speeds to calculate what the emissions would be over that cycle. The calculated values are fit, for each class of vehicles, to a simple function of route speed and normalized to form correction factors. It is assumed that the factors, which strictly speaking only apply to hot start emissions, can be used to correct cold start values. Uncertainties can arise from this assumption, from differences in actual and calculated emissions, and from unexplained variance in the fitting process. The latter is the smallest source, amounting to only about 3-4% for HC and CO, and 5-6%⁸ for NO_x. This goodness of fit may just be an artifact of the modal model, however, which would produce numbers lacking the "noise" found in real data. Unfortunately, discrepancies between calculated and actual emissions are about the same size as those between replicate testings, and so it is hard to estimate this uncertainty.

Temperature correction factors attempt to compensate for the substantial change in emissions as ambient temperature varies over the range encountered in use. Current factors³ are based on only a small number of vehicles--26--and are unreliable at high temperatures. From the raw data⁹ the uncertainty in the correction factor can be estimated. The uncertainty in these factors, and the hot/cold start correction, will be presented in the final version of this report.

2.3 Uncertainties in Predicting Aggregate Emissions

The accuracy with which emission can be predicted for future years depends strongly on the time period of the forecast. For short run predictions, the uncertainties are roughly the same as the measurement uncertainties, while for long run forecasts, they become difficult to estimate since the configuration of future vehicles is simply not known. Assuming a given, fixed set of standards which are met by new vehicles, the following factors (each with its associated uncertainty) are needed to calculate aggregate emissions:

- 1) emissions of current cars in future years (i.e., the deterioration of current emissions)
- 2) base emission factors of future new cars (since they usually will exceed the standards when new)
- 3) deterioration of emissions in use of future new cars
- 4) deficiencies of the 1975 FTP for use with future vehicles (e.g., temperature, driving cycle, altitude effects)
- 5) number of vehicles of each model year in the vehicle fleet; (characteristic driving patterns for each age of vehicle)

Each of the above factors is uncertain, but lack of data (caused by a lack of future vehicles in hand to test) makes estimation of the uncertainties very difficult. In addition, some of the individual uncertainties are functions of the forecast period. Predictions are often made for 10 to 15 years into the future. Uncertainties in these

predictions are strongly dependent on uncertainties about cars which do not yet exist, and could easily be a factor of two or three. For example, after less than one year of operation (the average odometer reading was 8,800 miles) 1975 vehicles were emitting 22.9 g/mi of CO on average (Denver and Los Angeles excluded)⁴ as compared to the 49 state standard of 15 g/mi. Considering that manufacturers normally target low mileage emissions at about 1/2 of the standard in order to assure themselves of meeting the durability requirements of the certification test (intended to insure that the standards are still met after 50,000 miles), this is rapid deterioration. It was not predictable, nor, in the context of current certification procedures, preventable.

Similarly, there are problems, based on limited data, with the sensitivity of catalyst-based emission controls to ambient temperature, cold starting, maintenance, etc. In each case, the conclusion would be that the effects are large but differ among specific systems. What is clear from this was also implicit in the analysis of current emissions: meeting a given standard based on the certification test procedure provides only a limited amount of information about what emissions actually will be in use. Too many important things happen in the field that are not captured by the test procedure. It is dangerous to assume that a numerical standard implies that cars will emit at that level. Though that may be the best estimate available, it is not likely to be a very accurate one.

While uncertainties in current aggregate emissions are limited mainly by the size and scope of surveillance activities, it seems that in the medium to long run predictions of aggregate emissions are inherently much more uncertain. When emission standards become fixed and control

systems settle into a few predictable configurations, this may no longer be true. However, that day still lies far in the future.

Given the large uncertainty in emissions from future vehicles, it appears that extensive field surveillance combined with mechanisms for fast and flexible regulatory response are necessary to assure, on a continuing basis, that emissions goals are being met in use. A good portion of the surveillance should be devoted to further investigating the known weaknesses of the 1975 FTP in reflecting actual emissions. In addition, an ongoing program of exploratory research should be established to probe for test procedure problems not now recognized and to look for unanticipated side effects of the regulations--e.g., sulfate emissions.

3. Uncertainties in Measuring and Predicting Trends in Air Quality

This section will investigate uncertainties in measuring and predicting changes in air quality (links 1 and 4 of figure 1). First, some of the complications involved in realistically defining "air quality" will be discussed. Next, some of the general limitations on ability to measure air pollution trends will be examined. A more detailed case study of trends at one site is deferred until Section 4. Finally, uncertainties involved in using models to predict air quality will be investigated. Since it is by far the most widely used of these models, the rollback model will be singled out for detailed study.

3.1 The Definition of Air Quality

Prior to analyzing "air quality," it is necessary to decide what the term means. Obviously "air quality" is closely related to air

pollutant concentrations in an area, but it is far from obvious exactly what characteristic of a set of concentration data one might single out as directly reflecting air quality. A good part of the difficulty arises from the fact that there are many questions that one might want to answer based upon the data, and each kind of question gives rise to a different summary statistic. For simplicity, attention can be limited to two questions:

- 1) What is the estimated rate of improvement (or degradation) in public health?
- 2) What is the estimated rate of reduction in emissions?

As was previously noted, the first question gives rise to some fundamental problems--those of assessing population exposure and dose-response relationships--for which there is no neat answer. For regulatory purposes, air pollution standards have arisen as the baseline against which health impact is assessed. Since the degree of compliance with the standards is such a widely-used method for summarizing air quality data, it is important to understand exactly what the standards are and how data are interpreted with respect to them. The following subsections summarize this information.

The legal definition of air quality

The Clean Air Act¹⁰ as amended by the Clean Air Amendments of 1970¹¹ required the Administrator of EPA to promulgate, within 120 days, national, primary and secondary air quality standards for all "criteria pollutants"--those for which air quality criteria had been previously published. Primary standards were to be those:¹²

"...the attainment and maintenance of which in the judgment of the administrator based on such criteria and allowing an adequate margin of safety, are requisite to protect the public health."

Secondary standards were to specify a level of air quality¹³

"...the attainment and maintenance of which...is requisite to protect the public welfare from any known or anticipated adverse effects associated with the presence of such air pollutant in the ambient air."

On April 30, 1971, the ambient air quality standards were promulgated.¹⁴ The standards refer to average concentrations over various time periods. For averaging times less than 1 year, the standards are "...not to be exceeded more than once per year,"¹⁴ while for annual averages they are presumably not to be exceeded at all. In these regulations, "ambient air" is defined as:¹⁵

"...that portion of the atmosphere, external to buildings, to which the general public has access"

The regulations also specify in detail the acceptable "reference methods" for measuring pollutant concentrations, while allowing the administrator to approve "equivalent methods."

To determine compliance with the standards, some form of monitoring is of course required. Congress left the details up to the judgment of the EPA and the states, via the state implementation plan mechanism, requiring that the implementation plan include:¹⁶

"...provision for establishment and operation of appropriate devices, methods, systems, and procedures necessary to monitor, compile, and analyze data on ambient air quality."

This was spelled out in more detail by EPA in the regulations for state implementation plans.¹⁷ The regulations set forth requirements (for each pollutant) for the number of monitoring sites in a given region, the measurement method, and the sampling frequency. It was also required that "At least one sampling site must be located in the area of estimated maximum pollutant concentration."¹⁸ The clear presumption was still that standards were not to be exceeded anywhere in a region's ambient air--and that compliance was to be determined by monitoring at the most highly polluted locations.

As written, the standards seem to provide an absolute level of protection--no one is to be exposed to levels at which there are detectable health effects. To the extent that people do not frequent the most highly polluted areas, the populace enjoys an extra margin of safety. The thorny problem of estimating population exposure--how many people get exposed to how much pollution for how long--seems to be moot. But is it really? The site of "maximum estimated concentrations" is not well-defined. Somewhere between, say, the exhaust pipe of an automobile and a sidewalk or park bench the polluted air mass becomes "ambient air" as legally defined. But exactly where this point occurs is primarily a matter of judgment, and in practice a site would not be placed too near, for example, an expressway on the grounds that such a site would not be "representative." So, in fact, the oft-heard remark that "if you can't meet the carbon monoxide standard at one station site, try moving it across the street" has some reflection in practice, and the seemingly simple link between meeting

standards and achieving public health goals is really not so simple.

Another conceptual difficulty arises in insuring that the standards are met. Legislative intent here is clear; implementation plans were to:¹⁹

"...include(s) emission limitations...as may be necessary to insure attainment and maintenance of such primary or secondary standards..."

But air pollution levels are governed by both emissions and meteorology, and the latter is for all practical purposes a stochastic quantity. It is impossible (short of reducing emissions to zero) to insure that standards will be met; one can at best make some statistical statement as to the likelihood of this event.

This is recognized to some extent in practice. Various schemes exist for estimating how the frequency distribution of pollutant concentrations will change in response to controls. But once it is recognized that measured levels represent one realization of a dynamic stochastic process, it is natural to go one step further and try in some way to estimate the parameters of the underlying process. While perhaps some statistic such as frequency of exceedance of the standard can be used to estimate the parameters, its lack of robustness makes it suspect. Thus, even if the air quality standards are chosen as the baseline for interpretation of air quality data, the choice of appropriate summary statistics is worth detailed consideration.

EPA Interpretation of Air Quality Standards

To complete the discussion of air quality standards, it is worthwhile to outline the way field data are actually interpreted

with respect to the standards. This operational definition, laid out by EPA guidelines²² forms the real basis for understanding the detailed significance of meeting or not meeting the standards in practice, and in it can be seen an attempt to deal with some of the problems previously mentioned. Since the points are technical and individually somewhat minor, they will be briefly summarized.

- 1) Averaging time. The standards do not specify whether compliance is to be based upon moving averages, or averages over specific time intervals. For data processing convenience, guidelines specify that the clock hour be the finest subdivision of measurement, that all 24 hour measurements be computed on the basis of a calendar day, and that the term "day" refer to a calendar day, and that the term "year" refer to a calendar year (both for averaging and for assessing compliance with standards "not to be exceeded more than once per year"). The exceptions to this general principle are SO_x and CO, which have a 3 hour and 8 hour standard, respectively. For these pollutants, a running average is used.
- 2) Scope of compliance. For assessing compliance, the question arises whether compliance is to be based on each site individually or on aggregate data from an entire urban area. EPA bases the assessment on all of the sites in an area, but "once per year" standards are deemed to be violated only if a given site has more than one excursion over the standards.

- 3) Sampling frequency. Especially for those pollutants monitored every sixth day, it is possible to predict (at some level of statistical confidence) that the numerical standard is being exceeded more than once per year even though no violation was actually detected. The guidelines here indicate that, while noncompliance will not be declared on such a basis, neither necessarily will compliance, and more frequent sampling may be required.
- 4) Siting. Beyond the requirement of one monitor at the site of "estimated maximum pollutant concentration," site location is left up to the states. In late 1975, EPA published guidelines,^{20, 21} which provide a unifying framework for site selection. The guidelines in general go into the factors which influence site selection for different pollutants and different monitoring objectives, and they are not binding upon state officials. Siting historically has been primarily a subjective and highly judgmental decisions, and to date there has been little uniformity in site selection. In general, monitors tend, however, to be located at "representative" sites, and care is taken (for long-term monitoring) to avoid sites which are subject to a heavy "local" influence--such as the previously mentioned expressway.

The need for better measures of air quality

The intent of this discussion is to show that it is necessary to be very careful, both in evaluation models and in analyzing trends, in deciding exactly what is meant by "air quality." Measured pollution levels will not reflect the worst that can be found or the best, but

some indeterminate middle ground. It is not feasible to provide the entire population with the level of protection which seems, on the face of it, to be guaranteed by meeting the air quality standards. Pollution patterns in a real city are complex, and are not well summarized by a few peak readings at a handful of sites.

In addition, it should be noted that some confusion arises because the term "air quality" is often used in different senses when discussing prediction and measurement. In the former case, air quality is of necessity a statistical quantity, a statement about what the underlying process governing air pollution levels is like. For the latter case, however, often the summary statistic chosen as measure of air quality is interpreted as directly indicating air quality, rather than merely being an estimate (with its own associated uncertainty) of the parameters of the underlying stochastic process. The confusion disappears when air quality is recognized throughout to have a random component.

Once it is realized that there is a strong random component in air pollution measurements, it is clear that for a given site the choice of a particular summary statistic is important. Peak values, even aside from the danger that they are erroneous (e.g., result from an instrument malfunction), are probably not appropriate for use in evaluating performance of predictive models or estimating trends. If one were to propose that air quality be measured by recording ambient pollutant concentrations for a randomly selected one hour period during each calendar year, most people would reject the proposal immediately. Most scientists and engineers, trained to work with data have the same reaction to usage of peak values. In each case the reason is qualitatively the same--the statistic has a large random component and could be dangerously

misleading if used uncritically.

3.2 Uncertainties in Measuring Air Quality Trends

Uncertainties in measured air quality trends are a function both of the data and the methodology used. "Data" here means some summary statistic used as a measure of air quality, and as previously discussed it is not easy to determine a realistic measure. A review of the literature, summarized in Table 5, reveals the divergence of opinion on this point. A wide variety of summary statistics, with different scales of spatial and temporal averaging have been used in analyzing air pollution data. In addition, there seems to be no consensus on the "right" methodology to use; some authors are content to deal with an annual summary statistic as a simple function of time, while others use more complicated statistical techniques.

To examine in general the uncertainties in trend estimation, it is necessary to pick both a summary statistic and a methodology. The methodologies used by EPA in its annual trends report^{23,24,25,26} are not useful for this analysis, since they are not "statistical"--they neither provide one with a numerical trends estimate nor with an estimate of uncertainty in the conclusion about trends. Choice of summary statistic is difficult; however, since it is deterministic trends (caused by emissions controls) which are of primary interest, and since under certain conditions (to be discussed in the section below) concentrations are linearly related to emissions, the average concentration over some time period is an appropriate choice. If desired, methods exist⁴⁰ to relate this statistic to the probability of exceeding the standard.

Table 5: Trend Analysis Methods and Empirical Models

Ref.	Aggregation	Pollutant	Summary Statistic	Method
26	National	TSP, SO ₂	<ul style="list-style-type: none"> 10th, 50th, and 90th percentiles of quarterly site means and maxims Number of sites reporting change > 10% 1970 70 to 1972-74 	<p>Plots of data and quarter moving averages vs time</p> <p>Histogram of number of sites showing decrease, no change, increase.</p>
	Regional	TSP, SO ₂	Quarterly means	Plots of data vs. time
	Selected cities and states (5)	CO	<ul style="list-style-type: none"> Average over sites of 2nd highest 1 hour concentration Average % of annual 8 hr. levels > 8 hour standard Annual average of daily max 1 hour concentration 	<p>Table of statistic vs. year</p> <p>Table of statistic vs. year</p>
	Regional (2: L.A. & S.F. Bay)	O _x	<ul style="list-style-type: none"> Average over sites of 2nd highest 1 hour level for year Average yearly no. of values above 1 hour standard Average daily max 1 hour concentration 	<p>Plots of data and 4 quarter moving average vs. time</p> <p>Plot of statistic vs. time</p> <p>Table of statistic vs. year</p> <p>Plots of data and 3 year moving averages vs. time</p>

Table 5 (continued)

Ref.	Aggregation	Pollutant	Summary Statistic	Method
		HC	Annual average daily 6-9 A.M. concentrations	Plots of data and 3 year moving averages vs. time
	City (3)	NO ₂	<ul style="list-style-type: none"> Annual average 6-9 A.M. level Annual average daily max 1 hour concentration 	<ul style="list-style-type: none"> Plots of data and 3 year moving averages vs. time Plots of data and 3 year moving averages vs. time
27	Unspecified	any	Averages or max's suggested	Plot of moving average; Spearman rank correlation; regression vs. time; analysis of variance of % of levels \geq standard
28	site/city	SO ₂	Average over all conditions and sites of ratio to base year of seasonal (summer/winter) average	Table of data vs. year
29	Site	CO	<ul style="list-style-type: none"> % of time level \geq 8 hour standard Monthly max 8 hour average hourly average 	Table of data vs. year
30	Site (L.A.,	O ₃ , CO NO _x , SO ₂ , HC _x	Log of bi-weekly average daily max 1 hour concentration	<ul style="list-style-type: none"> Trend envelope Multiple linear regression Box-Jenkins Model

Table 5 (continued)

Ref.	Aggregation	Pollutant	Summary Statistic	Method
31	Site	O _x , CO, NO ₂ , HC	Log of weekly average daily max 1 hour con- centration	Box-Jenkins Model
32	Site (L.A.)	O _x	Log of weekly average daily max 1 hour con- centration	linear regression on log of meteorological variables and (sometimes) lagged O _x
33	Site (7)	CO	Monthly Average	(0,0,1) x (0,1,1) ₁₂ with intervention term As above, with summer and winter trend terms
34	City (7 sites)	SO ₂	Monthly average	(1,0,1) x (0,1,1) ₁₂ with 2 trend and 1 intervention term
35	City (10 sites)	SO ₂	. 5 year average of week- day (Mon.-Thurs) levels by hour of day	Nonlinear regression on wind speed and exponentially smoothed traffic
36	Site (30 in St. Louis)	SO ₂	. As above, but summer daytime	as above, with inversion height; traffic not smoothed
			Winter Average	linear regression on tempera- ture, rain, wind speed
			Winter Average	As above; also linear fit to time with correction term for meteorology
			Daily Average	principal component descrip- tion of spatial variation; re- gression on Met factors to fit

Table 5 (continued)

Ref.	Aggregation	Pollutant	Summary Statistic	Method
37	Site	SO ₂	Daily Average	temporal variation nonlinear model with Met data and lagged SO ₂ term

Estimates of both aggregate automotive emissions and average emissions per vehicle show that a downward, nearly linear trend is expected for automotive pollutants for the period 1970-1980. An efficient trend estimate in such a situation is simple linear regression against time. It is desirable to eliminate the effects of meteorological variations (which affect dispersion) from the trend estimate, so that emissions trends are distinguished from random climatological trends. In what follows, it will be assumed that this has been done in some appropriate way (e.g., by regression on principal components).

Assume that by some model, characterized by a goodness of fit R^2 , the available concentration data has been reduced to the form:

$$y_i = \alpha t_i + e_i \quad (1)$$

where:

y_i = the residuals of the model with $E(y) = 0$ and $V(y) = \sigma_y^2$

α = the trend term

t_i = the time of the i^{th} observation, with $E(t) = 0$

e_i = a series of independent identically distributed random shocks with mean zero and variance σ_e^2

The least squares estimate for α is

$$\hat{\alpha} = \Sigma y_i t_i / \Sigma^2 t_i \quad (2)$$

and the variance of this estimate is

$$v(\hat{\alpha}) = \sigma_e^2 / (n V(t)) \quad (3)$$

where n is the number of data points. If the data are complete for a time period, T , then it can be shown that

$$V(t) = \frac{(n+1)}{12(n-1)} T^2 \quad (4)$$

Let a measure of significance (i.e., accuracy of trend estimate) be defined as

$$k^2 = \alpha^2 / V(\hat{\alpha}) \quad (5)$$

and note that

$$V(y) = \alpha^2 V(t) + V(e) \quad (6)$$

Combining the above four equations and assuming $n \gg 1$ yields:

$$\alpha = \frac{\sqrt{12} \sigma_y}{T(1 + n/k)^{1/2}} \quad (7)$$

or alternatively,

$$\alpha = k \frac{\sqrt{12} \sigma_e}{\sqrt{n} T} \quad (8)$$

The above equations contain five parameters, -- α , k , n , T , and either σ_y or σ_e -- four of which must be specified to solve for the fifth. In practice, two questions are of interest:

- (1) What is the minimum measurable trend, for given significance level, time period, etc.?
- (2) What is the minimum necessary time period over which data must be collected to measure a trend of the expected magnitude, all else given?

To facilitate the computation, let two levels of significance be defined. Let trend detection be characterized by $k = 2$ -- which corresponds, if the e_i 's are distributed roughly normally, to a trend significantly different from zero at the 95% level of confidence. Similarly, let $k = 10$ be defined as the desired level for trend measurement -- i.e.,

an estimate accurate to about $\pm 20\%$ with 95% confidence. Furthermore, restrict attention to 24 hr. average concentrations, and assume that all data are available so that $n = 365T$ with T in years. With these simplifications, it is possible to answer the first question in terms of the minimum "signal to noise ratio," represented by α/σ_e , required to either detect or measure a trend. Results of this calculation are shown in Table 6.

As can be seen, several years of data are required merely to detect a trend, unless the trend is comparable in size to the variability of the original data (less that portion due to the trend). And detection, as defined here, really only amounts to a determination of the sign of the trends--upward or downward. Of more interest is the size of the trend, evidently this is much more difficult to determine.

The above discussion can be made more concrete by way of an illustrative example. For Kenmore Square in Boston during 1974, average CO was about 6 ppm with a sample standard deviation (based on 24-hour averages) of 1.6 ppm. Based upon aggregate automobile emissions of CO, about an 8% per year decline in ambient levels would be expected over this period, or about .5 ppm. This corresponds to $\alpha/\sigma_y \approx .3$, which, from Table 6, should be detectable (assuming that it exists) in one year but which would require over three years to measure to a nominal $\pm 20\%$ accuracy.

In general, similar results would be expected for most cities. Pierrard et al.⁽²⁹⁾, in examining CO data from 6 CAMP cities, found that the sample standard deviation (based on 8-hour averages) tended to be about one-half of the annual average. Expected trends are rarely more than -10%/year, and so a typical ratio of trend to noise is .2-- which would imply that at least four years of data are required for trend

Table 6: Minimum Signal to Noise Ratios for Trend Measurement

T years	$(\alpha/\sigma_e)^*$ detect (yr ⁻¹)	$(\alpha/\sigma_e)^*$ measure (yr ⁻¹)	$(\alpha/\sigma_Y)^{**}$ detect (yr ⁻¹)	$(\alpha/\sigma_Y)^{**}$ measure (yr ⁻¹)
1	.363	1.813	.256	1.606
2	.128	.641	.128	.601
3	.070	.349	.070	.334
4	.045	.227	.045	.219
5	.032	.162	.032	.158

* from equation (8), with $n = 365T$

** from equation (7), with $n = 365T$

measurement.

An element not yet considered is the model used to go from the original data to the y_i . So far, it has been implicitly assumed in the numerical examples that the model did not eliminate any of the original data variance, but naturally much more processing could have been done. This would lead to two effects: a σ_y which is less than that of the raw data, and the possibility of collinearity between time and some explanatory factors in the model. If the latter possibility holds, then the confidence interval around the trend estimate will be even wider than heretofore assumed, since, loosely speaking, explanation of the trend "signal" would have to be shared among the factors correlated with time. In addition, the trend estimate would have to be altered. However, to take all of this into account would require a given model and some detailed knowledge of the factors, and generality would be lost. The calculations done here should, therefore, be viewed as generally indicative, rather than rigorously quantitative. In actually estimating a trend for a given site, of course, the fitting would be done in a single stage and all of these effects would be properly accounted for. Section 4 will provide an example of this.

It is instructive to look roughly at how the goodness of fit of the model affects the accuracy of the trend estimate, since there is a limit to how well a model can, in fact, fit the data, which is of limited accuracy. If model fit is summarized by the R^2 statistic, then an effect of removing exogenous influences from the data is to reduce σ_y by a factor of $\sqrt{1 - R^2}$. To reduce by half the amount of time necessary to detect a trend would require that $R^2 \approx .75$. This is a

fit characteristic of the best current predictive models; actual models for use in this context are likely to be significantly less good. The best possible fit is limited by inaccuracies in the measured data, which for CO are typically ± 1 ppm⁴¹. If this is interpreted as a 95% confidence interval, and taken as applicable to 24-hour average data, then $(\sigma_y)_{\min} = .5$ ppm, which for the 1974 Boston data implies $(R^2)_{\max} \doteq .9$. Even for a perfect model, a trend is therefore probably not measureable without more than one year of data.

Another factor which can vary is the averaging time. Use of daily averages is convenient for data analysis purposes, since systematic diurnal variations due to, for example, traffic patterns, are suppressed. The autocorrelation and noise structure of the data are also more complex at short averaging times, leading to modeling difficulties. If most of the contribution to variance by the noise were concentrated at high frequencies, then long averaging times could be advantageous since sample variance would then tend to decrease faster than would be the case with independent data points. Unfortunately, the opposite is true. Larsen⁴⁰ has shown empirically that maximum concentrations decrease with averaging time more slowly than would be expected from serially independent data. Others⁴² have found that the noise spectrum fits a Markovian form, implying an exponentially decaying autocorrelation function with noise concentrated at low frequencies. Short averaging times are therefore advantageous, since the improvement in resolution from use of more data points more than offsets the associated larger sample variance.

In summary, a record containing several years of data will usually be required to measure trends of the size expected. Use of a model relating the data to exogenous factors may or may not reduce this requirement. A model is, of course, necessary if the trend is to be interpreted as due to changes in emissions or due to changes in average meteorology. There are definite limitations on how well a physically realistic model can describe measured air quality, since the measurements are themselves subject to random error. However, this limitation is well beyond the performance of the best current predictive models, whose performance is, in turn, better than could be reasonably expected from a simple statistical model. In principle, use of an averaging time shorter than one day could allow improved trend estimation; in practice, the drastic increase in model complexity necessary to eliminate exogenous influences, to deal with the discrete nature of existing data at short averaging times, and to whiten the residual series probably makes this option unrealistic.

3.3 Uncertainties in Prediction of Air Quality

There is a multiplicity of models for predicting air quality, and this is due largely to the multiplicity of questions which the models address. Several applications can be listed:⁴¹

1. research
2. environmental legislation
3. implementation planning
4. impact assessment
5. transportation and land use planning
6. episode control systems

Some models are even more specialized, dealing with only a single pollutant in a specific kind of situation.

In looking at model uncertainties, it is necessary to be clear about both the kind of uncertainty and the sources of the uncertainty. For simplicity, uncertainty can be divided into two kinds: descriptive and predictive. Descriptive uncertainty is a measure of the error of a model in estimating pollutant levels for a past year, commonly using the known meteorological conditions for that year and usually "calibrated" based upon an empirical fit of the estimates to what was actually measured. Predictive error, on the other hand, is the uncertainty in a prior estimate of future air quality. Obviously, since future weather, emissions, etc. are now known with certainty, the forecast error will be larger than the descriptive error. This is important to recognize, since most model validation is done in terms of descriptive error.

Related to this is the fact that not all of the error, either descriptive or predictive, is due solely to the model itself. Emissions inventories, the basic input for any model, are far from perfect, either in their estimation of the spatial distribution of emissions or in that of how emissions change with time. Meteorological data, used extensively by some models, do not usually provide adequate resolution and are often taken at a site outside of the core city, raising questions of their representativeness. Pollutant concentration data, used to evaluate the errors, are not perfect either. Problems arise from instrument inaccuracies, from changes in calibration and maintenance procedures, from poor choice of sampling probe location, and so on. Finally, there may be inherent limits on the accuracy with which pollutant levels can be estimated. On the theoretical side, pollutant dispersal is governed by turbulence, which implies that there will be random fluctuations in concentration at short enough time scales. Only statistical properties of these fluc-

tuations can be estimated, at best. On the practical side, stochastic variations in source emissions are prohibitively expensive to monitor, and of course are not obtainable for future years.

All of these sources of error are over and above the error due to model deficiencies, and so in a way are inherent to any model. A fair comparison of models would either account for these influences directly, or eliminate them indirectly by basing evaluations on the same set of given data. Although the former approach is the more difficult, the latter approach has not been attempted on a large scale.

Almost all applications of models to regulation of automotive emissions, however, have used the rollback model. The original 1975 standards were based upon it⁴³, and it still remains the primary tool for analyzing regulatory impact^{44,45}. Since it is so widely used, the rollback model will be examined in detail here.

The basic premise of the rollback model is simple: that the level of pollution measured in the air is proportional to the rate of introduction of pollutant emissions into the air. This intuitive approach, however, fails to totally take into account the many complexities of pollutant dispersal and transformation--the relative effects of nearby and far away sources, effects of meteorology, effects of photochemistry, etc. There is thus some uncertainty in rollback projections, and it is worthwhile to attempt to quantify these uncertainties to see if they are indeed important.

A reasonably general form of the rollback equation is usually used.^{46,47} The resultant model is of the form:

$$\frac{X_j - \beta}{X_0 - \beta} = \frac{\sum_{i=1}^n (K_i Q_i) (G_{ij} F_{ij} T_{ij})}{\sum_{i=1}^n (K_i Q_i)} \quad (9)$$

where:

X_j is the projected air quality for year j

X_0 is the base year air quality

β is the background air quality

Q_i is the base year source strength for source category i ;
mass emissions per unit time

K_i is the emission source height factor category i , correcting
for stack height

G_{ij} is the growth factor for category i in year j , equal to
fractional increase in number of category i sources over
base year

F_{ij} is the emission factor ratio for category i in year j

T_{ij} is the transportation control factor (if applicable)
for source category i in year j , correcting for reductions
in aggregate vehicular emissions resulting from control
strategies aimed at, for example, reduction in total vehicle
miles traveled.

Although equation (9) seems complicated, most of this complication arises from the explicit calculations on the right hand side of an "effective" aggregate emission rate from its components. Since this disaggregation is unimportant for the present discussion, the following simplified version will be used where possible:

$$X_j - \beta = (1 - R_j) G_j (X_0 - \beta) \quad (10)$$

where:

R_j is the fractional reduction in emissions per unit source
(from base year values)

g_j is the appropriate aggregate growth factor in number of sources
for year j as compared to the base year

X_j is the predicted air quality in year j

Of course, X_0 , β , and S must be expressed in terms of the same averaging time. Often X_j is set at the level of the appropriate air quality standard, S , and eq. 10 is solved for the necessary emission reduction R_j .

To avoid confusion, it should be noted that several authors^{43,49,41} use the formula:

$$R_j = \frac{g_j X_0 - S}{g_j X_0 - \beta} \quad (11)$$

This equation, while apparently widely used, is not consistent with eq. 9 and leads to values of R_j somewhat higher than those derived from eq. 10. Something like eq. 11 could be derived by assuming that the background concentration β grows in time in a specific way, but it seems better to separately examine the uncertainty due to the assumption of a constant background level, and so for consistency this discussion will be based on eq. 10.

Rollback derives its validity from four assumptions about the situation to which it is applied. They are:

- (1) The relative distribution in space and time of emissions is approximately constant.

- (2) The pollutant is either inert in the atmosphere or decays via a first-order chemical reaction.
- (3) Background pollutant concentrations are either constant or negligible.
- (4) Important meteorological factors do not vary over the projection period.

Each of these assumptions will be discussed in turn.

1. Errors due to inadequate treatment of spatial distribution.

If all emissions in an area were reduced by a uniform percentage then, all else held constant, pollutant concentrations would drop by a like amount. However, a uniform reduction is not the typical situation, and so it is necessary to consider the possible effects on a rollback calculation of a redistribution of relative emissions among sources at different distances from the receptor and different heights above the ground.

Carbon monoxide is a good pollutant to use in looking at the effect of spatial redistribution, since it is primarily automotive in origin and most of the nonautomotive sources are near ground level. Many investigators have argued that CO is a "local" pollutant, and measured levels mostly reflect the influences of nearby sources. They usually argue that the growth factor g in eq. 10 should be adjusted downward to reflect the fact that traffic is saturated in the core areas of cities, where the highest CO levels are measured. Pierrand et al.²⁹ suggest a constant value $g=1$ should be used in place of the usual assumption of a 5%/year (noncompounded) growth rate based on vehicle registrations. Chang and Weinstock,⁵⁰ based on a mesoscale diffusion model and a simple growth model, suggest a worst-case growth rate of half that of the aggre-

gate growth rate. A National Academy of Sciences panel report⁵⁷ recommends a growth rate of 3%/year (noncompounded) for the period 1970-1988. The Federal task force report⁴⁴ uses a value of 1%/year for light duty vehicles; EPA⁴⁵ uses city-specific values ranging from .5% to 3% per year.

It seems reasonable to consider separately the uncertainties due to uncertainty in the aggregate growth rate--which are reflected in uncertainties in the aggregate emissions estimate--and those due to effects of spatial redistribution of relative emissions. The latter uncertainties will be examined in three stages. First, an illustrative calculation will be done to show the size of the error incurred by not correcting for redistribution. Next, the uncertainty in a typical calculation with this correction will be estimated. Finally, the complications introduced by a "street canyon" situation will be discussed.

An estimate of the error resulting not accounting for spatial redistribution can be obtained using the simple model of Gifford and Hanna^{39,52}. For a receptor interior to an area source of constant source strength, their model can be expressed:

$$X = (2/\pi)^{1/2} r^{1-b} \{a(1-b)\}^{-1} Q/\bar{u} + \beta \quad (12)$$

where:

X is pollutant concentration at the receptor

r is the distance from the receptor to the edge of the area source

a,b are constants, which may vary somewhat with atmospheric stability

Q is the area source strength

\bar{u} is the average wind speed

β is the background air quality

A simple approach is to approximate the city as circular, with two zones: a high emitting central core (with radius r_1) and a lower emitting suburban ring (with radius r_2). The base year area source strengths are designated as Q_1 and γQ_1 respectively. If growth were to occur uniformly throughout the city, application of eq. 12 would result in the ordinary rollback equation eq. 10. The opposite case is that of all growth occurring external to the central core. In that case, application of eq. 12 yields

$$\frac{1}{(1-R_j)} \frac{X_j - \beta}{X_0 - \beta} = 1 + (g_j - 1) \frac{(\alpha^k - 1)}{(\alpha^2 - 1)} \frac{[1 + \gamma(\alpha^2 - 1)]}{[1 + \gamma(\alpha^k - 1)]} \quad (13)$$

where

$$\alpha = r_2 / r_1 \quad (14)$$

$$k = 1 - b \quad (15)$$

To apply this formula, some typical numbers are needed for γ , α , g_j , and k . The following statistics were extracted for Boston in 1970:⁵³

<u>Study Area</u>	<u>Emission Density</u> (kg/mi ² /day)	<u>Area</u> (mi ²)
Boston Core	19,269	5.18
East Boston	16,186	.47
Inner City	4,206	48.38
Inner Suburbs	3,508	195.76

Lumping together "Boston core" and "East Boston" to represent the core area, and "inner city" and "inner suburbs" for the surrounding area results in $\gamma \doteq .192$, $\alpha \doteq 6.65$. Gifford and Hanna³⁹ give values for k which vary with meteorological conditions as follows:

<u>Meteorological Conditions</u>	k
very unstable	.09
unstable	.14
neutral	.20
slightly stable	.25
stable	.29

The choice of a value for g depends upon the time scale of the projection. For 10 and 20 year projections, values of $g_{10} \doteq 1.5$ and $g_{20} \doteq 2.0$ are appropriate Table 7 shows the results of evaluating eq. 13 for various values of α , g , and k .

As is evident, "local" emissions are the most important in determining air quality at a central business district (CBD) site. Growth in surrounding areas cannot be neglected, however, even under assumptions most favorable to that simplification, without incurring an error of roughly 5% - 10% in the direction of underestimating ambient concentrations. Ignoring the correction for source-receptor distance would lead to much larger errors in the other direction, of about the same size as $g-1$ (50% and 100% roughly for 10 and 20 year projections).

It is clear that to be accurate rollback calculations must complement site-specific measure of air quality used in the formula with a site-specific emissions inventory at a sufficient level of disaggregation to permit correction for source-receptor separation. The early calculations⁴³ upon which the auto emissions standards were based did not account for this

Table 7: Worst Case Evaluation of Effects of Nonuniform Growth for Typical Parameter Values

γ	α	k	g	growth deflator*	effective growth factor**
.2	7	.1	1.5	.045	1.023
.2	7	.1	2.0	.045	1.045
.2	7	.25	1.5	.123	1.062
.2	7	.25	2.0	.123	1.123
.2	3.5	.1	1.5	.038	1.019
.2	3.5	.1	2.0	.038	1.038
.2	3.5	.25	1.5	.099	1.050
.2	3.5	.25	2.0	.099	1.099

* equal to the coefficient of $(g_j - 1)$ in eq. 13

** equal to the right hand side of eq. 13

effect; more recent ones⁴⁵ do to some extent. The size of the correction depends upon meteorological circumstances and so the appropriate choice of correction factor depends upon the base statistic chosen to represent "air quality" (e.g., maximum concentrations or long term averages). In the simple example given, a variation of about a factor of 3 in the correction resulted in only about a 10% change in effective growth rate, so uncertainties in the size of a similar correction using more detailed data are unlikely to result in errors of more than about $\pm 5\%$.

A complicating factor which has not yet been touched upon is the possible existence of a "street canyon" effect. The argument is that the trapping effect of tall buildings along an urban street result in street level concentrations much higher than would be predicted by assuming unconstrained diffusion, and so measured concentrations at street level strongly reflect emissions that are "local" on the scale of a few blocks. Ratios of "rooftop" to "curbside" CO concentrations of 10% - 40% have been cited.⁴⁷ Other investigators⁵⁴ have found that the relationship changes with time and meteorological conditions and that, based on limited data, no gradient of CO concentration with height existed at the time of highest CO readings. On the other hand, more extensive measurements in St. Louis street canyons⁵⁵ have shown that quite high (20-30 ppm) street level CO concentrations can be associated with low (.4-7 ppm) levels at 115m, but that the relationship changes from day to day. Several mathematical models^{55,56,57} predict a substantial impact on street level concentrations due to the canyoning effect, with the size of the effect directly related to meteorology and details of sampling probe locations.

The significance of this street canyon effect depends on the specific site and on the particular statistic used to characterize air quality. Typical urban sites at which high levels of CO are measured often will reflect this effect, however, and so the question is how to compensate for it. If the air quality statistic to be projected is a long term average or the like, the effect of canyoning is to reinforce the notion that only local emissions matter, where local is now defined as a region of a few blocks around the receptor rather than a few miles. Then, to the extent that long term meteorology fluctuates little and that local traffic growth is the same as that in the rest of the CBD, a growth factor the same as that of the CBD is appropriate. Emissions from sources outside the CBD, when corrected once for distance and again for the canyoning effect, are negligible for most realistic cases, and the correction can be roughly estimated.

A crude estimate of the uncertainty in projected air quality resulting from the uncertainty in the size of the street canyon correction can be obtained by going back to the distance corrections and factoring in an additional street canyon correction. If the effect of growth outside the CBD is to increase projections based upon the CBD growth rate by 4% to 12%, and if average rooftop to curbside concentrations range from .1 to .4 then the net correction would range from .4% to 4.8%. This is a high estimate of the effect of ignoring emissions outside the CBD; if the correction is made, uncertainty in the predicted air quality due to uncertainty in the correction factor should not exceed $\pm 3\%$.

If a short term statistic such as maximum 8 hour concentration is used as the measure of air quality, then the uncertainty in correcting for source-receptor distance and street canyon effects becomes greater.

The appropriate correction factors for a short term measurement are specific to the meteorological conditions existing at the time of the measurement, and are inherently more uncertain than average corrections. While in principle the requisite detailed measurements and modeling could be done to estimate the correction, it is unrealistic to suppose that this will ever be done on a routine basis. As a result, it is difficult to say just how much a high concentration is reflective of purely local conditions (e.g., a traffic tie-up caused by streetwork), how much is due to core city emissions (e.g., as a result of a low level inversion lasting several days), and how much is due to regional emissions (e.g., from a shopping mall which happens to be directly upwind of the site for a full 8 hours).

There are really three separate factors at work here. The first is that the appropriate corrections themselves are more uncertain in this case. If this extra uncertainty were a factor of 2 for each of the corrections, then the resultant uncertainty would be roughly +12% (based upon the previous estimate of +3%). The second factor is the autocorrelation of pollutant concentrations over short time periods (less than a few days). In simple terms, this just means that pollutant levels at short averaging times are related not only to emissions at that time but to pollution "left over" from previously. (Other factors also contribute to the autocorrelation). Rollback can be derived in a number of ways from the physics of pollutant dispersal, but always with the proviso that the averaging time be "long enough." Averaging times less than a few days probably are not; and so it is not clear conceptually how to account for the influence of the spatial distribution of sources on the size of a short term air quality measure.

Another problem with using a short term statistic such as a maximum value to represent air quality is the lack of robustness. A given reading may be high simply because of instrument malfunction, or because of an abnormal traffic situation, or due to any of a number of mistakes or unusual circumstances which make the statistic an inappropriate one for use in projecting air quality. There is no way to tell how often high measurements are in fact erroneous, and so this uncertainty cannot be quantified. It is widely acknowledged, however, that quality assurance is currently a significant problem at many existing monitoring sites.

The third dimension in the spatial distribution of sources is the height of the source above the ground. Emissions from smokestacks are diluted considerably before reaching the ground, and do not have the same impact on street-level air quality as a similar amount of emissions at ground level. These emissions are characterized as "point source" emissions rather than the more continuous "area source" emissions appropriate for describing vehicular emissions in an urban area. The air quality impact for point sources can be estimated by a Gaussian model:

$$X = \frac{Q_i}{\pi \sigma_y \sigma_z u} \exp \left(\frac{-H^2}{2\sigma_z^2} \right) \exp \left(\frac{-y^2}{2\sigma_y^2} \right) \quad (16)$$

where:

- Q_i is the source strength
- u is the wind speed
- σ_y is the horizontal dispersion coefficient
- σ_z is the vertical dispersion coefficient
- H is the stack height (plus plume rise)
- y is the distance from the plume center line to the receptor

Equation 16 has implicit many simplifying assumptions. It is also not of a form which permits direct comparison with previous results, which were based on area source models. However, it seems plausible that for most cases the proper approach is to ignore the impact of smokestack emissions on street level air quality. The parameters σ_y and σ_z are dependent both on meteorology (stability) and source-receptor distance r . From reference 58, it can be seen that both are (very approximately) equal to $r/10$. If this is true than height correction becomes unimportant for $50 H^2/r^2 \ll 1$ or $r \gg 7H$. At smaller values of r , the effect is to discount the impact. At large values of r , the distance correction is large. In either case, the impact is much less than that of a nearby, ground level source emitting at the same rate.

2. Errors due to inadequate treatment of pollutant reactivity.

Additional errors can arise from rollback because of the implicit assumption that the pollutant does not react in the atmosphere. For CO, this assumption is generally acknowledged to be a good one; however, it is less so for NO_x and HC and obviously inappropriate for O_x .

NO_x poses a problem for a variety of reasons. First of all, the ambient standard is for NO_2 rather than NO_x ($\text{NO} + \text{NO}_2$ primarily) and so it is NO_2 that is reported as monitored in ambient air. However, most of the NO_x emitted from vehicles is in the form of NO, which is then converted in the atmosphere to NO_2 in a characteristic time which is highly variable with circumstances. Both the NO and NO_2 then can further react to form various nitrated compounds. The ratio of NO to NO_2 varies both diurnally and seasonally,⁵⁹ and so one cannot be used as a proxy for the other.

If rollback calculations are based on total NO_x , rather than just NO_2 , then all of the comments about CO apply with almost equal force to NO_x . Though stationary sources are more important for NO_x than CO-- accounting for roughly half of the NO_x emissions nationwide--the spatial distribution of major stationary sources is such that, in the central city, vehicular sources dominate. The dispersion of nitrogen compounds as a class from vehicular sources should be exactly similar to the dispersion of CO, including street canyon effects. Therefore, it would be expected that observed NO_x could be almost totally ascribed to nearby vehicular sources, as was the case with CO. Of course, this conclusion would have to be examined on a site by site basis, but it would generally hold true for most CBD sites. Inspection of CAMP data tends to confirm this. Peak concentration of NO_x usually coincide with or follow shortly after hours of peak traffic; the diurnal pattern shows little variation among weekdays but is different on weekends and holidays.⁵⁹

If projections of air quality in terms of NO_2 are desired, based on emissions of total NO_x and measured NO_2 , then there is an additional level of uncertainty arising from the implicit assumption that the ratio of NO_2 to NO_x is constant. Since the NO_2 standard is in terms of an annual average, rollback calculations would ordinarily be based on this averaging time, and so the seasonal and diurnal variations in the ratio would be averaged out. In order to investigate the remaining uncertainty, NO and NO_2 data from four arbitrarily selected CAMP sites were examined, as shown in Table 8. It can be seen there that the ratio of NO_2 to NO_x is reasonably stable for each site, with sample standard deviations ranging from roughly 5% to 10% of the respective mean.

This does not mean, however, that the ratio can confidently be expected to stay constant in time. For example, if new car emission controls for NO_x substantially altered the NO/NO_2 ratio in the exhaust, then conceivably the effect of the controls on NO_x would not be reflected in ambient NO_2 levels as measured at street level. Basically, this is another reflection of the problem of how to weight the impact of nearby versus distant sources, with the added complexity of needing detailed information on the NO/NO_2 ratio of emissions from each class of sources and on the chemistry of NO oxidation and other reactions as it travels from distant sources. Without such information, projections are uncertain by some additional amount which, though probably substantial, is not readily estimable. That it is risky to assume this problem away, on the other hand, can be seen from Table 8-- Philadelphia and Chicago both seem to exhibit a consistent trend in the NO_2/NO_x ratio, in opposite directions, and in contrast to Washington DC and Cincinnati where the ratio seems stable.

In summary, for central city sites reasonably far from major point sources of NO_x , rollback projections of total NO_x should be as accurate as those of CO, all else equal. Since most of the NO_x in this case is attributable to automobiles and has had a short residence time in the curbside atmosphere, total NO_x can be approximated well by NO_2/NO_x ratio. The former uncertainty can be as large as $\pm 20\%$ for annual averages (based on 95% confidence limits of ± 2 standard deviations); the latter depends on details of the future control strategy, and so cannot be estimated in a general way.

Another problem with NO_x is that different measurement methods for NO_2 give different results. Reference 20 shows results for several

Table 8: NO/NO_x Ratios at Several CAMP Cities
 \bar{x} is the average ratio, s is the sample standard deviation

WASHINGTON DC			
Year	NO	NO ₂	NO ₂ /NO _x
62	.030	.030	.50
63	.039	.034	.47
64	.034	.037	.52
65	.032	.035	.52
66	.036	.035	.49
67	.047	.043	<u>.48</u>
			$\bar{x} = .50, s/\bar{x} = .046$
CHICAGO			
Year	NO	NO ₂	NO ₂ /NO _x
62	.104	.043	.29
63	.097	.041	.30
64	.100	.046	.32
65	.096	.043	.31
66	.101	.057	.36
67	.077	.050	<u>.39</u>
			$\bar{x} = .33, s/\bar{x} = .12$

Table 8 (continued)

CINCINNATI			
Year	NO	NO ₂	NO ₂ /NO _x
62	.032	.031	.49
63	.032	.030	.48
64	.038	.032	.46
65	.031	.035	.53
66	.042	.036	.46
67	.032	.028	.47
			$\bar{x} = .48, s/\bar{x} = .057$

PHILADELPHIA			
Year	NO	NO ₂	NO ₂ /NO _x
62	.041	.039	.49
63	.046	.039	.46
64	.045	.038	.46
65	.049	.037	.43
66	.059	.039	.40
67	.063	.043	.41

cities where annual average NO_2 was measured by the Griess-Saltzman Method and the Jacobs-Hochheiser Method at the same site. On the average the former were three times as large as the latter even though "[b]oth methods have been carefully checked under laboratory conditions and are internally consistent."⁵⁹ The Jacobs-Hochheiser method was originally chosen as the reference method for NO_2 ; it has since been found to be in error in the direction of overestimating ambient NO_2 . Chemiluminescence techniques are likely to become the replacement measurement method, with the added advantage of ability to measure NO and NO_2 with the same instrument. However, the accuracy of this technique is about $\pm 10\%$,⁴¹ and interferences with peroxyacetyl nitrate and other nitrates do occur. To the extent that these interferences are not proportional to total NO_x , an error of uncertain magnitude is incurred.

The last problem with NO_x is that it is of interest for two reasons: for its direct health effects and for its relation to oxidant formation. The former effect is represented as well by curbside concentrations as any other pollutant is, but the latter is an entirely different effect from the one which has been discussed so far. Data do not commonly exist for other than ground level or near ground level sites. If projections are based on ground level measurements, the implicit assumption is that these are proportional to levels in the urban airshed. This is unlikely to be true; without street canyon effects or height corrections to lessen their apparent impact, stationary sources are likely to be much more important in determining levels in the air shed than those at street level. Also, removal mechanisms are more important for NO_x in the airshed, since the residence time is longer. It is not clear how to estimate the resultant uncertainty, however.

Hydrocarbons are reactive to varying extents, depending on the particular species. Methoane is virtually inert in the atmosphere, while other compounds react quickly in the presence of sunlight and NO_x to form ozone, PAN, and various other compounds. Transportation sources account for about 45% of total nationwide anthropogenic HC emissions, and a larger proportion in many urban areas. Based on the same reasoning used for CO and NO_x , local and street level sources would be expected to dominate measured HC at most CBD sites.

There is an air quality standard for hydrocarbons, but, since they have no known health effects at the levels normally encountered in urban air, ambient HC is important for its relation to O_x formation. In the past, this problem was dealt with by assuming a simple relation between rollback projections of early morning ambient reactive hydrocarbon levels and ambient oxidants, using for example the "Appendix J"¹⁷ curve. However, such an approach is inherently limited in that it ignores both the basic nonlinearity of the oxidant-forming reactions, the role of NO_x in the formation process, and the long range advection of oxidant precursors.

The importance of the latter effect is commonly recognized today; for example, one study⁶¹ in performing aerial measurements of oxidants in the air entering the New Jersey-New York-Connecticut area found that ozone at levels exceeding the federal standards was occasionally being transported into the tri-state area. Thus it seems that long range transport and perhaps natural production mechanisms are important in oxidant formation, and little confidence can be placed in modified rollback predictions for this pollutant.

3. Errors due to improper treatment of background levels.

The third assumption made in the derivation of the rollback equation is that background pollutant levels do not change in time. To some extent this is a matter of definition--background levels should reflect primarily emissions from non-anthropogenic sources that may be expected to change little over time. As a practical matter, some background levels are low enough in general to make the rollback calculation insensitive to any reasonable variation in their magnitude. Table 9 compares commonly accepted values for background concentrations with the air quality standards. Except for hydrocarbons and O_x , background levels are roughly an order of magnitude less than their standards.

Background level uncertainty is evidently important for HC and O_x . For HC, much of the problem could probably be avoided by measuring both methane and non-methane HC--most of the natural background is methane, and would be expected to fluctuate seasonally. O_x "background" levels are highly variable, and can approach the level of the standard. The relative roles of long range transport, stratospheric transport, and formation from non-anthropogenic HC are still areas of active research.

4. Errors due to meteorological fluctuations.

The final rollback assumption is that meteorological factors do not shift over the projection period. Table 10 shows several years of CO data for the Washington DC CAMP site. Even the annual average, a statistic which would be expected to be stable over years with near constant aggregate emissions, seems to vary randomly from year to year. It is hard to ascribe this to anything other than meteorology. To estimate

Table 9: Comparison of Background Levels
with Air Quality Standards

Pollutant	Air Quality Standard	Background ⁴¹
CO	9 ppm [*]	1 ppm
NO ₂	.35 ^{**}	.03 ppm
HC	.24 ppm ^{***}	.1 ppm
O _x	.08 ppm*	0-.1 ppm

* 8 hour average

** Estimated⁴¹ 1 hour average based upon air quality standard of .05 ppm annual average.

*** Average 6 A.M. to 9 A.M. (guideline only).

Table 10: Washington D.C. Emissions Inventory²⁹ and Air
Quality Data^{29,60}

Year	Aggregate CO Emissions (KT/yr.)	Annual Average CO (ppm)	24 hr. Max CO (ppm)	1 hr. Max Co (ppm)	5 min. Max CO (ppm)
1962	1596	5.3	15.	25.	30
1963	1629	6.9	23.	41.	44.
1969	1638	5.7	13.	32.	37.
1965	1626	3.7	10.	31.	49
1966	1617	3.3	15.	38.	47.
1968	1538	3.4			
1969	1479	3.0			
1970	1387	3.8			
combined*	1565	4.41	-	-	-
	1620	-	15.2	33.1	40.6
\hat{C}	-	.255	.257	.157	.159

* \hat{Q} and \hat{X} as discussed in text

the magnitude of this variability first appropriate "average" values for air quality and aggregate emissions over a period of years are needed. The method should account for changes in aggregate emissions from year to year, so that averaging is performed only over different weather patterns. The following statistics to aggregate yearly data were selected:

$$\hat{x} = \frac{\hat{Q}}{n} \sum (x_i/Q_i) \quad (17)$$

and

$$1/\hat{Q} = \frac{1}{n} \sum (1/Q_i) \quad (18)$$

where:

x_i is the air quality in year i

Q_i is the aggregate emissions in year i

\hat{Q} is the harmonic mean of aggregate emissions over the years considered

\hat{x} is the average air quality corresponding to \hat{Q}

n is the number of years

An appropriate measure of dispersion is the weighted sample coefficient of variation \hat{C} :

$$\hat{C} = (\hat{Q}/\hat{x}) \left(\frac{1}{n} \sum (x_i/Q_i - \hat{x}/\hat{Q})^2 \right)^{1/2} \quad (19)$$

This statistic is independent of aggregate emissions and so measures dispersion due to meteorological effects alone.

Values of \hat{x} , \hat{Q} , and \hat{C} for the Washington data are also given in Table 10. It can be seen that, for example, ± 25% variations in annual average CO concentrations would be expected despite constant aggregate emissions. This is a fundamental limitation on the accuracy of air quality prediction, since future weather is not predictable. How many years of baseline data are needed to achieve this sort of accuracy?

Assuming that each year's weather is independent of the previous year's (perhaps a strong assumption) then:

$$\sigma = \sigma_{\min} (1 + 1/n)^{1/2} \quad (20)$$

where:

σ is the standard deviation of the rollback projection error

σ_{\min} is the standard deviation of air quality (for constant aggregate emissions).

n is the number of time periods for which data has been collected.

Table 11 contains a comparison of σ/σ_{\min} versus n . It can be seen that, for the given assumptions, three years' worth of data provides a reasonably good baseline.

Table 11

n	σ/σ_{\min}
1	1.41
2	1.22
3	1.15
5	1.10
10	1.05

This section has looked at the various uncertainties in rollback prediction over and above those due to uncertainties in emissions. Obviously, most of the estimated uncertainties must be viewed as illustrative rather than rigorously correct, since each site in each city faces a different situation. In addition, for different pollutants different sources of error are predominant.

For CO, a nonreactive pollutant with low natural background levels, rollback works best. The major sources of uncertainty, namely uncertainty

in emissions and uncertainty due to future meteorological fluctuations, lie outside the model and apply equally to predictions of any model. A major source of error within the model is the use of measurements of air pollutant concentrations, which are stochastic quantities, as the estimate of air quality. Since almost all present day models are "calibrated" based on same kind of data, this source of uncertainty is also shared by these models. To calculate the total uncertainty requires detailed specifications of the source distribution, local topography, air quality statistics used, etc. If air quality is defined as the concentrations ultimately measured, though, it appears that the inherent limitations alone make predictive uncertainty (measured by 95% confidence limits, say) in excess of +30%--perhaps considerably in excess.

NO_x , a combination of pollutants, should behave very much like CO. While non-automotive sources contribute a larger portion of aggregate NO_x , they are for the most part power plants, with large smokestacks, situated far from the city center. The impact on ground level concentrations is thus substantially diminished. The street canyon effect, a fluid mechanical phenomenon, must apply to automotive NO_x just as it applies to CO. The extra uncertainty for NO_x comes from using it as a proxy for NO_2 . Since the ratio of the two is not constant, an extra 10% or 20% of uncertainty is added.

HC is of interest only for its role in O_x formation. Like CO, measured HC at CBD sites should often reflect street canyon effects. O_x , on the other hand, seems to be transported for great distances at concentrations approaching that of the standard. The relation between HC and O_x at a given site is tenuous, and still a subject of active research. Little

confidence can be placed in the accuracy of predictions of future O_x .

4. Analysis of Air Quality Trends at Kenmore Square, Boston

The general analysis of uncertainty in measurement of air quality trends, given in section 3.2, left several issues unresolved: how much data variance could be explained by exogenous factors, how inclusion of exogenous factors would affect accuracy of the trend estimate, and how reasonable is the assumption that the noise structure of the data is well behaved--i.e., consists of independent, identically distributed normal shocks. To investigate these issues, a trend analysis was performed⁶⁴ for an existing series of air quality data, and is summarized in this section. Though not conclusive, since it is based on a single site, the analysis is illustrative of the problems involved in obtaining a statistically sound measure of air quality trends.

A review of the literature⁶⁴ revealed little consensus about appropriate methodologies for trend analysis or about an appropriate summary statistic for air quality. In addition, little attention seems to have been given to the problem of distinguishing between deterministic trends (due to changes in emissions) and nondeterministic ones (due to, e.g., random climatological trends). Surprisingly, despite the substantial reduction which should have already occurred in automotive pollutants such as CO (see figure 3), no convincing demonstration was found that the expected improvement in ambient CO has actually happened.

In light of this, an attempt was made to devise an improved trend analysis methodology and apply it to data from a real monitoring site. Daily average CO was chosen as the summary statistic, since daily

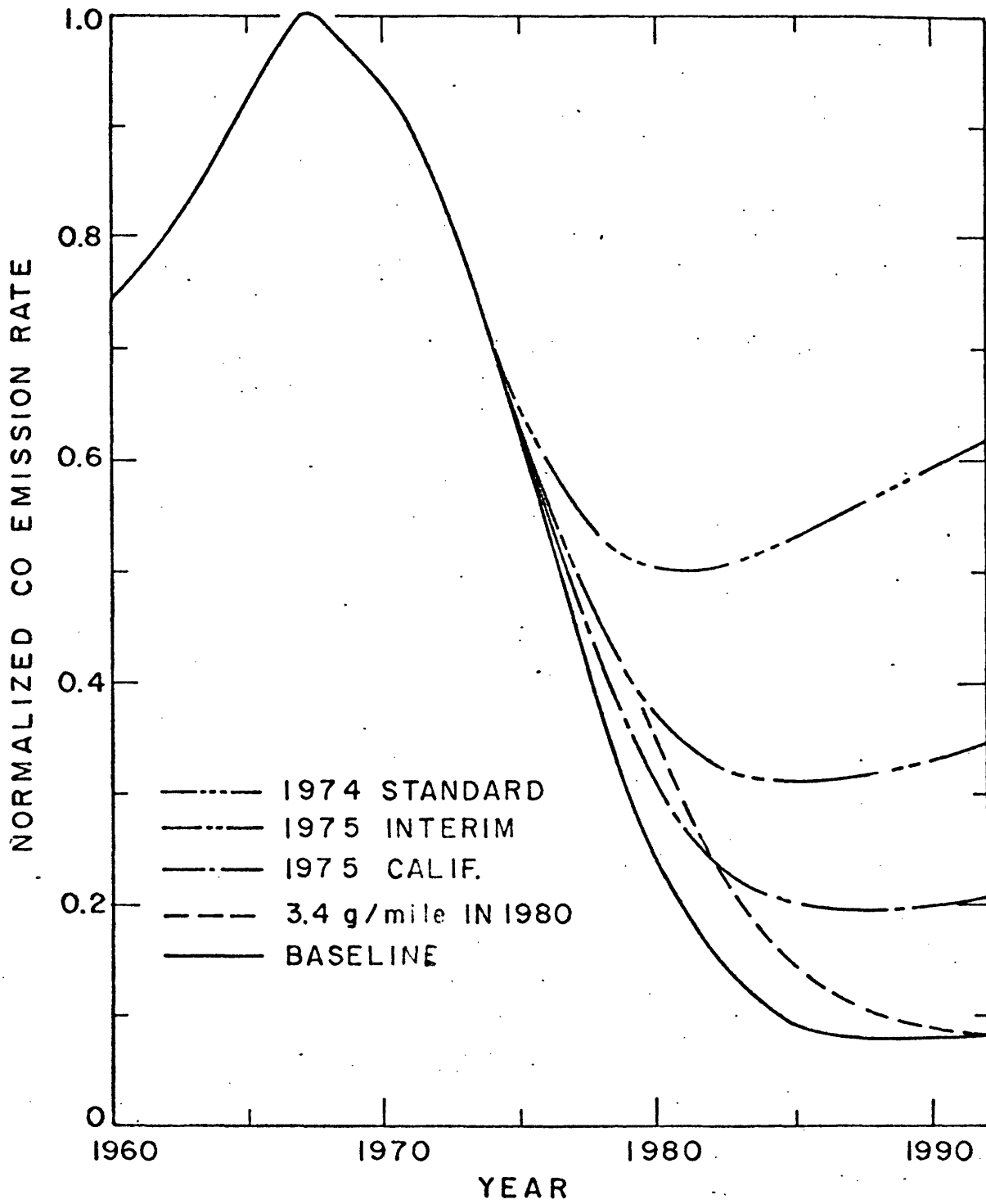


Figure 3. Aggregate CO Emissions Versus Time⁶⁵

averages eliminate the problems associated with the diurnal variation of shorter time averages. CO was chosen because it is primarily an automotive pollutant, and because good quality data spanning many years are available for this pollutant. The choice of a site in Boston was based simply on convenience in obtaining data.

The data used have three sources. The air quality data were supplied by EPA, and are available through the agency's SAROAD computer system. The meteorological data came from the National Climatic Center, U.S. Department of Commerce, and are published as monthly summaries entitled, "Local Climatological Data." The former data were obtained for a fixed site in Kenmore Square, Boston, and for CO, are available as hourly averages from June, 1969 (with numerous gaps). The latter were recorded at Logan Airport, Boston, and are available as ten minute averages for each hour of the day, though only published, by international convention, for every third hour. Stability was calculated for each hour using the method of Turner⁶² and summarized as a daily statistic by converting to the equivalent standard deviation of wind direction⁵⁸ and geometrically averaging over the daylight hours (arbitrarily defined as 0700 - 1900 LST). Nighttime stability was excluded since it is virtually always neutral or slightly stable.

The last meteorological factor considered was mixing height. Calculated mixing heights for Boston were also obtained from the National Climatic Center, but were the result of a special study, and so are not routinely available at low cost. The mixing heights were based on 1200 GMT (0700 local time) balloon soundings at the Chatham, MA station combined with surface data from Logan Airport. The morning mixing height was estimated from the intersection of the measured temperature profile

with the potential temperature profile based on surface pressure and a surface temperature 5° higher than the minimum from 0200 through 0600 LST. Maximum mixing height for the day was extrapolated based on the maximum surface temperature from 1200 through 1700 LST.

For the purpose of the exploratory development only a subset of the data--arbitrarily chosen as that for the year 1974--was used. This reduced the clutter on many of the graphs, and allowed aberrant points to be more easily identified. Also, use of a period too short for significant trend effects allowed easier identification of purely meteorological influences. Finally, it allowed the resulting tentative model to be validated over the full set of data.

Figure 4 shows the CO data plotted against time. It reveals no obvious trend and perhaps some seasonal effects. The logarithm of CO appeared to have roughly a normal distribution, with some discrepancies at the tails. The lack of fit at the lower end appeared serious, but was probably due to the lack of a "continuity correction" in combining and rounding measurements at the discrete (0, 1, 2 ppm) end of the instrument scale. Air pollution data often fit a lognormal distribution, after correction for continuity and for background levels.

The CO data were also autocorrelated. Although the autocorrelation seemed to persist beyond lags of one day, they were not significant at the 95% level. Inspection of the partial autocorrelations revealed that, as a pure time series, the data could be modeled as a first order autoregressive process. Correction must be made, however, for the fact that weekends (Fri. - Sun.) differ from weekdays due to the different traffic pattern. Monday through Thursday exhibited the behavior expected due to diurnal traffic variation, with morning and afternoon rush hour

DAILY AVERAGE CO VS DAY (1 JAN 1974-1)

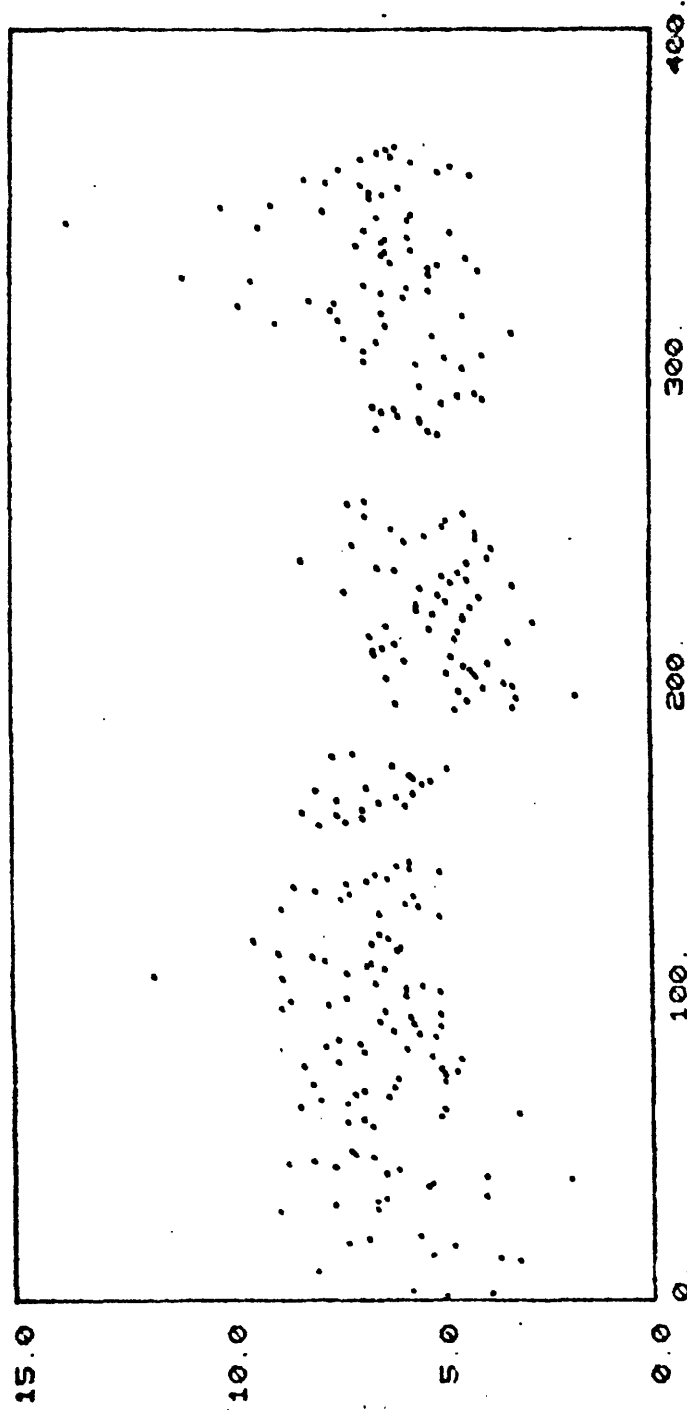


Figure 4. Daily Average CO Versus Time

evident. Friday was similar, with a higher afternoon rush hour peak and a buildup in the evening. Saturday and Sunday exhibited a different pattern, with pollution slowly building up through the day and reaching its peak at night.

Based on these observations, a simple linear model was fit to the data, as shown in Table 13. The weekend effects were approximated as simple shifts in expected CO, and so could be handled with dummy variables for those three days. The Saturday effect was not significant at the 95% level, since Saturday average CO happened to be about the same as that during the week. As expected, no trend could be detected. About 28% of the data variance was explained by this simple model; not all of the variables were significant at the 95% level.

To proceed, the CO data were plotted versus various meteorological factors. No clear patterns was evident in any of these graphs, but the following observations were made:

- (1) There was some correlation between CO and inverse wind speed, as expected. However, there were also several outlier points which did not seem to follow this relationship.
- (2) The expected downward trend of CO with temperature did not seem to hold at low temperatures. This may be due to the higher average wind speeds in winter, however.
- (3) There was a slight wind direction effect with high values for easterly winds and low ones for northerly winds. This is consistent with the geography of the Kenmore Square monitoring station, which has the Charles River basin about 100 meters to the North and a long, straight section of a major thorough-

Table 12: Nomenclature for Tables 13 to 17

<u>Symbol</u>	<u>Explanation</u>
NOB	Number of data points used in the regression
RSQ	The R^2 statistic for the regression (square of the correlation coefficient between predicted and observed)
F (n,m)	The F statistic for the regression, with n and m degrees of freedom
SER	Standard error of the regression
DW(n)	Durbin-Watson statistic
COEF	Label for column of estimated coefficients (by name)
VALUE	Label for column of estimated coefficient values
T-STAT	Label for column of coefficient t-statistics (estimate divided by standard error of estimate)
MEAN	Label for column of average values of explanatory factor associated with the coefficient
AO, A1, A2 B1, B2, B3, B4 C1, C2, C3	estimated coefficients

(continued)

Table 12 (continued)

<u>Symbol</u>	<u>Definition</u>
CO	Observed daily average CO (ppm)
TIME	Elapsed time from start of data series (years)
CO(-1)	Previous day's CO (ppm)
FRIDAY, SATURDAY SUNDAY, MONDAY	Dummy variables, equal to 1 if the corresponding observation took place on that day; 0 otherwise
U	Daily average wind speed (mph)
UINV	100/U
THLIL	$\bar{\theta}$, as defined in eq. 21
T	Daily average temperature (°F)
T.DEV	T - MEAN (T) (°F)

$$1. \text{ CO} = \text{A0} + \text{A1} * \text{TIME}$$

NOB = 272
 RSQ = 4.27E-04 F(1/270) = 0.115 SER = 1.5840
 DW(21) = 1.05

COEF	VALUE	T-STAT	MEAN
A0	6.07977	31.12730	1.00000
A1	-0.11099	-0.33714	0.51661

$$2. \text{ CO} = \text{A0} + \text{A1} * \text{TIME} + \text{A2} * \text{CO}(-1) + \text{B1} * \text{FRIDAY} + \text{B2} * \text{SATURDAY} + \text{B3} * \text{SUNDAY} + \text{B4} * \text{MONDAY}$$

NOB = 250
 RSQ = 0.284 F(6/243) = 16.064 SER = 1.3636
 DW(19) = 1.88

COEF	VALUE	T-STAT	MEAN
A0	3.24175	8.08058	1.00000
A1	-0.08016	-0.26345	0.52045
A2	0.47310	8.48669	6.04560
B1	0.62574	2.44524	0.15600
B2	-0.27235	-1.05232	0.15200
B3	-0.99619	-3.66353	0.13200
B4	0.46652	1.70143	0.13600

Table 13 Top: simple regression of 1974 data versus time.

Bottom: simple linear regression of 1974 data against
 endogenous factors.

fare (Commonwealth Avenue) to the East.

- (4) A tendency toward higher CO levels at low mixing heights also seemed to exist.

For the first two of these effects, physical reasoning leads to an expected form of the relation. For the wind direction effect, a crude piecewise linear function of two parameters was defined as follows:

$$\tilde{\theta} = \begin{cases} (\theta - \theta_1)/(\theta_2 - \theta_1) & \theta_1 \leq \theta \leq \theta_2 \\ (\theta_1 - \theta + 36)/(\theta_1 - \theta_2 + 36) & \theta > \theta_2 \\ (\theta_1 - \theta)/(\theta_1 - \theta_2 + 36) & \theta < \theta_1 \end{cases} \quad (21)$$

where:

$$\tilde{\theta} = \text{the wind direction effect term,} \quad 0 \leq \tilde{\theta} \leq 1$$

$$\theta = \text{the wind direction (tens of degrees)}$$

$$\theta_1, \theta_2 = \text{parameters (in tens of degrees)}$$

The parameters θ_1 and θ_2 were estimated by eye, and assigned the values 0 and 10 respectively.

Two models were fit to the 1974 data, as shown in Table 14. A linear model was fit first, to provide a benchmark for comparing the performance of the nonlinear model. Estimation of linear models is easier, both operationally and in terms of the theory, so a linear model which "works" almost as well as a nonlinear one is a useful thing. The linear model took the form:

$$\text{CO}_t = a_0 + a_1 t + a_2 \text{CO}_{t-1} + b_1 \text{Fri} + b_2 \text{Sat} + b_3 \text{Sun} + b_4 \text{Mon} + c_1 u^{-1} + c_2 \tilde{\theta} + c_3 T + \epsilon_t \quad (22)$$

$$3. CO = A0 + A1 * TIME + A2 * CO(-1) + B1 * FRIDAY + B2 * SATURDAY + B3 * SUNDAY + B4 * MONDAY + C1 / U + C2 * THTIL + C3 * T$$

NOR = 250
 RSQ = 0.42641 F(9/240) = 19.824 SER = 1.2281
 DW(19) = 1.73

COEF	VALUE	T-STAT	MEAN
A0	3.04111	6.01615	1.00000
A1	-0.07922	-0.28637	0.52045
A2	0.40193	7.71274	6.04560
B1	0.44629	1.92562	0.15600
B2	-0.50521	-2.14479	0.15200
B3	-1.07144	-4.37105	0.13200
B4	0.39111	1.56207	0.13600
C1	15.57060	4.72330	0.08695
C2	1.52061	4.52863	0.44366
C3	-0.02582	-4.94573	50.90400

$$4. CO = A0 + A2 * CO(-1) + C1 * (1 + A1 * TIME) * (1 + B1 * FRIDAY + B2 * SATURDAY + B3 * SUNDAY + B4 * MONDAY) * (1 + C2 * THTIL) * (1 + C3 * T.DEV) * UINV$$

NOR = 250
 RSQ = 0.4621 F(9/240) = 22.909 SER = 1.1893
 DW(19) = 1.67

COEF	VALUE	T-STAT	MEAN
A0	2.74803	7.00414	1.00000
A2	0.34572	6.78426	6.04560
C1	0.05680	1.99001	21.30260
A1	0.24585	1.10294	0.55487
B1	0.22392	1.37928	0.22470
B2	-0.31872	-2.82896	0.25793
B3	-0.78841	-4.33795	0.17317
B4	0.09735	0.55580	0.17492
C2	3.44437	1.80618	0.20931
C3	-0.02469	-5.12621	2.95460

Table 14. Top: Estimates for Linear Mode (eq. 22)

Bottom: Estimates for First Form of Multiplicative Model (eq. 23)

$$4. \quad CO = A0 + A2 * CO(-1) + C1 * (1 + A1 * TIME) * (1 + B2 * SATURDAY + B3 * SUNDAY) * (1 + C2 * THTIL) * (1 + C3 * T.DEV) * UINU$$

NOB = 250
 RSQ = 0.45687 F(7/242) = 29.081 SER = 1.1901
 DW(19) = 1.66

COEF	VALUE	T-STAT	MEAN
A0	2.79463	7.36556	1.00000
A2	0.33863	6.79006	6.04560
C1	0.05766	1.92897	20.92100
A1	0.19869	0.96771	0.56597
B2	-0.37654	-3.89309	0.27872
B3	-0.79932	-4.76557	0.18513
C2	3.91634	1.71741	0.19245
C3	-0.02566	-5.31080	3.14275

Table 14 (continued). Estimates for Final Form of Multiplicative Model

The nonlinear model was of the form:

$$CO_t = a_0 + a_2 CO_{t-1} + c_1 (1 + b_1 \text{ Fri} + b_2 \text{ Sat} + b_3 \text{ Sun} + b_4 \text{ Mon}) (1 + c_2 \tilde{\theta}) (1 + c_3 T) (1 + a_2 t) u^{-1} + \epsilon_t \quad (23)$$

where:

CO_t = the daily average CO on day t

$\text{Fri, Sat, Sun, Mon}$ } = dummy variables for Friday, Saturday, Sunday, Monday

u = daily average wind speed (mph)

$\tilde{\theta}$ = the wind direction effect term [eq. 21]

T = the daily average temperature (°F)

t = time (years)

ϵ_t = the residual noise

The dummy terms for Friday and Monday were later dropped due to lack of significance.

The nonlinear model was found to fit the data slightly better, and so residuals from this model were plotted against explanatory factors to explore the possibility of further model refinement. The residuals were well behaved; they were roughly normal, were not significantly autocorrelated, showed no tendency toward heteroscedasticity, and no longer showed discernible seasonability. In fact, examination of the residual plots revealed no further structure at all--not even the expected weak relation to stability and mixing height.

To complete the exploratory analysis, a set of regression diagnostics was performed for the linear model, following the techniques set forth by Welsch and Kuh.⁶³ These techniques allow one to identify specific

data points which are "outliers" in any of a number of senses--in their effect on the overall fit, in their influence on the estimate of a parameter or on the variance of the parameter, or in their deviation from the fitted value after correction for the expected variance of the residual. Application of the diagnostics to the model turned up several aberrant points corresponding to holidays--Thanksgiving, Christmas, Labor Day weekend, and Easter weekend. Several other points were also identified as having an undue influence by all of the above mentioned criteria. However, neither removal of the holiday points nor removal of the additional suspicious points significantly altered either the fit or the parameter estimates, and so it was concluded that there was no need to continue the analysis using robust estimation techniques.

The next step was to fit the two models to the full range of data, which extended from June, 1969 through February, 1975. As a preliminary step, the raw data was plotted versus time (Figure 5) and a simple linear regression versus time was performed (Table 15). As is evident both visually and from the regression, there is a distinct upward trend in this data, of about .5 ppm per year.

Both models were then fit to the data, as shown in Table 16, and each seemed to explain the observations equally well. All fitted coefficients were easily significant at the 95% level, and had the expected signs. It is necessary to do some additional arithmetic to go from the coefficient of the time variable to an actual estimate of trend, due to the presence of the autoregressive term in the linear model and the additional necessity of scaling to proper dimensional units in the multiplicative model. When this is done, both models give similar trend estimates: .44 ppm/year (linear model) and .39 ppm/year (multiplicative

DAILY AVERAGE CO US TIME

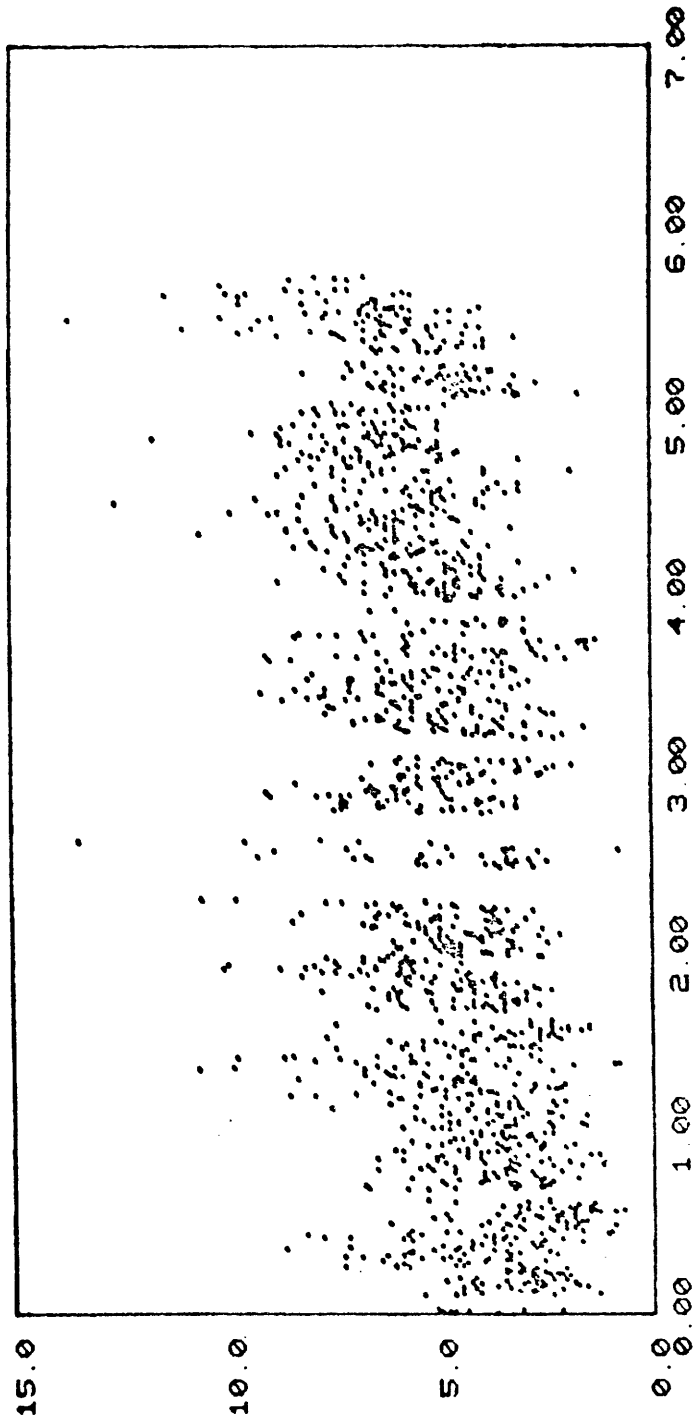


Figure 5. CO (ppm) versus time (years, starting 1 June 1969)

$$2. \quad CO = A_0 + A_1 * TIME$$

NOB = 1527
 RSO = 0 20657 F(1/1525) = 397.023 SER =
 1.6353 DW(116) = 1.01

COEF	VALUE	T-STAT	MEAN
A0	3 59862	44 23860	1.00000
A1	0.49457	19.92530	2.81033

Table 15. Simple Regression of CO Versus Time, All Available Data June 1969 through February 1975.

$$3. \quad CO = A0 + A1 * TIME + A2 * CO(-1) + B1 * FRIDAY + B2 * SATURDAY + B3 * SUNDAY + C1 * UINU + C2 * THTIL + C3 * T$$

NOB = 1384

RSQ = 0.53428 F(8/1375) = 197.176 SER =
1.2414 DW(102) = 1.58

COEF	VALUE	T-STAT	MEAN
A0	1.30114	8.28602	1.00000
A1	0.26160	11.51040	2.82890
A2	0.40901	19.54290	5.02514
B1	0.33603	3.38865	0.14234
B2	-0.65957	-6.68726	0.14740
B3	-1.04430	-10.72280	0.14884
C1	12.66680	12.46990	0.09445
C2	1.46231	10.05380	0.46040
C3	-0.01462	-6.73142	52.66180

$$4. \quad CO = A0 + A2 * CO(-1) + C1 * (1 + A1 * TIME) * (1 + B1 * FRIDAY + B2 * SATURDAY + B3 * SUNDAY) * (1 + C2 * THTIL) * (1 + C3 * T.DEV) * UINU$$

NOB = 1384

RSQ = 0.52351 F(8/1375) = 188.835 SER =
1.2556 DW(102) = 1.64

COEF	VALUE	T-STAT	MEAN
A0	1.38934	10.11350	1.00000
A2	0.43348	21.17720	5.02514
C1	0.06511	5.98663	21.85590
A1	0.24657	5.19964	2.39174
B1	0.16275	2.81746	0.22743
B2	-0.28269	-5.31477	0.23236
B3	-0.55628	-10.12700	0.23454
C2	1.22726	5.06201	0.42179
C3	-0.01586	-11.68550	5.60014

Table 16. Top: estimates for linear model, all data

Bottom: estimates for multiplicative model, all data

model).

Inspection of the residual plots revealed no serious deficiencies in the treatment of explanatory factors, but they did show significant problems with both the sample probability distribution of the residuals and their autocorrelations. Though the problem of a "heavy tailed" noise distribution can be handled by merely altering the estimation technique to a robust one or by suitable transformation of the data, the small but persistent autocorrelation of the residuals was a more serious problem, since it indicated that the model was misspecified.

To diagnose the problem, the model was refit without an autoregressive term and the time series nature of the resultant "corrected" CO data was reinvestigated. The differences residual series did not have significant autocorrelations beyond lag 2 and so a moving average process of order 2 was indicated.

Based on this noise model, all parameters were again estimated, as shown in Table 17. For simplicity, only the linear model was used; the resultant model was of the same form as eq. 22 (without the autoregressive term), but with the noise model

$$(1-B)\epsilon_t = (1-\theta_1 B - \theta_2 B^2)a_t \quad (24)$$

where:

B = backshift operator, $B a_t = a_{t-1}$

a_t = a series of independent, identically distributed random shocks

θ_1, θ_2 = parameters to be estimated

$$\begin{aligned}
B: \quad CO &= A0 + A1 * TIME + (1 - T1) * (CO(-1) - A0 - A1 * TIME(-1) - B1 * FRIDAY \\
&(-1) - B2 * SATURDAY(-1) - B3 * SUNDAY(-1) - C1 * UINU(-1) - C2 * THTIL(-1) \\
&- C3 * T(-1)) + (T1 * (1 - T1) - T2) * (CO(-2) - A0 - A1 * TIME(-2) - B1 * FRIDAY(- \\
&-2) - B2 * SATURDAY(-2) - B3 * SUNDAY(-2) - C1 * UINU(-2) - C2 * THTIL(-2) - \\
&C3 * T(-2)) + (T1 * T1 * (1 - T1) - 2 * T1 * T2 + T2) * (CO(-3) - A0 - A1 * TIME(-3) - \\
&B1 * FRIDAY(-3) - B2 * SATURDAY(-3) - B3 * SUNDAY(-3) - C1 * UINU(-3) - C2 * \\
&THTIL(-3) - C3 * T(-3)) + (T1 * T1 * T1 * (1 - T1) - 3 * T1 * T1 * T2 + 2 * T1 * T2 * \\
&T2) * (CO(-4) - A0 - A1 * TIME(-4) - B1 * FRIDAY(-4) - B2 * SATURDAY(-4) - B3 \\
&* SUNDAY(-4) - C1 * UINU(-4) - C2 * THTIL(-4) - C3 * T(-4)) + (T1 * T1 * T1 * (1 - T1) \\
&- 4 * T1 * T1 * T1 * T2 + 3 * T1 * T1 * T2 - 3 * T1 * T2 * T2 + T2 * T2) * (CO(-5) - A0 - A1 * \\
&TIME(-5) - B1 * FRIDAY(-5) - B2 * SATURDAY(-5) - B3 * SUNDAY(-5) - C1 * \\
&UINU(-5) - C2 * THTIL(-5) - C3 * T(-5)) + (T1 * T1 * T1 * (1 - T1) - 5 * (T1 * T1 * T2) + \\
&4 * T1 * T1 * T1 * T2 - 6 * T1 * T1 * T2 * T2 + 3 * T1 * T2 * T2 - T2 * T2 * 3) * (CO(-6) - A0 - A1 \\
&* TIME(-6) - B1 * FRIDAY(-6) - B2 * SATURDAY(-6) - B3 * SUNDAY(-6) - C1 * \\
&UINU(-6) - C2 * THTIL(-6) - C3 * T(-6)) + (T1 * T1 * T1 * (1 - T1) - 6 * T1 * T1 * T2 + 5 * \\
&T1 * T1 * T2 - 10 * T1 * T1 * T2 * T2 + 6 * T1 * T1 * T2 * T2 - 4 * T2 * T2 * T2 * T1 + T2 * T2 * 3) * \\
&(CO(-7) - A0 - A1 * TIME(-7) - B1 * FRIDAY(-7) - B2 * SATURDAY(-7) - B3 * \\
&SUNDAY(-7) - C1 * UINU(-7) - C2 * THTIL(-7) - C3 * T(-7)) + B1 * FRIDAY + B2 * \\
&SATURDAY + B3 * SUNDAY + C1 * UINU + C2 * THTIL + C3 * T
\end{aligned}$$

RANGE = 10, 59 TO 61, 100 TO 101, 111 TO 185, 194 TO 220, 259 TO 277, 291 TO 297, 311 TO 319, 334 TO 346, 354 TO 368, 381 TO 392, 402 TO 408, 419 TO 421, 441 TO 471, 481 TO 487, 501 TO 523, 538 TO 545, 556 TO 568, 577 TO 583, 619 TO 627, 636 TO 639, 650 TO 756, 766 TO 774, 783 TO 805, 824 TO 827, 836, 920 TO 927, 1021 TO 1061, 1077 TO 1080, 1089 TO 1099, 1110 TO 1126, 1179 TO 1193, 1202 TO 1208, 1217 TO 1220, 1230 TO 1237, 1247 TO 1254, 1266 TO 1297, 1306 TO 1312, 1332 TO 1333, 1342 TO 1350, 1383, 1408 TO 1412, 1463 TO 1496, 1508 TO 1521, 1540 TO 1579, 1593 TO 1598, 1615 TO 1620, 1630 TO 1645, 1673 TO 1678, 1718 TO 1722, 1737 TO 1790, 1803 TO 1813, 1832 TO 1841, 1869 TO 1911, 1921, 1961 TO 1963, 1991 TO 2022, 2031 TO 2043

NOB = 902

RSQ = 0.65011

F(9/892) = 184.156

SER = 1.0400

DJ(57) = 1.95

COEF	VALUE	T-STAT	MEAN
A0	0.46125	0.85754	0.16179
A1	0.40725	3.11520	0.45802
T1	0.60757	19.19760	-0.15602
B1	0.54570	5.79028	0.02333
B2	-0.33046	-3.33033	0.02690
B3	-0.97900	-10.32880	0.02325
C1	0.22378	17.67180	1.47956
C2	1.50942	10.17610	0.07358
C3	0.01083	1.90794	8.64730
T2	0.14840	4.84010	-0.15961

Table 17. Approximate Fit of Linear Model with MA(2) Noise Model

This upgraded model fits the data significantly better, accounting for about 65% of the variance of the CO data. The residuals (except for two points) followed a normal distribution reasonably closely, and were not significantly autocorrelated. Inspection of residual plots revealed no model deficiencies, although the wind direction effect term could perhaps be slightly improved.

Although the trend estimate of .41 ppm/year obtained from this model is consistent with earlier results, it is interesting to note that the accuracy of this estimate, as indicated by its t-statistic, is substantially less than would be expected based upon the previous discussion of data requirements for trend measurements, and it is worth further investigation. The basic structure of the final model used is

$$(1-B) (Y_t - \alpha_0 - \alpha_1 t) = (1 - \theta_1 B - \theta_2 B^2) a_t \quad (25)$$

It can be shown that, for the maximum likelihood estimate $\hat{\alpha}_1$, and for large samples:

$$V(\hat{\alpha}_1) = \sigma_a^2 (1 - \theta_1 - \theta_2)^2 / n \quad (26)$$

where $\hat{\alpha}_1$ implicitly has the dimensions of y_t divided by the time between observations. Defining an accuracy factor K similarly to eq. 5, and converting to a time dimension in years (assuming daily data) for compatibility with eq. 8, the above equation becomes:

$$\alpha_1 = 365K \sigma_a |1 - \theta_1 - \theta_2| / \sqrt{n} \text{ yr}^{-1} \quad (27)$$

The major difference between this calculation and that leading to eq. 8 is the noise model. If θ_1 and θ_2 were zero, then the model could be interpreted as predicting that observations will exhibit a random walk about the trend line. It is well known that random walks are characterized by long excursions from the expected value, which can lead to an apparent, nondeterministic trend when no deterministic trend exists. With independent noise, trend estimate accuracy increased with time period due to two effects--increased number of data points, which increases accuracy like $T^{1/2}$, and change in mean level of the series with time, which contributes a factor like T . Since a random walk is nonstationary in the mean, this second effect is not seen in eq. 27, and so accuracy increase like $T^{1/2}$ instead of $T^{3/2}$.

Equation 27 is disturbing, because it seems to imply that a very high signal to noise ratio is necessary to detect a trend. With K , σ_a , θ_1 , and θ_2 taken from the regression of Table 17, the result is $(\alpha_1)_{\min} = 9.2$ ppm/year. This is not what is observed, and the reason is that the derivation of eq. 27 is extremely sensitive to the assumption that the process is nonstationary. If instead it is assumed that:

$$(1-\phi B) (Y_t - \alpha_0 - \alpha_1 t) = (1-\theta B - \theta_2 B^2) a_t \quad (28)$$

then:

$$V(\hat{\alpha}_1) = \frac{\sigma_a^2 (1-\theta_1 - \theta_2)^2}{n (1-\phi)^2 V(t)} \quad (29)$$

$$V(\hat{\alpha}_0) = \frac{\sigma_a^2 \phi^2}{n (1-\phi)^2} \quad (30)$$

where it has been implicitly assumed that $E(t) = 0$.

Both of these variances become larger as ϕ approaches 1, and eq. 29 does not approach eq. 26 in the limit. The reason is that the maximum likelihood estimators are different in the two cases--in eq. 26 the estimator is a function of the differenced series. There is, of course, no analogue of eq. 30 for the nonstationary case, since the mean is undefined.

This explains why it was not possible to obtain a good estimate of α_0 , but not why an apparently significant value was found for α_1 . The reason is that the approximate fitting technique used is not so strongly sensitive to the assumption that $\phi=1$. Putting eq. 29 in the same form as eq. 8 gives:

$$\alpha_1 \doteq \frac{\kappa}{\sqrt{n}} \frac{\sqrt{12}}{\tau} \frac{|1 - \theta_1 - \theta_2|}{|1 - \phi|} \sigma_a \quad (31)$$

and this is obviously very sensitive to the exact value of ϕ .

For the Kenmore Square data, a value of $\phi \doteq .95$ would be sufficient to explain the estimated t-statistic of the final model. Inspection of the truncated autoregressive series used for the estimation shows that the weights change slightly for $\phi \neq 1$ but retain the same pattern; this small model misspecification is probably counteracted by adjustments in θ_1 and θ_2 in the fitting process.

At the Kenmore Square site, carbon monoxide is clearly increasing. It is possible only to speculate about the reason for this, other than to note that it is a deterministic trend and it does not seem to be due to climatological factors. Some aspect of the local traffic pattern, such as increased congestion resulting in a larger proportion of time at idle, could perhaps account for the trend. That emission controls could

perform so badly in use that emissions actually increased is not plausible, but if in fact the trend is due to traffic variables then apparently traffic control is more important than emission control in affecting air quality (at least for this pollutant). This approach to air pollution control has been little used to date, but may be the key to improving air quality at CO hot spots like Kenmore Square.

There is clearly room for improvement in the model developed here. That the residual series from the corrected, detrended CO data should be so close to nonstationary is unsatisfying and has no apparent physical basis. One approach to solving this would be to perform a full cross-spectral analysis with the exogenous explanatory variables, in the hopes that a transfer function model will eliminate the problem. For this it will be necessary, however, to find ways to cope with the numerous gaps typical of air pollution data. Another possible avenue of investigation is better modeling of the meteorological influences. Under some inversion conditions, for example, pollutants will accumulate in the airshed and so build up from day to day until the inversion is broken up. Under other conditions, there is essentially no carryover from one day to the next. This kind of effect can't be captured well by a time series model; some exogenous summary measure closely related to the expected effect is needed, and that in turn is best devised based upon physical reasoning and models. Lack of detailed and comprehensive meteorological data, however, limits the potential of this approach.

The interesting point to be made from all of this is that it is very difficult to demonstrate, in a statistically sound fashion, that the

calculated air quality trends are actually occurring. For at least this one site, the expected substantial decline in CO from 1969 to 1975 does not seem to have occurred. Though of course not conclusive, these results also seem to indicate that a substantial period of time--more than a decade--may be required to get a good measure of trends at a given site. This is clearly unsatisfactory, and indicates a pressing need for improved methods for analyzing air quality data.

5. Summary and Conclusions

This chapter has discussed the uncertainties which exist in regulation of automotive emissions, and the ways in which regulation in the presence of uncertainty should differ from regulation where there is no uncertainty. Two kinds of uncertainties were identified: predictive uncertainty, which is the limitation on ability to accurately forecast the future effects of regulation, and measurement uncertainty, which is the limitation on ability to accurately determine the present and past effect of regulation. The two are not totally independent, since often measurements are the basis for updating and improving predictions.

Given that uncertainty exists, there are a variety of ways of dealing with it. The ability of a regulatory structure to adapt flexibly in response to new knowledge is important, as is a mechanism for gathering information on the progress being made in achieving regulatory goals. For a risk-averse society, it is appropriate to incorporate a safety margin into regulations like numerical emission standards to account for the possibility of unforeseen problems. Finally, research can often be a useful response to uncertainty.

The gathering and analysis of data relating to progress in meeting regulatory goals plays a central role in regulation under uncertainty, since without such feedback, effects far different from those intended could occur undetected. This, in addition to the coupling between predictive and measurement uncertainty, is the reason that feedback is the most important of the responses to uncertainty. Feedback can be very useful, but is limited by several factors--measurement uncertainty, time lag for feedback, and relevance of the quantity being measured. These are interrelated, since time and measurement uncertainty can sometimes be traded off against one another (as in measuring trends in air quality) and since the time delay tends to become longer and longer as information gathered becomes more directly relevant to ultimate goals. For example, for automobiles, assembly line audits provide some information about emission performance, with a lag time of order one month. Field surveillance can provide better information on actual emissions, with a lag of several years. Analysis of air quality trends provides an even better indicator, but can require many years of data to provide meaningful information. Relating the trends to improvements in public health would probably require considerably longer.

Two forward links and two associated feedback links were examined in detail in this chapter, as indicated by the numbered links in Figure 1. Ability to predict and to measure both in-use emissions and air quality was analyzed in terms of the sources and magnitude of uncertainty for each. Since the circumstances of specific real situations vary widely, these calculations were intended to be more illustrative of

the size of the uncertainties than rigorously demonstrative. The major factor limiting accurate measurement of aggregate emissions is the number of vehicles tested. Assuming that current test procedures adequately reflect emissions on the road expansion of current surveillance efforts should be sufficient to allow good estimation of exhaust emissions on a city-specific basis. This expanded surveillance should include tests for evaporative and crankcase emissions. The current test procedure is not, however, a good direct measure of in-use emissions. Ambient conditions and, to a lesser extent, vehicle operating patterns, have strong effects on emissions. There is a need for expanded exploratory surveillance to quantify these effects accurately, to better relate the test results to in-use emissions.

Prediction of future aggregate emissions is more uncertain than measurement, since it is based in part on future emission from vehicles which exist now. Long range predictions are very uncertain--by a factor of 2 or so--since they depend heavily on estimates of emissions from vehicles which do not yet exist, even as conceptual designs. This is true even when it is assumed that the standards are met in certification testing. The test procedure simply does not indicate well what in-use emissions will be and it is necessary to apply large corrections to measured emissions in order to estimate actual emissions. The corrections are empirical and not well tested, and so much of the uncertainty can only be resolved by feedback.

Since there are uncertainties in predicting changes in air quality which are over and above those due to uncertainties in emissions, it is desirable to be able to directly measure changes in air quality. This is a statistical problem which has received little attention to date, and so most current methodologies for measuring trends in air quality are crude. Compounding the problem is the general lack (except in California)

of more than a few years of acceptable quality data. It appears that, for data with a well behaved stochastic component, roughly three or four years of data are necessary to measure a trend of around the expected size to within 20% accuracy.

Even given perfect prediction of emissions, it is difficult to forecast air quality. Air quality models have large uncertainties associated with their predictions. The rollback model is particularly worth examining because it is simple and widely used. If proper corrections are used for street canyoning and for source-receptor distance, rollback should be reasonably accurate for CO and total NO_x. This assumes that an adequate baseline is used to define present air quality, and does not consider the inherent uncertainty in projections of future air quality (by any model) associated with climatological fluctuations or uncertainties in emissions. Projections of O_x depend strongly on the relation of O_x to the primary pollutants, which is not well known. Little confidence can thus be placed in projections of O_x.

In the final section of this chapter, an attempt was made to develop an improved methodology for measuring trends in air quality. The intent was to measure trends in a way which was statistically sound and which properly separated deterministic trends (due to emissions) from nondeterministic ones (for example, those climatological variations). Based on analysis of data from Kenmore Square, Boston, it appears to be feasible to correct ambient pollution data for the effects of meteorology. However, the stochastic component of these data have an internal structure which made it more difficult to measure trends than would be the case with well-behaved data. If the Boston data were shown to be typical, then the a priori estimate of 3-4 years of data to measure a trend would have

to be substantially increased. This is an area where further research is needed.

Serious problems in our ability to actually know and understand the emissions of the in-use vehicle fleet and the present levels of pollution in the ambient air have been identified and estimated, and the even more serious difficulties in predicting these quantities for future years have been described and analyzed. Lack of recognition of these uncertainties can lead to undesirably inflexible regulation, to margins for error not compatible with the risk of missing regulatory goals, and to failure to adequately monitor success in meeting long term goals.

The analysis presented here focussed on the last problem, and the sorts of data collection and analysis which should be undertaken have been described. Without substantially increased effort to lower the degree of uncertainty in these areas, the national program for automotive emissions control will continue to proceed without any clear measures of its present or future effectiveness.

FOOTNOTES

1. "Air Quality and Automobile Emission Control," Report to the Committee on Public Works, United States Senate, September 1974 (GPO serial no. 93-24).
2. Williams, Marcia E., "Mobile Source Emission Factors--State of the Art and Future Programs," Presented at NAS, TRB (October, 1975).
3. "Compilation of Air Pollutant Emission Factors," EPA publication No. AP-42 (second edition, third printing, with supplement 1-5) (February 1976)
4. "Automobile Exhaust Emission Surveillance Analysis of the F.Y. 1974 Program," EPA report no. EPA-460/3-76-019 (September 1976).
5. "Automobile Exhaust Emission Surveillance--A Summary," EPA Publication No. APTD-1544 (May 1973).
6. Voelz, F.L. et al., "Survey of Nationwide Automotive Exhaust Emissions and PCV System Conditions--Summer 1970," Paper 710834, presented at SAE combined National Truck, Powerplant, and Fuels and Lubricants Meetings, St. Louis (October 1971).
7. "Automobile Exhaust Emission Modal Analysis Model," EPA Publication No. EPA-460/3-74-005 (January 1974).
8. "Development of Representative Driving Patterns at Various Average Route Speeds," EPA Publication No. EPA-450/3-76-023 (February 1974).
9. Ashby, H.A. et al. "Vehicle Emissions--Summer to Winter," Paper 741053, presented at SAE Automobile Engineering Meeting, Toronto, Canada (October 1974).
10. Clean Air Act (42 USC 1857, et seq)
11. Clean Air Amendments of 1970 (PL91-604)
12. Clean Air Act, Sec. 109 (b) (1) (a)
13. Clean Air Act, Sec 109 (b) (2) (a)
14. "National Primary and Secondary Ambient Air Quality Standards," Federal Register, Vol. 36, No. 84, Friday, April 30, 1971, (CFR, Title 42, Ch. IV, Part 410).
15. Ibid, Sec. 410.1 (e)

16. Clean Air Act, Sec. 110 (a) (2) (c)
17. "Requirements for Preparation, Adoption, and Submittal of Implementation Plans," Federal Register, Vol. 30, No. 158, 14 August 1971.
18. Ibid, § 420.17 (a) (2)
19. Clean Air Act, Sec. 110 (a) (2) (B)
20. "Guidance for Air Quality Monitoring Network Design and Instrument Siting," EPA Publication number OAQPS 1.2-012 (draft, September 1975).
21. Ibid, supplement A. "CO siting"
22. "Guidelines for the Interpretation of Air Quality Standards," EPA publication #OAQPS 1.2-008 (August 1974).
23. "The National Air Monitoring Program: Air Quality and Emissions Trends," EPA - 450/1-73-001 (2 vols., 1973).
24. "Monitoring and Air Quality Trends Report, 1972" EPA - 450/1-73-004 (1973).
25. "Monitoring and Air Quality Trends Report, 1973" EPA -450/1-74-007 (1974).
26. "Monitoring and Air Quality Trends Report, 1974," EPA - 450/1-76-001 (1976).
27. "Guidelines for the Evaluation of Air Quality Trends," EPA Publication OAQPS 1.2-014 (1974).
28. Clarenburg, L.A., "Is it Possible to Conduct a Rational Air Pollution Management Program?" Proc. Second International Clear Air Congress, paper AD 11B (1970).
29. Pierrard, John M. et al. "A New Approach to Setting Vehicle Emission Standards" Paper 73-75, presented at the 66th annual meeting, Air Pollution Control Association June 24-28, 1973.
30. Merz, P.H. et al. "Aerometric Data Analysis - Time Series Analysis and Forecast and an Atmospheric Smog Diagram" Atmospheric Environment, Vol. 6, pp. 319-342 (1972).
31. Chock, D.P. et al. "Time-Series Analysis of Riverside, California Air Quality Data" Atmospheric Environment, Vol. 9, pp. 978-989 (1975).
32. Tiao, G.C. et al. "Analysis of Los Angeles Photochemical Smog Data: A Statistical Overview," APCA Journal, Vol. 25., No. 3, pp. 260-268 (1975).

33. Tiao, G.C. et al. "A Statistical Analysis of the Los Angeles Ambient Carbon Monoxide Data 1955-1972" APCA Journal, Vol. 25, No. 11, pp. 1129-1136 (1975).
34. Box, G.E.P. and Jenkins G.M. Time Series Analysis Forecasting and Control. Holden-Day, San Francisco (1970).
35. Schmidt, F.H. and Velds, C.A. "On the Relation Between Changing Meteorological Circumstances and the Decrease of the Sulfur Dioxide Concentration Around Rotterdam."
36. Van Dop, H. and Kruizinga, S., "The Decrease of Sulfur Dioxide Concentrations Near Rotterdam and their Relation to some Meteorological Parameters during Thirteen Consecutive Winters (1961-1974)," Atmospheric Environment, Vol. 10, pp. 1-4 (1976).
37. Peterson, J.T., "Distribution of Sulphur Dioxide over Metropolitan St. Louis, as Described by Empirical Eigenvectors, and its Relation to Meteorological Parameters," Atmospheric Environment, Vol. 4, pp. 501-518 (1970).
38. Smith, F.B. and Jeffrey, G.H., "The Prediction of High Concentrations of Sulphur Dioxide in London and Manchester Air," Proceedings of the Symposium on Statistical Aspects of Air Quality Data, EPA Publication Number EPA - 650/4-74-038 (1974).
39. Gifford, F.A. and Hanna, S.R., "Urban Air Pollution Modelling," Proc. Second International Clean Air Congress 1146-1151 (1970)
40. Larsen, Ralph I., "A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards," EPA Publication No. AP-89 (1971).
41. "Air Quality and Automobile Emission Control, Volume 3--The Relationship of Emissions to Ambient Air Quality," U.S. Government Printing Office, Serial No. 93-29 (1974).
42. Shoji, Hikaru and Tsukatani, Tsuneo, "Statistical Model of Air Pollutant Concentration and its Application to the Air Quality Standards," Atmospheric Environment, Vol. 7, pp. 485-501 (1973).
43. Barth, Delbert S., "Federal Motor Vehicle Emission Goals for CO, HC, and NO_x Based on Desired Air Quality Levels," APCA Journal, Vol. 20, No. 8, pp. 519-524, (August 1970).
44. "The Report by the Federal Task Force on Motor Vehicle Goals Beyond 1980," (Draft 1976).
45. "Air Quality Impact of Alternative Emission Standards for Light Duty Vehicles," in "Implementation of the Clean Air Act--1975, Part 3 (Automobile Emissions)" Hearings before the subcommittee on public works, U.S. Senate Serial No. 94-H10, pp. 57-158 (1975).

46. "Modified Rollback Computer Program Documentation," Source Receptor Analysis Branch, Monitoring and Data Analysis Division, OAQPS - EPA (November 1973)
47. Morris, Roger and Hershel Slater, "Modified Rollback Models," in Proceedings of the Fifth Meeting of the Expert Panel on Air Pollution Modeling, NATO/CCMS publication no. 35, pp. 12-2-12-22 (1974).
48. De Nevers, Noel H. and J. Roger Morris, "Rollback Modeling--Basic and Modified," Presented at the 66th Annual meeting of the Air Pollution Control Association, Chicago, Illinois (June 24-28, 1973).
49. Larsen, Ralph I., "Determining Reduced Emission Goals Needed to Achieve Air Quality Goals--A Hypothetical Case," APCA Journal, Vol. 17, No. 12, pp. 823-829 (1967).
50. Chang, T.Y. and B. Weinstock, "Generalized Rollback Modeling for Urban Air Pollution Control," APCA Journal, Vol. 25, No. 10, pp. 1033-1037 (1975).
51. "A Critique of the 1975 Federal Automobile Emission Standard for Carbon Monoxide," Prepared by the panel on emission standards for the Committee on Motor Vehicle Emissions, National Academy of Sciences (May 22, 1973).
52. Gifford, F.A. and S.R. Hanna, "Modelling Urban Air Pollution," Atmospheric Environment, Vol. 7, pp. 131-136 (1973).
53. "Technical Support Document--Proposed Transportation Control Plan for the Metropolitan Boston AQCR," prepared by GCA corporation GCA Technology Division, Bedford, Mass. (1973).
54. Brunner, A.F. and K.B. Schnelle, "Air Pollution Patterns in an Urban Street Canyon," Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers, Vol. 100, No. EE2 (April 1974).
55. Ludwig, F.L. and W.F. Dabberdt, "Evaluation of the APRAC-1A Urban Diffusion Model for Carbon Monoxide," Stanford Research Institute (February 1972), available as NTIS publication, PB-210 819.
56. Nicholson, S.E., "A Pollution Model for Street Level Air," Atmospheric Environment, Vol. 9, pp. 19-31 (1975).
57. Hotchkiss, R. S. and F.H. Harlow, "Air Pollution Transport in Street Canyons," EPA Report No. EPA-R4-73-029 (June 1973).
58. Slade, David H. (ed.), "Meteorology and Atomic Energy 1968," U.S. Atomic Energy Commission, Division of Technical Information publication, TID-24190 (July 1968).

59. "Air Quality Criteria for Nitrogen Oxides," EPA publication AP-84, (1971).
60. "1962-1967 Summary of Monthly Means and Maximums-- Continuous Air Monitoring Projects," Available as NTIS publication, PB-194-764 (1969).
61. Wolf, George T. et al. (1975), "Aerial Ozone Measurements over New Jersey, New York, and Connecticut," APCA paper no. 75-58-6 presented at 68th annual meeting of the Air Pollution Control Association June 15-20, 1975, Boston, Mass.
62. Turner, D.B., "A Diffusion Model for an Urban Area," Journal of Applied Meteorology, Vol. 3, pp. 83-91 (1964).
63. Welsch, R.E. and E. Kuh, "Linear Regression Diagnostics," M.I.T. Sloan School of Management Working Paper No. WP923-77 (April 1977).
64. Martin, M.K., "Analysis of Trends in Ambient Air Quality," unpublished master's thesis, M.I.T. Sloan School of Management, June, 1977.
65. Heywood, J.B. and M.K. Martin, "Aggregate Emissions from the Automobile Population," SAE paper 740536 (1974).

Chapter 3

AVERAGE FUEL ECONOMY STANDARDS AND THE AUTOMOBILE INDUSTRY

by

E. Allen Jacobs and Lawrence H. Linden

1. Introduction

In an effort to lessen dependence on imports of petroleum by cutting gasoline usage, Congress in 1975 passed the Energy Policy and Conservation Act, mandating fuel economy standards for automobiles produced in (or imported into) the United States. This legislation is unique in the history of federal automotive regulation. It requires that every manufacturer (or importer) meet a specified average fuel economy for the fleet of vehicles it produces (or imports) in any model year. The EPCA regulations are in notable contrast to the regulation of automotive air pollution and safety, where each individual vehicle must meet a specified standard.¹ This difference is important because the automobile is not a homogeneous product. The principal characteristic which differentiates automobiles among the various segments of the automotive market is the size and hence the weight of the car, and weight is the crucial determinant of fuel economy. Technical changes alone may not bring the average fuel economy of a firm's production up to the fleet standard; altering the composition of the sales fleet may therefore become an important strategy for firms attempting to meet the standard. Thus, not only

are technical changes called for by the law, but, because of the flexibility inherent in a fleet average standard, changes in industry marketing strategies are likely to result as well. This may have important implications for the prices, profits, and structure of the industry.

The purpose of this paper is to examine the likely impact of the new regulations. In particular, we will focus on the industry's response to this law and the impact of the law on the structure of the industry.

It is particularly important that the level of policy discussion in this area be raised at the present time. The recently proposed "National Energy Plan"² would modify the structure of fuel economy regulation by imposing a tax/rebate scheme on top of the present regulatory system. Thus, the structure of the regulatory system itself will presumably come again to the attention of policy makers, and the opportunity for useful changes will arise. The difficulties with the present system arise in large part because Congress chose to impose measures only on the automobile suppliers, even though success in meeting the mandate may require changes in incentives to vehicle buyers. The tax/rebate scheme may be an important step to rectifying this difficulty.

This paper focuses on the situation where the automobile manufacturers are faced with violating the standards if traditional pricing and marketing strategies are used. It now appears that the trend in consumer purchases toward smaller cars, in combination with the

available technical improvements in automobiles, will very likely not keep up with the increasingly stringent fuel economy standards, thereby resulting in violations in the early 1980's. The analysis consists principally of a simple look at the production incentives in this circumstance, that is which production possibilities become more profitable and which become less so. Only the simplest assumptions about the structure of demands and costs are made. We do not review all the relevant data, and considerably more analysis would be required before our conclusions could be written quantitatively. However, we believe that our simple analysis provides sound evidence that there will be important and disagreeable incentive effects which were not foreseen by the drafters of the legislation.

The paper is structured as follows. After a brief review of the provisions of the EPCA, the paper turns to an examination of the situation where producers are faced with the possibility of violating the standards. Five aspects of this circumstance are examined. First, what are the strategies available to the manufacturers are described. Second, the simple profit-maximizing response of the firm in this situation, namely violating the standards and passing through the penalties, is examined. Third, the reasons why the manufacturers will likely strive to avoid violations are discussed along with the resulting pricing patterns. Then the impact of these corporate strategies on the structure of the industry is addressed. Finally, two possible solutions to the difficulties in the law are discussed. The paper will also set forth some illustrative calculations of price changes, fuel saved,

and the cost of this policy.

2. The Energy Policy and Conservation Act

The Energy Policy and Conservation Act³ mandates average fuel economy standards for major manufacturers and importers (above 10,000 cars per year) beginning with the 1978 model year, as shown in Table 1. The 1978 standard is 18.0 mpg. The standard progressively tightens until an average fuel economy of 27.5 mpg is required by 1985. The Secretary of Transportation sets the exact standards from 1981 to 1984, but within strict guidelines. The standard applies to each manufacturer and to each importer and refers to total cars produced or imported (rather than cars sold). The required average fuel economy is a harmonic average:

$$M = \frac{Q_1 + Q_2 + \dots + Q_n}{\frac{Q_1}{m_1} + \frac{Q_2}{m_2} + \dots + \frac{Q_n}{m_n}}$$

where

M = sales-weighted new car harmonic average fuel economy

Q_i = the number of model i vehicles produced

m_i = the fuel economy of model i vehicles

n = the number of different models

and each of these is for a given firm in a specified model year.

Alternatively:

$$M = \frac{Q_T}{\sum_{i=1}^n \frac{Q_i}{m_i}}$$

where

$$Q_T = \sum_{i=1}^n Q_i$$

= the total number of vehicles (for the given firm
in a specific model year)

The harmonic average fuel economy is the inverse of the (arithmetic) average fuel consumption:

$$M = \frac{1}{\sum_{i=1}^n \frac{1}{M_i} \cdot \left(\frac{Q_i}{Q_T} \right)}$$

The harmonic average fuel economy is used in the law in place of the simple arithmetic average because it more accurately reflects the fuel consumption of the new car fleet. Consider, for example, a fleet of cars of which half get 50 mpg and half get 15. The simple average fuel economy is 32.5 mpg ($\frac{1}{2} \cdot 50 + \frac{1}{2} \cdot 15$). However, for each gallon of gasoline consumed, the fleet will, on the average (and assuming equal distances driven by each vehicle), get 23.1 miles

$\left(\frac{1}{2} \cdot \frac{1}{50} + \frac{1}{2} \cdot \frac{1}{15} \right)$. Twenty-three and one-tenth mpg is the harmonic average fuel economy. While the direct reflection of the national goals is an appealing feature, we will see below that the nonlinearities introduced by use of the harmonic average results in some unusual effects.

Civil penalties are prescribed for a violator of the law -- \$50 for each average mile-per-gallon below the standard, multiplied by the number of cars produced or imported. Credits for production above

Table 1

Automotive Fleet Average Fuel Economy Standards
under the Energy Policy and Conservation Act

Model Year	Standard (mpg)
1978	18.0
1979	19.0
1980	20.0
1981-1984	To be determined by the Secretary of Transportation ^a
1985 and thereafter	27.5 ^b

Notes:

^a To be prescribed by July 1, 1977. Must be set at the "maximum feasible average fuel economy level, and ... [must] result in steady progress toward meeting the ... standard ... for model year 1985."

^b The Secretary may alter this to the "maximum feasible average fuel economy," but such action may be disapproved by Congress for levels below 26.0 or above 27.5 mpg.

the standard are calculated similarly, but can only be applied to the previous year's penalty. The Secretary of Transportation can reduce or modify the penalty of a manufacturer which applies for relief if: this is necessary to prevent bankruptcy; failure to meet the standard resulted from "an act of God, a strike, or a fire"; or the Federal Trade Commission decides that the normal penalty would lessen competition in the industry. The FTC will make such recommendations on a case-by-case basis for each firm which applies for relief. Manufacturers who make fewer than 10,000 passenger automobiles annually can apply for an exemption from the Secretary of Transportation (the "Checker Cab provision"); the Secretary of Transportation then sets alternative standards for such firms but does not alter the penalty structure.

3. Evaluating Potential Industry Responses

Industry responses to the fuel economy legislation may be categorized in three sets, depending on the length of time required for effectiveness, which in turn depends on the degree to which new technology and changes in vehicle designs and physical plant are required. Major new technological advances, such as advanced high-efficiency powerplants, new transmissions, or new lightweight materials, will not be available in quantity until the late 1980's at the earliest. Advanced power systems, such as the gas turbine and Stirling engines, have been under active investigation by the industry since automotive emissions became a major national issue in the late 1960's, and in some cases before then. It takes a minimum of 15-20 years for the research, development, introduction, and diffusion within the industry of such a

major technological innovation.⁴ Medium-run responses, using the technology available at any given time to modify the vehicle designs and physical plant, and thus new cars, can be implemented industry-wide in a period of 3 to 8 years. Massive programs utilizing the available technology to improve automotive fuel economy were initiated by the automotive industry in the 1972-74 period, and were accelerated after the price increases in gasoline in early 1974. They are now beginning to come to fruition in new automobiles such as the Chevette and the down-sized standard vehicles introduced in model year 1976. Technical improvements to the spark-ignition gasoline engine, utilization of the diesel engine in passenger cars, and various other weight-saving and minor fuel economy changes are included in this category.

Thus, the fuel economy of the vehicles offered to the American public is now improving rapidly and will continue to do so. Also over this period, vehicle buyers have continued their long-established trend toward the purchase of smaller, more efficient vehicles (although there are continuing short-term fluctuations).

It now appears that the combination of the available technological changes plus the trends in buyer behavior will suffice to meet the fleet fuel economy standards until about the early 1980s. There now exists a good possibility that at that time the standards will become binding on some firms, i.e., that, all else held the same, the standard would be violated.⁵ This circumstance, and the possible short-run responses of the automobile manufacturers, are the focus of this study.

The manufacturers have essentially two options. One is simply to treat the fines as added costs, and to absorb them or pass them through to consumers to the same extent as they would any other increase in costs. This is the short-run profit-maximizing strategy for either a monopolist or a competitive firm. There are, however, important non-economic considerations in the short-run which could have significant economic effects over the long run; thus a firm might instead try not to violate the standard.

The alternative to absorbing or passing through costs is to alter the sales mix between high and low fuel economy cars to raise the average fuel economy of cars produced, so as to approach or meet the standard. A firm can produce both large cars (with lower fuel economy) which fall below the standard and small cars which more than meet the standard. Switching buyers from large cars to small cars is not always possible. Real preferences currently exist for both large and small cars. This differentiates automobile demand according to fuel economy. Thus, buyers cannot easily be switched from purchasing low mileage cars to high mileage cars. Therefore, a firm which follows this strategy of attempting to meet the standard must use some combination of two tactics -- advertising campaigns and pricing decisions -- to influence buyer decisions. We will focus on pricing strategies.

4. Profit-Maximizing Strategy: Violating the Standards

The simple profit-maximizing response to this law is to treat the penalty and credits as an extra production cost. This cost will

vary among cars according to the fuel economy of that car. The penalty is a civil fine that is assessed from after-tax profits. Thus, the before-tax cost of this penalty (to be added to all other before-tax costs) is an adjustment of the actual penalty. The penalty thus is given by

$$P = \begin{cases} \frac{1}{1-Z} \cdot K(S-M)Q_T & M < S \\ 0 & M \geq S \end{cases}$$

where

P = total penalty

Z = the corporate profit tax rate (.48)

K = the civil fine per mile-per-gallon difference from the standard, per car produced (set to \$50 in the EPCA)

S = the standard (Table 1)

Then the average penalty (for $M < S$) on production of any car is

$$AP = \frac{1}{1-Z} K (S - M)$$

The marginal penalty for production of an incremental vehicle of model i is given by:

$$\begin{aligned} MP_i &= \frac{\partial P}{\partial Q_i} \\ &= \left(\frac{\partial P}{\partial Q_T} \right) \left(\frac{dQ_T}{dQ_i} \right) + \left(\frac{\partial P}{\partial M} \right) \left(\frac{dM}{dQ_i} \right) \\ &= \left[\left(\frac{1}{1-Z} \right) K(S-M) \right] - \left[\left(\frac{1}{1-Z} \right) KQ_T \right] \left[\frac{M}{Q_T} \left(1 - \frac{M}{M_i} \right) \right] \\ &= \left(\frac{K}{1-Z} \right) \left[S - M \left(2 - \frac{M}{M_i} \right) \right] \end{aligned}$$

It can be seen here that the average and marginal penalties are not dependent on the production total. The average penalty depends solely on a constant and the difference between the firm average fuel economy and the standard; the marginal penalty for a particular model depends on the constant, the standard, the firm average fuel economy, and the fuel economy of the particular model.

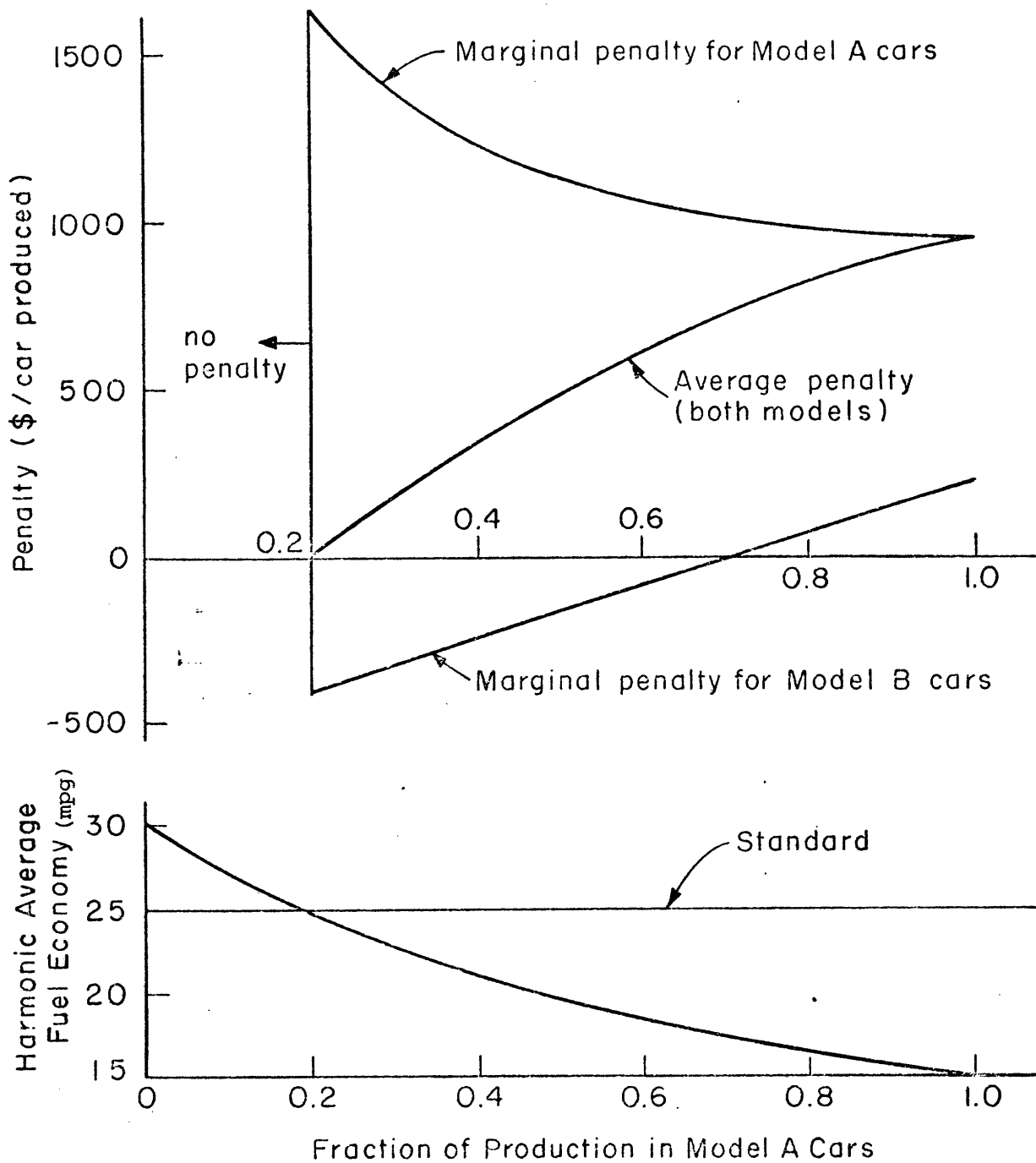
This marginal penalty will be treated by the profit-maximizing firm as a simple addition to the marginal cost of the i^{th} model. According to the usual tenets of economic theory, it will be absorbed partly by producers and partly by consumers, depending on the elasticities of demand and supply, and the degree of competition in the automotive market. Under perfect competition, firms attempt to pass on the full amount of the marginal penalty. Demand then adjusts, and prices follow. If all other costs (per car) are constant, a firm with monopoly power alters price and output until marginal revenue increases by the amount of the marginal penalty. Since a change in output changes marginal revenue more than it changes price (average revenue), the monopoly firm will absorb much of the cost and pass on the rest to the consumer. A firm within an oligopoly will increase prices by more than the monopolist but by less than the firm in perfect competition. The automotive companies will use the additional marginal cost as a guide for increasing a car's price.⁶ The initial price increase would be slightly less than this amount and would be based on simple judgment. This reflects confidence in their firms' market power, but an uncertainty about the extent of this market power. Further price changes would either be trial and error variations or would be left

solely to dealers.

The price change dictated by proper treatment of the marginal penalty may alter the firm's sales mix and bring the firm within the standard. Then the marginal penalty would be zero. In this case, the firm would settle on a smaller price change, one which maintains the firm's average fuel economy just at the standard.

A simple example illuminates the key characteristics of this unique regulation and its penalty structure. Several assumptions are required. First, over the relevant period, no (further) technological improvements in fuel economy are available. Second, the important distinction among automobiles produced is that some cars more than meet the standard fuel economy while others fall below this level. Assume that only two models are produced. Model A is larger and gets 15 mpg according to the official Environmental Protection Agency measurements used under the law; Model B is smaller and gets 30 mpg. Third, let the fuel economy standard be set at 25 mpg.

Now consider the case where the firm produces an equal amount of both models. Then the harmonic average sales-weighted fuel economy of this firm is 20 mpg, and the firm will be assessed an average penalty of \$250 per car out of profits, or about \$480 per car before taxes. The added marginal cost in penalty before tax of producing another "A"-type car (MP_A) is \$1,122. The marginal reduction in penalty before tax of producing another "B"-type car ($-MP_B$) is \$160. Figure 1 shows how the average and marginal penalties vary with the production mix between these two classes of cars.



Penalty Structure and Harmonic Average Fuel Economy for Two-Model Firm (Model A vehicles get 15 mpg, Model B vehicles get 30 mpg, standard is 25 mpg.)

Figure 1

Several results can be gleaned from this example. First, it is immediately apparent that a very important determinant of the short-run marginal cost of any one model will be the current overall sales mix of the firm, as it influences the firm's current average fuel economy.

An unusual feature of this penalty structure is that the more a firm is violating the standard the weaker are the incentives, at the margin, to comply. That is, at higher sales fractions of model A vehicles, the marginal penalty on model A is lower and the marginal credit for production of a model B car is reduced. In fact, as shown in Figure 1, the marginal penalty for the small car is positive under certain circumstances. That is, production of a car whose fuel economy is better than the standard increases the total magnitude of the penalty! This occurs when sales are dominated by large cars; the increasingly small improvement in average fuel economy of the firm producing another small car is offset by the increase in the total quantity of vehicles produced.

The unusual incentives implicit in this penalty structure may be seen in one far-fetched example. Assume that a different firm's harmonic fuel economy average were 15 mpg when the standard was 25 mpg. If this firm could produce a stripped-down model (or "souped-up" go-cart) which cost \$2,500 to make and got 90 mpg then this firm could make \$100 on each of the first few of these cars that were given away free! This follows from the marginal penalty being $-\$2,604$ at the introduction of such a car.

This perverse relationship of marginal penalty to average fuel economy is due to the use of the harmonic averaging procedure. Had a simple average fuel economy been specified, then the marginal penalties would be constant. As previously discussed, however, this would lead to inaccurate representations of fleet fuel consumption. Another alternative would have been to place the penalty on the difference between fleet average fuel consumption (gallons/mile) and a specified standard (say $1/S$); this again would give constant marginal penalties.

This simple two-car model offers clues to the law's possible impact on industry structure beyond the short run. By making several restrictive assumptions, possible scenarios for the evolution of the industry can be examined. First, assume that all firms have similar costs, other than penalties, and similar product technology. Second, each firm alters production capacity between the two models in an effort to increase total profits of the firm, that is, capacity is added where price is greater than cost and eliminated where price is less than cost. Third, the rate of production capacity changeover between these two models is limited to a fraction of total capacity. Fourth, the only differentiation of automobile demand is among the two sizes, so that a firm cannot maintain a price higher than a competitor's price on that model. Fifth, the two types of cars are partial substitutes and the own-price effect on total industry sales is the normal inverse relation. Finally, assume that the firm with the lowest marginal penalty for a given model will set the market price for that model at the level of normal cost plus marginal penalty. A justification for this behav-

ioral assumption about the industry is that most firms wish to increase their market share as long as normal profit margins are maintained. Alternative approaches might include specifying some degree of oligopolistic collusion or monopolistic competition with product differentiation of each model size. These approaches complicate analysis without adding new results.

It is possible to examine scenarios by assuming the existence of several firms (with hypothesized production mixes) which are in equilibrium at the initiation of the structure of penalties. Several conclusions are evident. Calculations refer to the two-model example discussed above. First, the firm which is breaking the standard the most will set the market price for large inefficient cars because their marginal penalty will be the lowest. This firm would also have the highest (algebraically) marginal penalty for small cars and so would begin to leave that business for increased large car production. Other firms will continue to get out of large car production, until they just meet the standard, because they are selling at a price less than the sum of normal costs, and marginal penalty. At this point, the "law abiding" firms will take whatever price (above normal cost) enables them to sell the maximum amount of large cars (Model A) while just meeting the fuel economy standard (four "B" cars to one "A" car in our example).

This circumstance may permit the existence of a firm producing all large cars. If the market-clearing price is more than \$961 above normal prices when all firms just meet the standard, then, in this example, there would be room for at least one firm to produce only large

cars and to make excess profits. The price of large cars would then be set by the firm(s) producing large cars and this price would lie between the full cost including the marginal penalty and the market-clearing price in the absence of this (these) firm(s). The small car price will be cut enough to maintain the sales mix of the firms at the standard.

Firms which have a sales mix that brings the average fuel economy well within the standard will be consistently undercut on small car prices. (This is not true in the case of a two-firm world, where one firm builds large cars only and the other small cars. If neither firm charges monopoly prices then this world is stable). The failure of such a firm to expand its large car production share until it just meets the standard will bring continual economic losses until the firm goes out of business. The latter result is a perverse penalty for specializing in the production of fuel efficient cars. Ultimately all firms either exactly meet the standard or produce only large cars.

If the purpose of this standard and penalty is to alter future gasoline consumption, then this penalty acts as a very erratic charge or subsidy upon anticipated gasoline use. The penalty may be viewed as a tax which is shared between firms and consumers. The result is demonstrated by returning to the initial example where a firm produces and sells equal amounts of Model A and Model B cars. The marginal penalty on Model A cars is equivalent to an average charge of 16.8 cents per gallon over the lifetime of the car. The marginal rebate on Model B cars is equivalent to an average subsidy of 4.8 cents per gallon.

5. Pricing to Meet the Standard

The preceding analysis has assumed that the economic incentives and penalties of this law were the only new influence on a firm's behavior. However, the penalties in this law are not taxes but civil fines which are assessed after a firm has been declared by a court to be in violation of federal law. In the eyes of many automobile industry executives, the public relations "cost" of being labeled a law-breaker dominates any careful consideration of penalty costs.

Another danger to firms which blindly follow the economic penalties and incentives of this law is that the law could easily change. Judging from past experience with Congress in federal automobile regulation (safety and emissions), automotive firms view future fuel economy standards with great uncertainty. One response is to design more flexibility into production capacity, at additional expense. Of more interest here, a second response is to view government intent as more relevant than the exact wording of the law. The intent of Congress in this law is clearly that the automotive companies should adhere to the fuel economy standard.

Thus, through a fear of bad public relations, or changes in the law, the largest automotive companies are in fact now planning to meet the standard in their production, even if this is more costly (both to themselves and buyers) than paying penalties. Calculating just what this additional cost is first requires the determination of the optimal strategy for a firm which constrains itself to meet the standard in the short run, i.e., given that all reasonable tech-

nological improvements have been made.

For a firm constraining itself to meet the standard in the short run, the optimal strategy is given by the results of the following simple constrained maximization problem. When the standard is constraining, the firm will want to meet the standard exactly (no safety margin is required in this simple model with no uncertainty). The constraint implied by exactly meeting the standard is given by:

$$M = S$$

or

$$\sum_{i=1}^n \left(\frac{S}{M_i} - 1 \right) Q_i = 0$$

For the simple example used earlier, this yields the requirement that the firm should produce 80% model B and 20% model A vehicles. Profits, to be maximized over all possible vectors $[Q_1, \dots, Q_n]$, are given by:

$$\Pi = \sum_{i=1}^n Q_i [R_i - C_i]$$

where ({...} refers to functional dependence):

R_i = price of model i cars

$$= R_i \{ Q_1, Q_2, \dots, Q_n \}$$

C_i = total cost of producing model i cars

$$= C_i \{ Q_1, Q_2, \dots, Q_n \}$$

Π = profits of a given firm

$$= \Pi \{ Q_1, Q_2, \dots, Q_n \}$$

If it is assumed that for one firm purchases of a given model depend only on the price of that model, profit maximization takes the following form, where λ is a Lagrange multiplier:

$$\frac{\partial}{\partial Q_i} \left[\Pi - \lambda \left(\sum_{i=1}^n Q_i \left(\frac{S}{M_i} - 1 \right) \right) \right] = 0$$

or

$$\frac{\partial \Pi}{\partial Q_i} = \lambda \left(\frac{S - M_i}{M_i} \right)$$

In this very simplified case, we see that the marginal profit (or loss) from production and sale of a given model must be proportional to the fractional amount by which that model's fuel economy is exceeded by (or exceeds) the standard. That is, when the firm is just meeting the standard the incremental production of a profitable large car must be countered by appropriate incremental changes in production of other cars; if the firm is at a profit-maximizing production mix, the marginal changes in profits will be in the same proportion as the differences between the model fuel economies and the standard. With a specification of market structure (e.g., perfect competition or monopoly as indicated in the demand function faced by the firm), this set of equations yields a single set of prices and quantities. It is extremely unlikely that any firm would attempt to calculate the optimal pricing and output strategy using this formula. The structure of demand is

volatile and estimates are very uncertain. However this framework conceptually expresses the pricing and output strategy that is implicitly sought by the firms through guesstimate, trial, and error.

With reasonable assumptions on demand -- that the price elasticities of each model and the cross-price elasticities between models are finite and of the expected signs -- then the solution to this problem is a strategy of cutting small car prices and increasing the price of large fuel inefficient cars. The impact on the industry structure of this behavior is clear. The sales mix of all firms will gravitate toward the point where each firm just meets the fuel economy standard. With similar technology, this means a similar mix by model size among all firms. Otherwise, firms selling more large cars would violate the standard, and firms selling primarily small cars would be undercut in price by firms subsidizing small cars in order to meet the standard.

The magnitude of possible price changes in this situation can be estimated by considering the necessary change in the sales mix for the industry as a whole. A fuel economy improvement in the neighborhood of 15% (3 mpg if current fuel economy is 21 mpg) may be achieved by increasing the small car production share by 50% (e.g., 40% to 60%). Using one recent set of estimates⁷ this might be done by dropping the price on small cars by 10% and increasing the price of large size and midsize cars by 25%. These data are very crude because the estimates used are based upon present market shares and not increases from some future arrangement of shares. Also, the cost of technological improvements

is very uncertain at present.

A crude estimate of the welfare cost to society that results from having prices altered from present prices is approximately \$1.7 billion annually.⁸ The largest share of this loss occurs from the increased purchases of small cars made when their price is cut below cost. The relevant model vehicles would consume something like 6 billion gallons less than otherwise over their normal lifetime of use. The average cost of reduced gasoline consumption would then be about 29¢ for each gallon saved.

6. Impact of the Law Upon the Automotive Industry

The likelihood that the standard will constrain production some time before 1985 is very great. Therefore, the average fuel economy standards now mandatory by law could dramatically alter the structure, sales, and pricing of the automotive industry as it has existed since World War II. We have discussed two possible strategies for the industry in dealing with the fleet average fuel economy standard. One would be to ignore the non-economic, or long-run, penalties of violating the law, and treat the fines as simply as additional costs. The other would be to meet the standard, at some net cost in profits foregone relative to violating the standards. Of course, the facts are most likely somewhere in between the extreme assumptions behind these cases -- the companies are certainly willing to pay something more than the amount of the fines in order to avoid breaking the law, but it is unclear how much. It also is not clear how much they will have to pay,

or when; this depends on changes in consumer tastes, the price of gasoline, other government policies, and other variables out of their control. However, whether or not the companies choose to meet the standard, the results will be qualitatively the same: the prices of small cars will be lowered, relative to their cost, and the prices of large cars will be raised. Thus the qualitative implications of the law are similar in either case.

Furthermore, this law pressures all firms toward a similar model mix. The production mix becomes one of the most important determinants of cost for a firm. Pricing margins which have not been set through explicit economic calculation, but through years of trial and error, will be rapidly and continually altered. The survival of small car producers or importers in the face of reduced prices for their product will depend upon their ability to break into the more rewarding large car market. Yet the barriers to entry into a new sector of the automotive market are notorious. Unless these firms make this move in the near future they will be faced with the difficulty of raising large amounts of capital for a new venture while taking losses on all small car sales.

The Big Three -- General Motors, Ford, and Chrysler -- have the most diverse product lines and produce most of the large cars, while the smallest firms produce mostly small cars. Thus, the pressure in this law to cut prices on small cars could lead to actions which the larger firms fear would be labeled as predatory pricing, an anti-trust violation. Thus, GM will attempt to avoid any quick or large

cuts in small car prices. The resulting competitive disadvantage might prompt GM to try a dealer policy designed to increase small car sales. Quotas based on fuel economy might be allocated to (or "forced" upon) dealers or incentives offered without significant changes in wholesale prices. This simply passes the price-twisting incentive on to dealers. However, each of the large manufacturers has more than one dealer organization. Each organization may sell only a portion of a manufacturer's output, making cross-subsidization within dealerships difficult. For example, Chevrolet sales must be somewhat subsidized by Cadillac sales, which cannot be done at the dealer level. (This would be corrected if each dealer organization moved toward selling the complete line of vehicle bodies as to some extent they are now doing.)

One possible outcome of such a situation is merger. Because each firm's production must meet an average standard there is a great incentive for firms to sell off the production of certain lines (perhaps while continuing to market the same cars) or to take over the production (but not marketing) of other car lines. It is doubtful whether the Justice Department would approve such activity by Ford or GM. However, a tremendous incentive exists for trade between a threatened importer such as Volkswagen and a large car producer such as Chrysler.

Another result of this legislation will be increased production costs for all American producers. The situation described in this paper implies a more volatile and unpredictable automobile market. Unforeseen actions by the government, consumers, and rival firms bring great

uncertainty for an individual firm. A more flexible and adaptable firm has a competitive advantage. Production facilities will be designed for quick changeovers or varied production at the expense of higher cost plants and slower production. Also, the installation of fuel-consuming extras -- such as air conditioning -- will likely be shifted from the factory to the aftermarket, so that the manufacturers would avoid the attendant penalties which apply only to automobile producers.

However, it is the impact on pricing, and thus profits, which will be the most important. The shake-up of the industry structure resulting from these standards may bring the end for companies recently in trouble, such as Chrysler. Industry observers predict that Chrysler would receive an exemption from these standards, were its existence truly threatened. The law permits (with no Congressional veto) the Secretary of Transportation to drop or modify a firm's penalty if this is necessary to prevent bankruptcy. This exemption along with Chrysler's current size might insure a permanent niche in the automobile market as a predominantly large-car producer. The Secretary can also reduce or drop a firm's penalty if the Federal Trade Commission rules that competition would be lessened otherwise. A manufacturer must first apply to the FTC for such relief. It is obvious that, even though it will not incur any penalty, American Motors Corporation is the company most seriously threatened by this legislation. Incredible as it may seem, it would be in AMC's interest (and would certainly be a legitimate complaint) to apply for penalty exemptions for Ford and General

Motors, so that the pressures on these large firms to reduce small car prices are alleviated. However, relief is only provided to violators who apply for it.

Finally, it should be noted that we have assumed implicitly that the present firms have similar access to new product designs and new product technology. In the dynamics of the adjustment to the standards, differences in these fronts could be crucial. In fact, it may well be the case that General Motors and Ford can adapt their product lines more rapidly and more effectively than Chrysler or American Motors, as the larger firms' greater profitability allows more and superior technical resources to be addressed to the problem. Thus either American Motors and Chrysler would be threatened, or Ford and General Motors would keep prices high relative to cost.

7. Possible Solutions

There is at least one policy action which would maintain the structure of the present law and its incentives for the industry as a whole, yet would mitigate the changes in the structure of the industry. This would be the creation of a market in the rights to produce fuel-inefficient cars, where the production would be counted in the average fuel economy of the buyer of the rights but not the seller. A firm such as American Motors could sell rights to produce fuel-inefficient cars up to the point where it still just meets the standard. A firm which will not meet the standard could either buy rights from a firm making efficient cars or pay the penalty. A firm which produced primarily fuel-efficient

cars then would not suffer losses. In fact, its competitive position would be better than the present situation. The real return of producing an incremental small car would be the price for which it can be sold plus the price of additional fuel economy rights which the firm could sell. Using the example cited earlier, assume that a second firm makes only model B cars (30 mpg). Producing one more B car would enable it to sell rights to the first company considered (firm average of 20 mpg) worth up to \$160.25.

Such a plan would tend to equalize the marginal penalty of the same car between the firms. That is, the marginal penalty or extra marginal cost would be less dependent on the production mix of any one firm and more dependent on the industry's production mix. Thus this plan would distribute the incentive across the industry. The perverse incentives faced by some firms and the extreme incentives faced by other firms would be "averaged." All firms within the industry would face the same incentive, yet fuel savings would be similar.

This market might resemble the "entitlements" market for price-controlled oil where manufacturers can buy and sell rights to lower-priced oil. However this market for fuel economy rights is not allowed within current law and would require action by Congress before the industry could engage in such activity.

Another possible solution is for the government to take over the cross-subsidization between model lines. This would require an explicit recognition that the present law places the burden for reducing gasoline consumption entirely on the vehicle producers. In fact, price controls

on crude oil encourage the purchase of relatively inefficient vehicles. Thus the government could impose a tax/rebate scheme, whereby efficient vehicles are subsidized and inefficient vehicles are taxed. The system could be arranged so that the net take to the government is zero. This is essentially the scheme proposed by President Carter. The analysis presented above has not been extended to this case, but a few simple conclusions are nevertheless quite apparent.

Most importantly, a tax/rebate scheme alleviates, in principle, the grossest of the industry impacts discussed above -- the pressures on producers to separate prices from costs, with a difference that depends on the model mix. A well-specified tax/rebate scheme would permit prices to producers to remain near production costs, while altering prices to consumers to reflect the social cost of gasoline consumption. In terms of the present structure of the industry, the larger firms would feel the burden of change, while AMC would have its vehicles subsidized. In essence, the government would take over the cross-subsidization, between vehicle classes, which will likely be handled very awkwardly by the automotive firms under the standard.

Of course a major difficulty is the determination of the tax/rebate schedule. The setting of prices -- and their tuning to fluctuations in the market -- is the forte of industry, relative to government, and results of a misplaced schedule may be severe. Given the present standards, the key parameter in the tax/rebate scheme is the degree of "tilt." If it is too low, the problems discussed above will be only partly eliminated. If it is too high, the present fuel economy law would be completely superfluous, because consumer buying

decisions would bring average fuel economies above standards.

An important issue raised in connection with the proposed tax/rebate scheme is its effect of subsidizing imports, since imported cars are presently more efficient, in average, than domestically produced vehicles. We have not analyzed this issue. It appears that the choice is a fundamental one, between free trade and aggregate gasoline consumption. However, it should be noted that the present system puts importers at a disadvantage relative to domestic manufacturers as prices to producers on small cars are forced down relative to costs. As discussed above a well designed tax/rebate scheme would merely correct this.

There are many other issues associated with the proposed tax/rebate scheme, but the first-order considerations discussed here would indicate that it is a useful supplement to the fleet average standards. However, further analysis is required.

8. Summary

Fuel economy regulations in the Energy Policy and Conservation Act of 1975 are a unique attempt to establish average standards across slightly differentiated markets. This law could dramatically alter the structure, sales, and pricing of the automotive industry as it has existed since World War II. It is likely that there will be some period (between technical improvements in present cars and major technological changes in the design of cars) when firms in the automotive industry

will be constrained by the increasing fuel economy standards. Indications are that this will occur in the early 1980s.

Various strategies of passing through penalties, altering prices and production mixes will be tried by the companies. One strategy is to treat the penalty and credits as extra production costs. The marginal penalty will vary perversely among firms according to a firm's production mix (lower for firms the more they break the standard). Another response is to use pricing and output strategies to meet the standard. All reasonable responses to this regulation include increasing the price of large fuel-inefficient cars, whose increased profits subsidize price cuts on small fuel-efficient cars which more than meet the standard.

This has wide-ranging consequences for the structure of the industry. Firms producing primarily small cars would tend to be undercut by others and would suffer losses -- a perverse incentive for producing small cars -- unless they expanded into at least some large car production. Most firms will face incentives to gravitate toward the same model mix. However, under certain situations, there will be an incentive for one firm to specialize in large car production. Also, a long list of inefficiencies result from this law.

One possible solution within the framework of this legislation would be to create a market in rights to producing fuel-inefficient cars. However, this option is not consistent with the present working of the law. It appears that a tax/rebate scheme whereby a tax on inefficient cars subsidized the purchase of efficient cars, could re-

tain much of the regulatory force of the present law while vitiating some of the negative features.

FOOTNOTES

1. This point is explored in detail in Chapter 1 of this report.
2. "The National Energy Plan," Executive Office of the President, Energy Policy and Planning, April 29, 1977; pp. 35-40 deal with automotive fuel economy.
3. Title V, Part A; 15 USC [2001-12].
4. See J.B. Heywood and L.H. Linden, "Alternative Automotive Power Systems and Energy Conservation," in Energy and Transportation (Society of Automotive Engineers, Warrendale, Pa., February 1976) for a discussion of the process by which major technical changes to the automobile are introduced into the passenger car market.
5. See H. Kahn, "Makers Uncertain if They Can Achieve 27.5 MPG by 1985," Automotive News, May 23, 1977, for a survey of submissions by the automobile manufacturers to the Department of Transportation and the Congress describing their strategies for meeting the standards through technical changes and their indications that they do not believe these changes to be adequate.
6. This was indicated to the authors in conversations with automobile industry executives.
7. Difiglio, C. and D. Kulash, "Market and Mobility," Report of a Panel of the Interagency Task Force on Motor Vehicle Goals Beyond 1980, (U.S. Department of Transportation, Washington, D.C., March 1976).
8. This is simply a rough estimate of the normal welfare loss in economic theory, approximated by:

$$\begin{aligned}
 & \left(\frac{1}{2}\right) (.25) \left(\text{avg. price of a large car}\right) \left(\text{size of annual reduc-}\right) \\
 & \quad \quad \quad \left(\text{under normal conditions}\right) \quad \left(\text{tion in big car sales}\right) \\
 & + \left(\frac{1}{2}\right) (.10) \left(\text{avg. price of a small car}\right) \left(\text{size of annual increase}\right) \\
 & \quad \quad \quad \left(\text{under normal conditions}\right) \quad \left(\text{in small car sales}\right) \\
 & = \$1.7 \times 10^9.
 \end{aligned}$$

Chapter 4

STANDARDS VERSUS TAXES: THE POLITICAL CHOICE

by

Frank Lerman

The growing concern for the environment in the 1960s and 1970s has produced a large increase in activity not only in the political arena but in the political science arena as well. Concomitant with the growth of environmental teach-ins, environmental lobbies, and environmental legislation has come an increase in the volume of scholarly writings on the politics of the environment. Some of this writing has been polemical and hortatory in nature, urging its readers to create new life-styles and new laws and political institutions; other works have been more analytical. In the latter category, scholars have examined the growth of environmental interest groups, the passage of environmental legislation by Congress, and the implementation of that legislation (or lack thereof) on the state and local levels. Students of the policy process have attempted to discover the relationships between the rise in public concern over the environment and the response by our executive, legislative, and judicial institutions, and the extent to which policy outcomes have reflected the strengths of the several interest groups concerned with this issue area.¹

Most of this analysis of environmental policy has treated the type or strength of the policy as the dependent variable, either explicitly or implicitly. In other words, policy analysts have attempted to use their knowledge of congressional behavior and interest group theory to explain

why a particular piece of environmental legislation was or was not enacted, or how effective the provisions were. But political scientists have ignored another equally important issue in the analysis of environmental policy--the question of why we have chosen to regulate the environment as we do. On the whole, the U.S. has chosen a regulatory policy based on the establishment of a system of standards of environmental cleanliness, and a system of penalties to enforce compliance with these standards. Of course, this is not the only solution to the problems of environmental regulation, either in theory or in practice. For example, economists who have long been concerned with the problems of regulating economic activity, often argue that market-type solutions, such as effluent taxes or the establishment of a pollution rights market, would achieve the goal of pollution reduction more efficiently (i.e., at less cost to society) than a system of standards and penalties.² This method of regulation exists not only in the realm of theory; a number of European nations, both in Western Europe and in the Socialist Bloc, have made use of effluent taxes in their pollution control programs.³

The question thus arises: Given a large body of opinion in favor of the use of market incentives to attain environmental goals, and given that the experience of other nations has shown that these methods attain some degree of success, why has the U.S. almost completely ignored market solutions in favor of a "standards and penalties" approach? Here we do not argue that the market solution is better than the regulatory solution or vice versa. This ground has been extensively covered by other professionals, and need not be rehashed one more time. Rather, the purpose here is to explore why we have chosen the approach that we have.

The exploration of this topic is divided into three sections. First will come a short description of the two methods of regulation--standards and penalties on the one hand, and market incentives on the other. Next I will develop several explanations for our choice of regulatory over market solutions, referring to a number of attempts to develop a system of market incentives. Finally, I will offer some conclusions about the most important factors in the regulation of environmental activity in the United States.

1. Methods of Control--Standards vs. Taxes

The argument for the use of some kind of emission tax to regulate pollution rests upon the analysis of the operation of competitive markets. If the industry is competitive and a few other conditions are satisfied,⁴ it can be shown that each firm in the industry, and the industry as a whole, will produce its output efficiently (i.e., at the lowest cost). But the appearance of externalities upsets this happy outcome: industry still produces at the lowest cost to itself but not at the lowest cost to society. For example, let us say that the production of some good entails air and water pollution. This pollution is certainly part of the cost of production; individuals suffer decreases in their welfare. But until recently the costs of pollution have not been paid by the polluting industry but by the general public in the form of reduced environmental quality. Because of this divergence between private and social costs the competitive firm will have an incentive to use production techniques that are inefficient from the standpoint of society; the firm will be economical in its use of those resources which it must provide, and profligate in its use of those resources--like air and water--which society provides at no cost.

In order to rectify this situation we need some method of forcing firms to internalize their externalities, i.e., to treat social costs in the same manner as private costs. This can be done by establishing a set of pollution taxes equal to the marginal social cost of each pollutant.⁵ This will force the firm to pay the true social cost of production, and production will once again be efficient.

The use of standards to regulate pollution is not based on any such elaborate chain of reasoning. If society decides that the present level of pollution is undesirable, it can set up a system of standards which will require economic actors to limit their output of pollution to whatever level is desired. This method seems much more simple and direct than a tax system, and it can achieve the same end--a reduction of pollution. In fact, in a world of certainty and perfect information the choice between standards and taxes is unimportant. Under these conditions society can attain any desired level of pollution and efficiency using either taxes or standards. With less ideal conditions, the choice is much more complex. We live in an uncertain world with imperfect information, and under these conditions it is often unclear whether taxes or standards will produce a better result.

Proponents of standards argue that the uncertainty of the linkage between the tax and pollution output makes a tax system too undependable. It is argued that even in a competitive industry, in which all firms respond immediately to price signals, we will be unsure of the actual amount of pollution reduction, since our information on the costs of pollution reduction will be scanty. The problems are multiplied when we consider that the connection between a tax system and pollution reduction is much more tenuous if the firms do not respond immediately to price signals. For example, under a condition of oligopoly rather than competition not only the tax system but the anticipated responses of all other firms in that industry would effect the behavior of each firm, leading to an indeterminate result.

Proponents of pollution taxes argue that a system of standards gives firms the opportunity to evade the pollution controls. One example frequently given is that of the auto industry. In 1970, Congress set tough auto emissions standards for 1975 and 1976, with a \$10,000 fine per car for failure to meet the standards. In addition, Congress, recognizing that the development of the necessary technology was uncertain, gave EPA the power to grant a one-year extension in the deadline. But the importance of the auto industry to the American economy made it highly unlikely that the fines would ever be imposed; the imposition of the fine would certainly result in a shutdown of the penalized company, producing disastrous consequences for the economy. As a result, opponents of the standards approach can argue that the auto companies had incentives to delay implementation of the necessary technology as long as possible, forcing the extensions to be given. And the extensions have been given--first by EPA in 1973, then by Congress in 1974, and again by EPA in 1975.

The purpose of this discussion is not to argue in favor of one method or the other, but merely to illustrate some of the arguments that are made. Actually, as people on both sides of the argument are now beginning to recognize, there is probably no general solution to the question of which regulatory instrument is best. The choice depends on the particular characteristics of the problem, the structure of the industry, the availability of information, the state of technology, and the losses incurred if regulation is too stringent or not stringent enough.⁶ The evaluation of these factors is quite a complicated undertaking and, a priori, one might expect that environmental policy would often be best implemented by a combination of taxes and standards. The choice of regulatory instruments in any given case would depend on the particular characteristics of the situation. But the empirical evidence contradicts

our a priori expectations. If we examine the methods which Federal and state governments have used to regulate pollution, we find that they consist almost exclusively of standards and penalties. Certainly policy analyses of the type described above have not been completely lacking, and the results have not been completely ignored, but even a cursory examination of the record reveals that often political decision-makers have used other criteria in developing environmental regulations. It is useful, then, to examine the values and attitudes of key environmental decision-makers to try to determine why we have chosen standards over taxes in regulating environmental pollution. The answer to this question should shed light on current difficulties in auto regulation, and provide some perspective regarding the changes that are likely to prove feasible in the future.

2. Alternate Theories of Societal Choice

Thus far, two methods of regulation of economic activity have been presented. Economic theory often leads in the direction of tax systems but, as pointed out, our choices in this area do not seem to revolve around economic models. Instead, we must look elsewhere for our explanations. The choice of standards over taxes in any particular instance might be explained by the way transaction costs happened to be distributed, or by analysis of interest group activity or the particular stakes of key Congressional leaders. But the topic of interest is not why a particular piece of legislation was passed or rejected. Rather, we should seek to explain the paucity of attempts to tax pollution and the general lack of receptiveness of members of Congress, environmental interest groups, and the public to the attempts that have been made. The intent here is to

move away from explaining policy decisions by factors unique to a particular piece of legislation, even though these unique factors will not be ignored. Instead the focus is on factors that seem to apply across the entire policy arena. Using this approach we can examine how the beliefs and attitudes of decision-makers have been affected by professional background and training, prior regulatory experience, relationships with constituents and interest groups, and basic notions of equity and fairness.

2.1 The Perception of Pollution as a Health Problem

If one examines the origins of our current concern with environmental pollution, one sees that it first came to public attention as a health problem. Any polemical speech or tract on the subject is likely to mention the 1948 Donora smog which killed twenty people, or the 1952 London smog which killed over 4,000, as prime examples of the dangers of pollution. This orientation can be clearly seen in the speech made by Senator Ribicoff in introducing the Clean Air Act. While he makes reference to the economic damage caused by air pollution, he overwhelmingly emphasizes the health effects of pollution.⁷ Certainly Senator Ribicoff's concern for health had a basis in his crusade against traffic deaths while Governor of Connecticut and in his tenure as Secretary of Health, Education, and Welfare. But the primacy of health effects in his speech was probably due to the general perception of the problem rather than idiosyncratic factors, as is made clear by numerous other references to air pollution from that time.⁸

The fact that environmental action had its origin in a concern for public health had several consequences for environmental regulation. First, the regulations were developed with the primary goal of protecting health, not of achieving economic efficiency or any other objective. Since it is

usually quite difficult to attain two distinct and sometimes conflicting goals with the same piece of legislation, it is not surprising that economic considerations were downplayed. Witness Senator Muskie's response to Senator Griffin's question about the cost of the 1970 Clean Air Act

Amendments:

Mr. Muskie: No. I said in my statement--I have not hidden anything--that our responsibility is to tell the industry what the public health requires.⁹

Second, since the intent of the regulatory effort has been to prevent damage to health, a large proportion of the expert witnesses who developed the regulations were doctors or other health professionals. The extent of the role of health professionals can be seen by examining the lists of witnesses at Congressional air pollution hearings. The 1963 Clean Air Act hearings were dominated by politicians and interest group representatives, but most of the expert witnesses who testified were health professionals, joined by some engineers. No economists, or others with a general concern for broad cost-benefit tradeoffs, were present. The same held true in hearings before the Subcommittee on Air and Water Pollution of the Senate Public Works Committee (the Muskie subcommittee) in 1964 and 1966. By 1967 the list of witnesses called during hearings on the Air Quality Act included one economist, representing the UAW, but he testified as an interest group representative rather than as a professional commentator on issues of efficiency, equity, and regulatory policy. One economist testified in 1969, but again he was called not as an expert witness but as an interest group representative. It was not really until the hearings on Water Pollution in the Spring of 1970 that economists were called as witnesses to present professional opinions on the costs and benefits of alternative regulatory instruments.

The strong association of pollution with health tends to place the issue in a well-defined historical context. Traditionally, health hazards have been regulated by determining their effects on health, then setting standards which, if followed, would render the hazard no longer dangerous. What is more, this approach seems intuitively desirable. If a small concentration of a toxic chemical in water or food can cause permanent injury or death, we must take special precautions to ensure that safe levels are not exceeded. It would seem that we can achieve this goal with far more certainty by mandating the level that is not to be exceeded, rather than by relying on the uncertain connection between financial incentives and the resulting level of emissions.¹⁰ As noted earlier, if we were in a world of certainty it would not matter which approach we used; the desired level of pollution could be obtained by either taxes or standards. But in the real world, we must examine the consequences of mistakes in setting the level of taxes or standards.

But when the danger from emissions is less immediate and the chances of economic dislocation due to improper control is greater, the pendulum begins to swing from standards to taxes. Certainly the danger from air pollution is not as acute as the danger from the chemicals in the example above. A detailed analysis might have shown a tax system to be more suitable than standards for the regulation of air pollution. But detailed analyses of this question were not done; rather, certain similarities between the regulation of pollution and the regulation of toxic chemicals became obvious to the health professionals involved in the development of regulation, and existing regulatory form was transferred from one field to the other. Other similarities that might have emerged from economic analysis were not brought to the fore, partially because economists were not prominent in policy debate before the Congress.

Certainly, then, this is one explanation (albeit a partial one) for the use of standards rather than taxes to regulate pollution: we regulate not so much by analysis as by analogy.

2.2 Professional Outlook--Lawyers and Economists

One explanation for the use of standards rather than taxes that seems plausible a priori is that lawyers dominate the policy process, both as members of Congress and committee staffers, and they would naturally tend to rely on regulatory instruments familiar to the legal profession. This is not to imply that such a choice is a deliberate attempt to create more jobs for lawyers. But a lawyer's professional orientation will probably lead him to look for solutions in certain directions rather than others; it is the extent of this effect that needs to be understood.

If one had to briefly characterize the central concerns of conventional economic analysis, then one could say that they are first overall productive efficiency, and second the broad distributional implications of each alternative action. Lawyers, on the other hand, are concerned with the rights of individuals. The primary focus of the legal profession is not upon determining the rational choice for society as a whole, but on those rights a person has and the enforcement of these rights through the legal system. As one lawyer puts it, "The basic priority is recognition by our courts that the public has a right to a salubrious environment."¹¹

Thus, if we recognize the regulation of pollution as a matter of individual rights, the proper way to enforce these rights is through the court system. But as has been shown in countless environmental cases, it is difficult for the courts to force an agency to enforce some kind of vague right to a healthful environment. In order for a right to be enforceable in court, ordinarily the violation of that right must involve

measurable (though not necessarily quantifiable) harm to some identifiable individual.¹² If some measurable harm is required to be found in order to enforce individual rights, it would seem that the logical procedure is to determine the level of pollutant necessary to cause harm, define in law the level of pollution that is harmful, and give the individual standing to sue in case he is harmed. In other words, use the standards and penalties approach. Thus, the logical process which the lawyer encounters during his training and throughout his professional life, and his natural focus on individual rights, lead to standards rather than taxes.¹³

2.3 The Congressional Process--Symbolic Politics

Despite the origin of pollution regulation as a health question, and the dominance of the policy process by lawyers, by 1970 the tax approach had become embodied in laws introduced before Congress. In the spring of 1970 hearings were held on S. 3181, a bill sponsored by Senator Proxmire and a bipartisan group of nine other senators which was designed to put into effect "a schedule of national effluent charges for all those substances other than domestic sewage which detract from the quality of the water..."¹⁴ Then in May 1970 the Nixon Administration proposed a tax of \$4.25/lb. on the lead additives in gasoline. This tax was designed to increase the price of leaded gasoline until it was greater than the price of unleaded gasoline of the same octane rating. Presumably oil firms would then have an incentive to produce unleaded gasoline and the consumer an incentive to buy it, significantly reducing the level of lead compounds in the atmosphere. Yet both of these measures went down to defeat, as did the sulphur tax of 1972 and a number of other proposed pollution taxes since that time. In order to understand the lack of acceptance of effluent taxes, it is helpful to examine the Congressional process through which this legislation passed.

Since its publication, Edelman's The Symbolic Uses of Politics¹⁵ has provided a guide to political scientists studying the regulatory process. Given its importance for a thorough understanding of the politics of the regulatory process, it is worthwhile to present his argument at some length. Edelman first observes that tangible benefits are usually not distributed to the general public as promised in the debate surrounding regulatory legislation. Edelman finds an explanation for this curious state of affairs in the observation that the population can usually be divided into two groups with regard to any given regulation. The first is a small collection of individuals and groups directly concerned with the regulated activity, for whom the consequences of any regulation are likely to be great, while the second of these is composed of the great bulk of the citizenry, for whom the costs and benefits of regulation are likely to be quite small. A citizen not directly concerned with the regulated activity has many more important things to which he must give his attention--his family, job, etc. To the extent he pays any attention at all to government activity in the policy area under consideration he will attempt to conserve his limited information-processing capability by looking for cues, for rules of thumb by which he can evaluate government regulatory policy.

In such a situation a politician can pursue a policy that will satisfy everyone. He can provide both the tangible benefits desired by those directly concerned with the regulated activity, and the rules of thumb needed by the mass public by providing strong, tough laws which are abrogated in practice to the benefit of special interests. The tough-sounding laws themselves stand as symbols of the government's concern for the well-being of its citizens, as evidence that the government is "doing something" about the problem, while appropriate administrative and budgetary

policies insure that the intentions expressed in the law are thwarted and the special interests are satisfied.¹⁶

It is important to note that it would not be possible to simply pass laws favoring special interests without providing symbolic reassurance to the public. The "tough" laws are necessary for the maintenance of stability in the political system. Edelman states:

There is, in fact, persuasive evidence of the reality of a political interest in continuing assurances of protection against economic forces understood as powerful and threatening. The most relevant evidence lies in the continuing utility of old political issues in campaigns. Monopoly and economic concentration, antitrust policy, public utility regulation, banking controls, and curbs on management and labor are themes that party professionals regard as good for votes in one campaign after another, and doubtless with good reason. They know that these are areas in which concern is easily stirred. In evaluating allegations that the public has lost "interest" in these policies the politician has only to ask himself how much apathy would remain if an effort were made formally to repeal the antitrust, public utility, banking, or labor laws. The answer and the point become clear at once.

The laws may be repealed in effect by administrative policy, budgetary starvation, or other little publicized means; but the laws as symbols must stand because they satisfy interests that are very strong indeed; interests that politicians fear will be expressed actively if a large number of voters are led to believe that their shield against a threat has been removed.¹⁷

Edelman portrays politicians as Machiavellian manipulators, but it is not necessary to accept this characterization in order to recognize the importance of symbols for the electorate. Even when a politician is genuinely interested in working for the public interest, the adoption of policies of benefit to the mass public might still arouse opposition if the policies did not appear to be in the public interest. Thus, a requirement for any policy instrument, whether intended to benefit the

mass public or special interests, is that the public perceive it to be in the public interest. The form of the law must provide symbolic reassurance to the electorate that the government is doing good.

One quality of effluent tax policies is that they appear not to provide the necessary symbolic reassurance, especially when compared to standards and penalties. First, there is a general perception of effluent taxes as an indirect solution, and standards as more direct. This attitude is not an isolated one, but often appears in the speeches and writings of environmental bureaucrats, lobbyists, and Congressional staffers.¹⁸ This perception on the part of those active in environmental affairs seems to have two components. First, many people are personally skeptical of taxes as being too indirect. In this view, the uncertainties of the tax-effluent level linkage, added to the uncertainties of the effluent level-health linkage, makes it far less likely that the policy will achieve its ultimate goal--the improvement of public health. Since it is difficult enough to determine what levels of pollution control must be attained if any given public health effect is to be achieved, we do not need to add the complexities of tax structure and the vagaries of polluter response.

Second, even those who intellectually accept the idea of a pollution tax, and admit that it may be a useful instrument, tend to agree that the public will not accept it. The concepts involved are simply too difficult to explain, and the level of economic sophistication among the public is too low for the case for pollution taxes to be made plausible.

Another, and perhaps more important symbolic barrier to the adoption of effluent taxes, is the notion of the tax as a "license to pollute." Once again there seems to be no survey data to show that this idea is widely held by the general public, but it is certainly present among the

elites who write articles, testify before Congress, and serve in the executive and legislative branches. The phrase "license to pollute" appears quite often in environmental rhetoric.¹⁹

The idea behind the slogan is that if pollution taxes instead of standards are used the rich will be free to continue to pollute as they wish, while the poor will be forced to stop. This stage of affairs is held to be wrong, since all citizens should share equally in the fight against pollution. At first glance it is difficult to understand why many people would support this concept. It is doubtful that many would be in favor of rationing food or clothing or land or Cadillacs in order to ensure that the rich do not have the ability to obtain more of these goods than do the poor. An economically rational person might reply that if the rich are able and willing to pay the true social cost of their pollution, society should be quite willing to let them pollute--just as society should be quite willing to let them have any quantity of any good as long as they are willing to pay the price. As is evident, the ability to pollute occupies a different place in the public mind than do those other goods.

Certainly there are other goods and services which cannot be bought and sold in our society, even if buyer and seller are in agreement. Individuals cannot sell themselves into slavery, nor can they sign restrictive covenants which forbid the resale of property to blacks. Perhaps more to the point, the sons of the rich who are drafted into the military cannot pay the poor to fight in their places, even though this practice was legal until after the Civil War. In each case an economic rationalist could well argue that the transaction should be permitted to take place, unless one of the parties could be shown to be mentally incompetent. After all, both parties are willing to enter into the

transaction, and economic theory tells us that society as a whole will be better off if all such transactions between willing parties are allowed. Of course, it is doubtful that many citizens, even those who are most scornful of the "license to pollute" argument, would wish society to permit the transactions outlined in the sentences above.

Clearly there is something other than notions of efficiency that enter into our judgments on the matter. Considerations of equity and morality are equally if not more important. If one feels that pollution is a threat to survival, one might then classify pollution abatement as a duty for all members of society, no matter what their station in life.

In summary, standards provide the public with symbolic reassurance that action is being taken to fight pollution, while taxes do not. Standards seem more directly connected to the problem than do taxes, and do not bear the stigma of providing the rich with "a license to pollute." Not only does a tax system fail to have the positive symbolic content of a system of standards, it has the negative connotations described above. If we examine the period 1969-1970, when anti-pollution fervor was at its peak, we can see that symbolic reassurance through forceful action seemed to be the primary demand of the public; given the importance of symbolic action, we would expect the public to receive standards more favorably than taxes.

It is clear that the public wanted "tough" action taken against polluters at that time: "public demonstrations and opinion polls projected a clear message to decision-makers: Do something about pollution."²⁰ The demand was for dramatic action, action that would provide the cues, the symbolic reassurance the public needed. Unfortunately, while many arguments to support the effectiveness of pollution taxes can be made,

even their strongest supporter must admit that a proposal to establish a set of pollution taxes that would achieve economic efficiency is hardly dramatic or symbolic of "toughness."

Another explanation of the importance of symbols for the electorate is given by Mayhew. Mayhew's argument is similar to Edelman's: the electorate pays little attention to most government activity, and responds to political figures and events by looking for small amounts of information which serve as indicators of government activity. As a result, congressmen enhance their chances for re-election by concentrating on the provision of cues for the electorate rather than on the shaping of substantive public policy. One of these types of cue-giving activity is position-taking, defined by Mayhew as "the public enunciation of judgmental statements on anything likely to be of interest to political actors."²¹ In other words, congressmen are judged by their positions on issues rather than the effects of their actions, and one type of position that will attract the attention of the electorate is the support of a simple, direct action that conforms to popular perceptions of effective public policy.

This is not to say that congressmen have no concern for substantive policy. Rather, the argument is that the provision of effective policy may conflict with the generation of electoral support; effective policy-making may call for indirect, complex solutions while electoral concerns may call for direct, simple solutions. If this conflict does exist, congressmen concerned with re-election have a powerful incentive to support legislation that provides symbolic cues for the electorate rather than legislation that produces effective policy.

Another incentive for congressmen to support policies that are effective symbolically rather than substantively comes from the desire of congressmen

for public prestige. Many congressmen are continually searching for issues with which they can become personally identified, and which can become the basis for increased political influence. According to Wilson:

Regulatory proposals emerging from this process are likely to have certain distinctive features. First, in order to ensure vital publicity and develop political momentum in the competition for attention in and around Congress, the bills will focus attention on an "evil," personified if possible in a corporation, industry, or victim. Second, the proposal will be "strong"-- that is, there will be little incentive in the developmental process to accommodate conflicting interests and thus little incentive to find a politically acceptable formula which all effected parties can live with. (To compromise the proposal would be to sacrifice the capacity of the bill to mobilize support by its moralistic appeal.) Third, though few substantive bargains will be struck many procedural ones will...²²

Thus, the desire of Congressional leaders to enhance their careers produces the same kind of legislation as does the need to symbolically reassure the public--legislation that is tough, moralistic, and dramatic, but that mollifies special interests through procedural concessions that are not noticed by the electorate.

2.4 The Congressional Process--Interest Groups

Another factor in the failure of effluent taxes has been the behavior of interest groups. It would not be too great an error to say that the only interest group that has consistently supported the effluent tax approach has been academic economists, a group whose political influence is minimal. Business and labor have opposed effluent taxes, while the support of environmental groups has been mixed. Thus the preponderance of pressure on this issues has come on one side only--against effluent taxes.

For example, an examination of the record shows business consistently opposed to the use of effluent taxes. The list of businesses and business groups opposing the 1970 lead tax proposal is a lengthy one: U.S. Chamber of Commerce, American Petroleum Institute, Standard Oil of Ohio, the National Petroleum Refiners Association, Independent Refiners Association of America, the Lead Industries Association, DuPont, the Ethyl Corporation, and others. In arguing against the tax, these parties attributed to the tax almost every ill effect which they felt would arouse Congress and the public against it. They argued that the tax would be inflationary; that it would discriminate against poorer people, who could not afford to buy new cars that did not need leaded gasoline; and that it would discriminate against independent refiners, who could not afford the new equipment necessary to make high octane, unleaded gas. Some groups argued that no definite connection had been made between health effects and lead in the atmosphere, while others admitted that regulation was necessary but urged that it be done through standards rather than taxes. Labor has shown a similar lack of enthusiasm for this approach to pollution control.

It is not the purpose here to analyze the motives of business and labor in opposing effluent taxes. We have no knowledge of the deliberations in corporate board rooms and union headquarters preceding the decisions to oppose emission taxes. However, it may be useful to indulge in some speculation about these motives. It is possible that business believes that the use of standards would result in less cost to them than would pollution taxes. They may believe that the stringent penalties often associated with standards are unlikely to be used, and the standards poorly enforced.²³ They also may fear that pollution taxes would quickly change from an instrument for reducing pollution to an instrument for

raising revenue.²⁴ By this reasoning, politicians would always be tempted to raise the taxes on a particular emission far above the social cost, forcing the industry to invest much more in pollution control equipment than necessary in the long run and in the short run increasing pollution tax payments. The labor unions apparently see a threat to employment and workers' disposable income.

Another, perhaps more important, reason for business opposition has to do not with possible economic damage to the firm but with symbolic damage. It has been suggested that companies feel they are being branded as criminals without due process of law if they are forced to pay pollution taxes; they look on the tax as a fine for violation of the law instead of a morally neutral incentive to promote a cleaner environment.²⁵ Moreover, some tax is likely to be due even after the firm's effort to control pollution. Standards, on the other hand, allow relaxation if it is demonstrated that the company is trying as hard as it can but still cannot comply. In this case the stigma is placed on the unreasonable law rather than the violating firm.

This argument about symbolic damage may be nothing more than rationalization, but there is evidence that supports it. In late 1974 U.S. Steel took the unusual step of closing down one of its older, dirtier facilities in Gary, Indiana rather than pay a court-ordered \$2,300/day fine to keep it open. It seems improbable that the payment of such a small fine would make the facility unprofitable and, at any rate, U.S. Steel did not cite financial pressure in its decision to close the facility down. Instead, U.S. Steel stated that it did not intend to pay "a daily tribute to the government"²⁶ in order to keep the steel mill open a little longer. What the firm seemed to object to was not the amount of the fine but the label of "lawbreaker" that went along with it.

If anything, business is exceeded in its opposition to emission taxes by labor. An example of labor's position is provided by the testimony from Andrew J. Biemiller of the AFL-CIO on the 1970 lead tax proposal:

The administration's request for a tax on lead additives should be rejected.

We are convinced than an excise tax is the worst way of raising tax revenues and probably the least effective way of reducing air pollution.

The tax would be paid by the consumer who feels that nonleaded gasoline might damage his automobile engine or cannot afford or would not wish to buy a new automobile.

In other words, consumers forced to use, or choosing to use, leaded gasoline could continue to pollute the atmosphere. It is clearly a license to pollute, if you pay the price.

Moreover, on the basis of past performance it is likely the oil companies would raise prices on all grades of gasoline and realize windfall profits.

We are flatly opposed to the use of leaded gasoline and any other atmosphere polluting additives. And the way to achieve that goal is through the air pollution legislation now before Congress... . This legislation would, in particular, empower the Secretary of Health, Education, and Welfare to remove from sale and interstate commerce all gasoline additives, including lead, which cause or contribute to air pollution which endangers the health and welfare of any person.²⁷

This statement is especially interesting because it includes many of the themes discussed earlier. Biemiller refers to the license to pollute argument, and seems not to believe that the economic system really works in the way described by economic theory. The references to health effects are there, as well as a strong statement in favor of the direct controls provided by standards.

Environmental groups have been divided in their approach to pollution taxes. Some have seen taxes as a device to allow pollution and its ill effects to continue, while others have seen them as a useful device for controlling pollution, given the imperfect world in which we live. Their division of opinion can be seen in their responses to proposals debated before the Congress. The 1970 lead tax proposal saw only one environmental group testifying before Congress--the Sierra Club. Though the Sierra Club representative supported the lead tax, he seemed to have little impact on the Committee; he was not asked a single question. But it was the absence of environmental support more than the presence of environmental opposition that hurt the proposal. In the presence of unified business and labor opposition, the lack of countervailing pressure doomed the lead tax.²⁸

By the time the 1972 sulphur tax proposal emerged, many environmental groups had changed their minds. A number of local, state, and national organizations formed the Coalition to Tax Pollution in order to coordinate efforts in favor of the tax. The support was lukewarm, however; environmentalists could simply not bring the same passion to an argument for effluent taxes that they brought to other fights such as that against the SST. Two characteristics differentiate pollution taxes from these other issues. First, environmentalists have tended to become most aroused over negative proposals--to stop the SST, or a dam or airport. In these cases there is specific damage to some area or group and it is clear how that damage can be prevented; it is much easier to focus attention than it would be if the benefits were more diffuse and the action less visible. Second, as noted earlier, it often is difficult to trace the connection from a tax proposal through the economic system to its ultimate effect on the level of environmental damage. As a result, it is difficult for people

to convince themselves that pollution taxes will be that much better than standards or the status quo. This lack of a clear focus and lack of belief in the efficacy of pollution taxes has led to a less intense effort by environmentalists in this area than in other areas of environmental concern.

As a result, though many environmental groups have come to look favorably on the pollution tax approach, they have not exerted themselves in its favor. In general, environmentalists have taken the position that pollution taxes may be useful as a supplement to standards, but not as a replacement for them. An example of this position can be seen in the testimony of Richard Ayres of the Natural Resources Defense Council in support of the Brown-Ottinger Bill, which proposed a system of emission fees on all emission of pollutants in excess of present state standards:

Such a proposal would supplement the present regulatory program rather than replace it. It would deprive those who have refused to comply of the unfair competitive advantages they have gained over those who have complied. It would reverse the present financial incentive for noncompliance, exerting substantial financial pressure to comply. The penalty would not be a tax, since it would apply only on emissions in excess of the legal limits, and would cease once the source had complied. Its purpose is to produce compliance, not raise revenue.²⁹

Thus, the support of environmentalists for effluent tax legislation is mixed. Most probably would rather rely on standards, since they seem a direct and more certain way of reaching environmental goals, but they recognize that standards have run into the problems described earlier.

Of course, the reluctant commitment to pollution taxes on the part of environmental groups means that they are far less effective in obtaining them than they are obtaining other goals to which they are more passionately committed. And clearly their commitment to pollution taxes is not strong enough to overcome the opposition of business and labor.

2.5 The Congressional Process--Other Factors

In addition to the reasons mentioned above, a number of other factors have contributed to the failure of the lead tax in particular and the effluent tax approach in general. Referring back to the statement by Richard Ayres quoted above, it can be observed that he takes pains to emphasize that the proposal is not a tax but a penalty. This, in all probability, reflects the desire of environmental groups to keep jurisdiction over pollution taxes out of the hands of the Senate Finance Committee and the House Ways and Means Committee, which are not considered to be dominated by interests favorable the environmental cause.³⁰ Note that this opposition is not based on philosophical disagreement with the concept of pollution taxes; rather, it is based on the realization that the implementation and administration of such a law are important features; to leave them in the hands of interests regarded as opponents of pollution control would be folly. In contrast, standards at least have the virtue of being implemented by their proponents rather than their opponents.

To be fair, congressmen are subject to limitations specific to the job. They must develop mechanisms which enable them to evaluate the vast flood of information on a myriad of issues that pass their desks. Not surprisingly, these cues are essentially the same as those used by their constituents: symbolic language, both in the law itself and in the public speeches and Congressional debate accompanying that law. Not only do Congressmen depend on symbolic language to reassure their constituents they also depend on it to reassure themselves.

The statements of individual Congressmen reveal the importance of pre-existing beliefs and values, and of the symbolic aspect of legislation. As discussed earlier, the perception of pollution as a health problem led

regulators in the direction of the standards used to regulate other health problems, rather than the taxes which might be more appropriate for this particular health problem. In fact, in the early stages of the environmental movement, activists often explicitly rejected the idea that economic and technological considerations should be a factor in setting pollution standards. In many cases Congressmen shared these opinions; to quote Senator Muskie at the time of the 1970 Clean Air Act Amendment Hearings:

"Here we have learned that the tests of economic and technological feasibility applied to these standards compromise the health of our people and lead to inadequate standards."³¹

In addition, the views of Congressmen on pollution taxes are often influenced by mistrust of the economic system. As examples of some of the opinions about effluent taxes held by influential Congressmen, three excerpts from the hearings on the lead tax are presented here. The first is from a statement by Representative William F. Ryan of New York, the second is from testimony by Representative Paul Rogers of Florida, and the third is an exchange between Representative Al Ullman of Washington and Acting Assistant Secretary of the Treasury John Nolan:

My support for the President's proposal is qualified because I have two chief reservations as to the taxation vehicle for ending the use of leaded additives. Firstly, I do not have the same faith as that registered by the Secretary of the Treasury Kennedy in his letter of July 30, 1970, to Speaker McCormack that the pressures of economic competition will suffice to achieve the desired goal of cessation of the use of leaded additives in sufficient time...

My view of the responsiveness of American industry to the public's needs is far less sanguine, to say the least. Competitive pressures will only work if there is, in fact, adequate competition. If all producers merely pass on the

tax to the consumer--and consumers are not sufficiently responsive to the added costs of leaded gasoline over the of nonleaded products--leaded gasolines will persist...

I do not believe that taxation is any more than a half-step to the action that is required--total abolition of leaded gasolines.³²

Therefore, we reached the conclusion in the committee that lead in gasoline should be removed as quickly as possible to eliminate first, health hazards of particulates, and secondly, to allow new and innovative devices to reduce other pollutants.

Because I feel lead should be eliminated from gasoline, I appear here to voice my concern over the administration proposal to tax gasoline...

What this recommendation represents, as has been stated to this committee, is a license to pollute. The administration is saying that you may pollute with lead as long as you pay the government a certain amount...

In addition, I think we have faulty logic when we think the petroleum companies are going to pay the tax alone. It will simply be passed on and it will be the drivers who will be paying the taxes.³³

Mr. Ullman. There is nothing in this proposal to keep the petroleum companies from spreading it across leaded and unleaded gasoline, so they can pick up revenues from their total sales to cover this additional tax?

Mr. Nolan. There is nothing contained in our proposal to prevent that legally but it simply has to be the fact that if a company can make more profit on one product than another--and they will make more profit on the sale of unleaded gasoline after the imposition of this tax--it is going to emphasize the production and sale of that product.³⁴

These statements reveal the attitudes and beliefs of the Congressmen involved, and bring to mind many of the themes discussed earlier. Health is given absolute priority over economic effects. The "license to pollute" argument is cited, again without any accompanying analysis, the intonation

of the ritual phrase apparently being enough to arouse the desired emotions. The statements also reveal a disbelief of the economic system and a mistrust of the motives of producing firms. According to the Congressmen, firms seek to maliciously thwart the goals of the legislation even if it would cost them less to comply.

Certainly, this is not to argue that every Congressman has opinions similar to those displayed above. But the support from Congress, like that from many environmentalists, is lukewarm. The issue of taxes vs. standards does not seem to arouse the public's interest, so there seems to be little incentive for a political leader to seize upon this issue as a means to greater influence. Even those congressmen who do understand and support the concept of effluent taxes do not see it as be-all and end-all, but merely as another useful tool in the fight against pollution. They simply do not bring the same passion, the same involvement to their cause as do those congressmen who oppose pollution taxes as the devil's work.

Another reason for the failure of the lead tax was the confusion caused by the Nixon Administration's presentation of it. Ostensibly, the major purpose of the tax was to provide financial incentives to encourage refiners and consumers to gradually make the transition to unleaded gasoline. But the tax also took on the appearance of a revenue-raising measure, partly because it was introduced at the same time as several other revenue measures and was coupled with them during hearings, and partly because the Nixon Administration explicitly presented it as a revenue measure as well as an anti-pollution instrument. Consider this dialogue between a House Ways and Means Committee staff member and Treasury Secretary Kennedy:

Mr. Watts. What I had in mind or what I thought you might have in mind anyway would be to set up an upper limit on the content in lead in a fuel and give the companies a reasonable number of years to reach that limit if that is what you claim again you hope to do with the tax.

I don't see any difference in doing it directly or indirectly by the tax.

Secretary Kennedy. We have the revenue item which will bring in \$1.1 billion in revenue at a time it is needed. This amount will phase out as lead disappears from gasoline.

Mr. Watts. Which is your principle interest to raise more revenue or to get rid of pollution?

Secretary Kennedy. Both.³⁵

The presentation of the lead tax as a revenue-raising device created a dilemma which was immediately recognized by Congress and repeatedly brought up throughout the legislative deliberations: if the tax were high enough to significantly reduce the use of leaded gasoline there would be little revenue, and if the tax were low enough to allow significant amounts of leaded gasoline to be consumed there would be more revenue but less reduction of pollution. The result was confusion among Congressmen as to the real intent of the tax and a belief that the proposal was not well thought out. Another problem was that many Congressmen as well as others concerned with the environment felt that even if the tax were intended solely as a device to reduce pollution, the availability of revenue would encourage the Administration to eventually look upon it as a revenue measure. Even Congressmen who were favorably disposed to the idea of pollution taxes were alienated by this confusion, thus putting one more nail in the proposal's coffin.

3. Conclusion

As we have seen, the consistent choice of standards over taxes to regulate pollution has a variety of causes. The health professionals and lawyers who have dominated the policy process are influenced by training and experience to look toward standards rather than taxes. Interest group activity and inadequate presentation have helped kill effluent taxes, as have the unfavorable attitudes of many congressmen. But in many of these respects emission taxes are no different than other innovative public policies that are adopted. Certainly the regulation of pollution by the standards and penalties approach has had to overcome opposition by business groups and the ingrained attitudes of many congressmen, and the struggle has not been an easy one. But standards have had one large advantage over taxes--the support of a large, vocal collection of groups and individuals who are passionately committed to that policy. In general, this difference in support does not appear to be due to a rational calculation of the costs and benefits of each approach, but to differences in their symbolic content. Taxes are seen as an uncertain method of regulation that fails to take tough action against those who despoil the environment, while standards provide the symbolic qualities necessary to arouse support in their favor.

FOOTNOTES

1. For examples of this literature, see L. K. Caldwell, Man and His Environment: Policy and Administration (Harper and Row, 1975); R. A. Cooley and G. Wandesforde-Smith, Congress and the Environment (University of Washington Press, 1970); J. Davies, III, The Politics of Pollution (Pegasus, 1970); C. O. Jones, Clean Air (University of Pittsburgh Press, 1975); S. Nagel, ed., Environmental Politics (Praeger, 1974); W. Rosenbaum, The Politics of Environmental Concern (Praeger, 1973); A. E. Hutton and D. H. Henning, eds., Environmental Policy: Concepts and International Implications (Praeger, 1973).
2. P. Bohm and A. V. Kneese, The Economics of the Environment (MacMillan, 1971); J. M. Buchanan and G. Tullock, "Polluters, Profits, and Political Response: Direct Controls vs. Taxes," American Economic Review, Papers and Proceedings, LXV (1975), pp. 139-147; J. H. Dales, Pollution, Property and Prices (University of Toronto Press, 1968); Davies, op. cit., pp. 176-180; P. B. Downing, ed., Air Pollution and the Social Sciences (Praeger, 1971); G. Hagevik, "Legislating for Air Quality Management," Law and Contemporary Problems XXXIII (Spring, 1968), pp. 369-98; A. V. Kneese and C. L. Schultze, Pollution, Prices, and Public Policy (The Brookings Institution, 1975); D. R. Mackintosh, The Economics of Airborne Emissions (Praeger, 1973); Karl-Goran Maler, Environmental Economics: A Theoretical Inquiry (Johns Hopkins Press, 1974); Larry Ruff, "The Economic Common Sense of Pollution," The Public Interest XIX (Spring, 1970), pp. 69-85; L. J. White, "Effluent Charges as a Faster Means of Achieving Pollution Abatement," Public Policy, XXIV, 1 (Winter, 1976), pp. 111-125; L. J. White, "The Auto Pollution Muddle," The Public Interest, XXXII (Summer, 1973), pp. 97-112; H. Wolozin, ed., The Economics of Air Pollution (W. W. Norton and Co., 1966).
3. W. A. Irwin and R. A. Liroff, Economic Disincentives for Pollution Control: Legal, Political, and Administrative Dimensions (Environmental Protection Agency, 1974), pp. 111-117.
4. The conditions include an absence of transaction costs, a "smooth" production function, and an absence of externalities, a point to be discussed later.
5. Problems arise with this simple scheme if the marginal social cost of pollution is not the same at all levels of pollution, but the complexities of this situation need not concern us here.

6. See L. G. Billings, "Pressure Cookery in Washington: The Politics of Pollution," in Rescuing Man's Environment, R. C. Axtmann, ed. (Princeton University Council on Environmental Studies, 1972); M. L. Weitzman, "Prices vs. Quantities," Review of Economic Studies, XLI (1974), pp. 477-91.
7. U.S. Senate, Committee on Public Works, Special Subcommittee on Air and Water Pollution, Hearings, Air Pollution Control, 88th Congress, 1st Session, 1963, pp. 37-40.
8. For example, see Davies, op. cit., 156; R. Martin and L. Symington, "A Guide to the Air Quality Act of 1967," Law and Contemporary Problems, XXXIII (Spring, 1968), pp. 239-241.
9. U.S. Senate, Committee on Public Works, A Legislative History of the Clean Air Amendments of 1970 (Washington: Government Printing Office, 1974), p. 238.
10. For a sound economic analysis supporting this view, see Weitzman, op. cit.
11. V. J. Yannacone, "A Lawyer Answers the Technocrats," Trial, August-September 1969, pp. 14-15.
12. G. J. Stigler and M. F. Cohen, Can Regulatory Agencies Protect Consumers? (American Enterprise Institute for Public Policy Research, 1971), p. 59.
13. Of course, this approach is not universal among lawyers. For example, see B. Ackerman, et al., The Uncertain Search for Environmental Quality (The Free Press, 1974); I. M. Heyman, et al., "The Challenge of Environmental Controls," Business Lawyer, XXVIII (March 1973), pp. 9-31; J. C. Krier, "The Irrational National Air Quality Standards: Macro and Micro Mistakes," UCLA Law Review, XXII (1974), pp. 323-342; C. J. Meyers, "An Introduction to Environmental Thought: Some Sources and Some Criticism," Indiana Law Journal, L (1975), pp. 426-53; A. J. Rosenthal, "The Federal Power to Protect the Environment: Available Devices to Compel or Induce Desired Conduct," Southern California Law Review, XLV (1972), pp. 397-449.
14. S. 3181, 91st Congress, 1st Session, Section 4 (a).
15. M. Edelman, The Symbolic Uses of Politics (University of Illinois Press, 1964), pp. 22-43.
16. A complementary view of this process, focused specifically on the Clean Air Act amendments of 1970, is provided in Chapter 5 below.
17. Ibid., 37.
18. For example, see J. E. O'Fallon, "Deficiencies in the Air Quality Act of 1967," Law and Contemporary Problems, XXXIII (Spring, 1968), p. 278.

19. For examples, see U.S. House of Representatives, Committee on Ways and Means, Hearings, Tax Recommendations of the President, 91st Congress, 2nd Session, 1970, pp. 135-139, 338; U.S. Senate, Committee on Public Works, Subcommittee on Air and Water Pollution, Hearings, Water Pollution--1970, Part 2, 91st Congress, 2nd Session, p. 540.
20. C. O. Jones, "Speculative Augmentation in Federal Air Pollution Policy-Making," Journal of Politics, XXXVI (May 1974), p. 463.
21. D. Mayhew, Congress: The Electoral Connection (New Haven: Yale University Press, 1974), p. 61.
22. J. Q. Wilson, "The Politics of Regulation," in Social Responsibility and the Business Predicament, J. W. McKie, ed. (The Brookings Institution, 1974), p. 146. The procedural bargains referred to at the end of the quotation refer to methods of undermining the strong language of the bill--the "fine print" which often sabotages the implementation of the simplest and most direct legislation.
23. White, "Effluent Charges as a Faster Means of Achieving Pollution Abatement," op. cit., p. 118.
24. John F. Burby, "White House Plans Push for Sulfur Tax Despite Strong Industry Opposition," National Journal, October 28, 1972, p. 1665.
25. Based on interviews with congressional staffers.
26. "Gary is Seen Facing Woes on Steel Mill," New York Times, January 2, 1975, p. 57.
27. 1970 Tax Hearings, p. 371.
28. John F. Burby, "White House, Activists Debate Form of Sulfur Tax; Industry Shuns Both," National Journal, October 21, 1972, p. 1647.
29. U.S. Senate, Committee on Public Works, Subcommittee on Environmental Pollution, Hearings, Implementation of the Clean Air Act--1975, Part 2, 94th Congress, 1st Session, 1975, p. 1508.
30. This motivation is by no means a weak one; several Congressional staffers and environmental lobbyists expressed this thought spontaneously immediately upon being asked their opinion of pollution taxes, and most others agreed with the sentiment when asked.
31. Legislative History of 1970 Amendments, p. 226.
32. 1970 Tax Hearings, pp. 451-54, passim.
33. Ibid., pp. 337-34, passim.
34. Ibid., p. 70.
35. Ibid., p. 61.

Chapter 5

ANOTHER VIEW OF THE POLITICS OF AUTO EMISSIONS CONTROL

by

Howard Margolis

There is a distinctively "American way" in the business of regulating automobile emissions, followed in no other country.¹ Six years ago, Congress wrote into the law very tough standards, to be met on a very tough time schedule, and did so explicitly without regard for cost or technical feasibility. Once the standards came into effect, the automakers were to be subject to fines of up to \$10,000 for each car not meeting the standard, should they produce any such cars. This was known as "forcing technology," and "holding industry's feet to the fire."

Since one can guarantee that "holding industry's feet to the fire" will produce cries of anguish, but one cannot guarantee that the desired technology will be forced to emerge, an escape clause is necessary. The details of the escape clause have varied from year to year. But the basic scenario has (through 1976) remained unchanged. After a suitable period in which changes of bad faith and incompetence are exchanged among industry, environmentalists, Congressional committees, and the EPA, the standard is delayed once again. No one is happy with this outcome. To industry it is a continuation of a disruptive and inefficient way of handling the problem. To environ-

mentalists it is an annual sellout to industry pressure. But to each side the stand-off has been preferable to a clear-cut victory for the other.

It is possible that this pattern will be broken in the coming year. The situation was left in an exceptionally confused state when Congress adjourned in October 1976, without acting on a bill providing a further extension of the standards. A new Administration now is in office, with a generalized environmentalist commitment, but without any specific commitment to or need to defend past policies. Meanwhile, industry has proceeded with plans to produce 1978 autos to 1977 standards, on the grounds that it is too late to do anything else--as it seems to be, though this presumption that the law will be bent to meet industry's dilemma does not sit well with the Congress.

So the 1970 Act may face a critical turning point in the coming year. Something must happen to resolve the confused and untenable position left at the end of 1976. One can sense a certain amount of hope, possibly justified, that a solution will be found that will end, in one way or another, the annual battles over delay. It is a good time, therefore, to look back and try to understand what has been going on.

The purpose of this chapter is to provide a view of what we may call the "social politics" of this recurrent battle. That is, we are directly concerned not with the details of particular decisions, but with the broader social/political context which produced what is, after all, a very peculiar way of handling the problem of auto emissions controls.

We will see that the key to understanding what has happened is in one's view of how the original legislation came to be written in a way that virtually guaranteed the confrontations that have followed.

Naturally, there are parallels between the handling of auto emissions and the treatment of other environmental problems, and (only partly overlapping) parallels with the American handling of business regulation generally. But it will be convenient to defer comment on these parallels until the concluding section of the chapter.

1. A Conceptual Framework

It is a fact of political life that (other things equal) a small number of people, each of whom has a substantial stake in the outcome of a political decision, will commonly exert more political influence than a large number of people, each of whom has but a small stake in the decision. So if a particular choice will bring \$1 worth of benefits to a million people, but will cost one thousand people \$1,000 each, we ordinarily have no doubt about how the decision will go. Indeed, even if there is a large excess of concentrated benefits over diffused costs, or the reverse, we still are not surprised to learn that the concentrated interest has won out over the diffused interest. The principle exception to this rule occurs when an issue has been made dramatically prominent. Scandal (as in GM's investigation of Ralph Nader's private life) or tragedy (as in the recent Kepone episode) often plays a precipitating role. Then the weight of numbers of an aroused public can override the concentrated interest.

James Q. Wilson has used these familiar observations to frame a simplified but useful typology of business regulation.² We can summarize his analysis with the use of Figure 1.

What of Case IV in Figure 1? This case, in which both benefits and costs are widely diffused, is not one that seems to have been much discussed in the political literature on business regulation.

		<u>COSTS</u>	
		Concentrated	Diffused
<u>BENEFITS</u>	Concentrated	I Affected parties compete for influence over the regulatory policy. E.g., NLRB	II Regulations serve Industry's interest with minimal (esp. symbolic) concessions to broader interests. E.g., CAB, Texas RR Commission
	Diffused	III Aroused public demands reform. E.g., FDA	IV

Figure 1 Wilson's Typology of American Business Regulation

For it is very generally assumed (as in Wilson's article, for example) that the regulated business activity either enjoys concentrated benefits or endures concentrated costs. Wilson lists the auto legislation, along with other environmental laws as examples of Case III. But in the case of auto emissions (and typically, though not always, for other environmental programs) we are really dealing with a Case IV situation in which the great bulk of the costs are as widely diffused

as the benefits. There are concentrated costs, of course. It is not hard to understand industry's unhappiness with tight regulation against tight deadlines; the howls we hear from the auto industry are motivated by real pain. Further, although it is hard (or impossible) to clearly identify these concentrated costs, there are presumably real dollar losses (or foregone opportunities for profit) that have accompanied the program. But from the point of view of overall impact of emissions controls on society, the costs concentrated on the industry are quite trivial compared to those dispersed among the users of automobiles. (Whatever the impression the general public may have had, the legislative history leaves no room for doubt that everyone directly involved understood this.³)

None of this argues that industry ought to pay, or could afford to pay, any substantial fraction of emissions control. It is the use of automobiles that creates emissions, and it seems appropriate that the users of automobiles should pay the costs of control. But clearly, the "concentrated costs/diffused benefits" model is an inaccurate description of the real situation, and (unsurprisingly, therefore) might lead to incorrect inferences about the effects of the legislation, and the interests of the various actors.

Now on the face of things, the law has conspicuously failed to do what, by its own language, it set out to do. We have not yet met the standards specified for 1975/6; and there are grounds to doubt that we will ever meet them. Indeed, as shown in Table 1, we have not yet (except in California) met the standards which were scheduled

for 1975 before the 1970 Amendments to the Clean Air Act were passed. Some other U.S. and foreign standards are included for comparison.

Of itself, the failure to achieve these legislated standards is not sufficient to deny the soundness of the effort. In the first place, it is hard to doubt that deficient as the result has been in terms of the legislated objective, nevertheless emissions are lower today than they would have been in the absence of the 1970 legislation. In any event, the concern with fuel conservation that arose so dramatically following the oil embargo in the fall of 1973 cut sharply against the single-minded concern with controlling emissions.

But it is hardly credible to place all, and perhaps unreasonable to place most of the blame on the unforeseeable problems of carrying out the program. What we will try to explain is how a policy likely to prove either unnecessarily expensive or ineffective swept through Washington with the near unanimous support of conservatives and liberals alike in the Congress, and with the effusive endorsement of a conservative, business-oriented administration. We will do so largely in terms of the interests of the active participants (politicians, environmentalists, and the industry).

We will stress two factors that facilitated this outcome. First, there is the variety of grounds on which a policy choice of this kind can be rationalized, so that misunderstandings could easily arise, and so that in any event, no one had compelling reason--or need have compelling reason today--to feel that his actions were selfish or stupid. Second, we will stress the fragmentation of expertise, authority, and responsibility which left the public at large, and even

those closely involved, confused about what was happening; and which assured that everyone had someone else to blame if things went badly.

Table 1. Selected Auto Emission Control Levels

Standard (gms/mi)	HC	CO	NOx
'63 (pre-control) actual	8.7	87.0	4.0
'65 (initial control)	6.5	65.3	not controlled
'68 (initial federal)	5.9	50.8	"
'70 (federal)	3.9	33.3	"
Nixon '75a	.93	16.2	1.4
California '75 to date	.9	9.0	2.0
Actual '77b	1.5	15	2
Muskie '75/'76c	.41	3.4	.4
Canada (thru 1980)d	2.0	25.0	3.1

Notes:

- a) Announced, February 1970 (includes upward adjustment for changes in federal test procedure);
- b) Actual '75/'76 was identical for HC & CO, but higher (3.1) for NOx.
- c) Statutory standards, drafted August, '70 by Muskie Subcommittee. HC and CO intended for '75; NOx intended for '76.
- d) Similar to those of Sweden and Australia. Common Market is substantially less stringent. Japan is numerically identical to U.S., but differences in test procedure make exact comparison difficult.

Especially in such circumstances, we can expect that those with an active role in shaping the decision would tend to follow a course

which fit comfortably with their private preferences. We need not assume that those private interests were purely selfish; indeed for the purpose of the argument of this paper, we need not assume they were selfish at all. But they were never identical to anything that could be mistaken for the interests of the broad public which would both pay the vast bulk of the costs of the program, and receive the bulk of the benefits. In particular, the mixed bag of elements which finally became law (it would imply much more conscious bargaining than actually took place to call it a compromise) is easy to question as the wisest public policy choice that could have been made. But once it did become law, new kinds of vested interests were created which made it difficult to change.

So, we will argue, what might casually appear as an example of Case III (concentrated costs, dispersed benefits: the auto industry vs. the public) resolves itself into an example of Case IV (dispersed cost and dispersed benefits), with the actual choices made under the conditions of Case I (concentrated costs; concentrated benefits). Bureaucrats, environmentalists, politicians, and to some extent even auto industry managers had substantial grounds for feeling satisfied with the outcome-diverse and conflicting as their interests were. What is most in doubt was how well the interests of society at large were served.

2. Reviewing the History

At the turn of the century, three powerplants were contenders for the dominant position in the infant automobile industry. Of the

three leading competitors (steam, electric, and gasoline) the gasoline-driven internal combustion engine had by far the most serious potential for emissions problems. For only in this engine did the motive power come directly from the fuel combustion chamber. Hence the designer could not optimize the burning of the fuel for low emissions without penalizing the design of the balance of the engine for the performance and cost features which more directly affect the car owner.

But for the infant auto industry--and indeed for a long time after infancy--emissions were not an issue. Half a century passed, during which facilities and know-how for designing, manufacturing and maintaining this engine steadily grew. Only then would question be raised as to whether the engine that dominated the industry was the one that would have been chosen if emissions control were a consideration. It would be an overstatement to flatly assert that, had emissions been a concern from the outset, then the ICE would not have been chosen. All we can flatly assert is that by the time questions were raised about emissions, a vast infrastructure, reaching from the top management of the automakers to the corner garage mechanic, had been built around the conventional ICE engine.

On the other hand, the absence of prior concern meant that, until significant pressure for controls began to develop in the 1960s, only incidental attention to emissions had gone into the design of these engines. We thus entered the period of major concern about emissions in the following mixed situation:

1. Substantial (on the order of 60%) reductions in emissions due to the incomplete burning of fuel (HC and CO) were possible at very little cost: on the order of \$25 per car.⁴
2. The situation would become more complicated to the extent it was necessary to control NOx, as well as emissions of incompletely burned fuel. (NOx comes from heating the air during combustion, not from incomplete burning of the fuel, so that steps to decrease HC and CO often increase NOx.)
3. In any event, at the margin, tighter standards become increasingly difficult and expensive. At some levels of emission-control stringency, therefore, it would presumably become more efficient to change the basic technology of the industry than to try to meet the requirements. (But whether the Muskie '75/Nixon '80 standards reach that breakpoint is in dispute. Both in 1970 and today, there are knowledgeable observers on both sides of the question.)

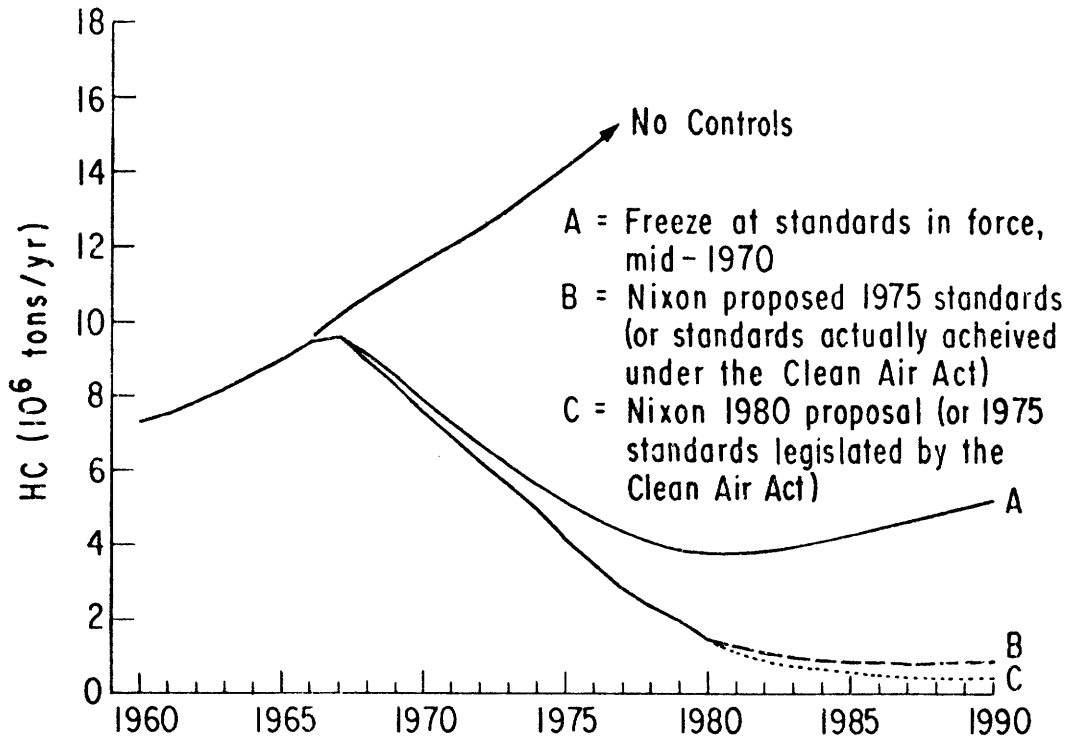
Now by 1970, the strategy implicit in point 1 had been exhausted. Model Year 1970 cars (which had gone on sale in the fall of '69) met federal standards which reduced HC and CO emissions by about 60%. Indeed, by the summer of 1970, when the Muskie subcommittee began to write its bill, public policy had already moved well beyond the "easy cleanup" strategy.⁵

In February, the Nixon Administration had proposed regulations (see Table 1) for MY '75 cars which extended the controls to NOx, and which simultaneously sharply tightened the existing standards for HC and CO. Further, goals for 1980 emissions had been published, and the Administration had proposed an R&D effort to develop an inherently clean engine technology by 1975. This set of proposals was a hedge against the dual possibility that (a) stricter standards would be desirable for the 1980s, and (b) these stricter standards could not be realistically met with the clean-up of the conventional engine which industry leaders openly treated as their preferred alternative.⁶

All of this was described in Nixon's February 1970 message to Congress. The only aspect of the program that was in any way controversial was the proposal for controlling fuels--in particular to provide no-lead gasoline. A month before the Nixon program was formally announced, the President of GM had described the forthcoming emissions standards to the Society of Auto Engineers, and said that the industry could meet them, provided the oil companies supplied gasoline which did not damage the devices (for GM, the catalytic converter) which industry planned to use. Since the oil companies were asserting that conversion to the new fuels would cost them (and eventually, the public) around \$5 billion, the auto companies would appear to have committed themselves to tight emissions controls in a most unambiguous way.⁷ For, having made their intent to meet much tighter emissions standards the basis for the Administration's proposal to control fuels, the auto companies had no basis for then arguing against the installation of the control devices.

Against this background, consider the projected impact of various emissions standards and timetables. The graph on the following page projects the impact of the Nixon Administration's proposals for 1975 and 1980, as submitted to the Muskie subcommittee in May, 1970.⁸ The bill produced by the subcommittee in August essentially moved the Nixon 1980 goals ahead to 1975.⁹ Note that Curve B represents both the Nixon proposal for 1975 standards and the standards in effect in 1975 (and since extended at least through 1978). In other words, the standards in effect today are approximately the standards for 1975 that the industry it could and would meet before the new legis-

Fig. II Projected effects of various Hydrocarbon Emissions Standards



Source : Legislative History of the Clean Air Amendments of 1970, Vol. II pp. 1083-1087

lation produced by the Muskie committee was ever conceived. Since this may strike some readers as surprising, it is worth emphasizing that the point is in no way controversial. As will be illustrated later on, everyone who has been close to the matter understands that this is how things have worked out. (More specifically, the current standards almost match the Nixon '75 proposal for HC; but fall short on CO and NO_x, though not by enough to make an important difference in projections such as that shown for HC.)

Similarly, Curve C in the charts represents both Nixon 1980 goals and the tougher program actually enacted. Standards are essentially

the same for these two cases, the difference being that under the program legislated (Muskie 1975 standards), the standards would come into effect five years earlier than under the Nixon proposal. The reason the five year speed-up makes little difference in the projections--so slight a difference that we can let Curve C represent both programs--is that throughout the 1970's the main factor influencing aggregate emissions is the gradual replacement of high emissions old cars (the average life of a car is 10 years) by relatively low emission new cars. The details of the standards--whether they reduce emissions by 90% or only, say, by 80% makes rather little difference. Or, put another way, had the Muskie standards actually been met, then new cars produced after 1975 would have been much more strictly controlled than under the Nixon proposal. But even under the Nixon 1975 standards, new cars would have had low emissions compared to the cars typically being replaced, and the difference between the two standards in terms of average emissions of all cars on the road (hence the difference in the aggregate level of auto pollution) would have been slight.

In summary, then, although particular dates have taken on great symbolic importance, the situation has always been that someone just looking at graphs to see the effects of changing the dates would wonder why such a fuss was being made.

This does not mean that there was no substantively significant difference between the Muskie and Nixon programs. The Nixon 1980 program represented goals subject to further discussion. The Muskie program made--or at least was intended to make those goals into a strong commit-

ment. Whether this difference would be worth its cost was open to much debate.

Thus the 1970 legislation had two quite distinct facets: in terms of the graphs, (1) firming up the commitment to Curve C (vs. Curve B), and (2) speeding up by 5 years the effective date of the Curve C standards, a speed-up with minimal effects on the environment for the reasons we have now reviewed. But it has been the speed-up in application of the standards, despite the difficulty of seeing much effect from doing so, that has been the root of the persistent controversy since 1970, and at the heart of the unhappiness with the law of both industry (which says the law is unreasonable) and environmentalists (who think it is the industry which is being unreasonable). We would like to understand the basis for this crucial policy choice.

At one point during the debate, the industry's sole defender (Sen. Griffin of Michigan) asked, "Does the Senator from Maine know what it would cost?"

Mr. Muskie: No. I said in my statement--I have not hidden anything--that our responsibility is to tell the industry what the public health requires.¹¹

How was it possible to justify this intense commitment? In the mood of the moment, when cleaning up the environment had very much the tone of a moral crusade, no analytical defense of the Act seems to have been required, and certainly none was offered. The important thing was to get clean air, and get it fast. Doing so was defined as reducing auto emissions by 90% (from 1970 levels). It was Congress' role to write into law the kind of strict controls the public was

said to be demanding; it was industry's task to "get the job done." Congress was willing to contemplate the possibility that the job might turn out to be technically infeasible on the time schedule demanded. But it was not willing to compromise in advance of a demonstration of infeasibility. "... we have learned," said Sen. Muskie, who was merely echoing the language of the committee's unanimous report, "that the tests of economic and technological feasibility applied to these standards compromise the health of our people and lead to inadequate standards."¹²

Now, making some allowance for political rhetoric, it might be possible to frame a strict cost/benefit argument justifying such a position. For if the health risk to the public were sufficiently severe, then one might well choose to put the whip to industry, push emissions down as rapidly as possible, and worry later about moving to better technology (possibly to a basically different technology) which could do the job more cheaply.

However, while conceptually one could frame such an argument, empirically there was no way to justify it. For the fact was, and remains today, that no one could make a solid scientific case that even totally eliminating auto emissions (possible only by totally eliminating automobiles) would have any detectable effect on the nation's health. Nearly all the usually-cited evidence for striking effects due to air pollution--the London "killer fogs", for example have no connection with automotive pollutants.¹³ When pressed, EPA will come up with estimates of health damage due to auto emissions. But these are hedged by so many caveats, provisos, and qualifications

that no one can claim they are anything better than a moderately informed guess.¹⁴

The situation was aptly described by an OECD expert committee in 1972 (with the U.S. represented by EPA): "Ultimately, the case for limiting vehicle emissions ... rests not on incontrovertible evidence of physical harm but rather on the growing feeling that it is in the public interest to minimize all emissions suspected of having adverse effects on man's health, well-being, and environment."¹⁵ But since it is always possible to achieve lower emission by paying a higher cost, "to minimize" emissions means, as a practical matter, looking at data such as that displayed in Figure 1 in the light of the costs and such indications of benefits as might be available, and reaching a judgment on how tightly to set the regulations. In such terms, it would never be possible to make a rational case for the crash program the 1970 legislation set in motion.¹⁶ So it was not at all surprising that the rationale of the program was framed in terms of moral imperatives which were presumed to transcend calculations of costs and benefits. In the process of slipping past the question of whether the standards were worth their costs, other important questions were also ignored: notably whether, given the marginal effects of changes in deadlines, it made sense to impose deadlines so tight that the industry would have good reason to feel forced to do just what it openly preferred to do anyway, namely concentrate on fixing up the conventional engine rather than aggressively explore major changes in engine technology.

But if a program provides invisible benefits, it is likely to go unchallenged only so long as it is paid for by invisible costs.

In 1970, benefits were highly visible--in the sense that the public clearly highly valued "doing something" about the environment. But the costs were invisible; not only did they lay in the future, but since they would be paid for through unspecified increases in the cost of owning a car, not through public spending, there was no need to state what they might be. Quite predictably, over time this balance shifted: public fervor could not be maintained at the high pitch of 1970; costs became more visible, while the benefits could be no more definitely established. (At one point, pushed by an environmentalist law suit, the EPA felt it had no choice but to issue regulations that would have effectively shut down the city of Los Angeles.¹⁷)

The prospective battle over strict enforcement turned most sharply against the environmentalists when the oil embargo and subsequent economic difficulties provided reasons, or at least excuses, for delays without directly challenging the wisdom of the basic legislation. In 1975, the UAW could--without noticable embarrassment--join the auto companies in recommending a prolonged postponement of the standards which the same UAW had very actively joined the environmentalists in demanding in 1970.¹⁸

To sum up, then, at the time of its passage the program was never analyzed, much less defended, in terms of the kind of rational balancing of costs and benefits which some would like to see govern public policy choices. Hence, it is not entirely surprising that over time, the costs became increasingly visible to the voting public, but not so the benefits. Further, the language of the Act, in particular through the stress on "good faith" of the industry, guaranteed a more than usual degree of acrimony in the subsequent arguments over relaxing the standards.

On top of all this, unforeseeable events--notably the oil embargo--added a heavy burden to the problems of the enforcers. This part of the complex history is a considerable inconvenience for the purposes of this chapter, for there is no way to prove that the program would have run into difficulty absent the embargo. Nonetheless, However, a listing of some of the peculiarities of the legislation, and comparison with the procedures in other countries, suggests that a prudent observer might have suspected that the program would turn out to be hard to implement:

- Save only Japan (which has particularly acute pollution problems due to the density of its population combined with its high rate of economic activity; and which sells roughly half of its total automobile production to the U.S.), no country in the world has found it prudent to limit maximum emissions within a factor of five of those in the U.S.¹⁹
- No other country has taken the view that cost and technical feasibility are not to be considered in setting standards.
- No other country has taken standards beyond demonstrated technical feasibility and known cost and frozen them into law (as opposed to regulation subject to administration revision).
- No other country has based its policy on the scientific fiction that a zero effects level of auto pollutants can be identified,²⁰ or on the political fiction that having done so, the public ought to be willing and will be willing to pay any cost to meet those standards.
- Finally, in the U.S. all this was done without hearings at which affected parties and expert witnesses could comment on the proposal before it was fixed in law. (For the central features of the bill--for automobiles, the 90% reduction by 1975--emerged from the committee in August, months after hearings had been held on a far milder bill, under which auto emissions standards would continue to be set administratively with due regard for costs and technical feasibility.)

Yet the Act swept through Congress virtually without debate, to be warmly embraced by the Administration, which claimed it as the response of Congress to the President's leadership on environmental matters.²¹

The easy answer to this puzzle is that, in the crusading mood of 1970, anything carrying the environmentalist banner was politically sacred. But that answer is not really enough. For the fact is that the Act is full of compromises. For examples: industry was given an extra year (to MY '76) to meet the NOx standard; the warrantee requirement was cut to 50,000 miles only, though the average life of a car is over 100,000 miles; a provision permitting EPA to consider a one year delay was added to the bill; the EPA was told to consider costs in setting an interim standard in the event a delay was granted; a National Academy of Sciences study of the technical basis of the standards was ordered; and so on.²²

There is a certain amount of political naivety in supposing that any bill would have swept through Congress without even a serious attempt to fight it. The auto industry was not politically powerless even in the magic year of 1970. And on this issue, where the interests of auto dealers--who are politically significant in virtually every Congressional district--were much the same as those of the automakers, the industry was not totally lacking the resources to mount a serious fight. Yet the public support the industry was able (or chose) to muster in Congress was limited essentially to a single speech by Sen. Griffin of Michigan. For three months after the initial draft was circulated in mid-August, the industry was sufficiently muted in its criticism of the bill for three months after the initial draft was circulated in mid-August that in mid-November press stories appeared about the industry's change of strategy when its spokesman began making some strong statements against the bill.²³

So there is a puzzle about why the legislation took the form it did which cannot be resolved simply by noting that the public mood guaranteed (as indeed it did) some sort of strong new environmental legislation in 1970. This explains why there was a bill, but not why the particular bill, rather than some alternative, swept through Congress. Nor does it explain why the industry was so quiet in the face of legislation which appeared to threaten enormous penalties should the industry fail to do what it asserted it did not know how to do.

3. Who Got What

We will try to show that the bill which passed did so with little controversy because it provided each of the groups most centrally involved--politicians, environmental activists, and the auto industry--what it most valued within the range of outcomes plausibly consistent with the political realities of 1970. I will argue that none of the difficulties described in the previous section received serious consideration because none of the key actors found it in his interest to press the issue; that, despite all the talk about the forcefulness and clarity of the bill, it was in fact sufficiently ambiguous in its likely effects that each could hope for an outcome that he could live comfortably with; and, further, that the bill had the appealing character that if the wider public (which would pay the bulk of the costs) grew disillusioned, everyone would have someone else to blame.

In this context, the incentives faced by two of the principals--the politicians and the environmentalists--are easily analyzed. We will consequently focus most of our attention on the auto industry.

Among the politicians, the two principals were President Nixon and Senator Muskie, the man expected (particularly after the mid-term election in November, and the final bill did not emerge from conference until December) to be the Democratic nominee in 1972. Senator Muskie was the Congress' Mr. Environment. Nixon had every incentive to try to seize this issue from him. Muskie then had every incentive to try to outdo whatever Nixon proposed. There is no need, nor any reason, to deny either man a genuine interest in cleaning up the environment. We need only assume that, in the course of doing what was right by the environment, everyone also paid some attention to his political interests. So we had a strong and, in hindsight, very sensible set of legislative proposals from the President in February of 1970; a still stronger bill came out of the Muskie committee in August. (In the interval, there were several important events: in April, a massive nationwide demonstration called Earth Day; in May, a stinging attack on Muskie from the Nader study group; in July, a highly publicized air pollution alert in Washington and other East Coast cities; and in August, GM's payment of \$425,000 to Nader to settle his invasion-of-privacy suit.²⁴) The President, and Republicans generally, then declined to be the people who were against cleaning up the environment. No Republican voted against the bill. (On a technicality even Sen. Griffin withdrew his negative vote.) Mr. Nixon then presided over an elaborate bill signing ceremony, celebrating the response to his environmental initiative.

Various detailed accounts of the history differ in emphasis and points of interpretation. But the essentials are always very much as just outlined.²⁵

For environmental activists, the basic politics of the situation were even simpler. Clearly, 1970 was a year of opportunity for anyone who wanted to see a major national commitment to cleaning up the environment. The stronger the commitment, the more clear-cut and ambitious the goals, the more specific the deadlines, the better the bill would be. It was perhaps possible to go too far. But if compromises had to be made, it was tempting to believe that these could be best made in future years: tempting to believe that any eventual compromises that had to be made would be more attractive from the environmentalist point of view, the stronger the commitments that could be gotten across while popular enthusiasm was at its height. In any event, environmental activists found it congenial in terms of maintaining their own constituencies, and entirely reasonable, to suppose that if anything in the bill was demonstrably unsound, then industry could be counted on to make that case and lobby for it. Certainly to the environmental activists, industry neither needed nor deserved help from environmentalists on that score.

But what were industry's incentives? In particular, what were the auto industry's incentives in seeking the outcome that was least disruptive, least likely to involve costs that could not be passed on to auto buyers, and least likely to add to the industry's public relations problems?

The outcome which posed the most severe threat on the first two of these desiderata would be a forced abandonment of the internal combustion engine around which the entire industry was built. How difficult the changeover would be would depend very much on how rapidly the change was pushed. A crash effort would be difficult and expensive

indeed. A gradual switch over a period on the order of 15 years would involve much less in the way of painful readjustment or extraordinary expense. (That the industry itself undertook exploratory work--even before strong political pressure arose to do so--on such radical alternatives as the gas turbine seems to show that industry management did not perceive a gradual and unforced changeover as forbiddingly difficult.)

In this context, one could make some pretty confident inferences about the industry's policy preferences, assuming no more than that the auto companies exhibit the usual sort of policy preferences that we see in large, long-established organizations in every sphere, from churches to armies. For example, we would expect the industry to be reluctant to be convinced that auto emissions were harmful; if emissions were judged harmful, industry would hope that a relatively cheap cleanup of the conventional engine would be an adequate response; if a drastic cleanup were unavoidable, we would expect industry to look carefully for some technical fix that would permit this change without requiring a basic technology change; and, if a new technology had to be brought in, then we would expect the industry to be anxious to avoid an obligation to make the changeover under severe time pressure.

The importance to the industry of the various stages of this set of choices escalated with the severity of the choice. A minor cleanup would be a minor matter. Emissions could be reduced by an impressive percent with almost no effect on car prices. A major cleanup, but short of drastic technology change, might be expensive enough to have a nontrivial effect on the demand for cars. It presumably

would have real effects and industry-borne costs (perhaps quite unequally distributed among the various automakers). But the effects could be expected to be too subtle to clearly identify and measure, and comparable in scale to the kinds of uncertainties and risks managers ordinarily live with. A gradual change in technology, without severe time pressure, might be a comparable problem.

But a forced change of technology, under time pressure, presented risks on an entirely different scale. It presented, as the earlier alternatives did not, real nightmares for the industry. It meant, almost inevitably, massive disruptions for the auto-makers, and opened up some very unpleasant questions. If the government could force the abandonment of the conventional engine, might it not also insist on choosing the alternative technology to be used? And would a political choice be biased towards "space age" technology without a real understanding of the practical difficulties of adapting very sophisticated technology to a form suitable for the skills of the corner garage mechanic?

Might political pressures and price elasticity combine to prevent industry from fully passing on the costs of forced technology change? What would happen if the government changed its mind about what should be done part-way through such a process? What would happen to the independence of the industry if the weaker companies were unable to manage the changeover without public subsidies?

Was it reasonable for the industry to actively worry about the possibility of a forced change in technology under severe time pressure? Perhaps it was. Consider the events during the first six months of

1970: the California state senate voted to ban the internal combustion engine by 1975; legislation was proposed in both House and Senate to do the same nationally; the President proposed a research effort to develop a non-polluting engine by 1975; the Senate passed, with Administration support, a bill requiring the federal government to pay premiums of up to 100% to purchase low-pollution cars (the bill was based explicitly on the premise that cleaning up the conventional engine was an uneconomical way to deal with auto pollution).²⁶

In general, a perusal of the New York Times Index for 1970 shows that not only was the automobile the special focus of escalating concern about air pollution, but the industry's commitment to the internal combustion engine was often perceived as the heart of the problem.²⁷ As early as March, the Wall Street Journal was reporting that although the industry was "bowing to the inevitable" and planned to meet the Nixon '75 and '80 standards "regardless of cost or effect on performance," nevertheless "Detroit wouldn't give up the internal combustion engine without an enormous political and public relations fight."²⁸

In July, the Nixon Administration sponsored a conference on alternatives to the internal combustion engine to kick off its program to develop a non-polluting engine by 1975. It received extensive press coverage, and was followed almost immediately by a joint statement by the United Auto Workers and the major environmental organizations calling on the Muskie committee to adopt "standards so strict as to ban the internal combustion engine within 5 years."²⁹

Meanwhile, in the background, the spectacular landings on the moon which were then climaxing the Apollo space program not only helped

stimulate popular concern about the environment of "spaceship Earth," but gave enormous political appeal to the notion of setting ambitious technological goals, to be met within the decade. The Times was very much in the spirit of the moment when it editorialized, after the Muskie committee reported its bill, that "a nation that can put a man on the moon in less than 10 years can clean up its engines in half that time."³⁰

Viewed against this record, the bill that emerged from the Muskie subcommittee in August was by no means the worst outcome the industry had reason to fear. That bill merely demanded that industry do what it felt to be impossible. There is a certain comfort in being asked for the impossible. You know you will not actually have to do it. And to the extent that the Muskie standards could in fact be approached, automobile owners not the automobile industry would have to pay the price. Far worse, from industry's interests, than taking Nixon's 1980 goals and writing them into law for 1975--a timescale that made it clearly impossible to push any alternative to maximum clean-up of the internal combustion engine--would be to take Nixon's 1975 goal of developing a non-polluting engine, and write into law a national commitment to replace the internal combustion engine "within this decade."³¹

So from industry's point of view, the August Muskie draft was a kind of Pandora's box. Certainly in an ordinary year, it would have been an exceedingly vulnerable bill; and even in 1970, it was far from an obviously invulnerable one. The bill required not only industry, but localities, to take drastic actions at unknown costs to meet strict health standards whose basis was exceedingly fuzzy.

All this was to be done against very tight deadlines. (The deadlines for implementing local transportation plans, which would have to be severe in many cities even if the auto companies were able to meet the emissions standards, were approximately the same as those imposed on the auto companies.) All of this was being done precipitously, without so much as a single hour of hearings on the bill. Few people seem to have had any clear idea of what was in the bill, beyond a general impression that the bill was tough and would make the auto industry clean up by 1975.

Simply on the procedural ground that hearings should be held before passing such far-reaching legislation, one might suppose that the industry could have won some allies even among those otherwise wholly unsympathetic. Indeed, in the mood of the moment, the procedural issue was the only issue on which the auto industry might reasonably have hoped for a sympathetic hearing. But pushing that issue might have bought time to make a concerted attack on the rigid standards and tight deadlines of the bill. So if the industry wanted to openly fight the bill, the place to start was on the procedural issue. The industry chose not to press the point. The only reference to the lack of hearings in the industry comments on the bill was an unprotesting mention of "time schedules which did not permit open hearings on new and important provisions in Committee Print No. 1" [the draft bill].³²

It becomes easier to understand this shyness on the part of the country's largest industry if we bear in mind that the bill could have been much worse from the industry's viewpoint--and conceivably might have become so had the issue been thrown open by a major industry

challenge. One need not argue that this was a likely outcome; one need only point out that only a few months earlier, the actual Muskie bill would have been regarded as wildly unlikely--even by Muskie himself, in view of the far milder legislation he was sponsoring earlier in the session.³³ Even a mildly risk-averse industry manager would have some doubts about the "unthinkability" of more explicitly anti-ICE legislation. (What actually happened was that, having passed a bill which "solved" the problem by mandating a clean engine in five years, the clean engine program was never funded even up to the inadequate levels proposed by the Nixon Administration.)

But the major influence we might reasonably impute to the industry's concern about attacks on the internal combustion engine comes not in interpreting its distinctly passive response to the draft legislation, but in seeing how far the industry had already gone, months before, towards accepting as reasonable the kinds of standards Congress now proposed to write into law. The Wall Street Journal article cited earlier reported that both GM and Ford already had running on their test tracks cars capable of meeting the Nixon '75 standards. GM's president, in the January speech cited earlier, had announced that GM already had demonstrated in its laboratories that the Nixon 1980 standards could be met with the internal combustion engine.³⁴ In general, industry sources had taken a quite aggressively optimistic view of what could be done by way of cleaning up the conventional engine. It would be far harder to understand this eagerness to please if not for the growing pressure to phase out the conventional engine, which gave the industry some incentive to claim that it had the technology close to hand to clean up without changing its basic technology. These

claims were made early in 1970, when there was no one in Washington proposing to write the Nixon standards rigidly into law, much less under a sharply speeded-up time-table. These statements then made it easy for Congressmen to believe that, with pressure, the industry might be able to deliver in five years what it had already promised to deliver in ten.

Seen in this context, the Muskie bill was a good deal less arbitrary than it has often been portrayed. For all its strong language, the bill was open to the interpretation that all that was intended was to push the industry to move as urgently as possible to meet standards that the industry itself had claimed that it could meet, and planned to meet. Indeed, Sen. Muskie repeatedly said as much during the debate:

I think the industry has an obligation to try to meet [the standards]...

... as one president of one auto company said, 'You can't put this in the record, but we are that close.' If we are 'that close' it seems to me we have to set the timetable and challenge them to meet it. They can always come back to Congress....

It is not necessary to say that any company is going to be closed on January 1, 1975.... Five years is a long time for the companies to make their effort, then to make their case, and then for Congress to consider a change of policy.³⁵

Five years later, Sen. Muskie was saying to the auto company presidents, who had come to ask for a further five-year extension of the standards:

The fact is that insofar as the automobile is concerned, you are in no different position as the result of the passage of the Clean Air Act than you would have been if the Clean Air Act hadn't been passed. The standards you are now meeting, you agreed to meet in 1969...So here we have gone through all of this, struggling to get the law on the books, struggling to get it implemented, yielding to your appeals that we have asked too much. We have come back to the point where we started. I am frustrated, frankly. I ask that not in the sense of putting you on the spot. But how do we get this thing moving? How do we get it off the ground?"³⁶

Now the transition from 1970 to 1975 was not so calm as the just cited set of quotations might (by themselves) suggest. Nor has there been calm since. But it is fundamental to an assessment of the interests of the auto companies in late 1970 to bear in mind that the most the auto companies could gain, even if they could succeed in killing the Muskie bill, would be a shift from a Congressional mandate to meet very tight standards (accompanied by repeated admissions that this might not be technically feasible and therefore might have to be relaxed) to a commitment to carry out their own claims that they could meet those standards on a more relaxed time schedule.

Yet the feasibility of meeting given numerical standards would depend a good deal on how those standards were operationally defined. The details of the regulations, and even the details of how regulations were going to be administratively interpreted, could have large effects on the feasibility of meeting given standards. For example, to what extent would field observations on real cars would be fed back into the certification process? Or was certification as by and large proved to be the case--to be based only tained by the companies' own expert staffs, not by ordinary drivers and mechanics? In this and a dozen other ways, the details of the legislation, and how it was likely to be administered, were of great importance to the companies. Moreover, these details were much easier to influence politically than a change in the basic standards and deadlines which had dominated public interest in the bill. Meanwhile, several of the key arguments against strict deadlines and rigid standards--particularly their possible costs, and the difficulty of assuring that the controls would retain their effectiveness during the life of the car--were very much

the kinds of arguments that could be turned against the industry by critics of the ICE, or which otherwise could turn out to aggravate rather than help the industry case.³⁷

So there was a real trade off between (1) a challenge to the basic provisions of the bill (the standards and deadlines), which even if successful would leave the Congress and the Administration under great pressure to be tough on the details, and (2) not fighting the basic standards, but quietly asking for help on the details of implementation, in view of how tough the basic provisions were on the industry.

All things considered, it does not seem surprising that the industry opted for the latter course. And while it did not get all it wanted, it got a great deal of what it wanted. Perhaps the most important of these changes were those which cut the warranty requirement (hence also the endurance test for pre-production certification) from 100,000 to 50,000 miles; and the provision giving EPA authority to grant a one-year extension. The most important failure for the industry was that, despite support from the Administration (withheld until after the election), EPA was not given authority to grant a series of one year delays.³⁸ After the first, the industry would have to come back to Congress. Nevertheless, the fact that the initial appeals for delays would come in the technocratic forum of EPA hearings, rather than the political forum of Congressional hearings, was a relief for industry--and no doubt also for the Congress, which thereby passed responsibility for the matter to EPA and the courts.

It is instructive to see how readily one finds illustrations of the propensity of the bill's proponents to value rhetoric, so that

the effect of changes was often to soften the effect of the bill without softening the declamatory ring of the language. One finds a number of oddities which are most easily understood on the hypothesis (suggested in Chapter 4 above) that the environmentalists had a particular attachment to language symbolic of commitments to environmental purity, while the industry was concerned with what the law would make them do. Thus the law does not say that cars must comply with the standards only for the first 50,000 miles (or 5 years) of use. Rather the certification and warranty clauses specify that cars must conform "for their useful life." However, elsewhere in the bill, "useful life" is defined as what, in practice, is half the useful life of the car.

The Senate version of the bill required American cars to meet the American standards, even if they were not manufactured for use in this country. This provision was killed in conference, not by the simple device of deleting it, but in a manner which might earn a "Catch 22" prize. The final bill retains the Senate language requiring that cars for export meet the standards. But it adds an "except" clause in the event that the country of export has standards different from the U.S. The report then completes the matter, by interpreting the different standards phrase to include the case where the export country has no standards. Thus the final bill included a provision requiring that cars for export meet the standards, but only in the event that they would be required to meet the standards in the absence of this provision.

A more important illustration of the propensities of the drafters is in a provision increasing tenfold (from \$1000 to \$10,000) the penalty for producing a car which violates the standards. At \$1000 (effectively

closer to \$2000 on the price of the car, since the penalty would not be deductible from corporate income tax), the fine was high enough for every practical purpose. But here was a provision which offered something for both environmentalists interested in symbolic matters and automakers interested in practical matters. At the \$1000 level, perhaps some purist would worry that plutocrats buying very expensive cars might indulge themselves in a "license to pollute." But it was also possible, with a fine of "up to" \$1,000 that fines would be levied proportionate to the extent to which cars failed to meet the standards, putting the penalties within the range of the excise tax the industry had been able to live with for many years. (The excise tax was rescinded in 1971, less than a year after the tougher emissions standards were enacted.) The government could then take a really tough stance on meeting standards, realistically demanding, comply or pay. But at \$10,000, fines pro-rated to the degree of compliance were certain to be prohibitive. The alternatives to compliance would be either to shut down the country's largest industry, or relax the standards.³⁹

Now in the light of this excursion into the background and history of the bill, we can see reasonably well how it came about that the most serious questions to be asked about the soundness of the 1970 bill were never pressed by anyone. The politicians had no desire to say a word that might be interpreted as lacking in environmentalist fervor in a year when environmentalist fervor was sweeping the land; the environmentalists themselves wanted the strongest possible commitments (with, dare one say, a certain amount of technical or political naivety showing up in a tendency to value rhetoric over substance); the auto companies, faced with the problem of having to get along

in the mood of the time, had less interest in questioning the intellectual foundations of the policy than in making it easier to live with. In the political mood of the moment, working quietly to soften details of the bill and ameliorate its administration could easily have looked more promising--particularly, given the public commitments they had already made earlier in the year--than fighting the basic premises of the bill.

4. What Went Wrong

Let us summarize the argument thus far: The 1970 legislation was based on a number of premises which were either dubious on their face (e.g., that it was prudent to pass legislation dealing with a complex problem, and certain to cost billions of dollars a year without bothering to hold hearings⁴⁰) or which even a superficial examination of available information would have shown to be certifiably false (e.g., that the five year speed-up in the application of the Nixon 1980 standards, which was the heart of the bill, would actually have some important effect on air quality). Under these circumstances, it is hardly surprising that the subsequent history has been marked by frustration and bad feelings among all concerned. The details of these difficulties were not predictable, since they depended on future events--such as the oil embargo. But that an expensive program based on false premises would somehow or other run into difficulty is hardly surprising. Nevertheless it is understandable that the bill passed in the form it did, since each of the major actors did what was natural, and even rational, from his point of view. That is, each (including the auto industry) acted to improve his own out-

come in a way that could seem (to that actor) entirely reasonable--even shrewdly calculated--given the overall political context.

This last point gives at least a first-order explanation not only of the original legislation, but of the subsequent controversy over implementing (or delaying implementation) of the standards. We have a pluralistic political system, in which no organized entity has full authority over policy, and under which, therefore, no organized entity feels fully responsible for the soundness of public policy choices. The various active forces are constrained by political power of competitors, and all are constrained by a hard-to-define but nevertheless very real force we call "public opinion"--which all the active forces seek to influence but none can predictably control. Under these circumstances, each actor perceives "the best as the enemy of the good," and works to get a better outcome, as perceived from his point of view. Indeed, a stronger statement is reasonable: each actor tends to resolve uncertainties in a way that suits his predilections, so that the perceived conflict between what he is doing and what, in an apolitical world, he ought to be doing, is minimized. This process of agreeable simplification is enhanced by the effortless device of not doing the work (and it would require work) to assimilate facts and arguments that might prove incompatible with one's predilections.

The output of this process may be far from socially optimal. But, once enacted, the program is difficult to change. Social and organizational inertia then favor continuation of the program as it exists. Highly visible aspects of the program may acquire symbolic or emotional significance which of themselves raise the social cost of a retreat

from those objectives. New vested interests are created which will resist change--notably a bureaucracy dedicated to administering the law as it exists. Those who supported the legislation (in this case, including virtually every member of Congress) quite understandably are reluctant to believe that the program may have been seriously misconceived.

If this were the whole story, one might be led to pessimism approaching despair over the prospects for dealing effectively with complex issues involving difficult tradeoffs affecting the lives of private individuals. Yet that kind of issue--the most salient currently involve energy--are certain to be extremely important in the years ahead. On this view, the most troublesome aspect of the auto legislation is not the difficulties with the air quality control program, but that the explanation of how the difficulties arose suggests that there may be a systematic propensity of American government, with its special emphasis on division of powers and growing emphasis on popular participation, to perform badly in dealing with precisely the sort of issues which are likely to dominate the public agenda in the years ahead.

But a closer look at the "what went wrong" question allows a somewhat more optimistic reading. Most important, the system has some capacity to learn--indeed more so than may be immediately apparent, given that political actors find it much easier to change the substance of policy than to change the rhetoric describing policy. The environmental issues being dealt with in the early 1970s represented a new class of problems for federal regulation of businesses, with important

differences in underlying character from those dealt with in the past. The very novelty of the issues made it easy, even for people who felt they had been spending a good deal of their time working on the problem, to err in their assessments of the substantive issues and to misjudge the appropriateness and likely effects of various proposed remedies. The constitutional division of power within the American government, which naturally creates some difficulties even as it guards against others, was aggravated not merely by the political difference between the Republican executive and the Democratic majorities in Congress, but also by the Vietnam-engendered weakening of confidence in the executive, and by Richard Nixon's peculiar capacity to incite distrust and bad feeling. With the passage of time, some of these difficulties have been cured, and others at least ameliorated. What has not been cured, or even adequately recognized as a defect in the policy process, is the structural flaw which, we will argue, lay at the root of the difficulties.

The general problem is that described in Section 1 of this chapter. Situations arise in which the active participants in a decision are erroneously taken to represent the full range of interests that have an important stake in the decision. But they do not do so, with the consequence that important considerations are easily left out of the policy process because none of the active participants are adequately motivated to raise them.

In the case at hand, the context was almost entirely perceived as the auto industry vs. the public interest in a cleaner environment. Thus when the Muskie committee completed its draft of the bill in

mid-August, it was not publicly released, but circulated privately to the auto companies (and to those industries affected by other portions of the bill). The committee knew that the environmentalists were happy with the bill; it therefore wanted to be sure that "the other side" had its say. But since "the other side" was not going to pay for the costs of meeting the standards--that would fall to the public--the automakers made only perfunctory objections to the features of the bill which would involve the greatest costs, and concentrated their efforts on improving the bill from the point of view of the industry. In other words, the industry invested its political capital in minimizing the likely costs to industry not in minimizing the total cost, most of which could be passed on to the public. But of course no one involved was motivated to think of the policy process in these terms, and certainly not to so describe the process in their public rhetoric.

It would be a mistake to suppose that this situation was unique, or even terribly unusual. The same sort of systematic error, possibly with far greater cost to the nation, could easily be made in the area of energy policy, to the extent that the energy problem is perceived as the big oil companies vs. the public.

There are several ways of looking at this dilemma. Here we will develop a view of the underlying difficulty as the political analogue of the economic problem of externalities--a connection which may be worth stressing since problems with political externalities are peculiarly likely to arise in contexts marked by important economic externalities. The argument goes as follows:

The reason that government must act to control auto emissions (and environmental pollutants generally) starts with the fact that the effects due to any individual car (or individual user of any product whose production or use creates pollution) is extremely inconsequential. Significant auto pollution is due to the combined effects of very large numbers of automobiles. Indeed, the individual auto user has no way of detecting the amount of controlled pollutants (HC, CO, and NOx) his car emits: only carefully instrumented tests can measure the emissions, which are invisible to the individual driver. So the individual driver has little motivation to consider emissions controls as something worth paying for. He cannot see that it makes any difference in his own car's exhaust; nor could the most careful instrumentation detect any difference in the total level of pollution in his community should he buy a clean-running car. So the pollution from his car becomes an "external" cost--meaning, as economists use the term, a real cost which the decision-maker is not adequately motivated to take into account. It is effectively external to his own judgments about how to spend his resources.

Since the consumer is but weakly motivated to take the costs of pollution into account, the firm making what the consumer buys is also weakly motivated. For the firm makes its profit by providing the consumer what he wants to buy (or at least, what he wants to buy influenced to the extent feasible by what the firm is able to persuade him he wants to buy). The result of all this is a social dilemma in which individual consumers, and individual firms, are not adequately motivated to take pollution costs into account, although all may have

a joint interest in having everyone do so. The solution to this dilemma is to have an agreement that everyone will do his share, on the condition that everyone else also does so. The mechanism for arranging this social agreement is to pass a law.

The reason a political analogue of this economic dilemma arises especially easily when economic externalities are at issue is that it will almost always be very inefficient--in fact, generally totally impractical--to have the law aimed at individual buyers. The number of firms which make a product will be very small compared to the number of buyers of the product, so it is easier to regulate (or provide incentives, as by some sort of tax scheme) on the producers rather than on the consumers. It is also generally, though not always, politically easier to regulate the producer. In this way the direct interests in the legislation (or the regulations coming out of an administrative proceeding, or the parties in law suits which shape policy) become the environmental interests vs. the producer interests. The consumer interest is not directly involved, and being diffuse and typically unorganized, it is hard to activate unless the legislative proposal is quite bluntly aimed at individuals. (Such would be the case, for example, if we sought to conserve petroleum by levying gasoline taxes comparable to those which prevail overseas. The situation naturally changes drastically if the law is aimed at individuals, though only subtly or indirectly so, but the costs of compliance eventually become clear at the level of individual impacts: as illustrated by the fate of the transportation controls in the Clean Air Act.)⁴¹

The net effect of this is that an important externality may arise which is political, rather than technological, in origin.

In the case at hand, the environmentalists perceived their role as maximizing benefits to the environment; if industry perceived its interest as that of minimizing total costs, then one could expect that out of this clash of competing interests some reasonable social balance would arise. But industry's interest is not to minimize total social costs, but to minimize those costs which must be borne by the industry--many of which are not even economic in nature, such as sheer distaste for having to deal with government bureaucrats, or having to do things which may not be terribly important in terms of industry-borne costs but which seem (to the industry, at least) foolish and offensive to the businessman's ethic of efficiency. (For both environmentalists and businessmen, one should not underestimate the power of ideological commitments--to cleanliness for its own sake for the environmentalists, regardless of whether one can make a good case for damaging consequences from a little more pollution; for efficiency and freedom from outside interference for businessmen, regardless of whether there is any substantial impact on profits.)

So we have a structural flaw, in which rational behavior by individual actors leads to a socially perverse outcome. It is not an inevitable problem: on many issues, no doubt, the clash of the directly involved interests produces a reasonable social result. Most of us would like to believe that that is generally the way the American political system works. But more and more, evidence accumulates (the most conspicuous cases are those involving regulatory agencies) that what I have been calling "political externalities" are a consequential problem. And it is not an easy problem to deal with.

It is not clear how institutional arrangements could be altered to create a new actor in the policy process whose own vested interest is to be the advocate for socially important considerations which are being ignored by others. It is not clear how one could operationally define something that approximates this goal; and, even if defined, it is not clear how it would attain enough power to become an important factor in the process, since others will naturally resist anything which dilutes their own power. (The "science court" proposal is apparently intended to deal with at least part of this problem. But the difficulty of defining what a "science court" would be, in a way that makes sense to technically and politically sophisticated skeptics, illustrates the difficulty of prescribing a remedy.) It is likely that new institutional arrangements will sooner or later evolve; but not likely that a workable all-at-once reform can be designed. The starting point is to recognize the problem.

FOOTNOTES

1. For more sympathetic accounts of the legislation than that given here, see Charles Jones, Clean Air: The Politics and Policies of Pollution Control, Chapter 7 (University of Pittsburgh Press, 1975), and Helen Ingram, "The Political Rationality of the Clean Air Act," paper prepared for the MIT Symposium on the Clean Air Act, December 1976 (forthcoming, MIT Press).
2. J.Q. Wilson, "The Politics of Regulation" in J.W. McKie (ed.), Social Responsibility and the Business Predicament (Brookings, 1974).
3. The first version of the bill circulated by the Muskie subcommittee specified that the costs of emissions control devices be listed on the new car price sticker. But this was deleted at the request of the industry. See Legislative History of the Clean Air Act Amendments of 1970, prepared by the Environmental Policy Division, Congressional Research Service, Library of Congress, for the Committee on Public Works, U.S. Senate, Serial No. 93-18, January 1974 (2 Vols.). Cited henceforth as LH. The "sticker price" provision appears at the end of Section 207c of the Committee draft (LH, p. 685); the industry's argument for its deletion appears at LH, p. 722.

I have not seen any estimate, by the industry or others, of the costs to the industry of compliance--that is, some estimate of the costs of compliance net of the costs passed on to auto buyers. Other things equal, an increase in the cost of cars can be assumed to reduce sales, but this does not necessarily imply a net loss to automakers, since the loss of sales is at least partly, and perhaps fully, offset by the increase in price per unit. The committee did assert (but did not write into the law) that the makers should not add a profit for emissions control devices. And no doubt one will not find in the any company's books a line labeled "profit from emission controls."

The industry has made estimates of the loss of sales (as noted above, not to be confused with loss of profits or lowered return on investment) due to government mandated emissions controls and safety standards. In terms of first cost to users, the latter are substantially more important. See, for example, the industry submissions at pp. 1216-1264, 1975 Hearings on Implementation of the Clean Air Act (henceforth cited as 1975 Hearings.) On projected sales losses, see, for example, Ford estimates at p. 339, 1975 Hearings.

4. The most recent effort to quantify total costs (including fuel penalties, etc.) is the Dewees paper for the MIT Clean Air Symposium.
5. See Figure 5, LH, p. 1243 (Model Year 1971 standards, which were promulgated prior to 1970).
6. See, Presidential Message (Nixon), LH, p. 1502.
7. New York Times, January 15, 1970, p. 1.
8. LH, pp. 1082-1087. The charts in the text were drawn by overlaying the separate charts for various standards submitted to the Committee.
9. There are actually differences among (a) the Nixon 1980 proposal; the Barth paper published some months later (June, 1970) intended to justify the Nixon 1980 goals (see Acting HEW Secretary Veneman's comments at LH, p. 738); and the 90% reduction from 1970 levels prescribed by the law. However, throughout the legislative history, the three sets of numbers are almost always treated as equivalent. See, for example, Sen. Muskie's comment, LH, p. 228.
10. See, for example, the discussion in F. Grad, et al. Note that the graphs refer to projected emissions, not to actually achieved air quality. The connection between auto emissions and air quality remains poorly understood, and since the relation between air quality and actual health effects is also poorly understood, any estimate of the health effects of marginal tightening of the standards becomes a matter of almost pure guesswork. Further, actual emissions measured on cars in use typically exceed the standards, but not by nearly enough to account for the relatively small improvements that have so far been noted in air quality. However, a glance at the data plots used in the Barth analysis (LH, pp. 768-9) shows the inherent difficulty of establishing clear relationships, which increases as tighter controls make auto exhaust emissions a smaller fraction of total pollution. The Barth paper, incidentally, based its calculations on, among other things, an assumption that cars in use would exceed the new car standards. (LH, p. 764). For more details on this issue, see Chapter 2 of this report.
11. LH, p. 238. See Chapter 4 of this report for more discussion of the role of health issues in the overall history of environmental regulation.
12. LH, p. 226.
13. Certainly the total elimination of auto pollution could be expected to have some effects, such as a major reduction of oxidant episodes in Los Angeles. But if we move beyond such effects, which few claim to themselves, represent a significant health problem, then

persuasive data seem hard to come by. See, for example, the discussion of the issue in Jacoby & Steinbrunner, Clearing the Air (Ballinger, 1974), Chapter 7. The strongest claims to measurement of aggregate public health effects are those of economists using regression analysis. These studies must be viewed with caution, for there are severe problems with multicollinearity and uncontrolled variables, notably on prevalence of smoking.

14. See, for example, 1975 Hearings, p. 25.
15. OECD, "Report of Committee of Experts of Automobile Pollution," 1972.
16. For detailed discussions of this issue, see the Mills/White paper for the MIT Clean Air Conference and Jacoby & Steinbrunner (op cit).
17. Jones (op cit), p. 270.
18. 1975 Hearings, pp.271 ff.
19. On the Japanese standards: although set at numerical levels like those of the U.S., the test procedure differs in a way that seems to make it harder to meet the HC and CO standards, but much easier to meet the NOx standard. See 1975 Hearings, pp 186-7 (noting that the Japanese standards are given in grams/kilometer not gms/mi). Not only is the Japanese test cycle very different from the U.S., and the endurance required only 18,000 miles (vs. 50,000 for the U.S.), but many other details of how the Japanese certify cars are something of a mystery to European and American car makers.

Aside from the U.S. and (as qualified above) Japan, the toughest standards are those of Canada, Sweden, and Australia, which are essentially the U.S. 1973 standards, and are about a factor of 5 below those set by the Clean Air Act. See Table 1 of the text.

20. The supposition that there is a scientific basis for identifying a zero effects level of pollution provides a nice example of the problems created by the fragmentation of expertise and responsibility. Required to produce such numbers, the technical people do so, fully aware that there is a large amount of arbitrariness in the procedure (starting with setting the air quality standards, then compounded in translating the quality standards into emissions standards). They protect their scientific conscience with numerous caveats in their reports. The numbers are then used by others as if they represented a scientific determination of levels below which public health is protected, and above which it is not. An EPA official who was closely involved in the establishment of these standards told me that the top level

response to the scientific misgivings about the whole business was typically that if the scientists couldn't figure out how to set the standards, then the lawyers would do it for them.

21. See Nixon's remarks at the bill-signing ceremony, LH, p. 105.
22. So far as I could find, there is no summary available of all the changes made in the bill between the original Muskie committee bill (LH, pp. 629ff) and the final bill. Of those noted here, the cut in the warrantee requirement and provision for a one year delay were added within the Senate committee; the others emerged from the House/Senate conference.
23. See for example, New York Times, November 18, p. 29.
24. New York Times, August 14, 1970.
25. See, for examples, Jones and Ingram (both cited, Note 1).
26. See comments on this point in the Senate Report, LH, p. 950.
27. New York Times Index (1970), pp. 34-39 and 159-162. Anyone interested in the legislation would do well to read through this readily available material, which has the great virtue that it simply reports what the Times was telling its readers in 1970, uncolored by the selection of material which narrative accounts cannot avoid.
28. Wall Street Journal, March 31, 1970.
29. New York Times, July 11, 1970 and July 12, 1970.
30. New York Times, August 23.
31. For the Nixon program, see his Message, LH, p. 1503, and comments by Administration officials reported in the Times account of July 11, 1970 (Note 29).
32. LH, p. 719.
33. Sen. Muskie's summary of the amendments to the Clean Air Act he introduced on March 4, 1970 (which along with the Nixon proposals, was the legislation on the table when hearings were held) appears at LH, pp. 1470-71. There is no mention of tighter standards for automobiles. Indeed, there is no mention of automobiles at all.
34. New York Times, January 15, 1970.
35. LH pp. 232, 234, 236.

36. 1975 Hearings, p. 398.
37. For a detailed discussion of the problems of maintaining the standards in use, see, for example, Jacoby & Steinbrunner, op. cit. See also, the material cited in Note 26 and the legislative hearings on that bill.
38. The Administration recommendations (Nov. 17, 1970) are at LH, pp. 211 ff.
39. On the 50,000 mile "useful life" see the Conference Report, LH, p.200, which also discusses the exemption for exported cars. The auto excise tax (then at 7%) was repealed 10 months after the passage of the Clean Air legislation (New York Times, December 11). The legislative history does not indicate a rationale for the increase in penalty from \$1000 to \$10,000.
40. When pressed on this point by Sen. Griffin, Sen. Muskie cited the various hearings his committee had held on related legislation (LH, p. 239). The point remains that no hearings were held on this legislation, and the hearings on which legislation was held contained no provisions comparable to those in this legislation. Why then could this procedural flaw be treated so casually by legislators, with their overwhelmingly legal backgrounds. I think that the answer is that as lawyers, they were accustomed to waiving all sorts of safeguards provided that the parties involved approved. If the issue was perceived, or misperceived, as the environmental interest vs. the auto company interest, then the lack of hearings was not a serious matter: neither of the parties objected to this short-cutting of normal procedure. (See the industry comment on this point, cited in Note 32.)
41. See, for example, New York Times, January 14, 1977, p. 1.