SANTA FE RESEARCH CORPORATION

AIR QUALITY BENEFIT ANALYSIS FOR LOS ANGELES AND SAN FRANCISCO BASED ON HOUSING VALUES AND VISIBILITY

FINAL REPORT

by

John Trijonis, Santa Fe Research Corporation
Mark Thayer, San Diego State University
James Murdoch, Northeast Louisiana University
Ronda Hageman, San Diego State University

January 1985

ARB Contract Number A2-088-32

Submitted to: Sylvia Champomier, Project Officer
Research Division
California Air Resources Board
P. O. Box 2815
Sacramento, California 95812

DISCLAIMER

The statements and conclusions in this report are those of the Contractor and not necessarily those of the State Air Resources Board. The mention of commercial products, their source or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products.

ABSTRACT

Benefits of improved air quality are estimated by examining the public's willingness to pay for visibility as revealed in home sale prices. The investigation is conducted for the 1978-79 period in two study areas, Los Angeles (four counties) and San Francisco (five counties). The analysis uses an extremely large, finely resolved data set for visibility levels, home sale prices, and house/community/location/socio-economic characteristics. The data are gathered from 39 airport weather stations (visibility), the Market Data Cooperative (home sale prices and house specific characteristics), as well as the Census and other data sources (other characteristics).

The economic analysis follows the hedonic price method. First, home sale price is related statistically to visibility and the house/community/location/socio-economic characteristics. The results (including the hedonic price for visibility) are statistically significant, have the expected sign, and are stable with respect to various sample sizes, visibility indices, model formulations, and functional forms. Next, the "hedonic price" for visibility is related to visibility levels and income levels in order to determine the economic (inverse) demand curve. This demand curve in turn yields estimates of benefits. It is found that, depending on assumed functional form, a 10% improvement in visibility would produce benefits of 250-617 million dollars per year in Los Angeles and 190-220 million dollars per year in San Francisco.

A demonstration is made of how to apply the results through an illustrative example -- benefits of controlling diesel particle emissions in Los Angeles. It is estimated that the elemental carbon component of diesel particle exhaust contributed 12% of Los Angeles' visibility reduction in 1980, and that diesel emissions will grow by a factor of 2.3 from 1980 to 1992 under a "no-control" scenario. It is found that 50% control of diesel elemental carbon emissions would yield benefits of one to five billion dollars over the 1980-1992 period. The benefits are substantially reduced by postponing and phasing-in controls.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1

1.0 INTRODUCTION 9

1.1 Economic Concepts and Approaches 10

1.2 Organization of the Report 11

1.3 References 12

2.0 DESCRIPTION OF THE VISIBILITY DATA 15

2.1 Weather Station Visibility Data 15

2.2 Data Compilation and Analysis 20

2.3 Visibility (Extinction) Indices 21

2.4 References 42

3.0 ESTIMATION OF BENEFITS FROM VISIBILITY IMPROVEMENT 43

3.1 Methodological Review 44

3.2 Data Specifics: Hedonic Housing Equations 54

3.3 Empirical Results: Hedonic Housing Equations 60

3.4 Empirical Results: Inverse Demand Equations 88

3.5 Concluding Remarks 97

3.6 References 97

4.0 ILLUSTRATIVE APPLICATION: DIESEL PARTICULATE CONTROL IN LOS ANGELES 99

4.1 Relationship Between Diesel Particulate Emissions and Visibility 99

4.2 Benefit Estimation 108

4.3 References 117

APPENDIX A: ECONOMIC BENEFIT ANALYSIS WITH THE 1973-74 DATA BASE . . . 119

EXECUTIVE SUMMARY

This study places an economic value on improved air quality by investigating the public's willingness to pay for visibility as revealed in housing market data. Specifically, we examine the relationship between housing values and visibility, discounting for the effects of other collinear variables (house/community/location/socio-economic characteristics). This relationship serves as the basis for determining an economic (inverse) demand curve for visibility which in turn allows estimation of the benefits associated with improved air quality.

In this project, emphasis is placed on what individuals perceive, as visibility is the object of the valuation exercise. The presumption is that all components of air quality valued by people are included in perceived visibility, and no attempt is made to separately value the independent components of air quality.

The analysis uses an extremely large, finely resolved data set for visibility levels, housing values, and house/community/location/socio-economic parameters. The study is conducted for the 1978-79 time period in two study areas, Los Angeles (four counties) and San Francisco (five counties). The report also demonstrates how the results can be applied through an illustrative example -- the benefits of controlling diesel particle emissions in Los Angeles.

The remainder of the Executive Summary is organized according to the three main chapters of the report: Description of the Visibility Data (Chapter 2), Estimation of Benefits from Visibility Improvement (Chapter 3), and Illustrative Application: Diesel Particulate Control in Los Angeles (Chapter 4).

CHAPTER 2: DESCRIPTION OF THE VISIBILITY DATA

Weather Station Visibility Data: Compilation and Analysis

One of the inputs required for the statistical analysis is information on visibility levels throughout the two study areas. The visibility data used herein consist of visual range observations made at airports and other weather stations. Specifically, data are compiled for three observations per day at twenty-one Los Angeles stations and eighteen San Francisco stations during each day of 1973-74 and 1978-79.

For the purposes of this study, the airport visibility data generally appear to be of good quality. The quality of the data was assured by conducting surveys at the weather stations and by scrutinizing the statistical distributions of the data. Because of the nature of reporting practices at airports, special techniques documented in the literature have been applied in determining cumulative frequency distributions and in calculating visibility statistics.

Visibility (Extinction) Indices

Maps of visibility in the Los Angeles area reveal that the lowest median visual range in the basin, about 7 miles, occurs in the San Gabriel and Pomona Valleys. Median visual range improves toward the west and east, to about 11 miles along the western coast and to over 25 miles in the San Gorgonio pass. Visibility is significantly higher in San Francisco than in Los Angeles. There are two notable hot spots in the San Francisco area where visual range is about 13 miles -- the San Francisco-Oakland urban core and the Southern Bay area around San Jose. However, the San Francisco area otherwise demonstrates median visibility of about 15 to 25 miles. Comparing maps for the two time periods indicates that a slight but not uniform improvement in visibility occurred in both study areas from 1973-74 to 1978-79.

For use in the economic regression studies, the visual range data are converted to atmospheric extinction data. The atmospheric extinction coefficient represents the fraction of light that is lost per unit distance as a light beam traverses the atmosphere. In a uniform atmosphere, a simple reciprocal formula relates extinction (B) to visual range (V): $B = k/V$, where "k" is a constant that depends on the observer detection threshold.

Extinction is generally considered preferable to visual range as a scientific measure of visibility. In addition, extinction seems preferable in this study for two practical reasons. First, economic regressions using extinction agree with the known principle that humans perceive a unit change in visual range much more at low visibilities (e.g. from 1 to 2 miles) than at high visibilities (e.g. from 100 to 101 miles). Second, airport data are known to be of relatively higher quality at lower visibilities than at higher visibilities; using extinction rather than visual range emphasizes the variations in the higher quality part of the data rather than the variations in the lower quality part of the data.

In each study area and each time period, three indices of extinction are compiled for use in the economic studies. These indices are median annual extinction for all hours, median annual extinction for hours excluding precipitation or fog, and median annual extinction with sea haze contribution subtracted out. It turns out that the first two indices are so highly intercorrelated that they yield identical results. The third index is not that highly correlated with the other two, but it nevertheless again leads to similar conclusions in the economic analysis.

CHAPTER 3: ESTIMATION OF BENEFITS FROM VISIBILITY IMPROVEMENT

Methodological Review

The economic analysis used herein follows the Freeman-Rosen approach for identifying demand curves of commodities not normally traded in markets. The essential element of the Freeman-Rosen approach, as applied to housing data, is the "hedonic" price function which relates the price of a home to its characteristics (structural, locational, neighborhood, and environmental aspects). This function allows determination of the implicit or hedonic price of each characteristic (i.e. visual air quality), which can be interpreted as the individual's marginal willingness to pay for that characteristic.

The individual's marginal willingness to pay for air quality depends upon other housing characteristics and the individual's characteristics, especially income. The second stage of the hedonic procedure is to estimate the relationship between marginal willingness to pay and these other characteristics. This latter relationship can be interpreted as the (inverse) demand curve for air quality, since it connects price to quantity and other shift variables. The benefits of a specific air quality change can be determined by integrating the inverse demand curve over the proposed improvement.

Until recently, the basic Freeman-Rosen framework has been widely accepted as a means of estimating the benefits of environmental improvements. However, the procedure has lately been criticized as being inappropriate under certain conditions. The criticisms have focused on two issues: (1) the functional form of the hedonic equation in the first stage, and (2) the identification of the (inverse) demand curve in the second stage.

With respect to the first issue, there are no clues as to the correct shape of the hedonic function. Therefore, sensitivity analysis is employed in this study to determine a range of benefit estimates. With respect to the second issue, various authors have questioned whether sufficient information exists for estimating the demand curve. Two possible solutions are utilized here: (1) combining data from multiple markets to yield information on how individuals respond to different price sets, and (2) using data from a single market but imposing further restrictions on possible functional forms.

Data Specifics

Implementation of the hedonic approach requires two data sets. The first data set includes the sale prices of numerous homes and their attributes (structure, neighborhood, community, and environment). The data on sale price and house structure were obtained from the Market Data Cooperative for the 1978-79 time period. Structural variables pertain to both quantity (square footage, number of bathrooms, etc.) and quality (pool, fireplaces, view, etc.). A very large number of observations were used to provide robust statistical estimation properties. Neighborhood refers to the surrounding census tract and includes the variables -- population, age, ethnic composition, distance to work, and distance to the beach. Community (city level) variables encompass density, school quality, crime rate, and others. The final variable included in the hedonic modeling is atmospheric light extinction, discussed in Chapter 2. The neighborhood, community, and light extinction data were matched with the household data using Thomas Brothers maps (4 x 4 km grid squares).

Once hedonic prices (marginal willingness to pay) for light extinction improvements have been determined from the first data set, the second step of the approach is to determine the shape of the inverse demand curve. This is done by relating the hedonic prices to light extinction and income. The data base for this second step is compiled at the community level.

Empirical Results: Hedonic Housing Value Equations

Given the data as described above, the hedonic technique consists of a two-stage estimation procedure. The initial task is to determine the relationship between home price and its characteristics. This procedure allows one to focus on the significance of light extinction while separating out the influence of other extraneous variables.

Benchmark hedonic housing value equations are estimated for both the Los Angeles and San Francisco areas. A number of aspects of the benchmark results are worth noting. First, the non-linear specifications are significant improvements over linear forms. Second, a large proportion of the variation in home sale price is explained by the independent variable set. Third, the independent variables generally possess the expected relationship to home sale price and are significantly different from zero at the one percent level.

However, the most important result from the perspective of this study is that the extinction variable is significantly different from zero and possesses the expected relationship to home sale price. Evidently, individuals are acting upon extinction information when making locational choices, with the action translated into a measurable hedonic gradient. This result is essentially invariant with respect to various sample sizes, extinction measures, model formulations, and functional forms.

The value of a hypothetical ten percent change in extinction ranges from approximately .7 - 2.1 percent of home price in the Los Angeles area and from 1.4 - 2.5 percent of home price in the San Francisco area. The specific value within an air basin is primarily dependent upon functional form.

Empirical Results: Inverse Demand Equations

The second stage of the hedonic price technique is to determine the inverse demand curve for light extinction. As discussed above, this step relates the individual marginal willingness to pay (hedonic price) to extinction level and income.

The inverse demand curves for the Los Angeles and San Francisco areas possess a number of noteworthy aspects. First, a large proportion of the variation in marginal willingness to pay is explained by the variables quantity (extinction) and income. Second, linear forms outperform non-linear forms. Third, the independent variables are generally significant at the one percent level. Finally, an interesting phenomenon occurs in the Los Angeles area only, where the inverse demand curves suggest that residents may be willing to pay even more for each unit decrease in extinction as air quality improves. This seemingly demonstrates what economists call "non-convex" preference patterns. However, since the hedonic housing equations are non-linear, no interpretation can really be attached to the inverse demand curves regarding convexity of preferences.

The magnitude of air quality benefits can be illustrated by calculating the annual basinwide benefits associated with a hypothetical ten percent improvement in visibility. These benefit figures are obtained by integrating the inverse demand curves over the proposed visibility change and summing over all households. The benefit estimates are dependent upon the functional form of the hedonic price equation. This is especially true for the Los Angeles area where benefits range from 250 to 617 million dollars per year (depending on functional form) for a ten percent change in visibility. The San Francisco results are not as dependent upon functional form, ranging from 190 to 220 million dollars per year for a similar ten percent improvement.

The range of benefit estimates discussed above utilize the classic hedonic approach as developed by Freeman and Rosen with an added functional form restriction. An alternative approach is to pool the data across markets and estimate one multi-market inverse demand curve. This approach requires the assumption that individual preferences be identical across the markets. The use of multiple markets adjusts the estimates in the manner anticipated. For instance, adding San Francisco households into an analysis of the Los Angeles area increases the benefit estimates since San Francisco home prices seem to show more sensitivity to visibility degradation.

CHAPTER 4: ILLUSTRATIVE APPLICATION: DIESEL PARTICULATE CONTROL IN LOS ANGELES

Relationship Between Diesel Particle Emissions and Visibility

The benefit analysis for a specific emission control strategy separates into two basic parts. First, one must determine the degree of visibility improvement associated with the control strategy (this subsection). Second, the benefits produced by the visibility improvement are estimated using the economic inverse demand curve (the following subsection).

The current visibility impact from diesel road vehicles (essentially heavy-duty trucks) is estimated using two models, an emission budget model and a lead tracer model. These two models indicate that heavy-duty diesels presently contribute about 20 percent of light extinction in the Los Angeles basin. This 20 percent is composed of 12 percent from the elemental carbon component of particle exhaust, 1 percent from other primary particle emissions, 4 percent from secondary products of NO_x emissions, and 3 percent from secondary aerosols due to SO_2 emissions. Because this example is concerned with directly-emitted particles (i.e. particulate emission standards), and because nearly all of the visibility effects from primary diesel particles are due to elemental carbon, the application study is restricted to only the contribution from elemental carbon particle emissions.

The lead tracer model calibrated against the average of the two models considered provides a detailed spatial distribution for the diesel visibility effects. In absolute terms, the greatest extinction contributions from current diesel elemental carbon emissions occur in a triangle from the San Fernando Valley to Long Beach in the west out to San Bernardino in the east.

The effect of future diesel emissions on extinction can be estimated by proportioning the current spatial distribution of diesel impacts according to emission changes. This can be done for emission growth increases, emission control decreases, or combinations of both.

For the purpose of control strategy analysis, elemental carbon emissions from diesel road vehicles are projected for a "no control" scenario covering 1980 to 1992. Without control, diesel emissions are forecasted to increase by a factor of 2.3 from 1980 to 1992. This rapid increase in emissions results from two factors, overall growth in highway traffic and partial conversion of gasoline cars and trucks to diesel power. It is noteworthy that, even in 1992, heavy-duty vehicles would still account for nearly 90 percent of total diesel fleet emissions. The benefit analysis considers various degrees of control applied to forecasted diesel emissions.

Benefit Estimation

A variety of control scenarios are examined for diesel particulate emissions, with calculated benefits representing the change from "no control" to the specified level of control. Benefits are computed over the 1980-1992 time period, at the community level, for various discount rates, with results expressed in terms of constant 1980 dollars.

Following the traditional Freeman-Rosen approach, the Los Angeles inverse demand equations are mathematically integrated over the visibility improvement to determine individual household benefits. Aggregate benefits then come from summation over the relevant population. The calculations are illustrated herein through a specific example. Results are presented for four alternative functional forms and two alternative discount rates.

In the first scenario, diesel emissions are assumed to be fixed over time, and a constant 50 percent particulate control level is imposed over the entire 1980-1992 time period. All subsequent scenarios are more realistic in that uncontrolled diesel emissions are allowed to grow over time. Scenario II considers 50 percent control of forecasted emissions starting in 1980. Fifty percent is approximately the degree of control being considered for state and national emission standards. Scenario III is similar to Scenario II, except that 80 percent control is assumed. Scenario IV considers the effect of phasing in controls by assuming that no emission reduction is imposed until 1985, and that control is phased in linearly up to 50 percent in 1992.

The net benefits, averaged over the four functional forms and the two discount rates, are as follows: $2.2 billion for Scenario I, $2.7 billion for Scenario II, $4.5 billion for Scenario III, and $.98 billion for Scenario IV. Scrutiny of the results leads to several general conclusions. First, the benefits grow disproportionately as control levels increase, but this effect is relatively minor for marginal changes in extinction. Second, the larger the control effect, the larger the corresponding benefits. The third conclusion, that the benefits of stringent control on basin-wide extinction should be extremely large, arises from a combination of the first two factors. This occurs because (1) the larger the control the larger the benefits, and (2) as pristine air quality is approached, benefits grow at an increasing rate due to the shape of the demand curve. The fourth feature is the substantial loss in benefits from postponing and phasing controls (i.e. Scenario IV).

1.0 INTRODUCTION

With increasing frequency, the cost effectiveness of public programs is being questioned. One of the largest environmental programs, especially in California, is the maintenance of air quality. The costs of the various programs which comprise air pollution control are reasonably well documented (for example, see Lloyd, 1979 and SCAQ, 1979), but benefit estimation is poorly understood. There exists confusion concerning the goods to value (health, aesthetics, etc.), the evaluation methods, and the relationship between them. In this project, an attempt is made to help fill this void by placing an economic value on changes in air quality. The analysis is completed for two air basins, the South Coast Air Basin (Los Angeles area) and the Bay Area Air Basin (San Francisco area).

The approach taken here is to examine the public's willingness to pay for improved visual air quality, as revealed by their preferences in the housing market. Specifically, we examine the relationship between housing values and visibility, discounting for the effects of other collinear variables (house/community/location/socio-economic characteristics). After being analyzed and tested according to economic and statistical principles, this housing-value/visibility relationship serves as a basis for estimating an economic (inverse) demand curve for visibility. This allows one to estimate the benefits associated with improved visual air quality.

The analysis is based on an extremely large, finely resolved data base for visibility levels, housing values, and house/community/location/socio-economic parameters. The investigation is conducted for the 1978-79 time period in two study regions, four Los Angeles area counties and five San Francisco area counties. An additional analysis for the 1973-74 time period is presented in an appendix.

In this project, emphasis is placed on what individuals perceive, as visibility is the object of the valuation exercise. The presumption is that all components of air quality valued by people are included in perceived visibility, and no attempt is made to separately value the independent components of air quality.

1.1 ECONOMIC CONCEPTS AND APPROACH

Environmental commodities such as visibility are different from most goods and services because they are not transacted in established markets. Hence, market prices do not exist, and value cannot be derived from an analysis of a market for these goods. In lieu of direct market revelation of value, economists have turned to alternative methods to value environmental goods. There are primarily two methods available: (1) analyze data from market transactions for goods and services related to the environmental good, and (2) ask individuals directly using survey instruments.

This analysis of visibility-related benefits uses the technique involving related market data. In particular, the "hedonic housing value" method is utilized to determine the value of visibility changes. This approach is the most common form of the hedonic price* procedure as developed by Rosen (1974), the basis of which is Lancaster's (1966) consumption theory. The procedure assumes that access to environmental (dis)amenities is capitalized in housing values. This assumption is based on the premise that households are willing to pay a premium for an otherwise identical home located in a clean air area versus one located in a polluted area.

This methodology was chosen for two reasons. First, it has received the greatest attention (theoretical and empirical scrutiny) and support from the economics profession. Among public goods which have been valued using the hedonic housing approach are air pollution (Anderson and Crocker, 1971; Harrison and Rubinfeld, 1978), social infrastructure (Cummings et al., 1978), noise level (Nelson, 1979), and ethnic composition (Schnare, 1976). Second, this approach is attractive because there is a wealth of readily available data for the required parameters. The necessary data include visibility levels, housing values, and housing/socio-economic/community/locational variables.

In employing the hedonic housing method herein, particular attention has been given to three specific issues. First, this project uses light extinction (inversely related to visual range) as a measure of air quality. Earlier studies have used suspended particules and oxides of nitrogen

*Hedonic prices are implicit prices of the characteristics which differentiate closely related items in a product class.

(Brookshire et al., 1982), carbon monoxide (Bresnock, 1980), and other measures to represent the overall pollution level. However, these pollutants are either imperceptible or correlate rather poorly with visibility (Trijonis et al., 1982). Thus, earlier studies can be questioned as to how people could perceive what was being valued. This study uses what people actually perceive. Furthermore, the results are amenable to policy applications since the transformation between light extinction and emissions levels can be determined.

The second issue concerns the hedonic housing value method. This approach is generally viewed as a multistage procedure (Rosen, 1974; Freeman, 1979). The initial step is to estimate the hedonic price gradient which explains home sale price as a function of its structural characteristics as well as the characteristics of the community and neighborhood in which it is located. The second step is to determine the implicit price of environmental change by differentiating the hedonic price gradient with respect to the environmental variable of interest. Subsequent steps include estimation of the inverse demand curve and integration to obtain benefit estimates.

The hedonic procedure as outlined above has been generally well received by the economics profession. Recently, however, a number of authors, including Brown and Rosen (1982), Mendelsohn (1980), and Palmquist (1981), have criticized the approach as not possessing sufficient information to identify the (inverse) demand curve in the subsequent steps. A possible solution to this under-identification problem is to use data from geographically separate markets to identify the inverse demand curve. This solution is examined empirically in this project by employing data from both the Los Angeles and San Francisco air basins. The multi-market results are then compared to the single market results.

The third issue concerns functional form. Bender et al. (1980) found that the value of air quality improvements was related to the assumed functional forms of the hedonic price equation and the inverse demand curve. Again, this issue is examined empirically in this study.

1.2 ORGANIZATION OF THE REPORT

The main body of this report consists of three chapters. Chapter 2 develops the visibility data base for the study. Visibility observations from 21 airports in Los Angeles and 18 airports in San Francisco are used to prepare maps of visual range for the two study regions. Three indices of atmospheric light extinction (inversely proportional to visual range) are then computed on a detailed spatial grid for each study area. Chapter 3 implements the hedonic housing value approach. A very thorough statistical analysis is performed to characterize the relationship of home sale price to visibility and home-specific/locational/neighborhood/social-economic parameters. This relationship allows estimating the inverse demand curve for visibility, which in turn serves as the basis for economic benefit calculations. Chapter 4 presents an illustrative policy application involving diesel particulate control in Los Angeles. Diesel particulate emissions are related to extinction (visibility) levels using a modified lead tracer model. The benefits of potential emission reductions are then estimated using the results of Chapter 3 and economic calculations of varying temporal complexity.

1.3 REFERENCES

Anderson, R. and T. Crocker, "Air Pollution and Residential Property Values," Urban Studies, 8, October 1971.

Bender, B., T. J. Gronberg, and Hae-Shin Havong, "Choice of Functional Form and the Demand for Air Quality," Review of Economics and Statistics, pp. 638-43, November 1980.

Bresnock, A. E., "Housing Prices, Income and Environmental Quality in Denver," University of Colorado, 1980.

Brookshire, D., M. Thayer, W. Schulze, and R. d'Arge, "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," American Economic Review, 72, March 1982.

Brown, J. and H. Rosen, "On the Estimation of Structural Hedonic Price Models," Econometrica, May 1982.

Cummings, R., W. Schulze, and A. Meyer, "Optimal Municipal Investment in Boomtowns: An Empirical Analysis," Journal of Environmental Economics and Management, 5, 252-67, September 1978.

Freeman, A. M., III, "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues," Scandinavian Journal of Economics, 81, 1979b.

Harrison, D., Jr. and D. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," Journal of Environmental Economics and Management, 5, March 1978.

Lancaster, K., "A New Approach to Consumer Theory," Journal of Political Economy, 74, April 1966.

Lloyd, Kenneth H., Cost and Economic Impact Assessment for Alternative Levels of the National Ambient Air Quality Standard for Ozone, USEPA, (February 1979).

Mendelsohn, R., "The Demand and Supply for Characteristics of Goods," University of Washington, 1980.

Nelson, J., "Airport Noise, Location Rent, and the Market for Residential Amenities," Journal of Environmental Economics and Management, 6, December 1979.

Palmquist, R., "The Demand for Housing Characteristics: Reconciling Theory and Estimation," North Carolina State University, 1981.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 82, January/February 1974.

Southern California Association of Governments and South Coast Air Quality Management District, Air Quality Management Plan, January 1979.

Schnare, A., "Racial and Ethnic Price Differentials in an Urban Housing Market," Urban Studies, 13, June 1976.

Trijonis, J. et al., "Analysis of Visibility/Aerosol Relationships and Visibility Modeling/Monitoring Alternatives for California," Prepared under contract #A9-103-31 for the California Air Resources Board, 1982.

2.0 DESCRIPTION OF THE VISIBILITY DATA

The economic benefit analysis in this report will be based on statistical relationships between housing values and visibility, determined for two study regions (Los Angeles and San Francisco) and two time periods (1973-74 and 1978-79). In order to develop the statistical relationships, we need a detailed characterization of spatial visibility patterns within the two study regions during the relevant time periods. This chapter explains how routine weather station (airport) visibility observations were assembled, processed, and analyzed in order to provide that characterization.

Section 2.1 describes the visibility observation procedures, indicates the study locations, and reviews data quality. Section 2.2 explains the special techniques that must be applied in order to estimate median visibilities from airport data. Section 2.3 discusses the three measures of atmospheric extinction (visibility degradation) that were compiled for use in the housing value study.

2.1 WEATHER STATION VISIBILITY DATA

The visibility data presented in this report consist of "prevailing visibility" recordings made by weather station observers. According to National Weather Service procedures, prevailing visibility is defined as the greatest visual range that is attained or surpassed around at least half of the horizon circle, but not necessarily in continuous sectors (Williamson, 1973). Daytime visibility is measured by observing markers (e.g. buildings, mountains, towers, etc.) against the horizon sky; nighttime visibility measurements are based on unfocused, moderately intense light sources. Because our experience indicates that daytime and nighttime observations are often incompatible, and the daytime data are usually of higher quality (Trijonis and Yuan, 1978; Trijonis, 1979, 1980), only daytime observations are used herein.

Weather observers usually perform visibility measurements each hour, but only the readings for every third hour are entered into the National Climatic Center computerized data base.* Although our previous studies of

*In the Pacific Time Zone, these hours are 1:00 AM, 4:00 AM, ..., 10:00 PM, standard time.

spatial visibility patterns were based on only one observation per day (1:00 PM PST in California), we decided to include three observations per day in this project so as to increase the statistical sample size. Because this study uses hard-copy weather station records rather than computerized NCC data (very few of our study sites had computerized data), there was complete freedom of choice in selecting the three hours. Nevertheless, we chose the computerized hours of 10:00 AM, 1:00 PM, and 4:00 PM in order to allow compatibility in any future projects that might employ computerized data bases.

For each study region, separate analyses were conducted for two different time periods, 1973-74 and 1978-79. Two periods were desirable for the economic studies in order to test the results for consistency over time. Two years of data for each period provided large, yet manageable, amounts of data.

2.1.1 Study Locations

Before sites were selected for this study, telephone surveys pertaining to data quality were conducted with visibility observers at the various weather stations. The main purpose of these surveys was to insure that each station had an adequate set of visibility markers for estimating visual range. In particular, we attempted to choose sites that had farthest markers located at distances at least as great as the visibility levels typical of the surrounding area. Because visibility is generally rather low in the study regions (especially the Los Angeles region), nearly all of the weather stations had adequate markers, and we acquired data for essentially every weather station in each region. In one case (Banning), the station did not have markers at distances exceeding the median visual range, so that we had to extrapolate the cumulative frequency distribution in order to estimate median visual range.

Figures 2.1 and 2.2 show the study sites in the 4-county Los Angeles region and the 5-county San Francisco region, respectively. Data were acquired and processed for 21 Los Angeles stations and 18 San Francisco stations. As discussed in the next section, we later decided to eliminate all Coast Guard stations, leaving 19 final study sites in Los Angeles and 14 in San Francisco.

At this point, it is worthwhile to mention the enormity of the task of

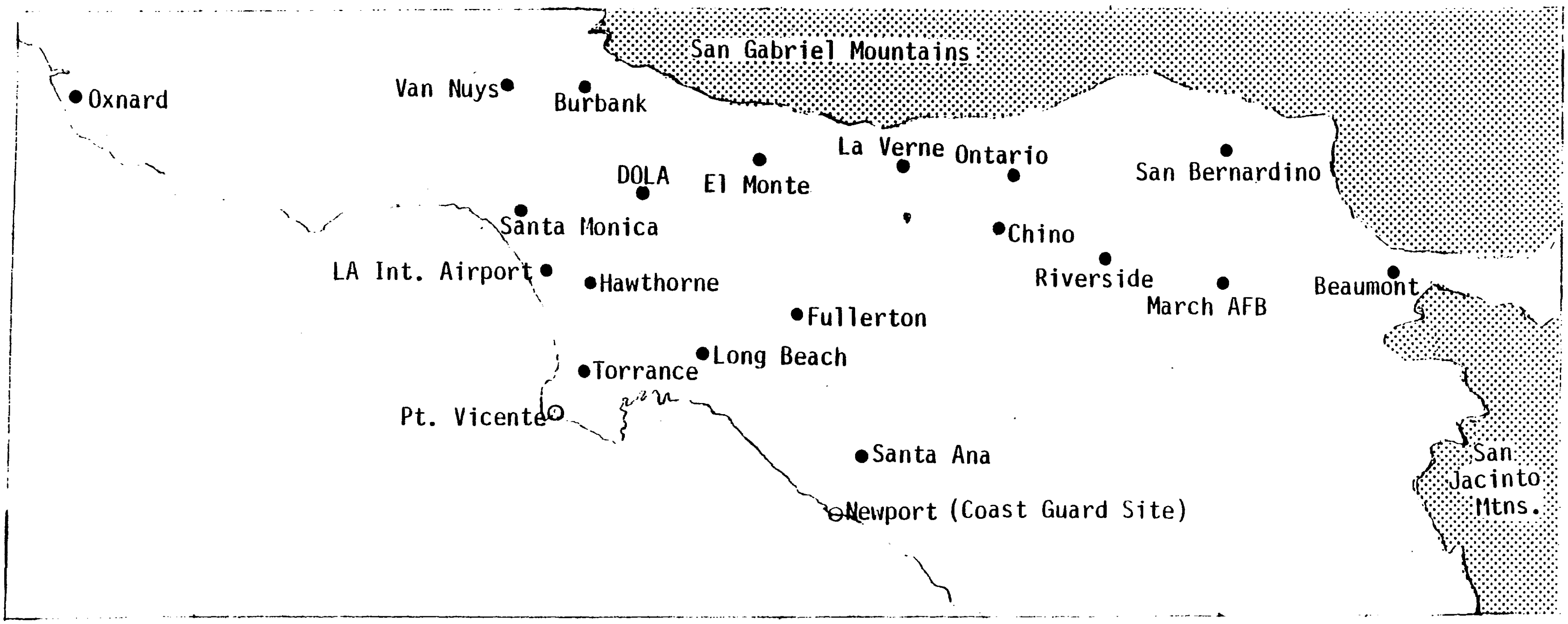

Figure 2.1 Weather stations providing visibility data for the Los Angeles study area.

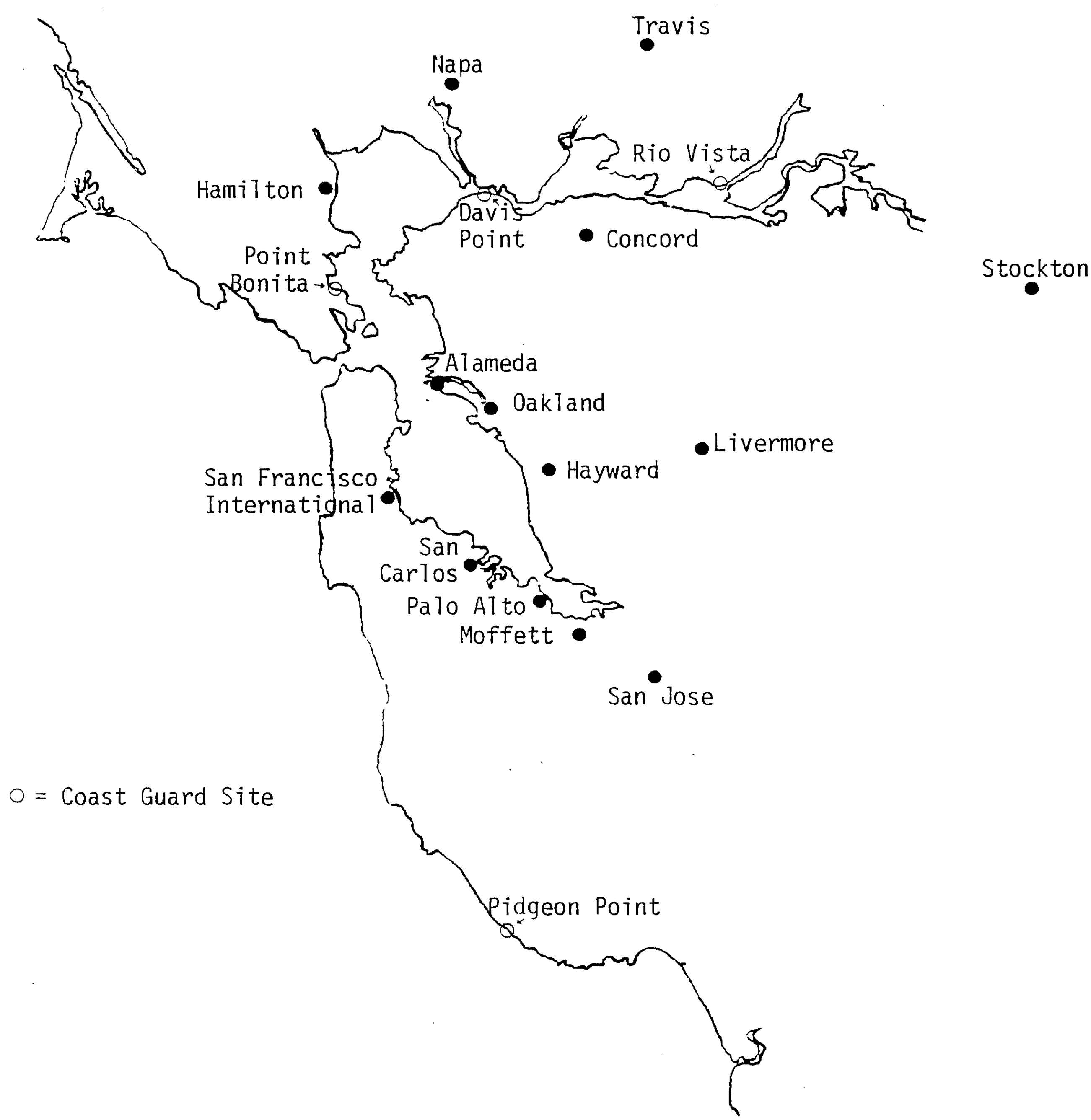

Figure 2.2 Weather stations providing visibility data for the San Francisco study area.

hand-processing visibility data for all the study locations. In all, we hand-transcribed and processed approximately 170,000 data points (39 sites x 4 years per site x 365 days per year x 3 observations per day).

2.1.2 Data Quality

Previous studies (Trijonis and Yuan,1978; Trijonis and Shapland,1979; Husar et al., 1979; Trijonis,1979, 1980) have found that airport data are of good quality for use in analyzing spatial visibility patterns. The quality of the data is indicated by the excellent consistency generally found in the observed spatial patterns of airport visibility. Specifically, the above cited studies have shown that median visibilities at neighboring airports tend to agree and that reasonable monotonic gradients often exist in passing from areas of poor visibility to areas of good visibility. Furthermore, Nochumson et al.(1983) has shown that the spatial patterns of airport visibility correlate highly with independent data sets for aerosol concentrations and relative humidity.

In our prior work with weather station visibility measurements, we have generally found that the best quality data come from sites operated by weather service personnel or by commercial airport personnel. This expectation was borne out in the present study by the fact that we observed no obvious anomalies in the data from such sites. Possibly because of personnel turnover or because of operational priorities, we have generally found that visibility data from Coast Guard, Navy, and Air Force facilities sometimes contain inconsistencies. All of our potential study sites were weather service or commercial airline operations except for six Coast Guard locations, two Navy air stations, four Air Force bases, and one APCD location. The data bases for these last 13 sites deserve special discussion.

The data from Coast Guard locations exhibited significantly lower visual ranges than the data from the other locations. The most likely explanation is that the Coast Guard stations are sited right on the water and are subject to more fog and sea spray. Another potential explanation concerns possible biases in the Coast Guard data -- from reporting procedures, marker types, or some systematic patterns in the rather large number of missing observations at Coast Guard stations. In either case, it seems best to exclude the Coast Guard data from our study. Even if the lower visibility at Coast Guard stations is a real effect (as we suspect), it appears to be extremely localized, and we cannot represent it adequately within the spatial revolution of our analysis (4 x 4 mile grid squares).

In processing and analyzing the data from the two Navy sites and four Air Force sites, we generally found these data sets to be reasonable. The only two exceptions were the Navy data at Alameda and Moffett during 1973-74. The median visibilities at these two sites in 1973-74 seemed anomalously low compared to median visibilities at nearby locations. We do not know the cause, but such an anomaly could have been produced by inconsistent reporting practices (such as one observer reporting a maximum visibility of 7 miles). This caveat regarding the 1973-74 data at Alameda and Moffett will be referenced later when we discuss the regression results for the San Francisco area during 1973-74.

The APCD data at Downtown Los Angeles appeared to be of good quality and consistency, but a bias occurred during 1978-79 because of the absence of weekend data for that period. We derived and applied a factor to eliminate that bias by quantifying weekend/weekday visibility differences at Downtown Los Angeles during 1973-74 as well as at Burbank and El Monte during all the years.*

2.2 DATA COMPILATION AND ANALYSIS

In practice, a weather station visibility recording of X miles usually means that visual range is <u>at least</u> X miles rather than visual range is exactly X miles. For example, at a station with farthest visibility markers of 40 and 30 miles, a recording of 40 miles would imply that visual range was at least 40 miles, and a recording of 30 miles would imply that visual range was between 30 and 40 miles. Because of this phenomenon, weather station visibility observations are most appropriately summarized by cumulative frequency distributions of the form "percent of time visibility is greater than or equal to X miles". These cumulative frequency distributions are determined by noting the percent of time that visibility exceeds the farthest reported value, and then adding percentages cumulatively as one proceeds toward the smaller reported values. In this process, it is very important to use only those visibilities that are <u>routinely reported</u>

*Visual range on weekends averages about 15% higher than on weekdays, implying that a full week average should be about 4% higher than a weekday only average.

by the weather observation team*; otherwise, artificial "kinks" will be produced in the cumulative frequency distribution. Summarizing the visibility data in the above way should make the data consistent from station to station even if the various stations have visibility markers at different distances.

The first step in processing the airport visibility data was to scan the readings in order to select the "routinely reported" visibilities by which the data should be organized. Next, the data were recorded on forms such as the one shown in Table 2.1. We used a separate recording form for each quarter (3 months) of data, even though the entire two-year period was aggregated in our statistical analysis. Compiling the data by individual quarters allowed us to detect and adjust for changes in the reporting practices (i.e. in the "routinely reported" visibilities). Also, organizing the data quarterly allowed for the possibility of seasonal stratification in potential future projects with these data sets.

Table 2.1 shows that we compiled the frequency distributions not only for "all hours" but also for "hours with precipitation or fog". By subtracting the latter from the former, we obtained the distribution for "hours excluding precipitation and fog". As discussed in the next section, both "all hours" and "hours excluding precipitation and fog" are among the visibility indices used in the housing value regressions.

After being recorded on the quarterly forms, the data for each site and each two-year period were summarized according to cumulative frequency tables (for example, Table 2.2). We also plotted the cumulative frequency distributions (as in Figure 2.3) to scan visually for anomalies. The median visual ranges over each two-year period were calculated by interpolating the cumulative frequency distributions between the routinely reported visibilities.

2.3 VISIBILITY (EXTINCTION) INDICES

This section explains how the median visual ranges for each two year period were summarized and formatted for direct use in the econometric studies.

*Referring to the previous example (with the 40 and 30 mile markers), if one member of the observation team occasionally decided to report 35 mile visibility rather than the routine 30 miles, then the 35 mile recordings should be lumped with the 30 mile recordings in the cumulative frequency distribution.

TABLE 2.1 EXAMPLE OF CHART FOR RECORDING VISIBILITY OBSERVATIONS (first quarter for Burbank, 1978).

Visual Marker (miles)	NUMBER OF OBSERVATIONS	
	Total Observations	Precipitation or Fog
50	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷	
40	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷𝍷	
25	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷	
20	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷	
15	𝍸 𝍸 𝍸 𝍸 𝍷	𝍷
10	𝍸 𝍸 𝍸 𝍷𝍷𝍷𝍷	
7	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	𝍸 𝍷
5	𝍸 𝍸 𝍸 𝍸 𝍷	𝍸 𝍷𝍷
2	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷𝍷𝍷	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷𝍷𝍷
1	𝍸 𝍷𝍷	𝍸 𝍷𝍷
0		

TABLE 2.2 EXAMPLE OF CUMULATIVE FREQUENCY DISTRIBUTION (Burbank, 1978-79)

Visibility (miles)	ALL HOURS: Number of Observations	ALL HOURS: Cumulative Number	ALL HOURS: Cumulative Percent	ALL HOURS EXCLUDING PRECIPITATION OR FOG: Number of Observations	ALL HOURS EXCLUDING PRECIPITATION OR FOG: Cumulative Number	ALL HOURS EXCLUDING PRECIPITATION OR FOG: Cumulative Percent
0	2	2190	100	0	1974	100
1	33	2188	99.9	1	1974	100
2	439	2155	98.4	309	1973	99.9
5	298	1716	78.4	269	1664	84.3
7	310	1418	64.7	297	1395	70.7
10	231	1108	50.6	227	1098	55.6
15	186	877	40.0	183	871	44.1
20	192	691	31.6	189	688	34.9
25	230	499	22.8	230	499	25.3
40	127	269	12.3	127	269	13.6
50	142	142	6.5	142	142	7.2
Median Visibility:		10.3 miles				12.4 miles

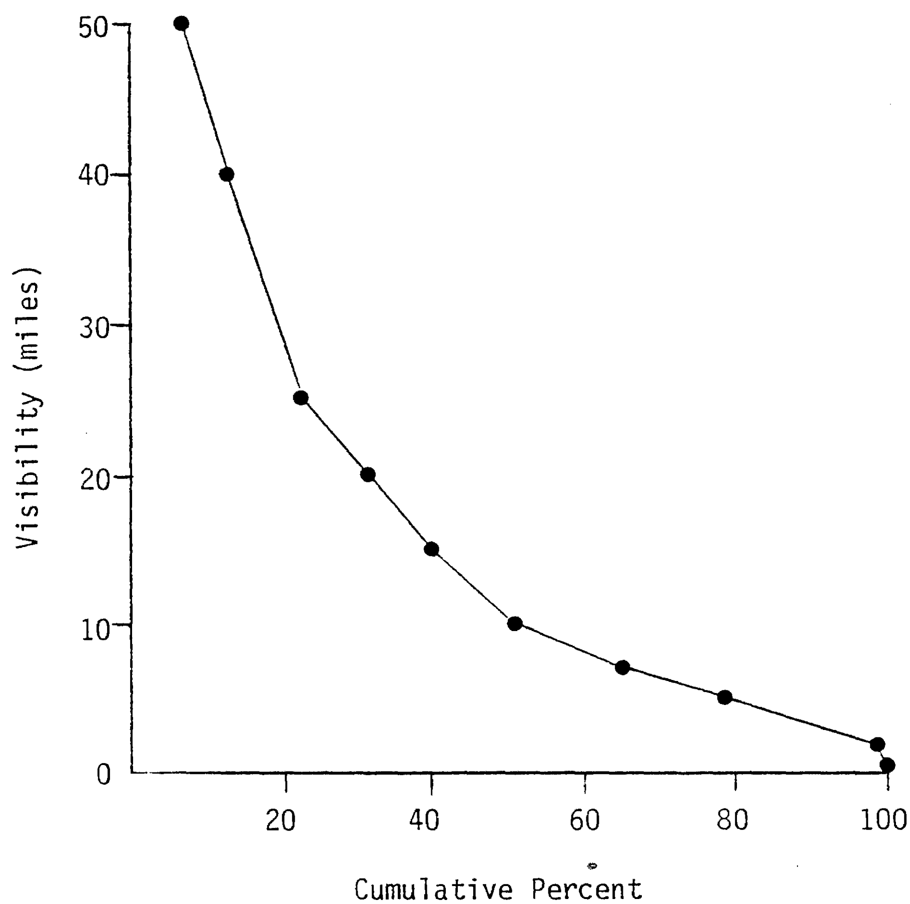

Figure 2.3 Example of cumulative frequency distribution plot (all data for Burbank, 1978-79).

The first major step in this final process involved the decision to base our studies on extinction indices rather than visual range indices. The atmospheric extinction coefficient represents the fraction of light that is attenuated per unit distance as a light beam traverses the atmosphere. In a uniform atmosphere, the extinction coefficient (B) is inversely proportional to visual range (V). The Koschmeider formula expressing this relationship is

$$B = \frac{k}{V}, \qquad (2\text{-}1)$$

where the constant "k" is usually chosen as 3.9 or 3.0, depending on whether one assumes a 2% or 5% contrast detection threshold for the observer. In this study, we have computed extinction based on a Koschneider constant of 3.0*, which is the appropriate value when using airport data (Trijonis et al., 1982; Allard and Tombach, 1980; Malm, 1979).

As a visibility index, extinction is scientifically preferable to visual range for two reasons: (1) all major indices of visual air quality (visual range, contrast, and discoloration) can be calculated in a straightforward manner knowing the spatial/spectral distribution of extinction and the optical specifications of the scene, and (2) total extinction is a simple linear sum of scattering and absorption from each particulate and gaseous component.** Extinction also seems preferable to visual range as an index of visual air quality in our housing value regressions for two reasons. First, we know that humans perceive -- and therefore value -- a unit change in visual range more at low visibilities (e.g. from 1 to 2 miles) than at higher visibilities (e.g. from 100 to 101 miles); linear regressions against extinction agree with this principle. Second, airport data are known to be of relatively higher quality at lower visibilities than at higher visibilities;

*We report our extinction values in the conventional units of $[10^{-4}m^{-1}]$. Because visual range is in units of [miles], the constant of k = 3.0 becomes transformed to a value of k = 18.7 when accounting for the switch in units.

**Total extinction consists of four basic components: natural light-scattering by air molecules (blue-sky scatter), light absorption by gases (essentially all from NO_2), light scattering by particles (usually dominated by fine particles in the 0.1 to 1.0 μm size range), and light absorption by particles (basically all from black carbon).

performing the regressions in terms of extinction rather than visual range emphasizes the variations in the higher quality part of the data rather than the variations in the lower quality part of the data.

Three indices of extinction were determined for each weather station and each two-year time period. The first is median annual extinction, calculated by applying Equation (2-1) to the median two-year visual range at each station. The second is median annual extinction for hours without precipitation or fog, calculated by applying Equation (1) to median two-year visual range for non-precipitation, non-fog hours. The third is median annual extinction discounting for sea haze. This last index is calculated at each weather station by subtracting an estimate of median ocean haze/fog extinction from median total extinction. We estimated the median ocean haze/fog contribution by studying shore to inland gradients of extinction (as reported by Trijonis, 1980, 1982) in rural coastal areas of California that are free from strong local pollution effects. Our general estimate of sea fog/haze extinction in California is

$$B_{sea} = 1 \cdot e^{-x/30\ km}, \qquad (2\text{-}2)$$

where B_{sea} is in units of $10^{-4}m^{-1}$, and "x" is the distance from the station to the coastline.*

Figures 2.4 through 2.6 present maps of the three visibility indices (specified as visual range rather than extinction) for the Los Angeles study area in 1978-79. Each map illustrates the visibility values at the various weather stations as well as isopleths drawn to those values. It is apparent from Figure 2.4 (all hours) that the lowest visual range in the basin, about 7 miles, occurs south of the San Gabriel Mountains in the Pomona and West San Gabriel Valleys. Median visual range improves toward the west and east to about 11 miles along the western coast and to over 25 miles in the San Gorgonio pass. Figure 2.5 (all hours excluding precipitation and fog) is very similar to Figure 2.4, except that the median visibilities are slightly higher when fog and precipitation are sorted out. Figure 2.5 (all hours, sea haze contribution subtracted out) is somewhat like Figure 2.4 except for

*For San Francisco sites, we chose "x" as a weighted average -- .75 times the distance to the outer coastline and .25 times the distance to the inner bay shoreline.

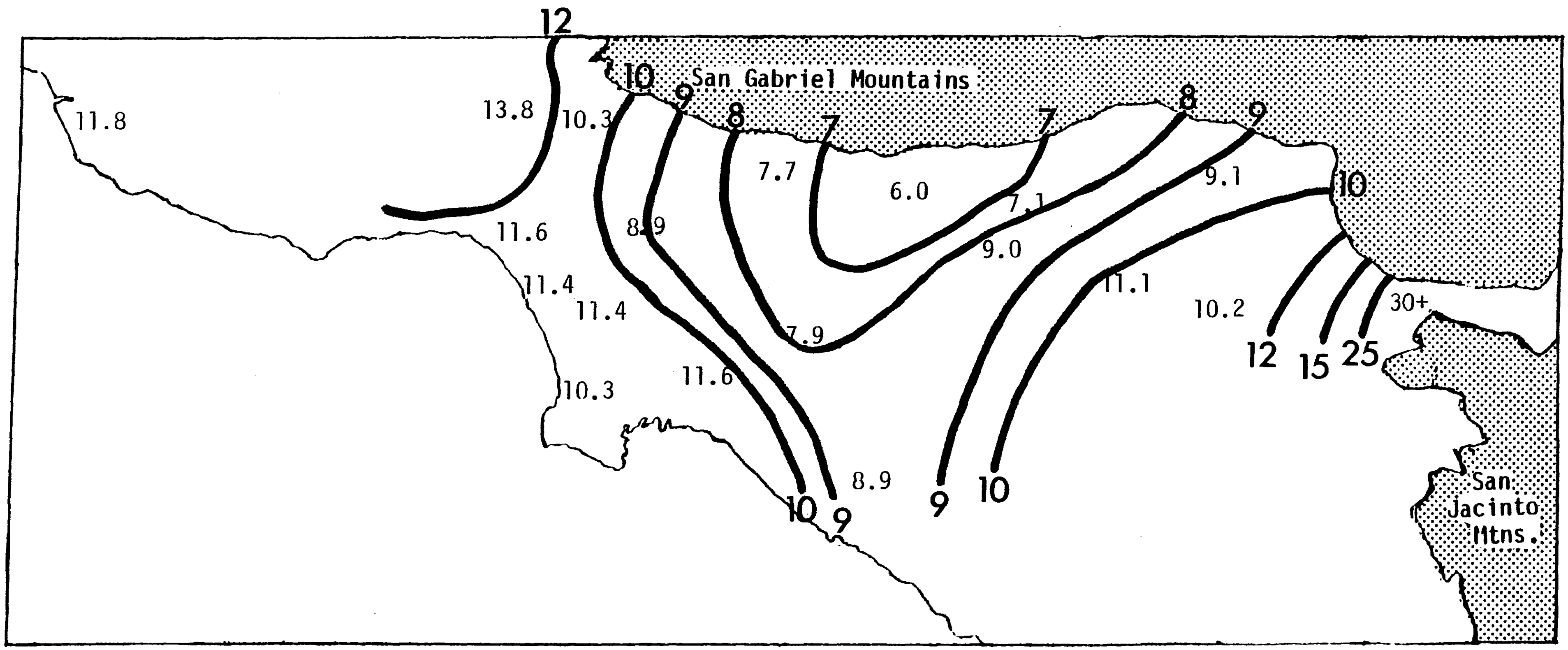

Figure 2.4 Isopleth map of median annual visibility for the Los Angeles area, all hours, 1978-79.

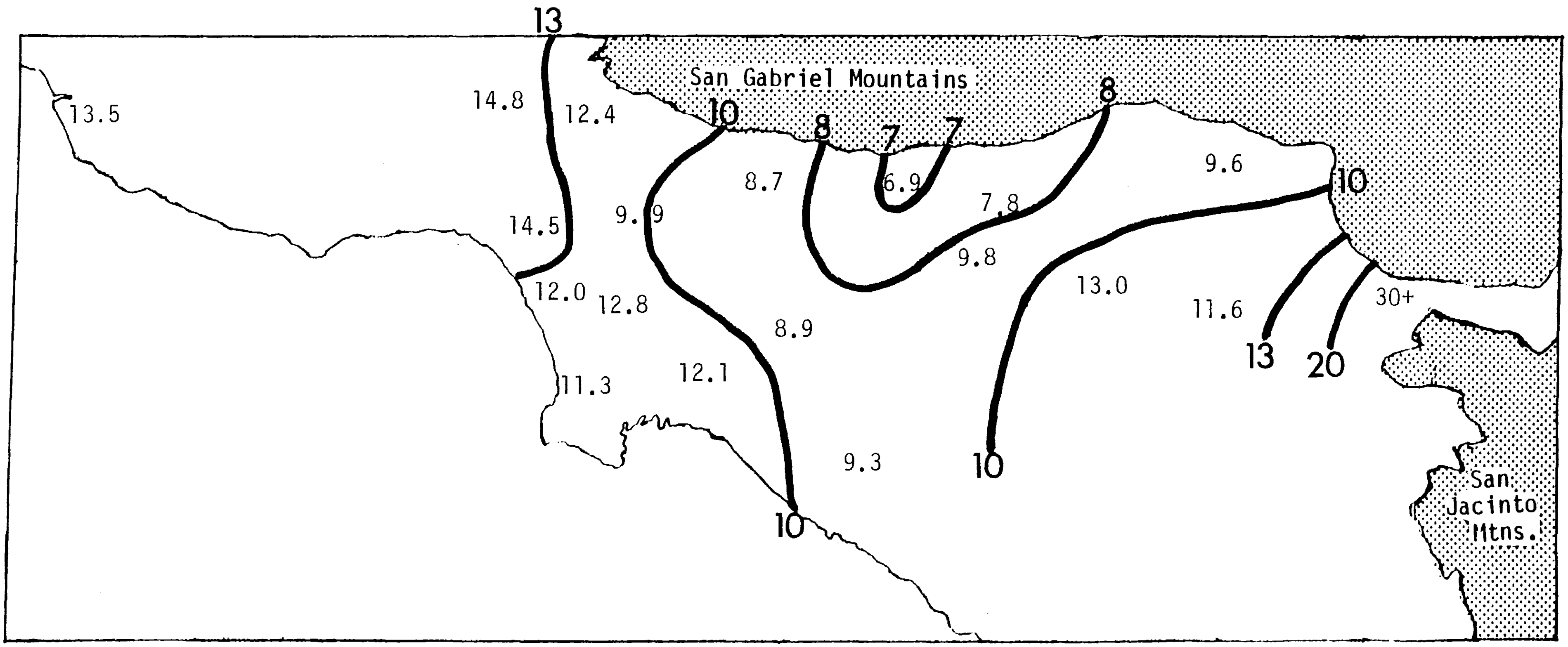

Figure 2.5 Isopleth map of median annual visibility for the Los Angeles area, all hours, excluding precipitation and fog, 1978-79.

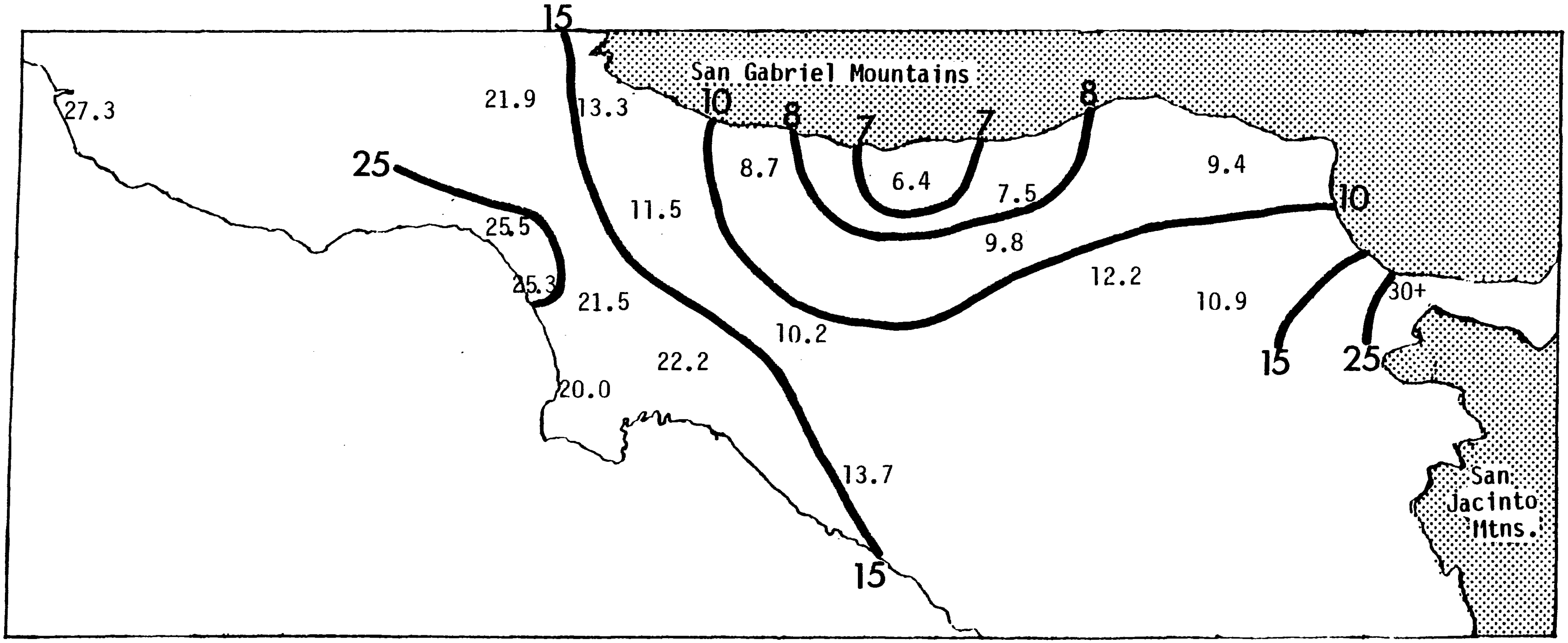

Figure 2.6 Isopleth map of median annual visibility for the Los Angeles area, all hours, sea haze contribution subtracted out, 1978-79.

the stronger gradient of improving visibility toward the coast; with sea haze discounted, median visual range along the west shore of the region is 20-25 miles rather than 10-12 miles.

Figures 2.7 through 2.9 present corresponding maps for the 1973-74 period. The spatial pattern during 1973-74 is generally similar to 1978-79, but the area of lowest visibility extended through the East San Gabriel Valley to Downtown Los Angeles in the earlier years. Generally, a slight improvement in visibility occurred from 1973-74 to 1978-79, with the largest increases in the Los Angeles County part of the region.

Figures 2.10 to 2.12 present maps of the three visibility indices for the San Francisco study area during 1978-79. It is obvious that visual range is significantly higher in San Francisco than Los Angeles. Figure 2.10 (all hours) reveals two notable hot spots in the San Francisco area where visual range is less than 15 miles. These are the San Francisco-Oakland urban core and the Southern Bay area around San Jose. Again, Figure 2.11 (all hours excluding precipitation and fog) is very similar to Figure 2.10 except for slightly higher median values. Figure 2.12 shows that, with sea haze eliminated, a band of very high visibility (visual range exceeding 35 miles) occurs along the western coast of San Francisco.

Figures 2.13 to 2.15 present San Francisco maps for 1973-74. The spatial patterns are similar to Figures 2.10-2.12, but again we find a general pattern of slightly improving visibility from 1973-74 to 1978-79.

For the purposes of our regression studies, we needed extinction indices for the Thomas Brothers grid squares shown in Figures 2.16 and 2.17. These were obtained by overlaying transparencies of Figures 2.16 and 2.17 on Figures 2.4 to 2.15 and estimating (interpolating) the visibility indices according to the isopleths. As noted previously, the visual range indices were converted to extinction indices for the purposes of the regression analyses.

We also assigned an "uncertainty value" to the data for each grid square. An uncertainty code of "0" denoted grid squares where we had the most confidence in the visibility data. An uncertainty code of "1" indicated squares where we had less confidence. Some of the uncertain squares are on the outskirts of the region covered by the weather stations. The other uncertain squares contain hilly or mountainous terrain (we thought that visibility might vary substantially with altitude within these squares, and

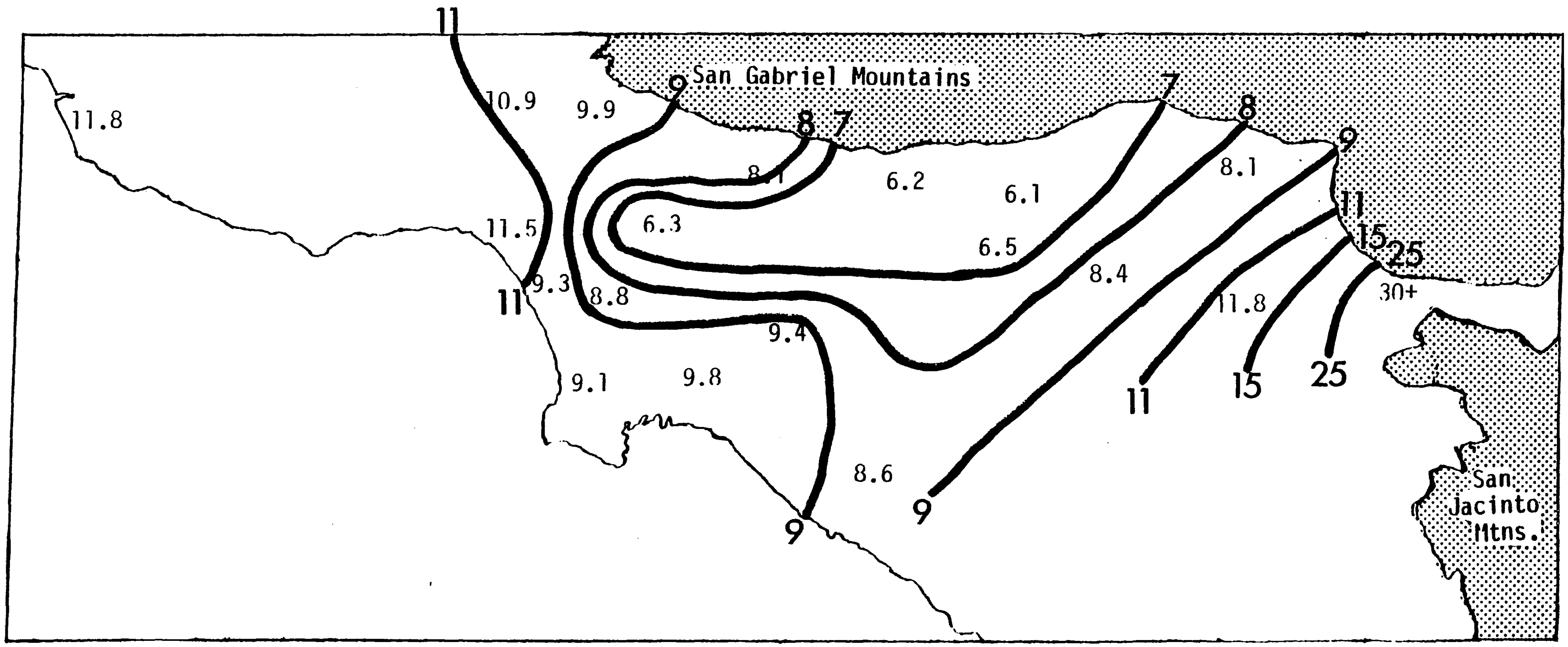

Figure 2.7 Isopleth map of median annual visibility for the Los Angeles area, all hours, 1973-74.

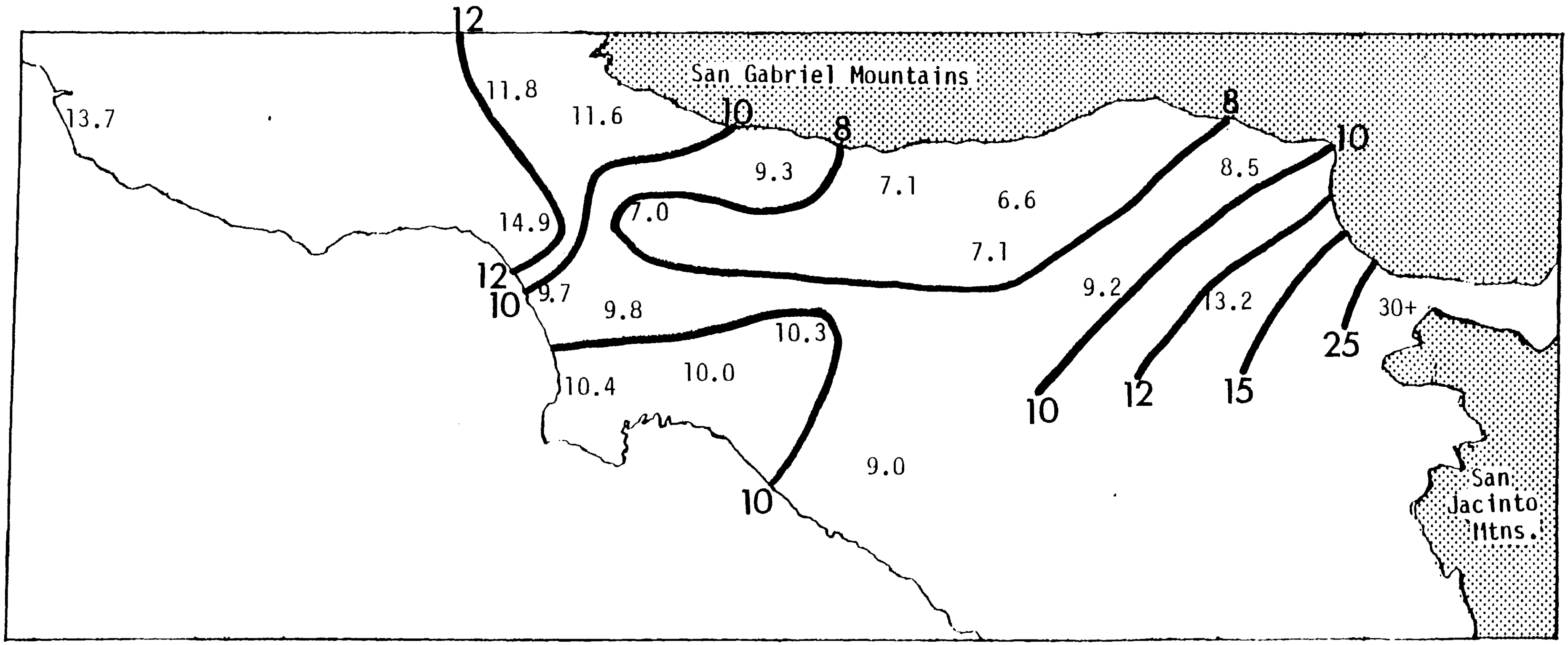

Figure 2.8 Isopleth map of median annual visibility for the Los Angeles area, all hours, excluding precipitation and fog, 1973-74.

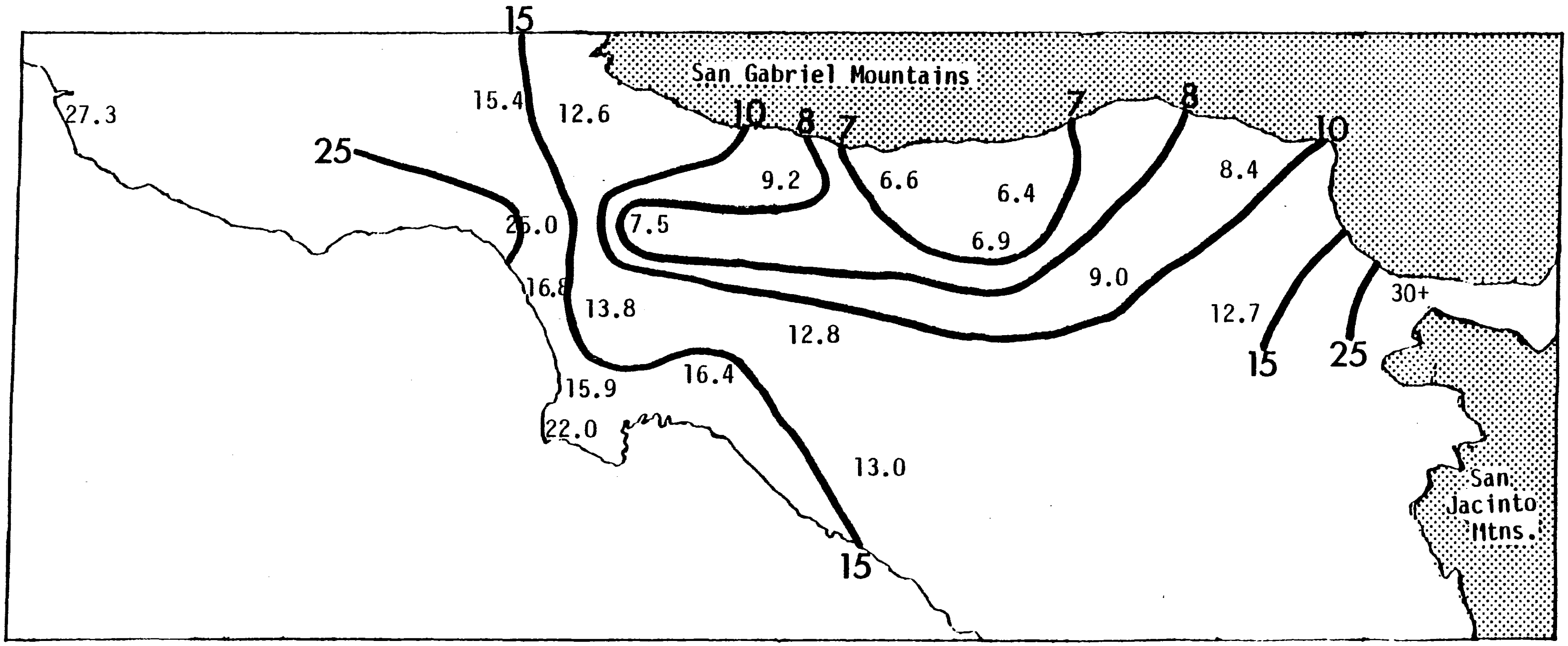

Figure 2.9 Isopleth map of median annual visibility for the Los Angeles area, all hours, sea haze contribution subtracted out, 1973-74.

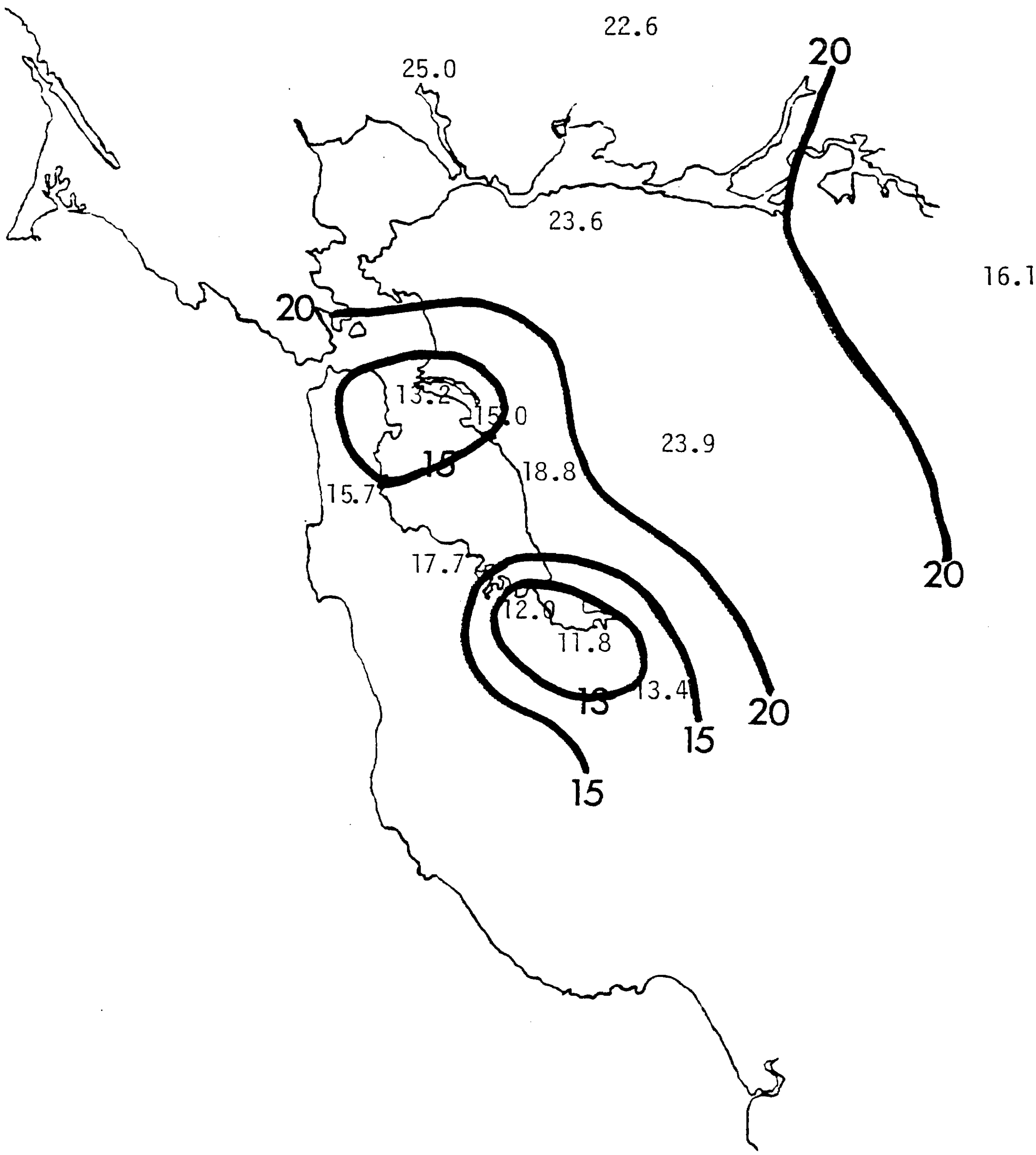

Figure 2.10 Isopleth map of median annual visibility for the San Francisco area, all hours, 1978-79.

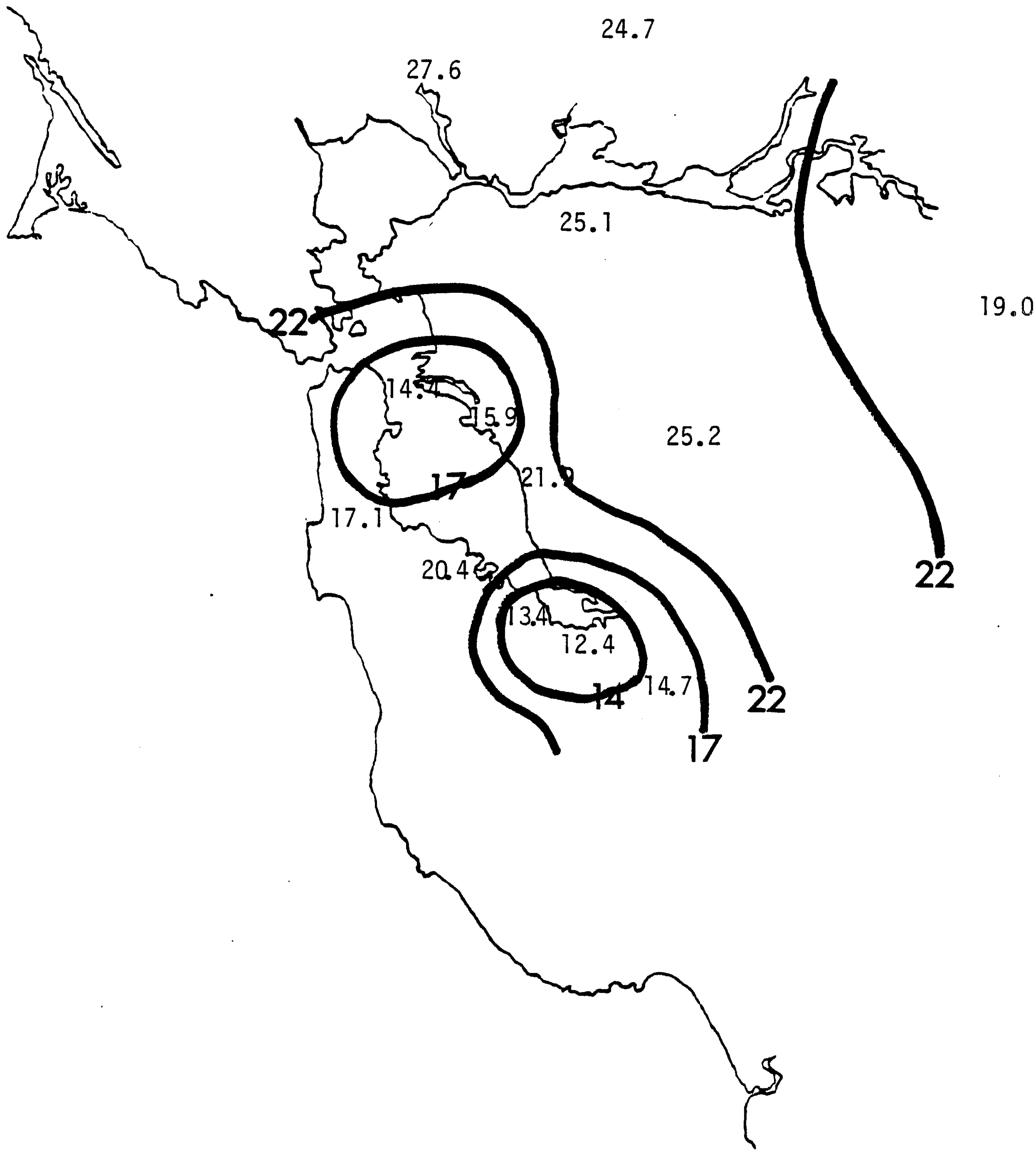

Figure 2.11 Isopleth map of median annual visibility for the San Francisco area, all hours, excluding precipitation and fog, 1978-79.

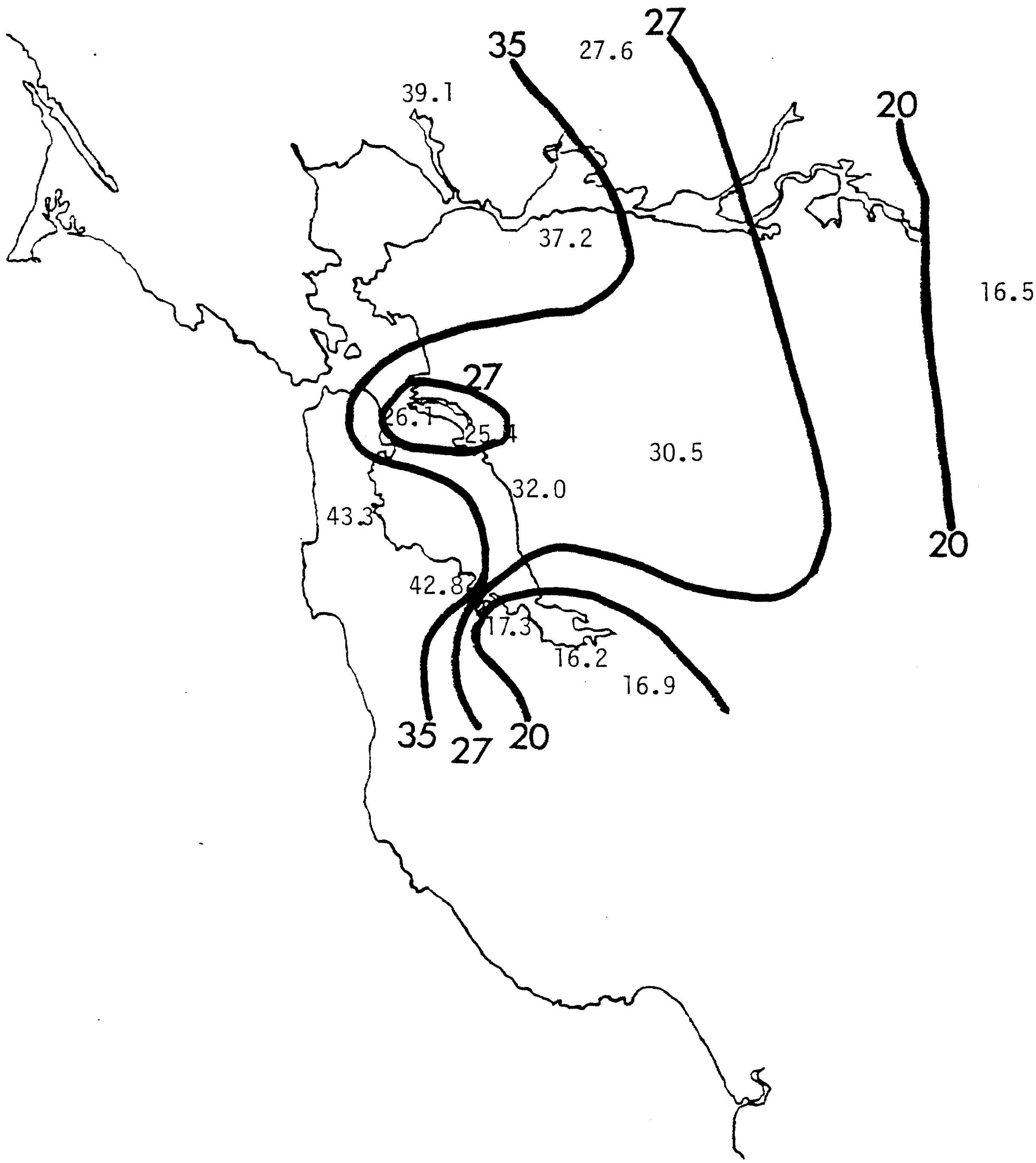

Figure 2.12 Isopleth map of median annual visibility for the San Francisco area, all hours, sea haze contribution subtracted out, 1978-79.

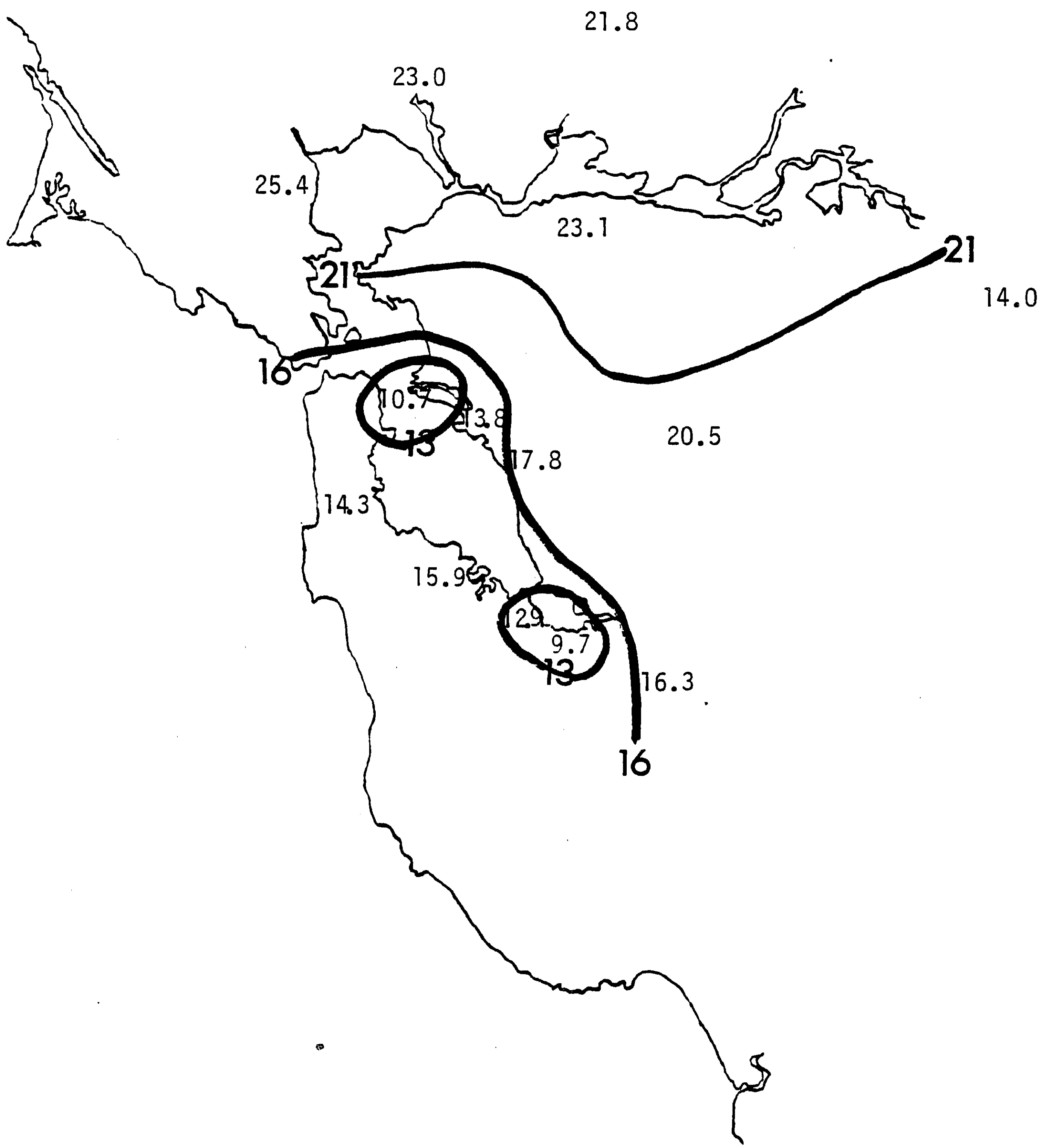

Figure 2.13 Isopleth map of median annual visibility for the San Francisco area, all hours, 1973-74.

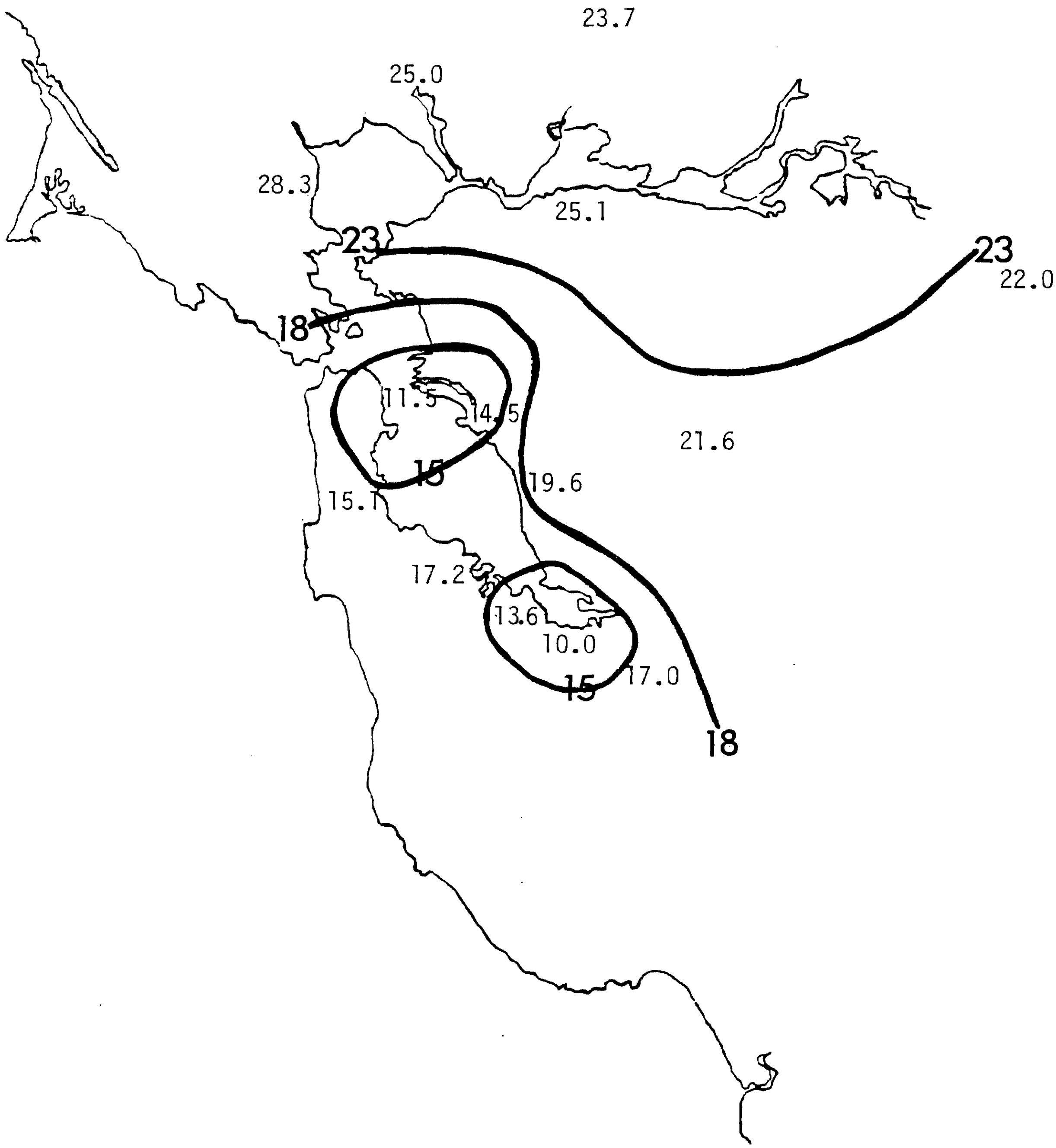

Figure 2.14 Isopleth map of median annual visibility for the San Francisco area, all hours, excluding precipitation and fog, 1973-74.

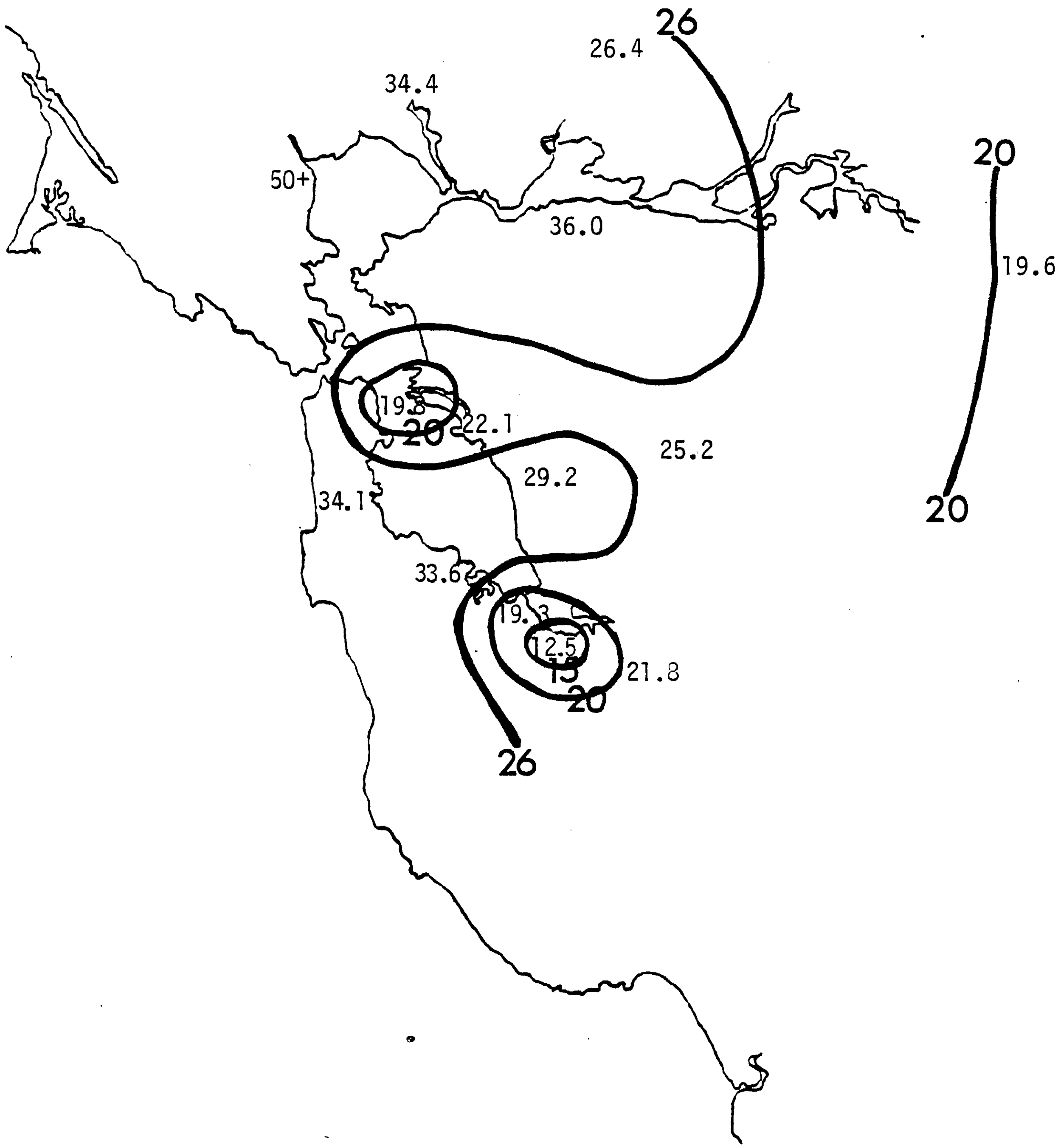

Figure 2.15 Isopleth map of median annual visibility for the San Francisco area, all hours, sea haze contribution subtracted out, 1973-74.

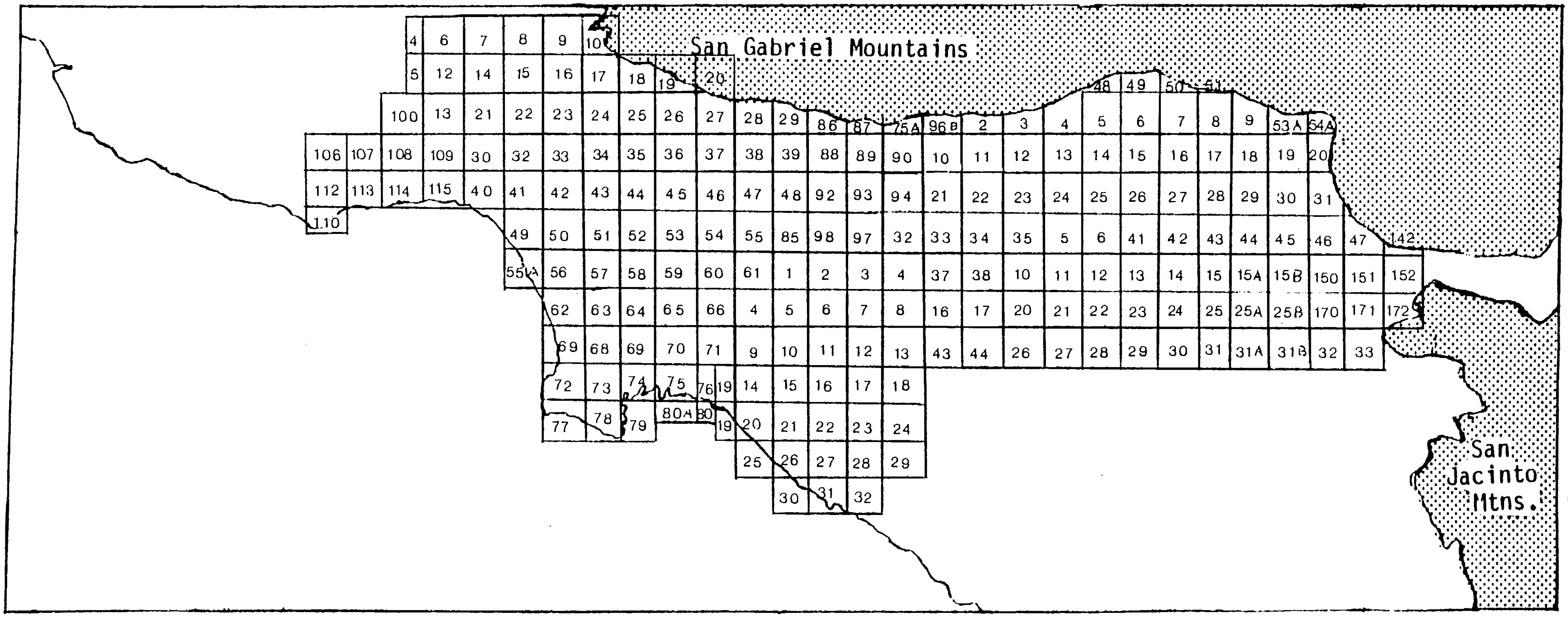

Figure 2.16 Thomas Brothers grid map system for the Los Angeles study region.

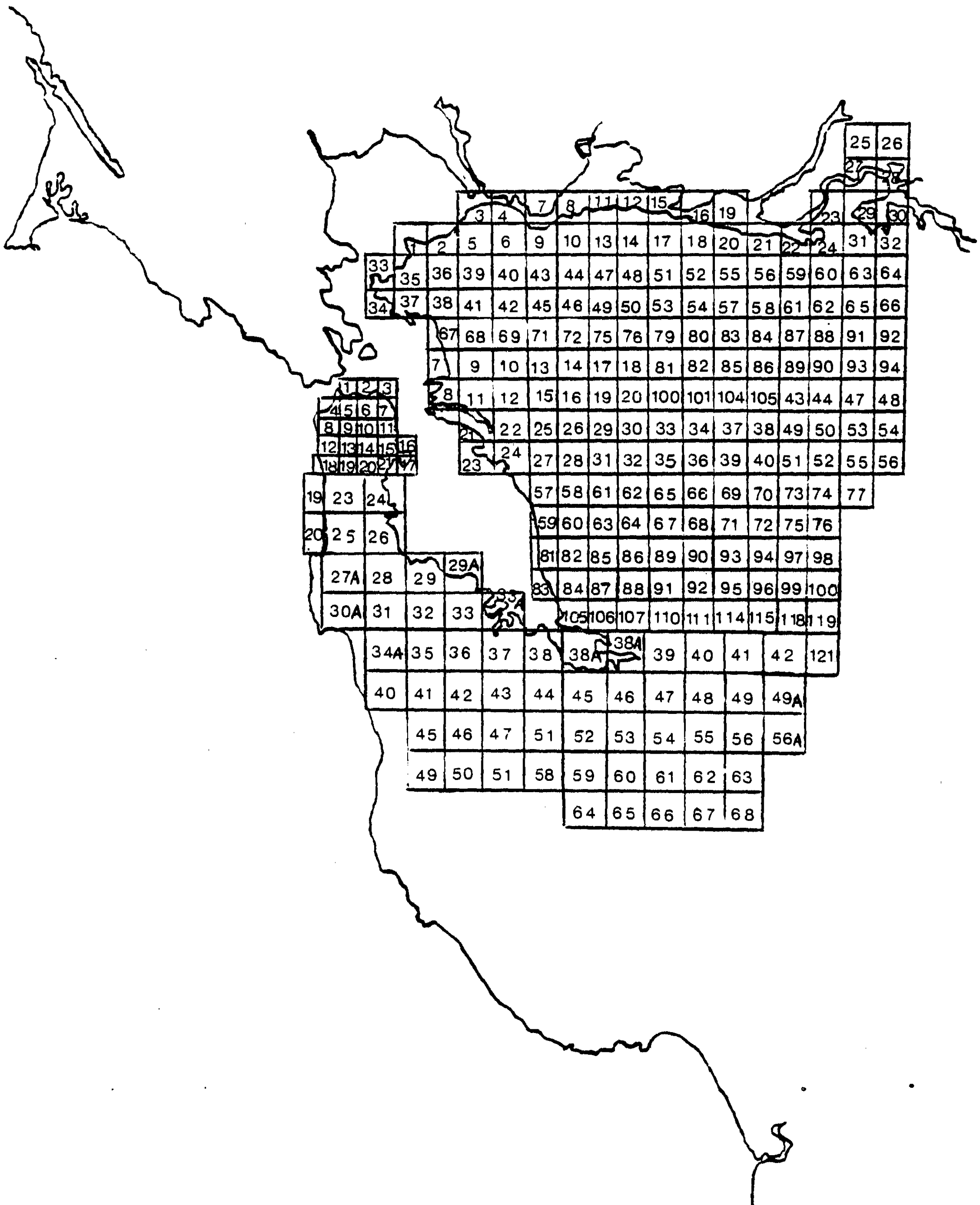

Figure 2.17 Thomas Brothers grid map system for the San Francisco study region.

that visibility might be valued differently in hilly areas than in flat areas). The housing value regressions were run with data from all the grid squares as well as with data only from the most certain grid squares. Little difference was found between the two sets of results. Only the regression runs using all the grid squares are reported in the next chapter.

2.4 REFERENCES

Allard, D. and I. Tombach, "Evaluation of Visibility Measurement Methods in Eastern United States," Presented at the 73rd Annual Meeting of the Air Pollution Control Association, Montreal, June 22-27, 1980.

Husar, R. B. et al., "Trends of Eastern U. S. Haziness Since 1948," Presented at the 4th Symposium on Atmospheric Turbulence, Diffusion and Air Pollution, Reno, Nevada, January 15-18, 1979.

Malm, Wm. C., "Visibility: A Physical Perspective," Proceedings of the Workshop in Visibility Values, Fort Collins, Colorado, January 28-February 1, 1979.

Nochumson, D. et al., "Potential Future Impacts on Visual Air Quality for Class I Areas," Presented at the National Conference on Applied Techniques for Analysis and Management of the Visual Resource, Incline Village, Nevada, April 23-25, 1979.

Trijonis, J., "Visibility in the Southwest - An Exploration of the Historical Data Base," Atmospheric Environment, Vol. 13, pp. 833-843, 1979.

Trijonis, J., "Visibility in California," Prepared under contract A7-181-30 for the California Air Resources Board, 1980.

Trijonis, J., "Visibility in California," Journal of the Air Pollution Control Association, February, 1982.

Trijonis, J. et al., "Analysis of Visibility/Aerosol Relationships and Visibility Modeling/Monitoring Alternatives for California," Prepared under contract #A9-103-31 for the California Air Resources Board, 1982.

Trijonis, J. and D. Shapland, "Existing Visibility Levels in the U. S., Isopleth Maps of Visibility in Suburban/Nonurban Areas During 1974-1976," EPA-450/5-79-010, 1979.

Trijonis, J. and K. Yuan, "Visibility in the Northeast: Long-Term Visibility Trends and Visibility/Pollutant Relationships," EPA-600/3-78-075, 1978.

Williamson, S. J., Fundamentals of Air Pollution, Addison Wesley, Reading, Massachusetts, 1973.

3.0 ESTIMATION OF BENEFITS FROM VISIBILITY IMPROVEMENT

The objective of this chapter is to attempt to place a monetary value on improvements in visibility (decreases in light extinction). The hedonic housing value approach is the methodology employed. The major results of this inquiry can be summarized as follows:

- In the hedonic price gradient estimation, the light extinction variable is a significant negative determinant of home sale price. This result holds in both air basins and for various sample sizes, functional forms, and light extinction measures.
- The monetary impact of a hypothetical ten percent change in extinction ranges from approximately .7 - 2.1 percent of home sale price in the Los Angeles Area and from 1.4 - 2.5 percent of home sale price in the San Francisco Area. The value within an air basin is primarily dependent upon functional form, consistent with the results of Bender et al. (1980).
- The inverse demand curve estimation which relates income and existing extinction levels to marginal willingness to pay (price) for extinction reductions indicates that both income and existing extinction levels are significant determinants of willingness to pay for decreases in extinction levels. In addition, these two variables explain a large proportion of the variation in this willingness to pay.
- The inverse demand results for the Los Angeles Area are somewhat surprising in that they apparently indicate non-convex preference patterns. The sign of the extinction variable seems to suggest that Los Angeles residents have a willingness to pay that is smaller for initial incremental improvements relative to subsequent incremental improvements. This would be contrary to standard economic theory but consistent with a growing volume of literature concerning environmental goods (see Crocker, 1981 for a summary). However, because the hedonic equations are non-linear, the sign of the pollution variables in the inverse demand curves cannot be predicted a priori (Bartik and Smith, 1984). Thus, the negative sign in the Los Angeles Area cannot really be interpreted as demonstrating abnormal convexity. it should be noted that the San Francisco Area results indicate consistency with more traditional economic theory. But again, no conclusive interpretation can be attached.

- Benefit estimates based on the inverse demand curves are variable, depending upon the functional form of the hedonic price gradient. This result is consistent with the Bender et al. (1980) findings. However, the benefit estimates are not dependent upon the functional form of the inverse demand curve. Given this background, a hypothetical 10 percent improvement in visibility would generate benefits of between 250 and 620 million dollars annually in the Los Angeles Area and between 190 and 220 million dollars annually in the San Francisco Area.
- The use of multiple markets adjusts the benefit estimates in the manner anticipated. For instance, adding San Francisco Area households into an analysis of the Los Angeles Air Basin increases the benefit estimates since San Francisco Area households seem to have a greater aversion to air pollution.

The remainder of this chapter is organized as follows. The hedonic housing value method is reviewed in Section 3.1. This section is somewhat technical and may be difficult for the lay person. However, it is a separate entity and the reader may neglect this material without a significant loss of understanding. Section 3.2 presents and discusses the housing data utilized in the estimation of the hedonic equations. The empirical results for the Los Angeles and San Francisco Areas are presented in Section 3.3. Section 3.4 describes the demand estimation procedures and results. Section 3.5 offers concluding remarks.

3.1 METHODOLOGICAL REVIEW

The benefits of improved visibility estimated herein employ a methodology derived from ideas originally proposed by A. Myrick Freeman (1974, 1979a, and 1979b) and Sherwin Rosen (1974). Their approach, referred to here as the Freeman-Rosen (F-R) technique, facilitates the identification of demand curves for commodities which are not normally traded in markets. Despite numerous professional comments (especially those appearing in the Review of Economics and Statistics), the basic framework of the F-R technique has become accepted by economists and applied to a wide range of problems. Several researchers have used this technique to estimate the benefits (whether marginal or total) of changes in various environmental commodities, among which are air pollution (Brookshire et al., 1982; Harrison and Rubinfeld, 1978; and Nelson, 1979) and shoreline (Brown and Pollakowski, 1977), indicating its applicability in the field of environmental economics.

However, the F-R technique has recently been criticized for being inappropriate under some very general conditions. Through the work of Brown and Rosen

(1982), Mendelsohn (1980), Palmquist (1981), and Quigley (1982), it has become increasingly clear that implementation of the F-R approach requires more assumptions and/or data than originally anticipated by Freeman and Rosen.

The purpose of this section is to review the F-R technique and reconcile it with these recent criticisms so that the methodology used in obtaining the benefit estimates reported in Section 3.4 can be completely understood. The major conclusion of the present chapter is that the benefits from improving visibility in the Los Angeles and San Francisco Areas can be estimated using the F-R framework and a couple of additional assumptions. Although more restrictive, these assumptions do not make the F-R technique unrealistically abstract or unusable. To the contrary, our benefit estimates correspond closely to our a priori expectations, which were based on previous work in this area.

3.1.1 The Freeman-Rosen Model

The fundamental importance of the F-R model is that it provides a methodology for estimating demands for the characteristics of certain commodities. For example, an automobile can be described by various characteristics, such as color, number of doors, type of seats, etc. The F-R methodology could, in theory, be used to determine the demand for, say, doors (e.g. two or four) on an automobile. Likewise, the technique can be used to determine the demands for the differing characteristics of homes. It is this application that we consider below, and, because our concern is with the environmental quality characteristics, much of the discussion focuses on it.

The key elements of the F-R model, as applied to housing markets, can be examined using the following notation. Let:

P = the price of housing.

S = a vector of site specific characteristics of homes. For example, living area, number of bathrooms, and the age of the home would be represented in S.

N = a vector of neighborhood characteristics of the home. These include, for example, age of the surrounding population, locational parameters, public services, and racial make-up.

E = the environmental quality associated with the home. For our purposes, E is light extinction (inverse of visual range).

X = a composite commodity. The variable represents consumption on all goods and services except housing. The price of X is set equal to one for simplicity.

Y = Income.

The measures in S, N, and E completely describe the housing prices provided by homes and therefore determine P for each unit. More formally, this relationship,

$$P = P(S, N, E), \tag{3-1}$$

is called the hedonic price function, and it is assumed to be continuous and twice differentiable. Since S, N, E, and P are observable during market transactions, Equation (3-1) is theoretically observable as well. Unfortunately, there are no clues to the shape of this function, requiring that its functional form be determined statistically through some type of estimation procedure. It is, however, improbable that the function will be linear in all of its arguments. This would imply, for example, that a home with 2000 square feet of living area would always be worth a certain amount more than one with 1000 square feet, an unlikely situation.

Equation (3-1) determines the total cost of a bundle of attributes represented by S, N, and E. The marginal cost due to an additional amount of some characteristic (e.g., E) is $P_E = \delta P/\delta E$. P_E is referred to as the implicit price of E or the hedonic price of E. A simple example will help clarify why P_E is, in fact, the implicit price of additional units of E. Imagine that P represents the total cost of a shopping basket containing various items represented by S, N, and E. If one of the items is soup, then we can calculate the change in P (the total cost) due to an additional container of soup, holding constant the other items in the basket. Obviously this is the same as the price of an additional container of soup. Similarly, P_E is the price of additional units of environmental quality.

In as much as Equation (3-1) is observed (or estimated) from data accumulated during market transactions, the implicit prices can be calculated. Thus, the existence of the hedonic price function necessarily implies that the implicit prices for the characteristics can be obtained. This is a major aspect of the F-R framework.

Next consider a consumer whose preferences over housing characteristics and other goods are represented by the following utility function:

$$U = U(X, S, N, E) \tag{3-2}$$

The behavior of the consumer is characterized by maximizing (3-2) subject to a budget constraint:

$$\text{Maximize: } U = U(X, S, N, E)$$

$$\text{Subject to: } Y = X + P(S, N, E).$$

The first order necessary conditions for utility maximization yield

$$U_E/U_X = MRS_{EX} = P_E \tag{3-3}$$

where subscripts denote partial differentiation. The implicit prices reveal marginal rates of substitution (MRS), a fundamental result of the F-R model, especially since E is a public good.

Now define W as the amount an individual is willing to pay for alternative amounts of S, N, and E given a level of satisfaction and some amount of income. W is an implicit function defined by:

$$U(Y-W, S, N, E) = \bar{U} \tag{3-4}$$

where $\bar{U}$ is arbitrarily fixed. Thus,

$$W = W(S, N, E, Y, \bar{U}). \tag{3-5}$$

The marginal willingness to pay for some characteristic (say E) is $\delta W/\delta E = W_E$ and

$$W_E = f(S, N, E, \bar{U}, Y) \tag{3-6}$$

is the consumer's compensated (inverse) demand curve for E. The outstanding feature of the F-R model is that in equilibrium, $W_E = P_E = MRS_{EX}$. P_E reveals the consumer's marginal willingness to pay for E, given the other characteristics, utility, and income. Moreover, data can be obtained for all the variables (except, of course, $\bar{U}$) in the equation. Under what conditions then, can (3-6) be identified empirically?

Following Freeman (1979) and Harrison and Rubinfeld (1978), a fair assumption is that the supply of E is exogenous or fixed, particularly in the short run. Given this, and a nonlinear Equation (3-1), there is variation in the price ($P_E = W_E$) and quantity (E) data, and applying ordinary least squares to Equation (3-6) should identify the inverse demand curve for E. On the other hand, if the supply of E cannot reasonably be assumed to be independent of P_E, then the demand and supply relationships should be estimated jointly (Nelson, 1979). In light of the fact that E is mainly determined by exogenous influences such as topography and wind patterns, we have chosen to ignore the supply side of E.

It appears as though the F-R model does provide a workable framework for estimating the benefits from discrete changes in E. An estimated version of (3-6) would be an ordinary inverse demand curve, even though the theory suggests that (3-6) is the utility compensated demand curve. This is because observations on $\bar{U}$ are not generally available, meaning that there is no way to empirically hold utility constant. If, however, the utility function is known a priori, or assumed, then compensated demand curves can be estimated. Quigley (1982) assumed a generalized utility function with constant elasticity of substitution and was able to identify compensated demand curves. In practice, when an ordinary demand curve is estimated, benefits are calculated as changes in willingness to pay and, when compensated demand curves are estimated, benefits are calculated as the measure of compensating variation. The difference between the two will be minor as long as the income elasticity is relatively small and the ratio of the consumer's surplus to income is small (Willig, 1976). To the extent that E is a relatively minor item for most individuals, the distinction between willingness to pay and compensating variation (or, for that matter, between ordinary and compensated demand curves) can be ignored.

As is the case whenever demand curves are estimated, it is necessary to assume that all individuals in the market are identical except for income and measurable taste shift parameters. The shift parameters are usually socio-economic variables such as education, sex, race, age, and political beliefs. Below, we have assumed that individuals within a market are identical except for differences in income levels.

3.1.2 Recent Criticisms and Comments

Brown and Rosen (1982), Mendelsohn (1980), Palmquist (1981), and Quigley (1982) have all illustrated a basic flaw in the F-R technique. Their argument is that the implicit prices (e.g., P_E) are endogenous in the model, rather than given to consumers. As a matter of fact, the consumer actually chooses P_E when making his locational decision. To see this, assume that the hedonic price function depends on only three arguments represented by S, N, and E. In general, then, implementation of the F-R approach requires the estimation of the following equations.

$$P = P(S, N, E)$$

$$P_E = f(S, N, E, Y)$$

$$P_S = g(S, N, E, Y)$$

$$P_N = h(S, N, E, Y)$$

where subscripts again denote partial derivatives. Since P_E, P_S, and P_N are deterministic functions of S, N, and E (according to the hedonic price function), it is impossible to estimate f, g, and h. Only when P_E, P_S, and P_N are exogenous will any new information be gained by estimating f, g, and h. Since this point is crucial to the implementation of the F-R technique, some additional comments and suggestions are warranted.

Following Mendelsohn, within a market (e.g., an SMSA) all individuals face the same set of prices for the characteristic under consideration. The price set given by P_E represents the array of prices faced by individuals when choosing optimal levels of E. The implication of this can be realized by comparing two individuals, A and B. Individual A chooses a different level of E than does B, only if his demand for E is different than B's (perhaps due to different income or tastes). Their quantity choices are not different because of differences in P_E. It seems as though the observed data reveal information about how different individuals respond to the same set of prices, rather than the desired situation of identical individuals responding to different prices. In essence, the data give us one point on each demand curve which, without some additional structure, is not enough information to estimate the shape of the underlying relationship between price and quantity.

Figure 3.1 visually highlights this issue. In the figure, P_E^1 is the implicit price set faced by all individuals in the market and W_E^A and W_E^B are the demand curves for A and B, respectively. The demand curves illustrate that different individuals choose different levels of E and, therefore a different P_E. The information revealed by the F-R approach is (P_1^A, E_1^A) and (P_1^B, E_1^B). Now, unfortunately, exactly the same information is revealed by the demand curves $\bar{W}_E^A$ and $\bar{W}_E^B$. And, in general, there will be no way to discern whether the shape implied by $\bar{W}_E^A$ and $\bar{W}_E^B$ is correct or the shape of $\bar{W}_E^A$ and $\bar{W}_E^B$ is the appropriate representation of reality.

Brown and Rosen have examined the econometric implications of the endogenicity of the implicit prices in greater detail. Say that the hedonic price function is estimated as the following polynomial:

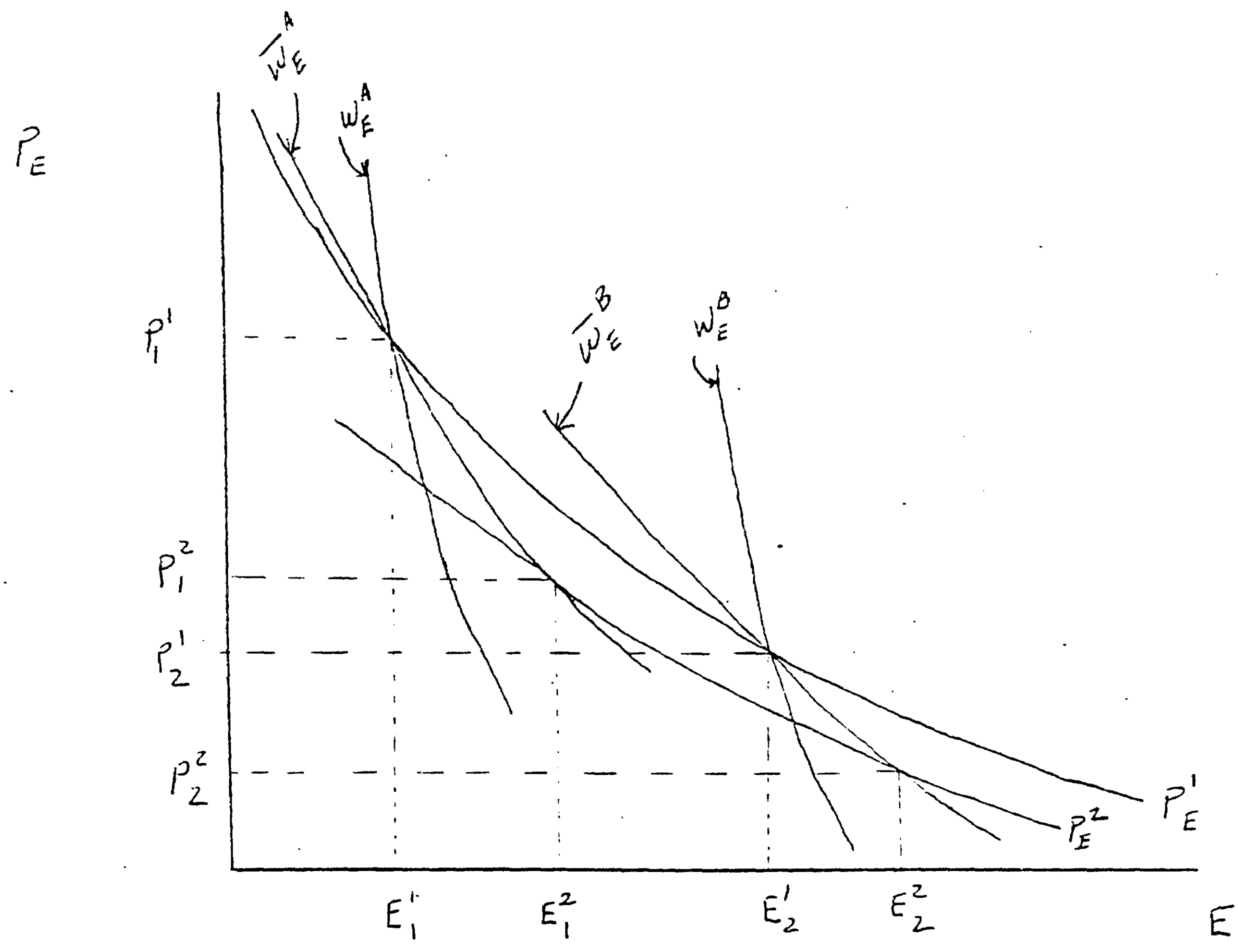

Figure 3.1 Alternate environmental demand curves for two individuals.

$$P = a_0 + a_1N + a_2S + a_3E + a_4E^2 \tag{3-7}$$

Then

$$P_E = a_3 + 2a_4E \ . \tag{3-8}$$

If we try to estimate a demand equation that is linear in E, say

$$W_E = B_0 + B_1E + B_2Y \tag{3-9}$$

we would find that the R-square is one, $B_0 = a_3$, $B_1 = 2a_4$, and B_2 is insignificant. Clearly, demand estimation cannot reveal any additional information other than that contained in (3-7) in this case. At first this may seem to be a fatal blow to the F-R technique; however, when the demand equation is not a deterministic function of the hedonic price equation, then estimation is still possible. The problem is a drawback, though, since it requires researchers to assume away the problem and avoid these situations.

Mendelsohn and Palmquist suggest that a way to overcome these difficulties may be to use hedonic price functions from several different markets. The effect of this is to add additional price sets into the problem and obtain information on how like individuals respond to different price sets. An example is illustrated by P_E^2 in Figure 3.1. P_E^2 is the set of implicit prices calculated from another market; with the additional information denoted by (P_1^2, E_1^2) and (P_2^2, E_2^2) it is possible to discover the appropriate price and quantity relationships. (In the figure, $\bar{W}_E^A$ and $\bar{W}_E^B$ reflect the true relationship between price and quantity.) Obviously, more precision will be gained by adding in more and more markets.

The use of multimarket data revolves around two issues; first, we need to determine what, if any, additional assumptions are required for estimation. Then, we need to be able to identify the different markets.

As noted above, demand estimation using the traditional hedonic approach requires the assumption of like preferences for individuals within the market. The multimarket approach requires an assumption of like preferences across markets. For example, individuals in Boston will have the same shaped demand curve for environmental quality as individuals in Los Angeles. This assumption describes how similar people respond to different price sets and, if appropriate, facilitates estimation of demand curves.

In order to identify the different markets, Mendelsohn suggests that a sufficient condition for hedonic functions to vary across markets is "that the underlying array of suppliers changes across the markets." An example would be different supply arrangements induced by building codes and realtor boards. Another sufficient condition noted by Mendelsohn is, "if the number of demanders in a market is independent of the market prices, the supply curves are not perfectly elastic, and the number of demanders varies across markets." This can result when the transportation costs between markets prohibit consumers from locating in either area (Palmquist). Therefore, we have some guidelines on defining different markets within a geographic region, which, coupled with the assumption of identical individuals across markets, enable us to implement the multimarket approach.

An alternative to the Mendelsohn and Palmquist suggestion is the approach taken by Harrison and Rubinfeld (1978) and formally suggested by Quigley (1982). In this case, we eliminate the endogenous nature of the implicit prices by taking the hedonic price function as given (or determined in a prior step) and use the nonlinear budget constraint to empirically determine preferences. The choice of these two procedures is examined below.

3.1.3 Assumption Choices

In order to implement the F-R technique, two different assumption sets can be imposed. The first requires the use of multimarket data, while the second, although operational with data from a single market, requires restrictions on the functional forms of the hedonic price equation and the demand curves. Even though there are good and bad points about each set of assumptions, we have concluded that the estimates from single market data will be more appropriate in this study. There were several reasons for this conclusion.

Foremost in our reasoning was the requirement that individual preferences needed to be identical across markets in order to use multimarket data. Indeed, since people tend to gravitate toward others who have similar preferences, we expect San Franciscans to be similar and Los Angeles Area residents to be similar, but there is no reason to suspect that the two groups are similar to each other. There does not seem to be an empirical test of this hypothesis, although we have found that hedonic price equations and the demand equations are different across the two areas. Thus, our initial empirical investigations seemed inconsistent with the multimarket approach.

Another consideration was the amount of data required for multimarket estimation. We have two viable markets but doubt that two is a sufficient number to adequately implement Mendelsohn's suggestion. More confidence could be achieved by obtaining data from more markets. Perhaps San Diego, Portland,

and Seattle would be reasonable choices to combine with Los Angeles and San Francisco.

As noted above, Brown and Rosen have illustrated that some functional form combinations must be eliminated. This type of assumption may not be overly restrictive. For example, if the demand curve is going to be linear in E, the only forms ruled out for the hedonic price equation are polynomials of degree two in E. Basically, this is what we have done below.

The approach used herein is essentially the same as that taken by Quigley. We assume that the individuals within each market are identical with the exception of income and that they take as given the nonlinear hedonic price function. Then, since the budget constraint is nonlinear, differences in income are sufficient to identify consumer's preferences. For a graphical presentation of this, see Quigley's Figure 1. Harrison and Rubinfeld appear to have used the same approach as well.

3.1.4 Methodological Summary

The following is a formal categorization of the procedures followed in obtaining the estimated benefits from changes in levels of visibility.

1. Determine the appropriate set of variables to be entered into the hedonic price equation. Included in this evaluation is the suitable measure of visibility.
2. Identify the functional form of the hedonic price equation. This can be very important -- several others have found benefit estimates to be sensitive to functional form (Bender et al., 1980).
3. Assume that all individuals within a market are similar except for income and that visibility (or extinction) is neither a substitute for nor a complement with other characteristics. Thus demand is given by

$$W_E = f(E, Y).$$

4. Assume that the demand equation is not a deterministic function of the hedonic price equation. Thus, all variation in P_E is not due entirely to differences in E.
5. Calculate P_E for each community and regress E and Y on these calculations. (The choice of community as the individual is arbitrary. See below.)
6. Use the results from (5) to calculate the benefits of various programs.

Implementation of this methodology is described in the remaining sections of this chapter.

3.2 DATA SPECIFICS: HEDONIC HOUSING EQUATIONS

The initial procedural step of the hedonic housing value approach is to estimate a hedonic housing equation which relates home sale price to the attributes of the home. Of particular interest is the relationship of home sale price to light extinction levels. The estimation is particularly concerned with testing the hypothesis of whether or not extinction levels are a significant determinant of home prices.

The study areas are within the South Coast and Bay Area Air Basins. Included in the first study area are Los Angeles, Orange, Riverside, and San Bernardino counties. The latter study area consists of San Francisco, Alameda, Contra Costa, San Mateo, and Santa Clara counties. The analysis is specifically confined to single family residences in these areas. Thus, not considered is the impact of visibility variations upon other structures (multiple family dwellings, mobile homes, commercial, etc.) or other ownership types (rental leasing, etc.). Therefore, within our sample, this research asks if households will pay a premium in price for single family homes located in clean air areas and what is the magnitude of that willingness to pay.

The data base was constructed to enable the testing of hypotheses concerning the impact of extinction levels on housing sale price. The dependent variable in the entire analysis is the sale price of owner occupied single family residences.* The independent variable set consists of variables which correspond to three levels of aggregation: house, neighborhood, and community. Table 3.1 describes further the data employed in the study.

The housing characteristics data, obtained from the Market Data Center (a computerized appraisal service centered in Los Angeles), pertain to homes sold in the 1978-79 time period and contain information on nearly every important structural and/or quality attribute.** Included in the list of available variables are those that pertain to both quantity (lot size, total number of rooms, square footage of living area) and quality (pool, view, number of fireplaces, parking,

---

*The sale price or the discounted value of the flow of rents rather than actual rent is used as the dependent variable. The two are interchangeable given the appropriate discount rate.

**The 1978-79 time period was chosen because it represented the latest years that housing data were available. It should be noted that this project was initially designed to also examine the 1973-74 time period. This effort constituted an attempt to examine preference shifts over time. However, the assumptions necessary to allow such a comparison were so restrictive as to make the effort meaningless. Therefore, the 1973-74 analysis is relegated to Appendix A.

TABLE 3.1 VARIABLES USED IN ANALYSIS OF HOUSING MARKET FOR 1978-79.

Variable	Definition (hypothesized effect on on housing sale price)	Unit	Source
Dependent:			
Sale Price	Sale price of owner occupied single family residences	($100)	Market Data Center
Independent-Housing:			
Sale Date	Month the home was sold (positive)	January 1978 = 1 December 1979 = 24	Market Data Center
Age	Age of home (negative)	Years	Market Data Center
Bathrooms	Number of bathrooms (positive)	Number	Market Data Center
Living Area	Square feet of living area (positive)	Square Feet	Market Data Center
Pool	1 if pool, 0 if no pool (positive)	0 = no pool 1 = pool	Market Data Center
Fireplaces	Number of fireplaces (positive)	Number	Market Data Center
View	1 if view present, 0 if not (positive)	0 = no view 1 = view	Market Data Center
Independent-Neighborhood:			
Distance to Beach	Miles to nearest beach (negative)	Miles	Calculated
Age Composition	Percent greater than 62 in census tract (positive)	Percent	1980 Census
Ethnic Composition	Percent white in census tract (positive)	Percent	1980 Census
Time to Work	Average time to employment from census tract (negative)	Minutes	1980 Census

TABLE 3.1 (Continued).

Variable	Definition (hypothesized effect on on housing sale price)	Unit	Source
Independent-Community:			
School Quality	Community's 12th grade math score (positive)	Percent	California Assessment Program (1979)
Population Density	Population per square mile in surrounding community (negative)	Persons/square mile	1980 Census, Thomas Brothers Grid Maps
Miles to Central Business District	Distance from census tract to dominant city in county (negative)	Miles	Thomas Brothers Grid Maps
Crime	Seven major crimes per 1000 people in surrounding communities (negative)	Crimes/persons	Summary Characteristics 1980 Census
Age	Median age of population in surrounding community (positive)	Years	Summary Characteristics 1980 Census
Race	White percentage of population in surrounding community (positive)	Percent	Summary Characteristics 1980 Census
Unemployment	Unemployment rate in surrounding community (negative)	Percent	Summary Characteristics 1980 Census
Education	Percentage of population in community with High School Diploma (positive)	Percent	Summary Characteristics 1980 Census
Poverty	Percentage of population in community below poverty level (negative)	Percent	Summary Characteristics 1980 Census
Home Density	Hundreds of people per square mile (negative)	Homes/square mile	Calculated
Population per Household	Persons per household (negative)	People/Home	Summary Characteristics 1980 Census

TABLE 3.1 (Concluded).

Variable	Definition (hypothesized effect on on housing sale price)	Unit	Source
Independent Air Quality:			
Light Extinction* (1)	Median annual extinction level (neg.)	$(10^4 \text{ meters})^{-1}$	See Chapter 2
Light Extinction* (2)	Median annual extinction level disregarding hours with fog or precipitation (negative)	$(10^4 \text{ meters})^{-1}$	See Chapter 2
Light Extinction* (3)	Median annual extinction subtracting sea haze contribution (negative)	$(10^4 \text{ meters})^{-1}$	See Chapter 2

*Light Extinction = 18.7/visibility, where visibility is in units of [miles].

stories, etc.) of each particular house. This list was pared to those variables presented in Table 3.1 in order to reduce collinearity problems. But note that both home quantity and quality are covered by the variables chosen.

It should be emphasized that housing data of such quality (e.g., micro level of detail over time) are rarely available for studies of this nature. Usually outdated data which are overly aggregated and collected irregularly (for instance census tract averages only in census years) are employed. Our data sets yield results relevant at the household (micro) level.

The Market Data Center provided data tapes listing all homes sold in the counties specified above during the 1978-79 time period. The number of entries was unmanageably large (in excess of 100,000 observations), so the data sets were reduced using a random number matching system. The selection criteria satisfied a desire to maintain: (1) a large data set (greater than 3000 observations), and (2) the relative proportions of homes sold in the counties. Table 3.2 lists the number of observations from each county, illustrating that 4765 and 3106 home sales were included in the Los Angeles and San Francisco Areas, respectively.

In addition to the immediate characteristics of a home, other variables which could significantly affect its sale price are those that reflect the condition of the neighborhood and community in which it is located. In order to capture those impacts and to isolate the independent influence of location vis-a-vis extinction differences, several neighborhood and community variables were included in the econometric modeling.

Neighborhood refers to the surrounding census tract and includes the variables -- population, age, ethnic composition, distance to work, and distance to the beach. Given the large number of census tracts (for example over 1500 in the Los Angeles Area) variations in this data are quite substantial. Pertinent community (city level) variables include density measure, school quality, crime rate, and others. However, in contrast to the house and neighborhood characteristics, there are only a limited number of communities. Thus, collinearity between community measures presents empirical difficulties (see following section).

The neighborhood and community data were matched to the household characteristic data using the transformation from Thomas Brothers grid maps to the relevant census tracts and communities. Thus, each household was matched with its corresponding neighborhood and community characteristics.

TABLE 3.2 NUMBER OF OBSERVATIONS FROM EACH COUNTY IN ANALYSIS OF HOUSING MARKET FOR 1978-79.

Los Angeles Area	4765
Los Angeles County	2682
Orange County	1399
Riverside County	187
San Bernardino County	497
San Francisco Area	3106
San Francisco County	580
Alameda County	908
Contra Costa County	461
San Mateo County	531
Santa Clara County	626

Summary statistics for the variables used in the hedonic housing equation estimation are presented in Table 3.3 (Los Angeles Area) and Table 3.4 (San Francisco Area). Most of the figures in the tables are self explanatory and require no further discussion. However, of special interest in this project are the light extinction variables. As is evident through a comparison of Tables 3.3 and 3.4, the San Francisco Area possesses significantly better average visibility than does the Los Angeles Area. Further, the range of the San Francisco Area extinction data (a factor of 2) is small compared to the range of the Los Angeles Area extinction data (a factor of 5). The Los Angeles Area would seem to be a better study area because the variations in the data are much greater.

A final point on the extinction data concerns the relationship between the various measures. The first and second extinction variables (annual median and annual median excluding hours of precipitation or fog) are nearly equivalent. Their means and standard deviations are quite close. Furthermore, the simple correlation between these measures is 0.98 in each study area. Therefore, in the empirical analysis that follows, the results for these two variables are treated as interchangeable. However, this is not the case with the relationship between the third extinction variable (annual median subtracting sea haze) and either of the other measures. Thus, extinction (3) is treated as a parameter which measures something different from extinction (1) or extinction (2). When either of these latter two variables are used it is just labelled extinction. Only with the third extinction variable is a distinction made by adding the parenthetic identification.

The data base assembled for the housing value study is appropriate to test the hypothesis outlined above for two reasons. First, the housing characteristic data are extremely detailed at the household level of aggregation and extensive in that a relatively large number of observations are considered. Second, a variety of neighborhood and community variables have been included to help isolate the specific effect of light extinction on housing values.

3.3 EMPIRICAL RESULTS: HEDONIC HOUSING EQUATIONS

The initial task in the hedonic housing value analysis is to determine the relationship between light extinction levels and home sale price. The underlying structure of this hypothesis test is an empirical equation which attempts to explain the variation in home prices located in the Los Angeles and San Francisco Areas for the years 1978-1979.* The estimated coefficients

*See Freeman (1979) and Maler (1979) for reviews of estimation techniques for hedonic housing equations.

TABLE 3.3 SUMMARY STATISTICS FOR VARIABLES USED IN ANALYSIS OF HOUSING MARKET FOR THE LOS ANGELES AREA.

	Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Housing	Home Sale Price	93060.00	60508.00	9000.00	725000.00
	Sale Date	11.63	6.45	1.00	24.00
	Age of Home	20.61	16.15	1.00	77.00
	Bathrooms	1.86	0.42	0.50	7.50
	Living Area	1524.00	608.90	371.00	9942.00
	Pool	0.13	0.34	0.00	1.00
	Fireplaces	0.72	0.61	0.00	5.00
	View	0.078	0.27	0.00	1.00
Neighborhood	Distance to Beach	14.53	11.99	0.13	59.80
	Age Composition	9.33	5.48	1.30	60.70
	Ethnic Composition	83.77	15.49	2.10	99.10
	Time to Work	23.33	3.40	10.00	38.00
Community	School Quality	66.07	4.56	45.60	81.00
	Population Density	58.65	22.13	1.38	190.00
	Miles to Business District	15.56	7.90	0.00	34.00
	Crime	48.06	13.28	15.47	212.62
	Age	29.87	2.68	22.00	49.00
	Race	75.63	13.93	6.22	97.49
	Unemployment	5.99	1.64	1.60	12.50
	Education	73.74	9.48	30.70	97.00
	Poverty	10.92	4.66	1.90	29.40
	Home Density	10.61	3.53	0.19	26.99
	Population Per Household	2.72	0.30	1.91	4.00
Air Quality	Light Extinction (1)	1.97	0.42	0.62	3.12
	Light Extinction (2)	1.80	0.37	0.62	2.67
	Light Extinction (3)	1.40	0.55	0.62	2.88

TABLE 3.4 SUMMARY STATISTICS FOR VARIABLES USED IN ANALYSIS OF HOUSING MARKET FOR THE SAN FRANCISCO AREA.

	Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Housing	Home Sale Price	89980.00	51410.00	15000.00	850000.00
	Sale Date	11.66	6.33	1.00	23.00
	Age of Home	25.46	18.34	1.00	79.00
	Bathrooms	1.69	0.65	1.00	5.50
	Living Area	1443.90	562.30	600.00	9600.00
	Pool	0.062	0.24	0.00	1.00
	Fireplaces	0.81	0.55	0.00	6.00
	View	0.10	0.31	0.00	1.00
Neighborhood	Distance to Beach	26.22	12.87	6.69	66.22
	Age Composition	10.11	6.47	0.70	84.30
	Ethnic Composition	74.75	21.84	2.70	99.70
	Time to Work	24.15	3.77	14.00	35.00
Community	School Quality	68.12	4.36	52.30	81.00
	Population Density	59.09	45.48	4.33	151.73
	Miles to Business District	11.13	7.64	0.00	27.84
	Crime	45.45	22.44	6.01	119.08
	Age	31.72	3.50	26.30	40.90
	Race	72.24	17.10	38.14	97.53
	Unemployment	6.42	2.24	1.20	12.80
	Education	78.62	6.72	58.70	97.50
	Poverty	9.05	5.03	2.20	21.00
	Home Density	24.49	20.84	1.35	66.81
	Population Per Household	2.58	0.34	1.74	3.49
Air Quality	Light Extinction (1)	1.11	0.19	0.78	1.56
	Light Extinction (2)	1.02	0.16	0.72	1.44
	Light Extinction (3)	0.66	0.20	0.45	1.10

of these hedonic equations represent the effects that changes in the independent variables have on sale price. In reference to the light extinction variable, this procedure allows one to focus on its significance while separating out the influence of other extraneous variables. Therefore, this analysis yields two outputs concerning the relationship of extinction differentials to housing price. The relative significance of location with respect to extinction is determined, and the estimated coefficient pertaining to extinction implicitly measures its monetary value at the margin.

This section is organized as follows. Estimated hedonic price gradients are presented for each of the study areas. The functional forms presented serve as benchmark results. The stability of those benchmark results is then analyzed by altering sample size, extinction measure, and functional form. In the latter case, both the independent variable set (other than extinction) and the mathematical form of the relationship are allowed to vary. The initial hedonic price gradients do not necessarily provide the best statistical fit of the data nor the most suitable relationship for subsequent analyses. Rather, only after all possible influences are analyzed do we choose the most appropriate relationship to utilize in the subsequent steps of the hedonic price method. Finally, it should be noted that the order in which we have chosen to analyze these potential destabilizing influences has no effect on the ultimate choice of the best estimated price gradients. Thus, for instance, sample size has essentially no effect whether it is analyzed first or last.

3.3.1 Benchmark Results

The estimated hedonic price gradients which serve as the base results are presented in Tables 3.5 and 3.6 for the Los Angeles Area and San Francisco Area, respectively. A number of aspects of the equations are worth noting. First, the independent variable set was chosen to account for all the different characteristics of a home. Thus, square feet of living area represents the quantity of a home, whereas pool, house age, and the number of bathrooms and fireplaces describe the quality. In addition, characteristics which reflect the immediate neighborhood (ethnic and age composition) and the location (time to employment and distance to beach) were included. Also, variables such as school quality and population density were included to represent overall community attributes. In this latter category, only a couple of available community variables were used because of collinearity difficulties. Collinearity is especially problematical because only a relatively few communities exist (112 in the Los Angeles Area and 51 in the San Francisco Area). There is insufficient variation to allow the

TABLE 3.5 ESTIMATED HEDONIC EQUATION (LOG-LINEAR)
FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ℓn (Sales Month)	0.110	27.64
ℓn (Age of Home)	-0.018	-5.39
ℓn (Square Feet of Living Area)	0.724	47.78
ℓn (Number of Bathrooms)	0.099	6.83
Number of Fireplaces	0.093	14.21
Pool	0.079	8.21
View	0.161	13.04
Neighborhood Characteristics:		
ℓn (Percent Greater than 64)	0.053	8.11
ℓn (Percent White)	0.269	26.55
Location Characteristics:		
ℓn (Distance to Beach)	-0.122	-30.21
ℓn (Time to Employment)	-0.030	-1.23
Orange County	-0.141	-15.28
Riverside County	-0.303	-14.51
San Bernadino County	-0.173	-11.27
Community Characteristics:		
ℓn (Population Density)	-0.017	-1.85
ℓn (School Quality)	0.498	8.15
ℓn (Light Extinction)	-0.202	-10.03
Constant	-1.65	-5.72
R-Squared	0.792	
Number of Observations	4766	

TABLE 3.6 ESTIMATED HEDONIC EQUATION (SEMI-LOG) FOR THE SAN FRANCISCO AREA.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.015	25.42
Age of Home	-0.0021	-7.59
Square Feet of Living Area	0.00039	35.26
Number of Bathrooms	0.053	5.40
Number of Fireplaces	0.094	11.63
Pool	0.105	6.55
View	0.0698	5.25
Neighborhood Characteristics:		
Percent Greater than 64	0.0043	6.41
Percent White	0.0062	27.72
Location Characteristics:		
Distance to Beach	-0.0067	-10.74
Time to Employment	-0.0036	-2.74
Alameda County	-0.315	-9.67
Contra Costa County	-0.443	-12.49
San Mateo County	-0.203	-6.83
Santa Clara County	-0.165	-5.51
Community Characteristics:		
Population Density	-0.0005	-1.86
School Quality	0.0097	8.23
Light Extinction	-0.228	-4.63
Constant	5.429	40.56
R-Squared	0.794	
Number of Observations	3106	

inclusion of more community variables. However, as is seen below, this problem can be successfully overcome using principal components analysis. The final variables of interest represent counties. These are zero-one dichotomous variables. In the Los Angeles Area, Los Angeles is the omitted county, whereas San Francisco county is omitted in the San Francisco Area. The coefficients yield information as to the home sale price differences between the omitted county and the included counties.

The second noteworthy aspect of the equations is that the non-linear specifications (log-linear for the Los Angeles Area, semi-log for the San Francisco Area) are significant improvements over linear forms.* As Rosen (1974) pointed out, this is to be expected since consumers cannot always arbitrage by dividing and repackaging bundles of housing attributes. Third, in each case, approximately 80 percent of the variation in home sale price is explained by the independent variable set ($R^2 = 0.79$). Fourth, with the exception of population density in both areas and the time to employment in Table 3.5,** all variables possess the expected relationship to home sale price and are significantly different from zero at the one percent level ($|t| > 2.326$). Population density is significant at the five percent level given a priori information whereas time to employment in Table 3.5 is insignificant. The latter result reflects the fact that most individuals travel approximately the same time to work. This is consistent with the small standard deviation around the mean (see Table 3.3).

However, the most important result from the perspective of this study is that the extinction variable is significantly different from zero and possesses the expected relationship to home sale price. These results indicate that individuals are acting upon extinction information when making locational choices and this action is translated into a measurable hedonic gradient. As is described below, this result is essentially invariant with respect to various sample sizes, extinction measures, model formulations, and functional forms.

Regarding the monetary impact on housing sale price of a change in an independent variable, the non-linear specification does not always allow straightforward interpretation because the effect of any independent variable

---

*The question of appropriate functional form is considered in detail in a subsection below.

**The insignificance of these variables does not present an overwhelming problem because in the final equations they are eliminated using principal components analysis.

depends upon the level of all other variables. However, the San Francisco Area results are particularly amenable to interpretation. This occurs because in the semi-log form, the coefficients represent the percentage change in home sale price given a one unit change in the independent variable.*

Thus, from Table 3.6 we can see that for the San Francisco Area in 1978-79, home sale prices were rising at 1.5 percent per month, 100 ft^2 of living area was worth 3.9 percent of price, and a pool had a value of 10.5 percent of price. In addition, a one unit change in extinction had a value of 22.8 percent of home sale price. Thus, a ten percent change in extinction would alter home sale price by 2.5 percent. Based on a mean home sale price of approximately $90,000, this latter figure translates into $2250.

In the Los Angeles Area the log-linear form is not as easy to interpret. However, if all independent variables are assigned their mean values, then a ten percent change in extinction would be valued at $1819. In relative terms, extinction is valued approximately equal to the presence of a view.

Given these benchmark results, their stability is the subject of the remainder of this section.

3.3.2 Stability of Results -- Sample Size, Extinction Measure

As an initial test of the stability of the results presented above, hedonic equations were estimated for various sample sizes. Table 3.7 presents a particular example for a smaller independent sample in the Los Angeles Area. As is illustrated, the results seem to be quite insensitive to sample size; that is, all the basic conclusions drawn above continue to be relevant. Although we present only this one example, insensitivity with respect to sample size is a general characteristic of our results.

Another issue is the appropriate light extinction measure. As indicated in the previous section, there is very little difference between the first two extinction measures (median for all hours, and median for hours without precipitation or fog); they are essentially interchangeable. The estimations in Tables 3.5-3.7 are based on these measures. However, the third extinction measure, which includes a subtraction of sea haze contributions, is significantly different from the others. For comparison, the hedonic equations were re-estimated substituting this sea haze adjusted extinction variable. The results of this estimation are presented in Tables 3.8 and 3.9. As is illustrated, the results are not radically altered by the choice of the extinction variable; extinction (3)

*See the subsection on functional form for a description of differentiation of non-linear forms.

TABLE 3.7 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE LOS ANGELES AREA WITH ALTERED SAMPLE SIZE.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ℓn (Sales Month)	0.106	21.73
ℓn (Age of Home)	-0.019	-4.86
ℓn (Square Feet of Living Area)	0.714	37.47
ℓn (Number of Bathrooms)	0.095	5.29
Number of Fireplaces	0.099	12.01
Pool	0.078	6.62
View	0.159	10.46
Neighborhood Characteristics:		
ℓn (Percent Greater than 64)	0.046	5.66
ℓn (Percent White)	0.259	20.50
Location Characteristics:		
ℓn (Distance to Beach)	-0.120	-23.80
ℓn (Time to Employment)	-0.052	-1.74
Orange County	-0.141	-12.35
Riverside County	-0.291	-10.89
San Bernadino County	-0.167	-8.80
Community Characteristics:		
ℓn (Population Density)	-0.018	-1.53
ℓn (School Quality)	0.569	7.54
ℓn (Light Extinction)	-0.216	-8.69
Constant	-1.738	-4.88
R-Squared	0.789	
Number of Observations	3096	

TABLE 3.8 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE LOS ANGELES AREA WITH SEA HAZE ADJUSTED EXTINCTION.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ℓn (Sales Month)	0.110	27.56
ℓn (Age of Home)	-0.017	-5.23
ℓn (Square Feet of Living Area)	0.725	47.56
ℓn (Number of Bathrooms)	0.102	6.70
Number of Fireplaces	0.094	14.25
Pool	0.081	8.30
View	0.163	13.08
Neighborhood Characteristics:		
ℓn (Percent Greater than 64)	0.060	9.15
ℓn (Percent White)	0.276	27.21
Location Characteristics:		
ℓn (Distance to Beach)	-0.114	-22.05
ℓn (Time to Employment)	-0.026	-1.07
Orange County	-0.159	-17.78
Riverside County	-0.275	-13.30
San Bernadino County	-0.192	-12.62
Community Characteristics:		
ℓn (Population Density)	-0.023	-2.47
ℓn (School Quality)	0.413	6.81
ℓn (Light Extinction) (3)	-0.090	-6.46
Constant	-1.444	-4.96
R-Squared	0.789	
Number of Observations	4766	

TABLE 3.9 ESTIMATED HEDONIC EQUATION (SEMI-LOG) FOR THE SAN FRANCISCO AREA WITH SEA HAZE ADJUSTED EXTINCTION.

DEPENDENT VARIABLE = $\ln$(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.015	25.36
Age of Home	-0.002	-7.56
Square Feet of Living Area	0.00039	35.25
Number of Bathrooms	0.053	5.37
Number of Fireplaces	0.095	11.68
Pool	0.105	6.55
View	0.071	5.31
Neighborhood Characteristics:		
Percent Greater than 64	0.005	6.63
Percent White	0.006	28.63
Location Characteristics:		
Distance to Beach	-0.005	-8.88
Time to Employment	-0.0029	-2.24
Alameda County	-0.289	-9.10
Contra Costa County	-0.403	-12.02
Santa Mateo County	-0.187	-6.27
Santa Clara County	-0.161	-5.29
Community Characteristics:		
Population Density	-0.0004	-1.54
School Quality	0.0096	8.09
Light Extinction (3)	-0.144	-3.16
Constant	5.177	45.87
R-Squared	0.794	
Number of Observations	3106	

is somewhat less important than the previous extinction parameters but is still very significant statistically. It should be noted that the Los Angeles results in Table 3.8 are somewhat deceiving because extinction (3) does not perform as well for other functional forms. The reason for this can be ascertained through an examination of a simple correlation matrix. The use of extinction (3) introduces a large amount of collinearity into the data set. The simple correlation between distance to beach and extinction (3) is 0.73, whereas the correlation between distance to beach and the other extinction measures is only 0.38. Thus, with a sea haze adjustment, extinction (3) and distance to beach are partly measuring the same phenomenon. Note that this is not the case in the San Francisco Area since distance to beach is not linear from the ocean but includes the bay impact. The conclusion for the Los Angeles Area is that extinction (3) does work for some functional forms, but that the other extinction variables perform better. On the basis of statistical performance, extinction (3) is eliminated. The results in the San Francisco Area are less variant with respect to the extinction variable, but for consistency extinction (3) is also omitted henceforth in San Francisco.*

3.3.3 Stability of the Results -- Principal Components Analysis

As indicated previously, the benchmark results contain only two community variables, population density and school quality. It could be argued that this limited number of variables does not account for all important community characteristics. If so, then the light extinction variable, in addition to representing air quality, could be serving as a proxy for one of the missing variables. In that case the coefficient of light extinction would be biased.

To test for such a bias, we should add in more community variables. However, additional community variables would introduce the problem of multi-collinearity. In order to overcome this latter problem, the method of principal components is utilized.

Principal component analysis is a method of transforming a given set of variables into a new set of composite indices or principal components that are orthogonal (uncorrelated) to each other. Because of the severe collinearity in this study, a transformation that yields uncorrelated variables is particularly useful. The transformation is accomplished by choosing the best linear combination of variables as the first principal component. In this context, "best"

*The sea haze adjusted extinction measure may also be questioned on a perception basis. That is, individuals are assumed not only to perceive visibility but also to divide visibility into its components.

implies that the combination chosen accounts for more of the variance in the data than any other linear combination of variables. The first principal component is therefore viewed as the single best summary of linear relationships exhibited in the data. The second component is defined as the second best linear combination of variables, given the condition that the second is orthogonal to the first. This continues until as much variation as possible is explained.

The principal component method can be expressed as:

$$N_i = a_{i1}F_1 + a_{i2}F_2 + \ldots + a_{iK}F_K \qquad (3\text{-}10)$$

where

N_i = the community variables included in the principal component analysis ($i = 1, 2, \ldots M$),

F_j = the principal components or factors ($j = 1, 2, \ldots, K$), $K < M$,

and a_{ij} = estimated coefficients.

If the number of factors equals the number of variables ($K = M$), then the entire variation in the variables is explained by the factors. However, it is the usual case to use fewer factors than variables because, if the two are equal, then the procedure is identical to not using principal components analysis (Johnston, 1972).

The estimated coefficients are important in that they indicate the relative importance of each factor. The importance of a given factor for a given variable can be expressed in terms of the variance in the variable that is explained by the factor. Mathematically this is the square of the estimated coefficient (a_{ij}^2). The total variation of a variable explained by all factors is obtained by summing the squared coefficients,

$$\left(\sum_{j=1}^{K} a_{ij}^2\right).$$

Given the relationships described in equation (3-10), the original variables are transformed into a set of composite scales or factor scores that represent the relative importance of the respective factors or principal components. In order to do this, the matrix of a_{ij} is transformed into a factor score coefficient matrix (b_{ij}). The composite scales or factor scores are then calculated as:

$$Z_j = b_{1j}(N_1 - \bar{N}_1)/\sigma_1 + b_{2j}(N_2 - \bar{N}_2)/\sigma_2 + \ldots + b_{Mj}(N_M - \bar{N}_M)/\sigma_M \quad (3\text{-}11)$$

where

Z_j = factor score representing the j^{th} factor ($j = 1, 2, \ldots, K$),

b_{ij} = factor score coefficient ($i = 1, 2, \ldots, M$),

N_i = original variable ($i = 1, 2, \ldots, M$),

$\bar{N}_i$ = mean of the i^{th} independent variable,

and σ_i = standard deviation of the i^{th} independent variable.

Note that the original variables are normalized to be nondimensional (Johnston, 1972).

The factor scores represent the transformed data set in which orthogonality is preserved. These new data are then inputs into the home sale price hedonic equation as explanatory variables. In essence, a set of highly correlated variables are replaced by a new set of uncorrelated variables which measure precisely the same information. However, it should be noted that the initial variables have been constrained to a linear relationship. Essentially, the procedure represents the imposition of a linear restriction, where the linear relationship is not based on a priori information but is chosen as the one which best fits the data.

In the semi-log form, the hedonic equation can be written as:

$$\ell n(P) = \beta_0 + \sum_{j=1}^{K} \beta_j Z_j + \sum_{\ell=j+1}^{L} \beta_\ell S_\ell + \beta_{L+1} E \quad (3\text{-}12)$$

where

P = home sale price,

Z_j = factor scores representing the principal components ($j = 1, 2, \ldots, K$),

S_ℓ = remaining explanatory variables (site influences) not included in the principal component analysis ($\ell = j+1, \ldots, L$),

and β_0, β_j, β_ℓ, β_{L+1} = estimated coefficients.

Since the principal components are linear combinations of other variables, no precise interpretation can be given to the factor score variables. However, one can still determine the relative effect of a change in a variable included in the principal component analysis by differentiating equation (3-12) with respect to that variable. For instance, consider the impact of N_1, a variable included in the principal component analysis. Substituting equation (3-11) into (3-12) and differentiating, we obtain

$$\frac{\partial P}{\partial N_1} = \frac{\partial P}{\partial Z_1}\frac{\partial Z_1}{\partial N_1} + \ldots + \frac{\partial P}{\partial Z_K}\frac{\partial Z_K}{\partial N_1} \tag{3-13}$$

Thus, although N_1 does not enter the hedonic housing equation directly its relative importance can still be determined.

In the particular situation under study, a severe collinearity exists between the community variables. Thus, it was decided to perform principal component analysis on these troublesome variables to transform them into a set of uncorrelated variables. In both study areas, eight community variables -- school quality, crime rate, unemployment rate, educational level, poverty rate, population per household, population per square mile, and miles to the business district -- were transformed into three factors or principal components.

The initial factor matrix (a_{ij}) is presented in Table 3.10 for each of the study areas. As is illustrated, the first factor or principal component largely explains school quality, crime, unemployment, education, and poverty. Because of the distribution of signs on these variables the expected relationship of Factor 1 to home sale price is negative. A similar analysis can be conducted for the other factors. Finally, in each of the two areas, the three factors explain approximately 83 percent of the variation in the variables.

As outlined above, the initial factor matrix is transformed in a factor score coefficient matrix. The relevant matrices are presented in Table 3.11. These factor score coefficients are used to compute factor scores or composite scales which represent the relative importance of each factor for each variable. This is accomplished in accordance with equation (3-12). The factor scores are input data (explanatory variables) into the hedonic housing equation.

The hedonic equations which include the three factors to account for eight community variables are presented in Tables 3.12 and 3.13. These results are quite consistent with the benchmark results. All variables remain significant, R^2 is essentially the same, etc.* Although there is little change in the overall results, the equations presented in Tables 3.12 and 3.13 are considered superior because they include more community variables. In addition, it is noteworthy that the coefficient on extinction is affected by the inclusion of more community variables. In each study area, a 10 percent change in light extinction is valued significantly less due to the inclusion of more variables ($1194 in Los Angeles and $1250 in San Francisco). Light extinction was evidently serving

*Note that time to employment has been left out in Tables 3.12 and 3.13. This variable was replaced by miles to the central business district and included in the principal component analysis.

TABLE 3.10 FACTOR COEFFICIENT MATRIX.

Los Angeles Area

Variable	Factor 1	Factor 2	Factor 3
School Quality	-0.828	-0.317	0.126
Miles to Business District	-0.595	0.402	-0.050
Crime	0.797	-0.302	-0.222
Unemployment	0.943	0.057	-0.122
Education	-0.799	-0.470	-0.093
Poverty	0.954	-0.055	-0.075
Population Per Household	-0.169	0.897	-0.034
Population Per Square Mile	0.396	-0.019	0.904

San Francisco Area

Variable	Factor 1	Factor 2	Factor 3
School Quality	-0.714	0.544	0.191
Miles to Business District	-0.587	-0.353	0.658
Crime	0.784	0.129	0.384
Unemployment	0.885	-0.332	0.081
Education	-0.818	0.414	0.026
Poverty	0.902	0.017	0.145
Population Per Household	-0.554	-0.660	-0.305
Population Per Square Mile	0.626	0.530	-0.205

TABLE 3.11 FACTOR SCORE COEFFICIENT MATRIX.

Los Angeles Area

Variable	Factor 1	Factor 2	Factor 3
School	-0.277	0.158	0.082
Miles to Business District	-0.032	-0.316	-0.074
Crime	0.165	0.269	-0.202
Unemployment	0.246	0.029	-0.078
Education	-0.248	0.274	-0.154
Poverty	0.214	0.105	-0.031
Population Per Household	0.159	-0.630	-0.021
Population Per Square Mile	-0.144	0.009	0.981

San Francisco Area

Variable	Factor 1	Factor 2	Factor 3
School	-0.372	0.290	0.083
Miles to Business District	0.049	0.289	0.839
Crime	0.069	0.443	0.284
Unemployment	0.301	0.029	0.113
Education	-0.328	0.096	-0.035
Poverty	0.144	0.226	0.064
Population Per Household	0.210	-0.577	-0.052
Population Per Square Mile	-0.130	0.139	-0.435

TABLE 3.12 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE LOS ANGELES AREA WITH PRINCIPAL COMPONENT COMMUNITY VARIABLES.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ℓn (Sales Month)	0.109	27.54
ℓn (Age of Home)	-0.018	-5.66
ℓn (Square Feet of Living Area)	0.725	48.15
ℓn (Number of Bathrooms)	0.095	6.64
Number of Fireplaces	0.088	13.52
Pool	0.078	8.10
View	0.156	12.73
Neighborhood Characteristics:		
ℓn (Percent Greater than 64)	0.041	6.47
ℓn (Percent White)	0.278	27.67
Location Characteristics:		
ℓn (Distance to Beach)	-0.1167	-28.75
Orange County	-0.134	-15.29
Riverside County	-0.274	-13.83
San Bernadino County	-0.173	-11.52
Community Characteristics:		
Factor 1	-0.027	-6.62
Factor 2	0.033	8.22
Factor 3	-0.0093	-2.16
ℓn (Light Extinction)	-0.133	-5.83
Constant	0.205	1.81
R-Squared	0.796	
Number of Observations	4766	

TABLE 3.13 ESTIMATED HEDONIC EQUATION (SEMI-LOG) FOR THE SAN FRANCISCO AREA WITH PRINCIPAL COMPONENT COMMUNITY VARIABLES.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.015	25.57
Age of Home	-0.0022	-8.06
Square Feet of Living Area	0.00039	35.60
Number of Bathrooms	0.052	5.31
Number of Fireplaces	0.091	11.33
Pool	0.097	6.11
View	0.068	5.19
Neighborhood Characteristics:		
Percent Greater than 64	0.0039	5.68
Percent White	0.0063	27.27
Location Characteristics:		
Distance to Beach	-0.0056	-8.21
Alameda County	-0.212	-11.38
Contra Costa County	-0.336	-15.34
San Mateo County	-0.121	-6.52
Santa Clara County	-0.097	-4.21
Community Characteristics:		
Factor 1	-0.051	-9.40
Factor 2	0.020	3.15
Factor 3	-0.0067	-1.08
Light Extinction	-0.127	-2.70
Constant	5.77	76.81
R-Squared	0.797	
Number of Observations	3106	

as a proxy for excluded community variables. This suggests that the previous benchmark results were likely biased due to specification error.

The principal component analysis is considered an improvement over the benchmark results. Thus, the equations in Tables 3.12 and 3.13 will be those that are analyzed further in subsequent sections.*

3.3.4 Stability of the Results -- Functional Form

As previously mentioned, a priori information about the functional form of the hedonic price equation is unavailable. This requires that some consideration be given to correct form, especially since the estimated benefits depend heavily upon the implicit prices calculated from the equation. As an aid in determining the appropriate functional form, the Box-Cox transformation has been employed. This has become a fairly standard approach and has been used by Bender et al. (1980), Quigley (1982), and others. Moreover, most researchers have found that benefits are indeed sensitive to functional form, adding importance to these considerations.

Our search ranged over six possible forms: linear, semi-log, log-linear, classical Box-Cox, extended Box-Cox, and semi-log exponential. A basic understanding of these forms and how they relate to each other can be gained by considering a simple four variable model. Let P be the dependent variable (home sale price) and S, N, and E the independent variables (a site variable, neighborhood variable, and an environmental variable, respectively). Then the linear form is stated as

$$P = a_0 + a_1 S + a_2 N + a_3 E. \tag{3-14}$$

By using ℓn to denote the natural logarithm of a variable, the semi-log form is given by

$$\ell n\ P = a_0 + a_1 S + a_2 N + a_3 E. \tag{3-15}$$

(The letters -- a_0, a_1, a_2, and a_3 -- are used to denote statistical coefficients in each equation. These are not necessarily equal between equations.) The semi-log form is commonly used to estimate an exponential function, equation (3-15) being equivalent to

*The effect of sample size and different light extinction measures were examined using the equations in Tables 3.12 and 3.13. Our previous conclusions were unaltered.

$$P = \exp(a_0 + a_1S + a_2N + a_3E). \tag{3-15a}$$

On the other hand, the log-linear form

$$\ell n\ P = a_0 + a_1\ \ell n\ s + a_2\ \ell n\ N + a_3\ \ell n\ E \tag{3-16}$$

is used to estimate the well known Cobb-Douglas equation;

$$P = e^{a_0} S^{a_1} N^{a_2} E^{a_3} \tag{3-16a}$$

The classical Box-Cox specification can be denoted by

$$P(\lambda) = a_0 + a_1S + a_2N + a_3E \tag{3-17}$$

where $P(\lambda) = (P^{\lambda} - 1)/\lambda$. This type of transformation is very convenient because when $\lambda = 1$, it is the same as the linear form (equation 3-14) and as λ approaches zero, $P(\lambda)$ approaches $\ell n\ P$, and (3-17)is the same as the semi-log form (equation 3-15). Thus, both the linear and semi-log specifications are special cases of the classical Box-Cox transformation. Furthermore, using advanced estimation techniques (the method of maximum likelihood), λ can be estimated simultaneously with a_0, a_1, a_2, and a_3. In effect, the Box-Cox transformation is a flexible functional form that allows the data to choose the best form. Also, after estimating λ, restrictions such as $\lambda = 0$ or $\lambda = 1$ can be tested statistically. This methodology rationalizes, to some extent, the search for appropriate functional forms.

The logical next step is to consider transformations on all variables in the model. (In practice, dummy variables and others whose values can be equal to zero are not transformed, since $\ell n(0)$ is undefined.) This is referred to as the extended Box-Cox specification and is written as

$$P(\lambda) = a_0 + a_1S(\lambda) + a_2N(\lambda) + a_3E(\lambda) \tag{3-18}$$

indicating that the Box-Cox transformation has been applied to each variable. The log-linear form is a special case of (3-18) where $\lambda = 0$, therefore, the log-linear form can be tested statistically with equation (3-18).

The last form examined is semi-log exponential. This is given by

$$\ell n\ P = a_0 + a_1S + a_2N + a_3E^{\phi}. \tag{3-19}$$

In equation (3-19), one of the independent variables is transformed by a power of $\emptyset$, while the dependent variable is entered as its natural logarithm. This form was used by Harrison and Rubinfeld (1978) who found it performed better than the semi-log or the log-linear. We have included it for completeness but have only examined two cases, with $\emptyset = 2$ and $\emptyset = 1/2$.

Equations (3-14) through (3-19) represent the range of functional forms considered for the 1978-79 data sets in both study areas. In order to choose among the alternatives, the value of the likelihood function for each form was computed (see Spitzer, 1982). The form that achieves the highest likelihood (i.e., that is most probable) is generally considered the most appropriate. In order to determine whether one value is statistically greater than another, a likelihood ratio test (Judge et al., 1980) can be performed.* Table 3.14 summarizes our findings in terms of the log of likelihood values. The most probable form was the extended Box-Cox in the Los Angeles Area, while in the San Francisco Area, the classical Box-Cox out-performed the others. Interestingly, in the classical Box-Cox case, the maximum likelihood estimate of λ from both areas was equal to -0.05. For the extended Box-Cox form λ equalled -0.05 in the Los Angeles Area, but was equal to -0.08 in the San Francisco Area.

For the Los Angeles Area, the three best functional forms are extended Box-Cox, log-linear (Table 3.12), and classical Box-Cox. The first and last of these are presented in Tables 3.15 and 3.16, respectively. In the Los Angeles Area, the semi-log form does not perform to the level of the other forms, but it is nevertheless presented in Table 3.17 for use in the next section.

In the San Francisco Area, the most probable functional forms are classical Box-Cox, semi-log exponential ($\emptyset = 2$), and semi-log (Table 3.13). The first two of these are presented in Tables 3.18 and 3.19.

In all of the tables alluded to above, the light extinction variable remains significantly different from zero and has the expected (negative) relationship to home sale price. Thus, the results are quite stable among functional forms in those respects. However, as discussed in the next section, the resulting benefit figures vary somewhat. For purposes of illustration, estimated demand

---

*The ratio of the likelihood values is distributed chi-square with one restriction. If the log of the likelihood values are used then two times the difference in these values is distributed chi-square. It should also be noted that all functional forms cannot be compared in this manner. For instance, the classical Box-Cox and extended Box-Cox cannot be compared using a chi-square test since one is not a restricted form of the other. Thus, classical Box-Cox should be compared to semi-log and linear whereas extended Box-Cox is compared to log-linear. Since some cross comparisons cannot be made statistically, we opt for the forms which have the largest likelihood values.

TABLE 3.14 LOG OF LIKELIHOOD VALUES FOR THE VARIOUS FUNCTIONAL FORMS BY AREA.*

	Functional Form	Los Angeles	San Francisco
1.	Linear	k-27696	k-17591
2.	Semi-Log	k-24640	k-15893
3.	Log-Linear	k-24631	k-15945
4.	Classical Box-Cox	k-24633	k-15880
5.	Extended Box-Cox	k-24621	k-15938
6.	Semi-Log Extinction Squared	k-24641	k-15891
7.	Semi-Log Square Root Extinction	k-24638	k-15894

*k is a constant common to each functional form

TABLE 3.15 ESTIMATED HEDONIC EQUATION (EXTENDED BOX-COX) FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = (HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)*

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month*	0.084	27.27
Age of Home*	-0.015	-6.03
Square Feet of Living Area*	0.741	47.91
Number of Bathrooms*	0.068	6.48
Number of Fireplaces	0.063	13.53
Pool	0.055	8.07
View	0.109	12.42
Neighborhood Characteristics:		
Percent Greater than 64*	0.030	5.97
Percent White*	0.238	28.00
Location Characteristics:		
Distance to Beach*	-0.087	-28.20
Orange County	-0.093	-14.81
Riverside County	-0.204	-14.50
San Bernadino County	-0.131	-12.21
Community Characteristics:		
Factor 1	-0.020	-6.99
Factor 2	0.243	8.42
Factor 3	-0.0059	-1.91
Light Extinction*	-0.103	-6.22
Constant	0.2414	2.52
R-Squared	0.793	
Number of Observations	4766	

* Indicates the variable is transformed using the Box-Cox transformation.

TABLE 3.16 ESTIMATED HEDONIC EQUATION (CLASSICAL BOX-COX) FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = (HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)*

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.010	28.84
Age of Home	-0.0014	-7.66
Square Feet of Living Area	0.00025	40.90
Number of Bathrooms	0.073	12.82
Number of Fireplaces	0.073	15.78
Pool	0.033	4.83
View	0.112	12.81
Neighborhood Characteristics:		
Percent Greater than 64	0.0041	8.70
Percent White	0.0049	28.86
Location Characteristics:		
Distance to Beach	-0.0095	-25.91
Orange County	-0.126	-19.45
Riverside County	-0.069	-4.34
San Bernadino County	-0.0077	-0.63
Community Characteristics:		
Factor 1	-0.027	-9.24
Factor 2	0.0199	6.72
Factor 3	-0.0078	-2.47
Light Extinction	-0.025	-2.51
Constant	4.79	187.53
R-Squared	0.792	
Number of Observations	4766	

* Indicates the variable is transformed using the Box-Cox transformation.

TABLE 3.17 ESTIMATED HEDONIC EQUATION (SEMI-LOG) FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.014	28.93
Age of Home	-0.0018	-7.07
Square Feet of Living Area	0.00036	41.58
Number of Bathrooms	0.104	13.02
Number of Fireplaces	0.101	15.64
Pool	0.048	4.99
View	0.162	13.19
Neighborhood Characteristics:		
Percent Greater than 64	0.0060	9.00
Percent White	0.0068	28.42
Location Characteristics:		
Distance to Beach	-0.013	-26.17
Orange County	-0.177	-19.54
Riverside County	-0.095	-4.26
San Bernadino County	-0.0068	-0.39
Community Characteristics:		
Factor 1	-0.037	-9.18
Factor 2	0.027	6.67
Factor 3	-0.117	-2.67
Light Extinction	-0.038	-2.70
Constant	5.44	152.11
R-Squared	0.795	
Number of Observations	4766	

TABLE 3.18 ESTIMATED HEDONIC EQUATION (CLASSICAL BOX-COX) FOR THE SAN FRANCISCO AREA.

DEPENDENT VARIABLE = (HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)*

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.107	25.64
Age of Home	-0.016	-9.28
Square Feet of Living Area	0.00027	35.04
Number of Bathrooms	0.037	5.40
Number of Fireplaces	0.067	11.53
Pool	0.067	5.90
View	0.047	5.05
Neighborhood Characteristics:		
Percent Greater than 64	0.0027	5.46
Percent White	0.0046	27.92
Location Characteristics:		
Distance to Beach	-0.004	-8.18
Alameda County	-0.152	-11.44
Contra Costa County	-0.243	-15.57
San Mateo County	-0.087	-6.60
Santa Clara County	-0.071	-4.35
Community Characteristics:		
Factor 1	-0.036	-9.30
Factor 2	0.148	3.28
Factor 3	-0.0062	-1.41
Light Extinction	-0.91	-2.70
Constant	5.026	93.94
R-Squared	0.795	
Number of Observations	3106	

*Indicates the variable is transformed using the Box-Cox transformation.

TABLE 3.19 ESTIMATED HEDONIC EQUATION (SEMI-LOG EXPONENTIAL) FOR THE SAN FRANCISCO AREA.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1978-79 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
Sales Month	0.015	25.59
Age of Home	-0.0022	-8.03
Square Feet of Living Area	0.00039	35.57
Number of Bathrooms	0.052	5.31
Number of Fireplaces	0.092	11.34
Pool	0.097	6.14
View	0.068	5.17
Neighborhood Characteristics:		
Percent Greater than 64	0.0039	5.68
Percent White	0.0063	27.18
Location Characteristics:		
Distance to Beach	-0.0058	-8.78
Alameda County	-0.214	-11.52
Contra Costa County	-0.336	-15.88
San Mateo County	-0.123	-6.67
Santa Clara County	-0.090	-3.97
Community Characteristics:		
Factor 1	-0.052	-9.57
Factor 2	0.020	3.17
Factor 3	-0.0053	-0.86
Light Extinction	-0.066	-3.38
Constant	5.72	115.17
R-Squared	0.797	
Number of Observations	3106	

curves (and benefits) will be presented for the two best functional forms from each area as well as for the semi-log and classical Box-Cox forms. The interpretability and performance of the semi-log and classical Box-Cox specifications made them logical choices to impose on both areas. Direct comparisons between regions can be made by examining the results for these latter two forms.

Table 3.20 presents the derivatives of the five forms that will be used to estimate demand curves. Because the Box-Cox transformations generate rather abstract functional forms, the derivatives of these forms are also quite abstract.

3.4 EMPIRICAL RESULTS: INVERSE DEMAND EQUATIONS

The final set of hedonic equations discussed in the previous section are the basis for determining the benefits of visibility improvements. In order to complete the benefit estimation procedure, the following steps are required. First, the hedonic equations are differentiated to determine the marginal willingness to pay for a change in extinction. The marginal willingness to pay is evaluated for each individual point in the data set. These values represent the implicit price of extinction for each individual and are dependent upon all the other characteristics of the home.* Given these implicit prices, an inverse demand curve** can be estimated by regressing price against quantity (extinction) and other household (homeowner) shift variables (income, etc.). Integrating under these inverse demand curves for any proposed extinction change yields the benefits attributable to the change.

In this section, estimated inverse demand curves are presented for both the Los Angeles and San Francisco Areas following the approach set out by Freeman (1974, 1979) and Rosen (1974). In addition, an alternative estimation to the Freeman-Rosen methodology is performed. In this latter instance, a multi-market demand curve is estimated using the data pooled from across the two study areas.

Because of data difficulties, the demand curve estimation is performed at the community level, rather than at the individual level. Specifically, because individual homeowner income is not available, the price and quantity data

*For instance, sale prices of homes with views exhibit a larger marginal willingness to pay than for those without views (in the Bay Area, the difference is approximately $800 over the life of the home according to equation (3-13)).

**Demand curves usually represent quantity as a function of price. In this case, price is a function of quantity, so the label is "inverse" demand curve.

TABLE 3.20 VARIOUS FUNCTIONAL FORMS AND THEIR DERIVATIVES.

Form	Derivative with Respect to E
1. Semi-log	
$\ell nP = a_0 + a_1S + a_2N + a_3E$	$a_3 \exp(a_0 + a_1 + a_2N + a_3E)$
2. Log-linear	
$\ell nP = a_0 + a_1\ell nS + a_2\ell nN + a_3\ell nE$	$a_3e^{a_0}S^{a_1}N^{a_2}E^{a_3-1}$
3. Classical Box-Cox	
$P(\lambda) = a_0 + a_1S + a_2N + a_3E$	$a_3[(a_0 + a_1S + a_2N + a_3E)\lambda + 1]^{1/\lambda-1}$
4. Extended Box-Cox	
$P(\lambda) = a_0 + a_1S(\lambda) + a_2N(\lambda) + a_3E(\lambda)$	$a_3E^{\lambda-1}\{[a_0 + a_1S(\lambda) + a_2N(\lambda) + a_3E(\lambda)]\lambda + 1\}^{1/\lambda-1}$
5. Semi-log Exponential ($\emptyset = 2$)	
$\ell n\ P = a_0 + a_1S + a_2N + a_3E^2$	$2a_3E \exp(a_0 + a_1S + a_2N + a_3E^2)$

must be aggregated to the community level. The demand curve estimation utilizes the following data. Marginal willingness to pay (in hundreds of dollars) is the implicit price of light extinction improvements per $10^{-4}m^{-1}$. It is the derivative of the hedonic equation evaluated for each data point, and it represents the average home sale price differential attributable to a unit extinction difference. The quantity variable is the initial average community extinction level. Income represents average community income in hundreds of dollars per year.

The estimated inverse demand curves for the Los Angeles Area are presented in Table 3.21. There are a number of noteworthy aspects. First, only quantity (extinction) and income are employed to describe the variation in price. A large proportion of the variation is explained (R^2 = 0.8 and above), so additional variables would be of marginal significance. Second, the linear forms presented for the inverse demand curve perform better than alternative non-linear forms. But this is not a crucial point since the resulting benefit figures are quite insensitive to the functional form of the inverse demand curve. The third aspect is the range of results for the various functional forms of the hedonic equations. There is a close similarity between the equations that use a left-hand side transformation only (semi-log, classical Box-Cox) and between the equations that use a transformation of all the variables (log-linear, Box-Cox extended). However, the difference between the two types is quite pronounced. Fourth, all variables are significant at the one percent confidence level.

In addition to the points mentioned above, also note that income possesses the expected relationship to marginal willingness to pay. Another interesting result of the demand estimation for the Los Angeles Area is the "unexpected" negative relationship of extinction to price. A positive relationship might be expected because this is an inverse demand curve for a "bad"*, and standard economic theory suggests that individuals will pay more to avoid an additional unit of "bad" at higher levels than at lower levels. Thus, the negative relationship might be contrary to conventional economic theory and might imply non-convex preferences. This would not be without precedent in the study of environmental goods, as Crocker (1981) has demonstrated in a recent paper. However, Bartik and Smith (1984) indicate that this is a mute issue because the non-linearity of the hedonic equation prevents a priori prediction of the sign on the pollution variable. Thus, although the negative sign seems to suggest non-convex preferences, there is really insufficient evidence to make such a conclusion.

*A "bad" is a commodity for which less is preferred to more.

TABLE 3.21 ESTIMATED LINEAR DEMAND CURVES FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = MARGINAL WILLINGNESS TO PAY FOR LIGHT EXTINCTION IMPROVEMENTS IN HUNDREDS OF 1978-79 DOLLARS

Functional Form of Hedonic Price Gradient	Independent Variable Coefficients (t -statistics)				
	Constant	Light Extinction	Income	R^2	Number of Observations
Extended Box-Cox	124.70	-55.10 (-12.30)	0.209 (17.88)	0.82	112
Log-linear	108.50	-43.51 (-13.72)	0.136 (16.48)	0.82	112
Classical Box-Cox	14.75	-3.84 (-2.39)	0.090 (21.39)	0.81	112
Semi-log	17.28	-3.89 (-2.30)	0.092 (21.06)	0.80	112

The inverse demand curves for San Francisco are presented in Table 3.22. As is illustrated, the San Francisco results are apparently consistent with standard convex preferences. However, as in the Los Angeles case, no a priori interpretation can really be attached to the sign of the pollution variable.

There are other aspects of the San Francisco Area inverse demand curves that are worth noting. First, extinction level is a relatively insignificant determinant of price.* This is not unexpected in the San Francisco Area, since the data demonstrate little variation at the community level. Only in the case where extinction is transformed in the hedonic equation (semi-log exponential) is it significant at the one percent level in the inverse demand curve. Second, the R^2 values are quite high, indicating a large proportion of the variation is explained by the independent variable set. Third, the linear equations perform better than non-linear equations.

A deeper understanding of the inverse demand relationships can be obtained by calculating the annual household benefits associated with a hypothetical ten percent improvement in visibility. These benefit figures are obtained by integrating the inverse demand curves over the proposed visibility change. The resulting benefit figures are presented in Table 3.23. As is evident the benefit estimates are dependent upon the functional form of the hedonic price gradient. This is especially true for the Los Angeles Area. The San Francisco results are not as dependent upon functional form. However, the variation increases as the size of the improvement increases. Thus, for non-marginal changes, the benefit estimates would be more dependent upon functional form.

An alternative approach to estimating separate demand curves is to pool the data across markets and estimate one multi-market inverse demand curve. The theoretical reasoning for this approach, outlined in Section 3.2, is associated with Mendelsohn (1980) and others. This approach requires the assumption that individual preferences must be identical across the markets. It is felt that this is a very unreasonable assumption because individuals tend to gravitate to their own kind. For example, those who are relatively adverse to pollution might not live in the Los Angeles Area. Our doubts notwithstanding, the multi-market estimation is completed for comparison purposes.

---

*Extinction is always significant at the ten percent level with a priori information concerning its sign. This latter condition allows the use of a one tailed t-test.

TABLE 3.22 ESTIMATED LINEAR DEMAND CURVES FOR THE SAN FRANCISCO AREA.

DEPENDENT VARIABLE = MARGINAL WILLINGNESS TO PAY FOR LIGHT EXTINCTION IMPROVEMENTS IN HUNDREDS OF 1978-79 DOLLARS

Functional Form of Hedonic Price Gradient	Independent Variable Coefficients (t -statistics) Constant	Light Extinction	Income	R^2	Number of Observations
Classical Box-Cox	-22.78	23.49 (1.29)	0.429 (15.50)	0.84	51
Semi-log Exponential	-130.89	135.48 (8.60)	0.408 (17.06)	0.90	51
Semi-log	-9.68	25.65 (1.73)	0.337 (149.69)	0.84	51

TABLE 3.23 ESTIMATED ANNUAL HOUSEHOLD BENEFITS FOR A HYPOTHETICAL TEN PERCENT IMPROVEMENT IN VISIBILITY.

Functional Form of the Hedonic Price Gradient	Los Angeles Area	San Francisco Area
Extended Box-Cox	152.6	--
Log-Linear	125.0	--
Classical Box-Cox	56.9	124.6
Semi-Log	62.2	128.4
Semi-Log Exponential	--	115.2

The multi-market inverse demand curves are presented in Table 3.24 for the common hedonic price equations. The equations seem to be dominated by the negative extinction result for Los Angeles. The coefficients are quite significant, but R^2 values are lower reflecting the difficulty of determining a common demand curve for diverse groups.

A comparison of the single market result to the multi-market results can best be completed by calculating benefit figures from each equation. Consider again a hypothetical ten percent improvement in average visibility over the entire Los Angeles Area. Integration of the inverse demand curves for this change yields the household benefits associated with the change.* The Los Angeles Area inverse demand curve based on the semi-log hedonic equation yields an average annual value of approximately $62 per household.** The corresponding pooled estimate is $99 per household annually. The difference may be partly attributed to the use of San Francisco individual preferences in evaluating a Los Angeles policy. Since San Francisco individuals are likely more adverse to pollution, they place a higher value on the improvement. This result confirms our reservations about pooling the data across obviously diverse groups.

In conclusion, the work described in this section was designed to determine the inverse demand curve for visibility improvements. This has been done using both the Freeman-Rosen methodology and a multi-market approach. The Los Angeles results seem unusual with respect to preference convexity but may not really be unusual in light of the work of Crocker (1981) and Bartik and Smith (1984). The San Francisco results conform to traditional theory. Finally, the multi-market approach alters the single market analysis in an expected manner.

3.5 CONCLUDING REMARKS

The analysis reported in this chapter was designed to determine the value that individuals place on air quality improvements. The information required to conduct this calculation was obtained from the market for single family residences using the hedonic price procedure. The analysis was conducted in two air basins, Los Angeles and San Francisco.

---

*The formula used in these calculations is:

$$\int_{\text{Extinction after}}^{\text{Extinction before}} (MWTP_i)\, d(\text{Extinction})$$

where $(MWTP_i) = f(\text{income, extinction})$.

**In order to determine basin-wide benefits one must multiply by the number of households in the basin. In 1980 this figure is 4,045,800. Total benefits are then approximately 250 million dollars annually.

TABLE 3.24 ESTIMATED LINEAR DEMAND CURVES FOR POOLED DATA.

DEPENDENT VARIABLE = MARGINAL WILLINGNESS TO PAY FOR LIGHT EXTINCTION IMPROVEMENTS IN HUNDREDS OF 1978-79 DOLLARS

Functional Form of Hedonic Price Gradient	Independent Variable Coefficients (t -statistics) Constant	Light Extinction	Income	R^2	Number of Observations
Semi-log	111.39	-60.06 (-10.80)	0.178 (8.82)	0.57	163
Classical Box-Cox	110.96	-63.62 (-10.02)	0.207 (8.98)	0.56	163

The hedonic equation estimation indicates that air quality, as measured by light extinction, is a significant determinant of home sale price. This implies that decreases in visual air quality cause housing values to decrease. Further, this result is independent of sample size, extinction measure, and functional form. However, the ultimate benefits are somewhat dependent upon functional form. For this reason, benefits cannot be precisely determined. Rather, a range of benefits dependent upon the functional form of the hedonic housing equation is provided.

Give the hedonic equations, inverse demand curves for air quality have been estimated. Integration of these demand curves over a proposed improvement yields the total benefit associated with the improvement. As an example, the annual basin-wide benefits of a hypothetical ten percent improvement in air quality in the Los Angeles Area range from approximately 250 million dollars to 617 million dollars. A similar ten percent improvement in the San Francisco Area would have an annual benefit range of approximately 190 million dollars to 220 million dollars. A type of confidence interval could be constructed around these benefit estimates by adjusting the extinction coefficient in both the hedonic equation and the inverse demand curve by the corresponding standard errors of the estimated coefficients. This would yield a measure of the uncertainity of the estimated benefits. However, this exercise is not undertaken since the uncertainity inherent in the range of functional forms is far greater than uncertainity around the estimated coefficients. These figures must be compared to the cost of obtaining such an improvement before the efficiency of the change can be determined.

3.6 REFENCES

Bartik, J.J. and U.K. Smith, "Urban Amenities and Public Policy," Vanderbilt University, 1984.

Bender, B., T. J. Gronberg, and Hae-Shin Havong, "Choice of Functional Form and the Demand for Air Quality," Review of Economics and Statistics, pp. 638-43, November 1980.

Brookshire, D., M. Thayer, W. Schulze, and R. d'Arge, "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," American Economic Review, 72, March 1982.

Brown, G. M. and H. O. Pollakowski, "Economic Valuation of Shoreline," Review of Economics and Statistics, 59, 1977.

Brown, J., and H. Rosen, "On the Estimation of Structural Hedonic Price Models," Econometrica, May 1982.

Crocker, T. P., "A Synthesis of Contingent Valuation Studies of the Value of Atmospheric Visibility," University of Wyoming, 1981.

Freeman, A. M., III, "On Estimating Air Pollution Control Benefits from Land Value Studies," Journal of Environmental Economics and Management, 89, Number 3, August 1974.

Freeman, A. M., III, The Benefits of Environmental Improvement, Johns Hopkins Press, Baltimore, 1979a.

Freeman, A. M., III, "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues," Scandinavian Journal of Economics, 81, 1979b.

Harrison, D., Jr. and D. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," Journal of Environmental Economics and Management, 5, March 1978.

Johnston, J., Econometric Methods, McGraw-Hill, New York, 1972.

Judge, G. G., et al., The Theory and Practice of Econometrics, John Wiley and Sons, New York, 1980.

Maler, K., "A Note on the Use of Property Values in Estimating Marginal Willingness to Pay for Environmental Quality," Journal of Environmental Economics and Management, 4, December 1977.

Mendelsohn, R., "The Demand and Supply for Characteristics of Goods," University of Washington, 1980.

Nelson, J., "Airport Noise, Location Rent, and the Market for Residential Amenities," Journal of Environmental Economics and Management, 6, December 1979.

Palmquist, R., "The Demand for Housing Characteristics: Reconciling Theory and Estimation," North Carolina State University, 1981.

Quigley, J. M., "Nonlinear Budget Constraints and Consumer Demand: An Application to Public Programs for Residential Housing," Journal of Urban Economics, 12, 1982.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 82, January/February 1974.

Spitzer, J. J., "A Primer on Box-Cox Estimation," Review of Economics and Statistics, 62, 1982.

Willig, R. D., "Consumer's Surplus Without Apology," American Economic Review, 66, Number 4, September 1976.

4.0 ILLUSTRATIVE APPLICATION: DIESEL PARTICULATE CONTROL IN LOS ANGELES

This chapter presents an example illustrating how the results of the previous chapter can be used to estimate the benefits associated with a given emission control strategy. The specific example considered is the control of diesel particulate emissions in Los Angeles. As noted by Trijonis (1983, 1984), diesel trucks are currently a significant cause of visibility reduction in Los Angeles and an expanded diesel highway fleet is projected to become the single most important source of visibility reduction in the region by the early 1990's.

The benefit analysis for emission control strategies separates naturally into two basic parts. First, we determine the degree of visibility improvement associated with the control strategy (Section 4.1). Next, the economic benefits produced by the visibility improvement are estimated using the results of the previous chapter (Section 4.2).

4.1 RELATIONSHIP BETWEEN DIESEL PARTICLE EMISSIONS AND VISIBILITY

Diesel vehicles affect visibility via three emission categories: sulfur dioxide (precursor of sulfate particles), nitrogen oxide (precursor of nitrate particles and NO_2), and primary (directly-emitted) particles. Because this chapter is concerned with control strategies for directly-emitted particles (i.e. particulate emission standards), we will consider the visibility impact for the third category only. It is known that about 90% of the visibility effect from primary diesel particulate emissions comes from elemental carbon, which constitutes about 70% of diesel particle exhaust (Trijonis, 1983). Because elemental carbon dominates the visibility effect of diesel particle emissions, and because some control strategies might affect elemental carbon differently from other emitted particles, we will focus exclusively on the visibility effects from the elemental carbon component of primary particulate exhaust.

The remainder of this section is organized as follows. Section 4.1.1 discusses the methodology for estimating current visibility impacts from diesel elemental carbon emissions. Our assessment of the current impacts also serves as the basis for the methodology to estimate the visibility effects of potential control strategies. Section 4.1.2 briefly describes the projected growth of diesel emissions for a "no control" scenario. Finally, Section 4.1.3 deals with a side issue -- the relevance of human perception of visibility changes in considering control strategy benefits.

4.1.1 Current Visibility Effects from Diesel Elemental Carbon Emissions

In this section, we used two approaches to determine the current (1980) visibility impacts from diesel elemental carbon emissions. The first approach, an emission budget model, provides an estimate of overall basinwide impacts. The second approach, a lead tracer model, yields not only basinwide effects but also the spatial distribution of the impacts. To arrive at a best estimate, we use the spatial distribution from the lead tracer model, but calibrate the predictions to the average of the two models.

The emission budget model is quite simple and direct. Several investigators have concluded that elemental carbon contributes about 22% $\pm$ 7% of light extinction (visibility reduction) in the Los Angeles basin (Conklin et al., 1981; Trijonis et al., 1982; Appel et al., 1983; Davidson, 1983). Furthermore, the emission inventory of Cass et al. (1982), modified to be consistent with the diesel emission factors of Trijonis (1983), indicates that diesel highway vehicles account for about 40% $\pm$ 10% of elemental carbon emissions in Los Angeles. The inference from these two results is that elemental carbon from diesel vehicles contributes about 9% (22% x 40%) of light extinction in Los Angeles. A reasonable uncertainty range for this estimate is $\pm$4%.

The lead tracer model for diesel elemental carbon concentrations is based on the equation:

$$[C]_{jy} = [Pb]_{jx} \bullet \frac{EC_y}{EPb_x}, \qquad (4.1)$$

where $[C]_{jy}$ = annual mean elemental carbon concentration from diesel vehicles at site "j" in year "y" (y=1980 for current impacts),

EC_y = total elemental carbon emissions from diesel vehicles in year "y",

EPb_x = total lead emissions from all sources in year "x",

and $[Pb]_{jx}$ = annual mean lead concentration at site "j" in year "x".

Trijonis (1983) has conducted a detailed analysis of statewide lead and diesel emissions for use in Equation 4.1. For example, his lead emission inventory includes stationary sources as well as light-, medium-, and heavy-duty gasoline vehicles, taking into account long-term trends in the percentage of non-catalyst (leaded-fuel) traffic and yearly changes in the Pb content of leaded gasoline. The predictions of Trijonis' statewide lead tracer model are adjusted herein by a factor of 0.76 = 3.5/4.6, because heavy-duty trucks contribute 3.5% of total traffic in Los Angeles as opposed to 4.6% statewide.

Once elemental carbon concentrations from diesels have been estimated by Equation 4.1, the computation of visibility impacts is straightforward. The extinction efficiency for fine elemental carbon is well documented: 12 ± 3 m^2/g, approximately 9 m^2/g absorption and 3 m^2/g scattering (see list of references in the 1983 Trijonis paper). The light extinction due to diesel elemental carbon emissions is simply the product of the elemental carbon concentration and the extinction efficiency. One can further compute percentage contributions of diesel carbon to total light extinction using the total extinction levels in Los Angeles given by Figure 2.4.

Using ambient lead data for 26 monitoring sites in the Los Angeles region, we calculate that -- for the average site in 1980 -- diesel elemental carbon contributes 15% of total light extinction. A reasonable uncertainty level for this estimate is ±6%.

Averaging the results of the emission budget model (9%) and lead tracer model (15%), we conclude that elemental carbon from highway diesels accounts for about 12% of light extinction in the Los Angeles region during 1980. To obtain the spatial distribution of the 1980 impacts we use the lead tracer model but multiply the predictions by 0.8 = 12%/15% to calibrate against the average of the two models. Figure 4.1 presents the resulting spatial distribution of estimated light extinction levels from diesel truck elemental carbon in 1980.

Using a simple proportioning procedure, Figure 4.1 serves as the basis for calculating extinction improvements associated with various degrees of diesel elemental carbon control. For example, if we assume 50% control from current levels, then the extinction changes are one-half of the values presented in Figure 4.1. Or, if we assume emissions grow by 60% to some future year with 50% control applied at that time, then the extinction changes are 80% of the values illustrated in Figure 4.1. For the purposes of the economic benefit analysis, we have digitized Figure 4.1 according to the Thomas Brothers grid squares using the same procedure as in Chapter 2.

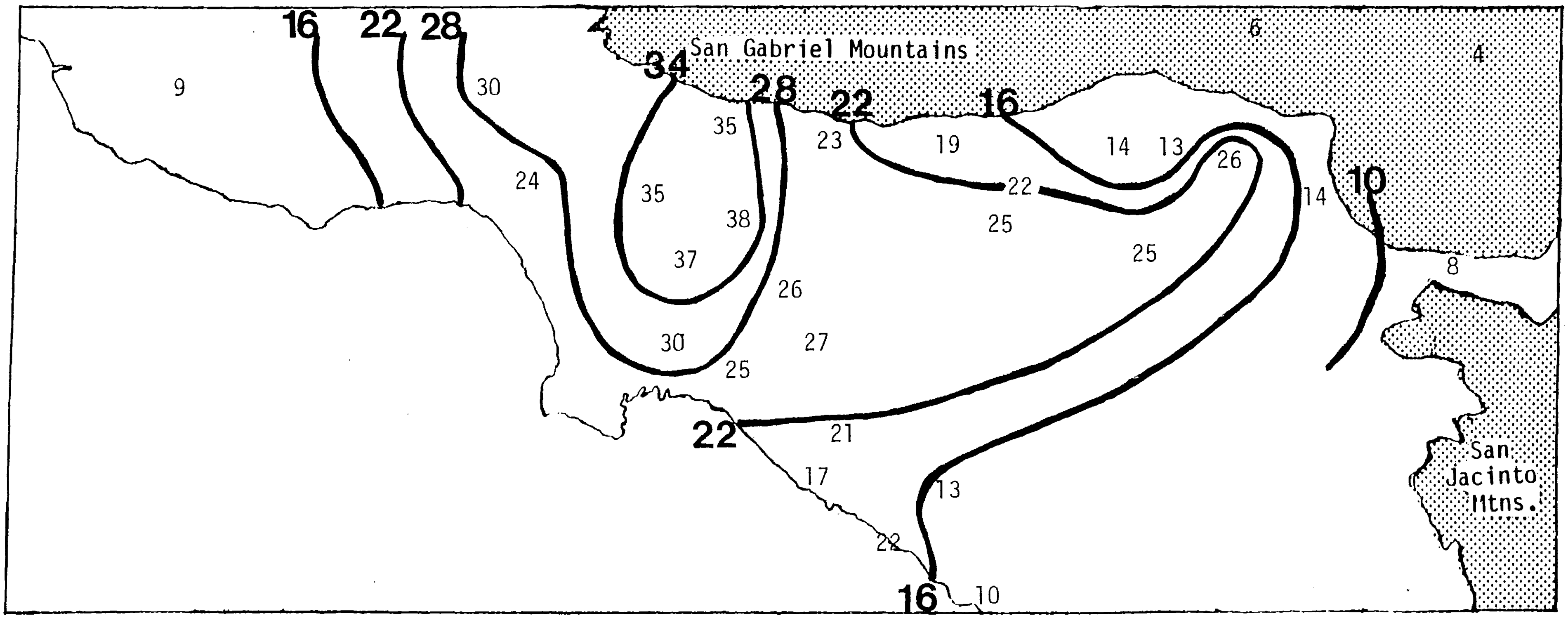

Figure 4.1 Current (1980) light extinction levels (Mm^{-1}) from diesel vehicle elemental carbon in Los Angeles.

As an aside, it is worthwhile restating that the above analysis does not include all of the visibility impacts from diesel emissions. In addition to the 12% light extinction contribution via elemental carbon, diesel vehicles currently contribute about 4% of total light extinction via NO_x emissions, 3% via SO_2 emissions, and 1% via other primary particle emissions (Trijonis, 1984). If one were considering a control strategy that reduced NO_x and SO_2 emissions as well as elemental carbon emissions (i.e. reduced diesel usage or methanol conversion), then the total visibility benefits would be significantly greater than those just from the elemental carbon reductions.

4.1.2 Projected Growth in Diesel Emissions

For the purpose of considering control strategy effects over time, we need to know future projections of diesel elemental carbon emissions. Trijonis (1983) has forecasted diesel emissions from 1980 to 1992 on a statewide basis for California. The projections used herein are based on his forecasts with two exceptions: (1) an update has been made on the assumptions regarding the fraction of gasoline vehicles that will be converted to diesel power, and (2) the statewide projections have been disaggregated to provide estimates of emission growth on a county basis.

Table 4.1 lists the fundamental assumptions regarding traffic growth and dieselization percentages. Basically, it is assumed that total traffic in California will increase by 2.6% per year from 1980 to 1992, with growth rates relatively higher in the heavier vehicle classes. Furthermore, it is assumed that 4% of the light-duty fleet, 10% of the medium-duty fleet, and 50% of the currently gas-powered heavy-duty fleet will become diesel-powered by the early 1990's. As noted in the table, these updated dieselization percentages differ significantly from earlier forecasts.

For the "no control" scenario, we adopt the following particulate emission factors compiled by Trijonis (1983):

light-duty cars and trucks . . . . 0.4 gm/mile,

medium-duty trucks . . . . 0.7 gm/mile,

heavy-duty-G trucks (smaller type that was gasoline powered in 1980) . . . . 1.3 gm/mile,

and heavy-duty-D trucks (larger type that was already diesel powered in 1980) . . . . .1.8 gm/mile.

TABLE 4-1. TRAFFIC GROWTH AND DIESELIZATION ASSUMPTIONS FOR PROJECTING STATEWIDE DIESEL EMISSIONS (Trijonis, 1983)

VEHICLE WEIGHT CLASS	CALIFORNIA STATEWIDE VMT* PROJECTION			DIESELIZATION PROJECTION (Fraction of VMT in vehicle class contributed by diesels)	
	1980 VMT by weight class	Average yearly VMT growth 1980 to 1992	1992 VMT by weight class	1980	1992
Light-duty cars and trucks (GVW* ≤ 6000 lbs)	88.5%	2.5%	87.0%	0%	4%[a] (10%)[b]
Medium-duty trucks (6000 < GVW ≤ 8500 lbs)	4.2%	3%	4.4%	0%	10%[a] (20%)[b]
Heavy-duty-G trucks (GVW > 8500 lbs, and were gasoline-powered type in 1980)	2.7%	4%	3.2%	0%	50%[a] (60%)[b]
Heavy-duty-D trucks (GVW > 8500 lbs, and were diesel-powered type in 1980)	4.6%	4%	5.4%	100%	100%
All vehicles	$150 \times 10^9 \frac{\text{miles}}{\text{year}}$	2.6%	$208 \times 10^9 \frac{\text{miles}}{\text{year}}$	4.6%	10.9%[a] (16.9%)[b]

* GVW = Gross Vehicle Weight, VMT = Vehicle Miles Traveled

a: Revised dieselization projections used herein. The revisions are based on updated information from the California Energy Commission (CEC 1984) and General Motors (Chock, et. al. 1983).

b: Original projection of Trijonis (1983).

For each vehicle class, we furthermore assume that 70% of these emissions are elemental carbon. The above "no control" emission factors pertain to average emission rates over the life of the vehicle and therefore include deterioration.

Based on the above assumptions, Table 4.2 presents projections for statewide elemental carbon emissions from diesel vehicles. It is forecasted that statewide diesel emissions will increase by a factor of 2.3, from 26.2 TPD in 1980 to 59.5 TPD in 1992. An elucidating way to view this projection is as follows: Of the 59.5 TPD of diesel elemental carbon in the early 1990's, 26.2 TPD (44%) already exists with the current 1980 heavy-duty diesel fleet. Another 16.7 TPD (28%) will be added from projected growth in use of these (ultra) heavy-duty vehicles. About 9.1 TPD (16%) will result from the 50% dieselization of heavy-duty vehicles that are currently gasoline powered. About 6.1 TPD (10%) will arise from the 4% dieselization of light-duty vehicles (GVW $\leq$6,000 lbs). Only 1.4 TPD (2%) will come from the 10% dieselization of medium-duty vehicles (6,000< GVW $\leq$ 8,500 lbs). It is especially notable that, even in 1992, the heavy-duty class will still account for 88% diesel fleet emissions.

For the purpose of the economic benefit calculations, we need to know the <u>spatial distribution</u> of diesel emission growth with the <u>Los Angeles</u> air basin. This has been obtained by factoring the statewide annual emission growth rate (7.1% from 1980 to 1992 from Table 4.2) according to county population growth relative to statewide population growth. This calculation indicates that diesel elemental carbon emissions (and related atmospheric extinction) will grow annually by 3.1%, 7.5%, 14.8%, and 14.5% in Los Angeles, Orange, Riverside, and San Bernardino counties, respectively.

As an aside, it is interesting to note the insensitivity of our economic benefit calculations with respect to the projected growth rate for diesel emissions. For example, we now forecast diesel elemental carbon emissions to increase by a factor of 2.27 from 1980 to 1992, whereas Trijonis (1983) previously forecasted an increase factor of 2.74 (because of his higher dieselization percentages). However, putting both forecasts into the economic calculations yields a difference in estimated control benefits of only about five percent! The reason is that projected emission growth is significant only in the future, but the future is discounted by appropriate interest rates in the economic calculations.

TABLE 4.2 FORECASTS OF STATEWIDE ELEMENTAL CARBON EMISSIONS BY DIESEL VEHICLE CLASS

VEHICLE CLASS	STATEWIDE ELEMENTAL CARBON EMISSIONS TONS PER DAY (PERCENT OF DIESEL TOTAL) 1980	1992
Light-duty cars and trucks	---	6.1 (10%)
Medium-duty trucks	---	1.4 (2%)
Heavy-duty-G trucks (were gasoline powered in 1980)	---	9.1 (16%)
Heavy-duty-D trucks (were diesel powered in 1980)	26.2 (100%)	42.9 (72%)
Total	26.2	59.5

4.1.3 A Side Issue -- Human Perception Thresholds

Experiments have demonstrated that a 5% change in light extinction is the approximate perception threshold for human observers viewing two pictures of the same scene simultaneously (one picture taken through a haze and one without the haze) (Malm, 1980; Trijonis, 1981). If, however, observers do not make simultaneous comparisons but must rely on memory, then the perception threshold corresponds approximately to a 10-13% change in light-extinction (Malm, 1979; Trijonis, 1981). Considering these results, it is evident that the forecasted average 1992 contributions of diesel elemental carbon to light extinction -- 27% based on the projections of Table 4.2 -- would be definitely perceptible. The current (1980) average visibility effect of diesel elemental carbon (12%) is also probably perceptible. If, however, one were considering 50% control of diesel particles in 1980, then the impact (6% extinction change) would be of questionable perceptibility.*

The above remarks skirt around an interesting side issue: "Does a control strategy have benefits if it fails to yield a perceptible change in visibility?" The correct answer is yes, a control strategy should be assigned its estimated benefits even if it does not yield a perceptible change. The alternative "No" answer leads to a logical inconsistency that the whole is not equal to the sum of its parts. For example, assume that 100 hypothetical pollution sources produce 1% each of the visibility reduction in a region. If we assign zero benefits to cleaning up a single source because the change is imperceptible, how could this be consistent with assigning a large benefit to cleaning up 20 sources? Each unit of emissions must be viewed as a non-zero part of the whole and must be assigned its "disbenefit" in proportion to its contribution. Another way of seeing the correct answer involves the basic interpretation of our hedonic estimation technique. The hedonic technique provides an estimate of underlying preferences and willingness to pay for visibility improvements as measured by integrated steps along an inverse demand curve. If the improvement in visibility is finite, then the step along the preference curve is finite, and benefits are finite.

---

*Actually, in this later case, because the impact of diesels varies significantly from day to day and place to place, there would be times and locations of perceptible impacts as well as times and locations of imperceptible impacts.

4.2 BENEFIT ESTIMATION

The diesel extinction contributions presented in the previous section are used in this section to estimate the benefits of diesel particulate control in Los Angeles. The benefit estimation depends upon the inverse demand curves developed in Chapter 3. Following the traditional Freeman-Rosen approach, these equations are mathematically integrated over the visibility improvement range to determine individual household benefits. Aggregate benefits are then obtained through summation over the relevant population.

A variety of control scenarios are examined, with calculated benefits representing the change from "no control" to the specified level of control. Benefits are computed over the 1980-1992 time period, at the community level, for various discount rates, with results expressed in terms of constant 1980 dollars. The time period corresponds to available data on the growth in diesel related emissions (see Section 4.1.2). The community level was chosen since it is the most disaggregated level at which accurate income data are available. Different discount rates are used to demonstrate the sensitivity of the results to interest rate assumptions. Constant dollars are used to eliminate the problem of forecasting future inflation levels.

4.2.1 A Preliminary Issue

Before proceeding to the actual benefit calculations, it is necessary to first dispense with a common misunderstanding concerning benefits derived from a hedonic analysis of housing values. The question that often arises in these studies is "How is it possible for everyone to experience benefits from a general air quality improvement?" That is -- "With no increase in region-wide income or population in-migration, how can everyone's house increase in value when air quality changes?"

In order to understand benefit estimation in the hedonic framework, one must first realize that the approach is <u>not</u> based on changes in house prices. Therefore, everyone's house price does not have to increase for benefits to exist. The common misperception is that air quality alters the demand for housing units (see Figure 4.2), causing the price of homes to increase (P_1 to P_2), and that this price change is the basis for benefit estimation. This is an <u>incorrect</u> perception of the hedonic technique.

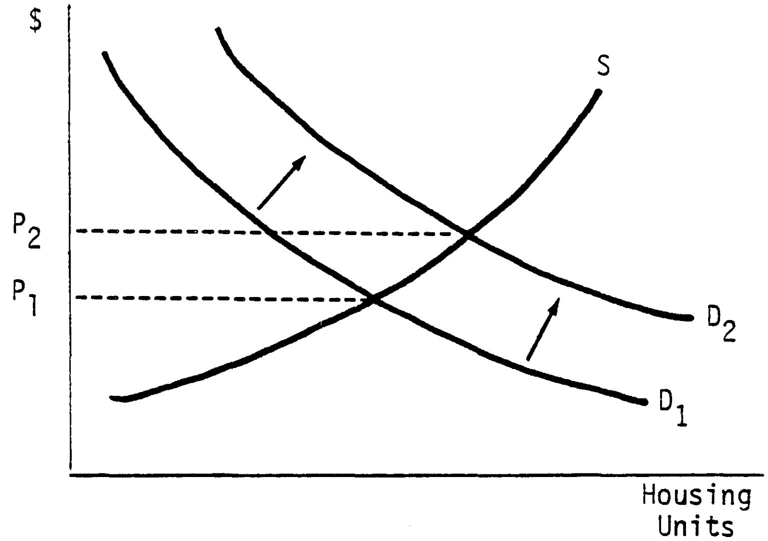

Figure 4.2 Demand and supply of housing units.

Rather than estimate the demand and supply of housing units, the Freeman-Rosen hedonic approach uses information on how consumers choose homes in order to reveal the parameters that determine consumer preferences. Thus, the hedonic approach is concerned with these underlying parameters, not the demand and supply of housing units.

The hedonic approach uses differentials in house prices to provide information on how consumers trade air quality for other goods (living area, location, etc.). With this information, the structure of the consumer's demand for air quality can be logically deduced (see Figure 4.3). Note that this is not the demand for housing units but the demand for only one aspect of the housing unit, the air quality aspect.

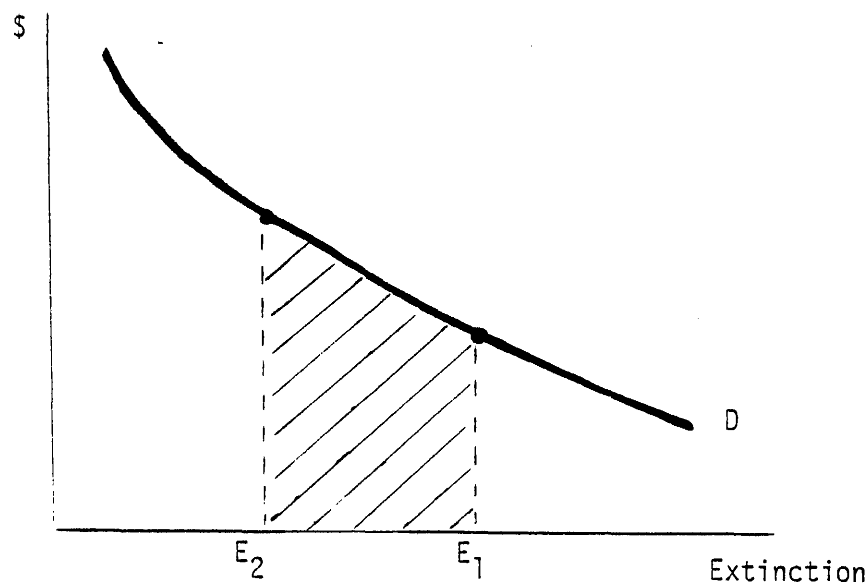

Figure 4.3 The demand for extinction.

The demand curve for air quality allows computation of the consumer's willingness to pay for changes in air quality. Willingness to pay is the accepted definition of benefits and is not the same as a change in house prices. In Figure 4.3, willingness to pay is taken as the area under the demand curve for a change in extinction (E_1 to E_2). Thus, even with an overall change in air quality, individuals are better off and are willing to pay positive amounts for this change. Our estimates of the (inverse) demand curves for extinction allow calculation of these benefits.

In conclusion, it can be stated that individuals are made better off even if home prices do not change. Improved air quality translates into a movement along the demand curve for air quality (movement to a greater satisfaction level). Of course, an increase in one's house price relative to all others would also make one better off. However, this is not necessary for a welfare improvement. Thus, an overall improvement in visibility with no increase in either income or population would produce positive benefits even though it might not affect prices.

4.2.2 Assumptions and Procedures

Household benefits are calculated herein by integrating the inverse demand curves presented in the previous chapter over the decrease in extinction for each control scenario. Since Los Angeles is the study area, the relevant equations are the Los Angeles inverse demand curves. Benefits are calculated for four different hedonic functional forms: extended Box-Cox, log-linear, classical Box-Cox, and semi-log.

The benefit equations (inverse demand curves) yield household benefits over the life of the house since capital values (house prices) are the units used in the hedonic analysis. The benefit figures for the first year of control are obtained by multiplying by the capital recovery factor.* Total first-year benefits are then determined by summing over all individuals in each community and then summing over all communities.

In the following year, benefits are determined in the same manner as the control measure continues. However, two adjustments are necessary. First,

---

*The capital recovery factor (CRF) is the rate which transforms a fixed amount into a series of equal annual payments which include principal and interest. $\mathrm{CRF} = \frac{r}{1-1/(1+r)^T}$ where r is the interest rate and T is the time horizon. For this study an interest rate of 9.5 percent and a 30-year time horizon were utilized.

growth in the region occurs, which increases the number of households over which benefits are summed. Second, because the benefits occur in the future, they must be discounted to reflect preferences for present over future returns.* Two different discount rates are used to demonstrate the sensitivity of the results. These are real discount rates rather than nominal (inflation affected) interest rates; inflation is not relevant, because benefits are expressed in constant 1980 dollars. Subsequent year benefits are calculated similarly. Finally, summing over all years yields the present value of the total stream of benefits.

A detailed outline of the specific calculation procedure is as follows. The first step is to calculate the benefits for the average household in each community. For instance, consider the inverse demand curve based on the log-linear hedonic equation. This can be written as

$$\text{MWTP} = 108.5 - 43.51 \times E + .136 \times Y, \qquad (4.2)^{**}$$

where MWTP = Marginal willingness to pay for extinction improvements in hundreds of dollars,

E = Extinction level in $(10^4 \text{ meters})^{-1}$,

and Y = Average household income.

The community level data required for benefit estimation under this equation are: (i) the existing (no control) extinction level; (ii) the new extinction level after control, and (iii) the average household income.

Benefits are determined by integrating Equation 4.2 over the change in extinction. The integrated form of Equation 4.2 can be written as

$$\text{TWTP} = 108.5 \times E - (43.51 \times E^2)/2 + .136 \times Y \times E. \qquad (4.3)$$

The total willingness to pay by the average household for an extinction improvement from E_1 to E_2 is $\text{TWTP}(E_1) - \text{TWTP}(E_2)$. As an example, if total extinction

---

*The discount factor (D) is calculated as $D = \frac{1}{(1+r)^t}$ where r is the discount rate and t is the time (years) from the beginning period.

**See Table 3.21.

is 2.0, diesel related extinction is .36, and proposed diesel control is 50 percent, then Equation 4.3 is evaluated at $E_1 = 2.0$ and $E_2 = 1.82$ ($1.82 = 2.0 - .5 \times .36$). Note that TWTP is actually a function of year "t", because uncontrolled diesel emissions are assumed to grow over time under some of the scenarios.

The calculated household benefits are in hundreds of dollars over the life of the home. The next step is to convert those to annualized figures and present values. This is accomplished by multiplying by the capital recovery factor (CRF) and discount factor (D_t), respectively. Thus, in year t,

$$\overline{TWTP}_t = TWTP_t \times CRF \times D_t, \tag{4.4}$$

where "$\overline{TWTP}$" represents an annualized, present value. Benefits for the entire community are then obtained by multiplying by the number of households in each community (NH_t),

$$\text{Community Benefits}_t = \overline{TWTP}_t \times NH_t. \tag{4.5}$$

The procedure outlined above is completed for each community (i = 1,...,N), and each year (t = 1,...,M). Then, the current value of total basin-wide benefits is determined by summing over all communities and all years.

$$\text{Basin Benefits} = \sum_{t=1}^{M} \sum_{i=1}^{N} \text{Community Benefits}_{it}. \tag{4.6}$$

4.2.3 Benefits of Diesel Particulate Control for Various Scenarios

In order to demonstrate benefit estimation within the hedonic framework, four different control scenarios are analyzed. In the first scenario, diesel emissions are assumed to be fixed over time, and a constant 50 percent particulate control level is considered over the entire 1980-1992 time period. All subsequent scenarios are more realistic in that uncontrolled diesel emissions are allowed to grow over time as estimated in Section 4.1.2. Scenario II considers 50 percent control of forecasted emissions starting in 1980. Fifty percent is approximately the degree of control being considered for state and national emission standards. Scenario III is similar to Scenario II, except that 80 percent control is assumed. An 80 percent reduction of primary particulate emissions might be achieved through very strict standards or through methanol conversion of diesels. (Actually, because methanol would reduce SO_x and NO_x emissions as well as particulate emissions, the total visibility benefits from a strategy of complete diesel methanol conversion would be nearly twice those of Scenario III.) Scenario IV considers the effect of phasing in controls by assuming that no emission reduction is imposed until 1985, and that control is phased in linearly up to 50 percent in 1992.

Scenario I

This scenario is the simplest of the four analyzed. The diesel related extinction levels presented in Figure 4.1 are assumed to remain constant over the period examined. Control is then set at 50 percent of these values. In Table 4.3, total benefits (present value) are presented for the 1980-1992 time period for two different discount rates and for the various functional forms. As is illustrated, total benefits range from approximately one to four billion dollars. The upper end of the range is the most reasonable, since extended Box-Cox and log-linear are the best functional forms (see previous chapter) and real rates of discount are likely closer to 3 percent than to 10 percent.

Scenario II

In this case, extinction is allowed to grow over the time as forecasted in Section 4.1.2. The control level in Scenario II is assumed to be 50 percent, starting in 1980 and maintained (at 50 percent of projected emissions) through 1992. As shown by the second column of Table 4.3, the estimated benefits range from 1.2 to 5.0 billion dollars. The benefits are greater than in Scenario I because greater amounts of forecasted emissions are controlled. Also, as expected, the relative differences between the scenarios are somewhat less for a 10 percent discount rate than for a 3 percent discount rate, because the growth in future emissions counts less with a 10 percent discount rate.

Scenario III

In this case, the implications of various degrees of control are demonstrated. As in the previous scenario, emissions are assumed to grow at the established rates. However, in this case a more stringent control program

TABLE 4.3 PRESENT VALUE OF BENEFITS FOR CONTROL OF DIESEL PARTICLE EMISSIONS IN LOS ANGELES ($ BILLIONS) 1980-1992

FUNCTIONAL FORM	SCENARIO I	SCENARIO II	SCENARIO III	SCENARIO IV
	no growth, 50% control	growth, 50% control starting in 1980	growth, 80% control starting in 1980	growth, 50% control phased in between 1985 and 1992
		Discount Rate = 3%		
Extended Box-Cox	4.01	5.03	8.45	1.92
Log-Linear	3.34	4.22	7.06	1.62
Classical Box-Cox	1.42	1.94	3.13	.80
Semi-Log	1.56	2.14	3.45	.88
		Discount Rate = 10%		
Extended Box-Cox	2.68	3.27	5.47	.96
Log-Linear	2.23	2.74	4.57	.81
Classical Box-Cox	.95	1.24	2.00	.40
Semi-Log	1.04	1.37	2.20	.44

is assumed, 80 percent rather than 50 percent.

As is expected, benefits increase with the more stringent control program. This occurs for two reasons. First, since more tonnage of emissions (more atmospheric extinction) is controlled, the benefits must increase. Because the control level is increased by 60 percent (50 percent to 80 percent), the benefits should increase by about 60 percent. Second, as more extinction is controlled, movement occurs to an area of the inverse demand curve with increasing willingness to pay for the marginal unit (see previous Figure 4.3). Will- Willingness to pay increases disproportionately with greater control because of the shape of the demand curve. The size of this effect varies with functional form. The log-linear and extended Box-Cox forms yield the largest effect, whereas the semi-log and classical Box-Cox forms yield a very small effect. For a 60 percent increase in control (50 percent to 80 percent), the log-linear and extended Box-Cox forms produce an increase in benefits of 68 percent, whereas the semi-log and classical Box-Cox forms increase benefits by 61 percent. Thus, the effect of moving up the demand curve accounts for approximately one to eight percent increased benefits over this range of the data. As one moves toward more pristine air quality, this effect would become even larger.

Scenario IV

This scenario represents an attempt to be more realistic than the previous cases. Two adjustments are made to the control assumptions of Scenario II. First, no control is imposed until 1985. This is consistent with the current situation, since there are no controls at the present time. Second, standards are assumed to have a minor impact initially with a subsequent larger effect. The control level is assumed to change linearly from 0 percent in 1985 to 50 percent in 1992. The estimated benefits for Scenario IV range from 0.4 to 1.9 billion dollars. As is expected, these values are significantly smaller than those of the other scenarios. This occurs for two reasons. First, the overall level of control for the 1980-1992 period is much less than in any previous scenario. Fifty percent control is not reached until the final year of the program, and in all prior years, controls are smaller than any other scenario. Second, discounting has a larger impact on Scenario IV. This occurs because the initial years, which are discounted least, have no attached benefits (zero control). When benefits do begin to occur, the discount factor is larger, thereby reducing the present value. The message is that earlier controls lead to significantly greater present value of benefits.

4.2.4 Concluding Remarks

Estimated benefits have been presented above for four different control scenarios. The various cases were chosen to demonstrate the applicability of the hedonic technique for estimating benefits of visibility improvements in the Los Angeles air basin. A number of discernible features can be identified in the results. First, the estimated equations indicate that increasing controls produces disproportionately larger benefits. However, as a comparison of the first two scenarios demonstrates, the disproportion is relatively small for marginal changes in extinction.

Second, the larger the control effort, the larger are the corresponding benefits. Of course, the costs of control also increase with the level of effort. Ideally, one would choose the optimal control effort through a comparison of the benefits and costs of each level of control.

The third result from the analysis arises from a combination of the first two factors. That is, the benefits of extremely stringent control should be very large. This occurs because 1) the larger the control the larger the benefits, and 2) as pristine air quality is approached, benefits grow at an increasing rate due to the shape of the demand curve.

The fourth feature is the substantial loss in benefits from postponing and phasing controls (i.e. Scenario IV). Under Scenario IV, the average control level over the 1980-1992 time period is substantially less than for the other scenarios. In addition, the controls of Scenario IV come mostly in the future, so that the present value of the benefits is reduced by the discounting procedure. The sooner the controls are in place, the larger the overall benefits.

In conclusion, the hedonic approach can be utilized to determine the benefits of air quality improvements under a variety of scenarios. Further, these benefits do not require either region-wide income increases or population in-migration, as they are unrelated to housing price increases. Rather, they are based on individual willingness to pay for improved air quality.

4.3 REFERENCES

Appel, B.R. et al., "Visibility Reduction as Related to Aerosol Constituents," Prepared for the California Air Resources Board, Interagency Agreement No. ARB A1-081-32, Sacramento, California, October 1983.

California Energy Commission (CEC), Personal communication with Leigh Stamets, Traffic projection for gasoline and diesel vehicles, December 1984.

Cass, G.R., P.M. Boone, E.S. Macias, "Emissions and Air Quality Relationships for Atmospheric Carbon Particles in Los Angeles," in Particulate Carbon: Atmospheric Life Cycle, G.T. Wolff, R.L. Klimisch, eds., Plenum Press, New York, 1982.

Chock, David P. et al., "Estimates of Diesel Particulate Concentration in Four Urban Areas under Different NO_x Emission Scenarios," GMR-4514, ENV #165, General Motors Research Library, Warren Michigan, October 1983.

Conklin, M.C., G.R. Cass, L-C Chu, and E.S. Macias, "Wintertime Carbonaceous Aerosols in Los Angeles: An Exploration of the Role of Elemental Carbon," in Atmospheric Aerosol: Source/Air Quality Relationships, E.S. Macias and P.K. Hopke, eds., ACS Symposium Series, 167, American Chemical Society, Washington, D.C., 1981.

Davidson, A., "Visibility in the South Coast Air Basin: Patterns, Relationship to Particulate Concentrations and Trends," Presented at the Annual Meeting of the West Coast Chapter of the Air Pollution Control Association, Scottsdale, Arizona, October 26, 1983.

Malm, Wm.C. et al., "Visibility in the Southwest," Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency, Las Vegas, Nevada, 1979.

Malm, Wm.C., EPA National Environmental Research Center, Las Vegas, Nevada, Personal communication of data, March 1980.

Trijonis, J., "Overview of Visibility Regulations," Prepared under contract to Los Alamos National Laboratory, 1981.

Trijonis, J., "Analysis of Visibility/Aerosol Relationships and Visibility Modeling/Monitoring Alternatives for California," Prepared for the California Air Resources Board, Contract No. A9-103-31, February 1982.

Trijonis, J., "Effect of Diesel Vehicles on Visibility in California," Proceedings of the Second International Conference on Carbonaceous Particles in the Atmosphere, Linz, Austria, 1983.

Trijonis, J., "The Effect of Diesel Vehicles on Visibility in Los Angeles," Presented at Methanol Symposium, Los Angeles, June 1984.

APPENDIX A: ECONOMIC BENEFIT ANALYSIS WITH THE 1973-74 DATA BASE

This appendix reports on the economic regression analyses for the 1973-74 time period. The primary motivation for examining this earlier time period was to study the possibility of preference shifts over time. The initial idea was to estimate demand functions for visibility (light extinction) in the two time periods and then to compare them to determine if a shift occurred. However, further inspection indicated that a simple comparison of these two demand equations would be insufficient to identify the source of the shift. Thus, the shift could be associated with a change in preferences, satisfying our initial hypothesis, or could result from a change in prices of the housing characteristics or changes in prices of goods outside our consideration. Without very restrictive assumptions concerning these other influences (i.e., all other prices constant) no definitive statement could be made. Since the 1973-74 analysis adds little to the information inherent in the 1978-79 data, this appendix has limited value in that the estimates provided are never used. However, in order to fulfill the terms of the contract, we do present hedonic estimates for the 1973-74 time period.

The definitions of the data utilized in the hedonic housing value equations are specified in Table 3.1 of the text. The estimated hedonic equations for the Los Angeles Area are presented in Tables A.1 and A.2. The former uses population density and school quality to represent community characteristics, whereas the latter uses principal components. For each of the equations presented the following generalizations hold. First, all variables except population density (significant at the 5% level) are significantly different from zero at the 1% level and possess the expected relationships to home sale price (HSP). Second, in each equation approximately 80% of the variation in HSP is explained by the variation in the independent variable set. Third, light extinction is significantly different from zero and is negatively related to HSP. This indicates that the 1973-74 data confirm the 1978-79 results. Also like the 1978-79 results, the 1973-74 results are stable across functional forms, sample sizes, and model formulations.

The estimated hedonic equation for the San Francisco area is presented in Table A.3. As in the Los Angeles case, the estimated coefficients generally possess the expected relationships to HSP and are significantly different from

TABLE A.1 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = ℓn(HOME SALE PRICE IN HUNDREDS OF 1973-74 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ln (Sales Month)	.0272	7.06
ln (Age of Home)	-.055	-17.04
ln (Square Feet of Living Area)	.627	47.21
ln (Number of Bathrooms)	.137	10.91
Number of Fireplaces	.085	14.91
Pool	.100	12.00
View	.183	13.11
Neighborhood Characteristics:		
ln (Percent Greater than 64)	.023	4.33
ln (Percent White)	.078	13.92
Location Characteristics:		
ln (Distance to Beach)	-.099	27.64
ln (Miles to Business District)	-.014	- 5.35
Orange County	-.156	-16.66
Riverside County	-.193	-11.11
San Bernadino County	-.172	-12.22
Community Characteristics:		
ln (Population Density)	-.016	- 1.71
ln (School Quality)	.579	11.96
ln (Light Extinction)	-.147	- 7.04
Constant	-1.1419	- 5.18
R-Squared	.80	
Number of Observations	4934	

TABLE A.2 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE LOS ANGELES AREA.

DEPENDENT VARIABLE = ℓn (HOME SALE PRICE IN HUNDREDS OF 1973-74 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ln (Sales Month)	.028	7.20
ln (Age of Home)	-.053	-16.34
ln (Square Feet of Living Area)	.632	47.66
ln (Number of Bathrooms)	.133	10.65
Number of Fireplaces	.084	14.67
Pool	.099	11.92
View	.187	13.36
Neighborhood Characteristics:		
ln (Percent Greater than 64)	.021	3.68
ln (Percent White)	.080	14.56
Location Characteristics:		
ln (Distance to Beach)	-.095	-25.92
Orange County	-.142	-15.49
Riverside County	-.166	-10.48
San Bernadino County	-.168	-13.50
Community Characteristics:		
Factor 1	-.033	-10.08
Factor 2	-.019	- 5.58
Factor 3	-.012	- 2.77
ln (Light Extinction)	-.162	- 7.63
Constant	1.139	11.93
R - Squared	.80	
Number of Observations	4934	

TABLE A.3 ESTIMATED HEDONIC EQUATION (LOG-LINEAR) FOR THE BAY AREA AIR BASIN.

DEPENDENT VARIABLE = ℓn (HOME SALE PRICE IN HUNDREDS OF 1973-74 DOLLARS)

Variables	Coefficient	t - Statistic
Site Specific Characteristics:		
ln (Sales Month)	.039	8.82
ln (Age of Home)	-.030	- 8.22
ln (Square Feet of Living Area)	.669	44.68
ln (Number of Bathrooms)	.115	8.83
Number of Fireplaces	.077	11.04
Pool	.058	4.31
View	.096	8.90
Neighborhood Characteristics:		
ln (Percent Greater than 64)	.062	10.12
ln (Percent White)	.139	14.43
Location Characteristics:		
ln (Distance to Beach)	-.161	- 9.87
Miles to Business District	-.00005	- .93
Alameda County	-2.96	-10.65
Contra Costa County	-2.88	-10.16
San Mateo County	-2.90	-10.31
Santa Clara County	-2.98	-10.50
Community Characteristics:		
ln (Population Density)	-.037	- 4.69
ln (School Quality)	.734	11.36
ln (Light Extinction)	.193	5.90
Constant	.754	6.47
R - Squared	.75	
Number of Observations	3527	

zero at the one percent level. The major exception is the light extinction variable; it is significant, but it has an unexpected relationship to home sale price. This result is contrary to our findings in Los Angeles and San Francisco in 1978-79. In addition, the poor performance of the extinction variables in San Francisco in 1973-74 is quite stable across functional forms, sample sizes, and model formulations. While this does not invalidate our 1978-79 result, it does suggest something is amiss.

There is no certain answer to why light extinction fails to perform properly in the 1973-74 time period in San Francisco. However, a number of possible reasons might be proposed. The first is the lack of variation in the light extinction variables in the earlier time period. For instance, the third extinction variable (sea haze adjusted) has a mean of .72 with a standard deviation of .14 in 1973-74. This contrasts to a 1978-79 mean of .66 with standard deviation of .20. Thus, there is more relative variation in the data in the later years. Second, inspection of the correlations between the visibility variables across the two time periods provides further evidence of data problems. For instance, the correlation between extinction (3) in 1973-74 and the same variable for 1978-79 is approximately .91 in Los Angeles. The corresponding figure for San Francisco is only .80. This latter figure suggests that either a shift in the visibility pattern occurred or the earlier data are poor and do not reflect the actual situation. It is our opinion that the earlier San Francisco extinction data may not reflect the true situation since values at two critical locales (Alameda and Moffett) seem out of character (see discussion in Section 2.1). In conclusion, the evidence suggests that the 1973-74 San Francisco extinction data may not provide a true test because of low inherent variation and, possibly, because of poor data quality.

The next step in the analysis would be to estimate demand curves for the 1973-74 time period. However, these results would not provide any further information than that gleaned from the above hedonic equation estimates. Again, the 1973-74 results would strongly support the 1978-79 results in Los Angeles but not in San Francisco. Our reasoning on why would remain the same, so this exercise is omitted. Therefore, we conclude that, since the assumptions necessary to comment on the presence of preference shifts over time are too restrictive, the 1973-74 results have use only as qualitative confirmation of the 1978-79 results. This occurs in the Los Angeles study area but not in San Francisco. The latter finding may be attributable to lower variation and poorer quality in the San Francisco data during 1973-74.