RISE | RICE INITIATIVE *for the* STUDY *of* ECONOMICS

RISE Working Paper 14-027

"Real Estate Values, Air Pollution and Homeowner Perceptions: A Hedonic Study"

by

Peter Hartley

RICE Department of Economics
Baker Hall, MS22
6100 Main Street, Houston, Texas 77005
https://economics.rice.edu

Real Estate Values, Air Pollution and Homeowner Perceptions: A hedonic study

Introduction

Among the class of major metropolitan areas in the United States, Houston, Texas has a reputation for being one of the most polluted.[1] Also, when compared to those in other metro areas in the U.S., the housing prices in Houston are below average.[2] While there could be several reasons for the lower prices including the lack of zoning laws, topography, or hot weather, we investigate whether ozone pollution within the city may be another reason for its relatively low housing prices. We do this by using a hedonic pricing model to investigate whether price differentials across neighborhoods reflect differences in pollution levels—an evaluation of pollution patterns indicates that the concentration of pollutants can vary widely across the city. Also, because we recognize that homeowners may not appreciate the spatial distribution of pollution, we conduct a survey to collect homeowners' expectations about pollution in their neighborhoods and others. The subjective measures are used in the same hedonic model to determine which calculation offers the better explanation of housing prices.

The main tenant of hedonic pricing theory is that "goods are valued for their utility-bearing characteristics." (Rosen, 1974) For example, the value of a home may depend on several structural factors such as square footage, number of bedrooms and bathrooms, quality of the heating/cooling system, or whether or not the house has a

[1] One could argue that the pollution problem may have been exaggerated—although Houston has recently displayed among the highest number of ozone exceedence days of all U.S. cities, it has been in compliance on all other pollutants monitored by the Environmental Protection Agency.
[2] According to the ACCRA Cost of Living Index, housing costs in Houston are 15 percent below average among the largest 23 metropolitan areas in the United States. www.accra.org.

garage, fireplace, or pool. While the list of structural characteristics of a house is conceivably endless, one must also consider other more nebulous characteristics that people value such as quality of the neighborhood public school, proximity to shopping districts, and air quality. The hedonic pricing method can be used to evaluate any characteristic of the composite good, in our case, housing.

The main focus of this paper—the effect of pollution levels on housing prices—is a common application of the hedonic pricing method.[3] Anderson and Crocker (1971), Wieand (1973), Deyak and Smith (1974), and Smith and Deyak (1975) perform early studies of the effect of air pollution on residential housing values. The results range from pollution having no effect on value to a large negative effect. Smith (1978) strays from traditional methods by estimating the premium paid for one location over another, while Nelson (1978) uses a two-step method suggested by Rosen for determining the demand for urban air quality in Washington, D.C. Diamond (1980) uses implications of bid-price theory to solidify the relationship between land values and amenities (including air pollution), while Krumm (1980) develops a spatial model of household and firm demand and supply for amenities.

Data and Model Specification

The data set of the present study consists of 4,235 homes from 10 neighborhoods in different geographical areas of Houston. The different neighborhoods were chosen because of their close proximity to ozone monitoring stations[4]. Price data, as well as the structural characteristics of the homes, were collected from the Harris County Appraisal

[3] In fact, Smith and Huang (1995) perform a meta-analysis of thirty-seven hedonic property value models that estimate the effect of air quality.
[4] The exercise could be completed with homes that are further from monitoring stations, but a spatial estimation technique would be required to accurately estimate the ozone level in the neighborhoods. As

District (HCAD). The structural characteristics of the homes in the data set are as follows:

size	square feet of living space
lot size	acres of land
age	age of the house in 2002
bedrooms	number of bedrooms
bathrooms	number of bathrooms
total rooms	number of primary rooms
recreation room	dummy=1 if there is a recreation room
stories	number of stories
A	grade is A[5]
B	grade is B
C	grade is C
D	grade is D
F	grade is F
very poor	condition is very poor[6]
poor	condition is poor
fair	condition is fair
average	condition is average
good	condition is good
very good	condition is very good
excellent	condition is excellent
pool	dummy=1 if there is a pool
shed	dummy=1 if there is a shed
canopy	dummy=1 if there is a canopy
carport	dummy=1 if there is a carport
detached garage	dummy=1 if there is a detached garage
fireplace	number of fireplaces
contemporary	dummy=1 if style is contemporary
ranch	dummy=1 if style is ranch
traditional	dummy=1 if style is traditional
aluminum vinyl	dummy=1 if exterior is aluminum vinyl
asbestos	dummy=1 if exterior is asbestos
brick veneer	dummy=1 if exterior is brick veneer
brick masonry	dummy=1 if exterior is brick masonry
frame or concrete block	dummy=1 if exterior is frame or concrete block
masonry frame	dummy=1 if exterior is masonry frame
shake shingle	dummy=1 if exterior is shake shingle
stucco	dummy=1 if exterior is stucco
stone	dummy=1 if exterior is stone
slab	dummy=1 if foundation is slab
crawl space	dummy=1 if foundation is crawl space
partial basement	dummy=1 if foundation is partial basement
a/c only	dummy=1 if there is only air conditioning
central heat and air	dummy=1 if there is central heat and air
central heat	dummy=1 if there is only central heat

long as homes within roughly 1.5 miles of the station are selected, there is no need to perform a spatial estimation of ozone level.

[5] A number value is assigned to each grade.

[6] A number value is also assigned to each condition. Both grade and condition are assigned by the assessor.

central air	dummy=1 if there is only central air
no heat or air	dummy=1 if there is no heat or air
attic	dummy=1 if there is an attic

The non-structural characteristics can all be viewed as measures of neighborhood quality, which is also a utility bearing characteristic and presumably lends to the value of a house. The non-structural characteristics we consider are as follows:

crime level[7]
elementary school quality
middle school quality
high school quality
distance to downtown[8]
proximity to neighborhood parks
proximity to schools
ozone level

A large data set was available for determining school quality—so large in fact, that the number of independent school variables exceeded the number of neighborhoods. Therefore, it was necessary to reduce the number of school variables while still retaining a measure of school quality. Principal component analysis was used to construct a quality index for elementary, middle and high schools.[9] For each type of school, the first two principal components are used. The first two principal components explain 46.3, 65.8, and 55.8 percent of the variation in elementary, middle, and high schools variables respectively. Given how the variables load into the first factor for the high schools, the first principal component seems to pick up white and Asian populations, high SAT scores, and participation in gifted programs. The second principal component for high

[7] Crime data was obtained from the APB/CrimeCheck rating system that was developed by Crimes Against Persons Index Inc. The ranking varies for different ZIP codes and takes into account the level of violent crime.

[8] The distance to downtown was calculated using the Yahoo Maps web site and the proximity to public parks and schools was determined by using the Houston Key Map. Specifically the park and school dummy variables denote homes that are within one block of a park or school, although alternative measures such as absolute distance and adjacency to parks and schools were also tested.

schools tends to pick up the Hispanic population and participation in "English as a Second Language" (ESL) programs. The same pattern holds true for the middle school principal components. However, the elementary school principal components are basically reversed in order. The first principal component picks up the Hispanic population, percentage of limited English proficiency students and participation in the ESL programs, while the second principal component picks up the white and Asian populations, as well as the participation in gifted programs. Incidentally, the second principal component for elementary schools only explains 16 percent of the variation in the elementary variables, and is dropped from every hedonic specification used later in the paper.

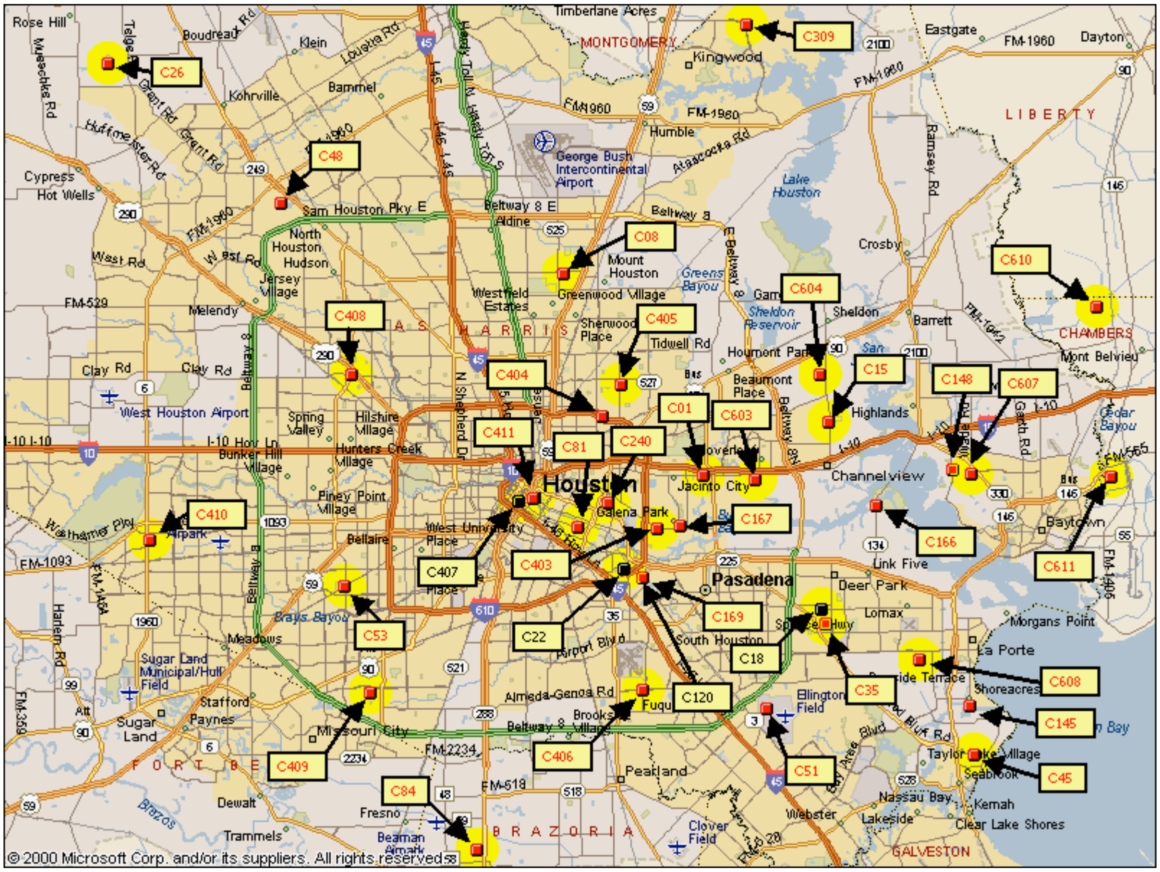
The values for ozone level are measured at various air quality measuring stations in the Houston metro area.[10] The system of stations in Houston is quite extensive, as can be seen in the following map. Most of the stations are concentrated in the industrial eastern side of Houston in order to carefully monitor the pollution emitted by petrochemical and petroleum refining plants in the area.

---

[9] The following variables were used in determining the index: attendance record, racial makeup, percentage of students in special programs, student-teacher ratio, years of teacher experience, percentage of economically disadvantaged students, and standardized test scores.

[10] Measurements for every hour of the day can be found at the Texas Commission for Environmental Quality web site, http://www.tceq.state.tx.us/.

Figure 1. Pollution Monitor Locations



Source: Texas Commission for Environmental Quality

The four monitoring stations used for this paper are C01 in east Houston, C08 in north Houston, C53 in southwest Houston, and C403 in southeast Houston. The various monitoring stations measure a number of variables including weather conditions and pollutants. However, not all stations monitor all weather and pollution conditions. Therefore, for the hedonic study, it was necessary to consider stations that were not only near residential areas, but also those that offered a measure of ozone concentration.

We consider ozone as the pollutant measure for Houston for two reasons: 1) Ozone is the pollutant for which Houston is most frequently in violation, and 2) residents in Houston tend to be most aware of ozone pollution because of the attention from the local press, as well as environmental groups.[11] Once the pollutant itself has been selected, it is important to choose the specification of that pollutant. Many studies use average monthly or yearly pollution levels, while some use the second highest reading as the measure. It might also be interesting to consider the daily average, maximum daily reading from a station, or perhaps the average value at a certain time of day. Since ozone levels in Houston generally peak from one o'clock to four o'clock in the afternoon, we also measure pollution as the average value at two o'clock on weekday afternoons. For this study, ozone readings from the four stations were available from March 1998 to November 2002.

Through a survey of the literature, we found that the variables used for the current exercise offer more detail than the sets employed in most studies. The only study containing a more detailed set of structural variables was Mieszkowski and Grether (1974), which included information about the type of flooring, plumbing, electrical systems, storm windows, insulation, and size of the garage. Mieszkowski and Grether obtained their data from a multiple listing service, while the data in the current paper comes from the county appraisal records.

The set of neighborhood characteristics used in the current paper also is very thorough when compared to other studies. However, there are some neighborhood variables that this paper does not include. For instance, Krumm (1980) and Linneman (1980) both include measures of the quality of streets, noise and visibility of garbage. Other studies include demographic variables such as the percentage of non-white

[11] Ozone watches—days for which ozone concentration is forecasted to exceed 125 parts per billion—are well-publicized in Houston. Television stations and newspapers include the "ozone forecast" along with the weather forecast. Also, road signs along the freeways alert drivers on high ozone days. Resident awareness regarding ozone levels will play an important role in the survey of pollution expectations.

residents (Smith and Deyak (1975), Nelson (1978)), average income (Wieand (1973), Deyak and Smith (1974)), population density (Nelson (1978), and percentage of workers participating in various types of jobs (Palmquist (1984), Cropper et al. (1988)).[12]

As with any econometric study, it is crucial to select the appropriate functional form for the regression. Bender, Gronberg, and Hwang (1980) show that the choice of functional form can greatly affect the hedonic price estimation and the resulting demand analysis. However, after testing various specifications, including linear, log-linear, and semilog, we use a Box-Cox (Box and Cox, 1964) specification in which both independent and dependent variables are transformed. The Box-Cox functional form allows for transformation of dependent and/or independent variables into the following form,

$$g^{(\lambda)}(x) = \frac{x^{\lambda} - 1}{\lambda},$$

where x is the variable in question and λ is determined by maximum likelihood estimation. Despite the value of λ, the Box-Cox transformation allows us to express the function in an intrinsically linear form, $g(y) = \beta x + \varepsilon$. The Box-Cox used in this paper is expressed

$$v_i^{(\theta)} = \alpha + \beta x_i^{(\lambda)} + \gamma z_i + \varepsilon,$$

where $v_i^{(\theta)}$ is a vector of transformed house values, $x_i^{(\lambda)}$ is a vector of transformed structural and neighborhood characteristics, and z_i is a vector of untransformed structural

[12] As we noted above, the total number of neighborhood variables that can be included in the study is limited by the number of neighborhoods represented in the data set. In addition, one could argue that the school variables we have included are likely to capture some of the neighborhood demographic and socio-economic variables.

and neighborhood characteristics.[13] The specification in which both left-hand and right-hand side variables are transformed is used because the values obtained for both θ and λ are statistically different from 0 and 1, therefore suggesting that the linear and logarithmic specifications can be improved upon.

While many studies have used a flexible Box-Cox functional form, the Box-Cox is restrictive as it supplies only a local approximation to the true function. Perhaps more importantly for our purposes, it restricts the environmental variable to enter the function in the same form as all other structural variables (Palmquist 1991, p. 88). Cassel and Mendelsohn (1985) state that since the environmental quality variable is likely to explain only a small part of the house price, it will therefore play a small part in determining the parameter in the Box-Cox form. They suggest that the Box-Cox will force a simple environmental variable into a more complex function, thus resulting in untrustworthy estimates. However, in support of the Box-Cox method, Cropper, Deck, and McConnell (1988) test the effectiveness of different functional forms with both correct and incorrect specifications. While the linear and quadratic Box-Cox functions produced the lowest errors when all attributes of houses are observed, the linear and linear Box-Cox functions perform best when attributes are unobserved or replaced by proxies – a situation that is likely, even with excellent data collection.

<u>Results of the Hedonic Model</u>

As was mentioned in the previous section, there are several choices for the ozone measure to be used in a hedonic pricing model. We use the average daily value, as it

[13] The z_i vector includes all binary variables, as well as variables that may have negative values, such as the school principal component values. In order for the transformed variable to be defined for all values of λ, the initial value, *x* must be strictly positive.

offered the highest log likelihood of the five measures in the Box-Cox specification. Table 1.1 reports the results of eliminating insignificant variables from the Box-Cox specification. In the interest of space, only neighborhood variables are presented—the coefficients of the structural housing variables show the expected signs, and remain consistent in various models tested.[14] Note that the Box-Cox model in Table 1 solidly rejects the hypotheses that θ and λ are equal to negative one, zero, or one, suggesting it would be less appropriate to model left-hand or right-hand side variables as reciprocals, natural logs, or linear values.

Table 1 Box-Cox Model for House Value – Insignificant Variables Excluded

Number of Observations	4235
LR chi-squared	13285.21
Prob>chi-squared	0.000
Log likelihood	-42666.22

Transform Value	Coefficient	Standard Error	z	P>|z|
θ	.1886	.0198	9.55	0.000
λ	.4387	.0445	9.85	0.000

Nontransformed variables	Coefficient	chi-squared	P>chi-squared
hspc1	.3068	30.21	0.000
hspc2	.4238	40.69	0.000
mspc1	-.2354	91.75	0.000
mspc2	-.606	143.79	0.000
espc1	-.374	98.4	0.000
park dummy	-.258	30.62	0.000
constant	15.5		

Transformed Variable	Coefficient	chi-squared	P>chi-squared
downtown	-2.37	360.07	0.000
average ozone	4.01	343.04	0.000

Hypothesis Tests	chi-squared	p>chi-squared
$\theta = \lambda = -1$	2692.7	0.000
$\theta = \lambda = 0$	133.64	0.000
$\theta = \lambda = 1$	2671.52	0.000

[14] We also tested the effect of leaving the ozone variable untransformed, but the resulting log-likelihood

The dummy denoting homes that are within one block of a park has a negative coefficient, which at a first glance, is a somewhat unexpected result. However, this may not be anomalous, as parks can be a nuisance because of the people and cars they may attract. All of the school principal component terms stay in the regression, except for the second principal component for elementary schools. This is not surprising, since this component only explained 16 percent of the variation in the school variables. By applying the coefficients to the principal components associated with each school we find that all high schools except McArthur (in north Houston) lend positively to the value of the homes in the neighborhoods. Bellaire High School (southwest Houston) has the largest effect on house value, adding $4,596 to the value of homes zoned to the school. However, the middle schools in the sample do not increase the value of the neighborhood homes, with the exception of Fondren Middle School in southwest Houston, which adds $1,657 to the value of the homes zoned to the school. For elementary schools, four of the nine schools add to house values in their respective neighborhoods. The elementary school associated with the largest positive effect on housing values is Deer Park Elementary, which adds $2,976 to house value. The elementary school zoned for southwest neighborhoods also had a positive effect, which means that the southwest neighborhood is the only one for which all three schools lead to an increase in house value.

Of the transformed variables, the distance from downtown Houston displays the expected negative coefficient. However, the coefficient for the average ozone level has a positive and significant coefficient, which is opposite what we would normally expect.[15]

---

indicated that it should be transformed with the other variables.

[15] This result obtains regardless of the measure used for ozone.

A positive and significant coefficient on ozone level was also found when land value was regressed on the neighborhood characteristics and size of the lot. There are a few possible explanations as to why the ozone variable has this positive sign.

First, homeowners in Houston may be unaware of the ozone levels in their neighborhoods. The "ozone alerts" publicized through the media are done so on a city-wide basis, that is, alerts are not issued for particular parts of the city even though ozone levels in various areas can differ significantly depending on traffic patterns or wind direction. In addition, it seems to be a common opinion among residents in the western part of Houston[16] that the east and southeast parts of the city are "dirtier", and therefore are perceived to be more polluted than other parts. This is most likely due to the presence of petroleum refining plants in the east and southeast that produce byproducts that sometimes have an unpleasant odor. It is possible that residents mistake such byproducts for ozone pollution, and therefore believe that neighborhoods without such odors are cleaner – at least in terms of ozone pollution.

However, ozone is a secondary pollutant that is created in the reaction of nitrous oxide pollutants (most commonly produced by cars) and volatile organic compounds (commonly produced by a single tree in small amounts and also by vehicles, but produced in concentrated amounts by petrochemical plants) in the presence of solar radiation (sunlight). If there are many VOC producing trees in an area, there will be a higher background level of VOCs.

The west side of Houston, which is a high volume traffic area (and the location of many high-price neighborhoods), therefore tends to exhibit ozone levels as high as the southeastern areas. A lack of awareness of residents regarding ozone level combined with the fact that the ozone tends to form on the west side of the city may then explain the positive coefficient on the ozone measure in Table 1.

Another interesting potential explanation of why high-priced neighborhoods may exhibit higher ozone levels than low-priced neighborhoods involves the number and types of trees in the neighborhood—oak trees release isoprene, a naturally occurring substance that contributes to ozone production.[17] Essentially, the nitrous oxides released from cars react with volatile organic compounds (VOCs), both natural and man-made, and sunlight to create high ground level ozone concentrations. In Houston, it has been estimated that 67 percent[18] of total VOCs are naturally occurring, rather than man-made, with the main producers being oak trees.[19]

Thus a feature that might positively contribute to the value of a neighborhood – the abundance of trees – may indeed lead to higher ozone levels in the neighborhood. However, Matt Fraser, a faculty member in the Civil and Environmental Engineering Department at Rice University, downplays the contribution of trees to the ozone problems in Houston.[20] He states that while emissions of biogenic hydrocarbons from trees do elevate the background concentration of ozone, biogenic emissions alone do not

---

[16] This will be apparent when we present the homeowner survey.

[17] For the chemistry-minded, see Zhang and Zhang (2002), Zhang et al. (2002), Wesely and Xu (2002), and Lamb et al. (1993).

[18] Galveston-Houston Association for Smog Prevention (1999). However, the national figure computed in Lamb et al. (1993) estimates that approximately half of total VOC emissions in the U.S. are attributable to natural sources.

[19] However, before we chop down all of our beautiful oak trees, it is important to note that VOCs are necessary but not sufficient for the formation of smog. Controls on automobile emissions would remove another essential ingredient for smog formation. Furthermore, modeling of natural VOCs is not yet at exact science – often times the modeled levels are higher than measured levels for various reasons. Additionally, trees can play a positive role in reducing pollution. A Department of Forestry study of Chicago determined that a single tree can actually remove carbon dioxide, ozone, sulfur dioxide, particulate, and nitrogen dioxide pollution from the air. An Urban Ecosystem Analysis (UEA) study of Austin, Texas, found that tree cover saves millions of dollars in pollution-control devices.

[20] Dr. Matt Fraser, Department of Civil and Environmental Engineering, Rice University, personal communication, January 30, 2003.

cause the violations of EPA standards and do not enter into discussion of policies to reduce ozone levels around the city. This is primarily because the emissions from trees are dispersed throughout the city and surrounding areas, while emissions from petrochemical plants are concentrated in certain parts of the city. Petrochemical plants are the dominant producers of VOCs that lead to ozone formation. Nonetheless it is important to consider characteristics of neighborhoods that could directly lead to higher ozone levels while at the same time raising house values.

There is related possible interpretation of the unexpected positive coefficient on ozone levels. There could be an underlying utility-bearing characteristic of neighborhoods that *indirectly* leads to higher ozone levels, which has been omitted from the current data set. This characteristic may increase the popularity of a particular neighborhood, resulting in two outcomes: 1) prices will increase, and 2) traffic in the area will increase, presumably leading to higher ozone concentrations in the neighborhood. In the case of Houston, this is not an unreasonable scenario. Growth tends to be toward the west side of the city, resulting in high prices and heavy traffic, as well as high ozone levels.

Of course, one of the primary objectives of any regression model is to determine the economic impact on the dependent variable of a change in an independent variable. In the case of the Box-Cox transformation in Table 1, the elasticity for ozone level, evaluated at the mean of house value and average ozone level, is 1.903. Thus a one percent increase in the average ozone level is associated with a 1.903 percent increase in house value. Alternatively, at the mean values, an increase in average ozone level of .24

parts per billion is associated with a $1,659 increase in house value (less than one standard deviation).

Determining Homeowner Awareness

To evaluate the hypothesis that individuals do not appreciate the spatial distribution of ozone pollution in Houston we conducted a survey of homeowners in the southwestern part of the city—one of the neighborhoods evaluated in the hedonic pricing model.[21] The survey attempts to determine the awareness of residents about the probability of ozone alerts in various neighborhoods given weather conditions. We shall also use it later to measure the relationship between subjective expectations of ozone pollution and property values.

A problem with asking subjects to answer survey questions with probability values is that people often provide statistically incoherent responses. A simple example illustrates the concept: Suppose that a subject predicts that during a summer day in his or her neighborhood the probability of an ozone alert is .8. Then suppose that the same subject predicts that the probability that the temperature will be above ninety degrees *and* that there will be an ozone alert is .85. The subject, knowingly or not, has assigned a higher probability to the conjunctive statement than to the simple statement. This is inconsistent with the laws of probability and the set of probability statements is said to be *incoherent*. The problem commonly arises with conditional probability statements. These are useful, however, since people often have stronger intuitions about conditional probabilities than about elementary events. Nevertheless, when probability statements are incoherent, there is no unique way to represent subjective expectations. For the perceptions of the participant to be meaningful, the statistical incoherence must be corrected. In this paper we utilize an optimization method introduced by Batsell et al. (2002) to transform the incoherent observations into coherent ones with minimal distortion.

The positive and significant coefficient on ozone level led us to hypothesize that residents may be unaware of the relative levels of ozone pollution in Houston neighborhoods. Two measures of predictive accuracy will be employed in an attempt to ascertain the degree of awareness, if any, possessed by homeowners. We then will use subjective expectations for ozone alerts as another potential variable to explain house prices. With results of both hedonic regressions in hand, it is possible to compare the objective value of air quality versus the subjective value for determining housing prices. The insights to be gained from such a comparison are potentially fascinating. For instance, if the imputed effect of air quality on housing values is negative, the conclusion is that residents are concerned about perceived pollution levels and the potential effects, but their expectations are nevertheless at variance with reality (even though, as we will later observe, residents indeed have some knowledge about pollution in various neighborhoods). However, if the imputed effect of subjective expectations is higher than the objective expectations, then something else that is correlated with ozone levels must be driving housing prices and overwhelming any negative impact of higher perceived pollution levels.

Survey Method and Results

Using the survey, we want to test subjects (homeowners) on the probability of an ozone watch in various neighborhoods given particular weather conditions. Subjects were either surveyed regarding conditions at 2:00 p.m. on May 16th (18 subjects) or July

---

[21] Research supported by NSF grant ISS-9978135.

16th (23 subjects). Specific dates were chosen so that the prediction of the subjects could be compared against the "truth" for those days, allowing us to assess how knowledgeable subjects appear to be. Ten *simple statements* (henceforth called "variables") are associated with each date.

For May 16th, 2:00 p.m., the ten variables are as follows:

1. There will be an ozone watch in your neighborhood.
2. There will be an ozone watch at Hobby Airport.
3. There will be an ozone watch at the Medical Center.
4. There will be an ozone watch at the ship channel.
5. It will be sunny or mostly sunny at Hobby Airport.
6. The temperature will be at least 80 degrees at Hobby Airport.
7. The level of relative humidity will be at least 60% at Hobby Airport.
8. The wind speed will be at least 10 miles per hour at Hobby Airport.
9. Wind at Hobby Airport will be blowing primarily from the northwest.
10. Visibility will be 1 mile or LESS at Hobby Airport.

For July 16th, 2:00 p.m., the ten variables[22] are as follows:

1. There will be an ozone watch in your neighborhood.
2. There will be an ozone watch at Hobby Airport.
3. There will be an ozone watch at the Medical Center.
4. There will be an ozone watch at the ship channel.
5. It will be sunny or mostly sunny at Hobby Airport.
6. The temperature will be at least 85 degrees at Hobby Airport.
7. The level of relative humidity will be at least 50% at Hobby Airport.
8. The wind speed will be at least 7 miles per hour at Hobby Airport.
9. Wind at Hobby Airport will be blowing primarily from the northwest.
10. Visibility will be 1 mile or LESS at Hobby Airport.

Each subject is asked to assign a probability to each of the simple statements above (depending on the month for which he or she is selected). Then the subject is asked to assign probability to thirty-six *complex statements* (events) that may take any of six forms. If *p* and *q* are each variables, the complex statements are forms show in Table 2.

---

[22] The variables differ slightly in the two months because of different weather conditions. Note that the threshold value for temperature in July is 85 degrees instead of 80 degrees. Similar changes exist for humidity and wind speed.

Table 2. Descriptions of and Notation for Complex Events

Description	Notation
Conditional statements of the type *p* assuming *q*	$(p \mid q)$
Conditional statements of the type *p* assuming *not q*	$(p \mid \neg q)$
Conjunctions of the type *p* and *q*	$(p \wedge q)$
Conjunctions of the type *p* and *not q*	$(p \wedge \neg q)$
Disjunctions of the type *p* or *q*	$(p \vee q)$
Disjunctions of the type *p* or *not q*	$(p \vee \neg q)$

Since there are ten simple statements in our survey, there are ninety (10x9) possible complex statements of each type listed above. The subject is asked to assign probability to a *random selection* of thirty-six complex events, that is, six of each type. After combining the thirty-six complex events with the ten elementary events, each subject was asked for forty-six probability judgments.

Subjects were given a handout explaining the survey and were paid $25 to complete the online questionnaire, which took about thirty minutes. Residents from only one neighborhood were surveyed, and that neighborhood is located in southwest Houston. The web site itself reminded the participants of information included in the handout and then proceeded to present the forty-six events for probability estimation. All participants were constrained to select probability values on the interval [0,1].

For this study, we are primarily interested in the homeowners' evaluations of ozone levels in their neighborhood and others. We survey participants about four areas of the city: their own neighborhood (southwest Houston), Hobby Airport (south

Houston), the Medical Center (close to downtown), and the ship channel (southeast Houston).

Since we want to determine how a homeowner perceives pollution in his or her own neighborhood as compared to others, it is meaningful to look at the probabilities that subjects assigned to the simple events involving ozone levels. Table 3 provides the average probabilities assigned by homeowners in a southwest Houston neighborhood (Meyerland). Standard errors are always reported in parentheses next to the mean, unless otherwise stated.

Table 3. Probabilities Assigned to Simple Ozone Events

Event to which probability is assigned	May Subjects	July Subjects
There will be an ozone watch in Meyerland.	.31 (.241)	.432 (.314)
There will be an ozone watch at Hobby Airport.	.424 (.293)	.517 (.288)
There will be an ozone watch at the Medical Center.	.366 (.236)	.488 (.276)
There will be an ozone watch at the ship channel.	.434 (.284)	.579 (.275)

For the May observations, we find that no average probability assigned to an ozone event differs significantly from any other average probability assigned to another ozone event. According to the t-values, the comparison that is closest to displaying a significant difference in means is the comparison of Meyerland to the ship channel ($t(34)=1.4$). This is not surprising since—at least anecdotally—residents seem to view the ship channel area as the “dirty” part of the city, while they view the southwest part as more “clean”.

For the July subjects, the average probability assigned to the simple event that an ozone watch will occur in the ship channel is significantly higher than the average probability that an ozone watch will occur in Meyerland (t(44)=1.693, p<0.1). However, for the other ozone events there are no significant differences in the average probabilities assigned.

However, there are seven types of events, including the simple event (*p*), for which subjects are asked to assign probability. The average probability assigned by subjects to each type of event is reported in Table 4, along with standard deviations in parentheses. In fact, standard deviations are always reported in parentheses next to the mean, unless otherwise stated.

Table 4. Mean Probabilities Assigned to All Types of Events

Type of event	May subjects	July subjects
p	.499 (.3)	.584 (.321)
$(p \mid q)$	.455 (.309)	.548 (.31)
$(p \mid \neg q)$	.443 (.314)	.466 (.307)
$(p \wedge q)$	.42 (.308)	.49 (.293)
$(p \wedge \neg q)$	.439 (.325)	.399 (.267)
$(p \vee q)$	.558 (.3)	.64 (.294)
$(p \vee \neg q)$	.501 (.303)	.544 (.291)

We would expect, in accordance with the principles of probability, that probability estimates for simple events would be higher than estimates for conjunctive events and smaller than those for disjunctive events. T-tests for the May subjects reveal that the average probability assigned to the simple event is significantly higher than the average probability assigned to the conjunctive event (t(286)=2.12, p<.05), while there is no significant difference between the average estimate for the simple statement and the average estimate for the disjunctive statement (t(286)=1.55). This is an initial indication of the incoherence that infects subjects' judgments. For July subjects, again, the average probability assigned to the simple event is significantly higher than the average probability assigned to the conjunctive event (t(368)=2.81). For this sample, the average estimate assigned to the disjunctive event is significantly higher than the estimate assigned to the simple event (t(366)=1.67, p<0.1).

We are also interested in the statistical coherence of the subjects' responses. We use the following four laws of probability to judge whether a particular probability assigned to a conjunctive or disjunctive event is coherent.

1. $\Pr(p) + \Pr(q) - 1 \le \Pr(p \wedge q) \le \min\{\Pr(p), \Pr(q)\}$

2. $\Pr(p) - \Pr(q) \le \Pr(p \wedge \neg q) \le \min\{\Pr(p), 1 - \Pr(q)\}$

3. $\max\{\Pr(p), \Pr(q)\} \le \Pr(p \vee q) \le \Pr(p) + \Pr(q)$

4. $\max\{\Pr(p), 1 - \Pr(q)\} \le \Pr(p \vee \neg q) \le \Pr(p) - \Pr(q) + 1$

The average number of "misses", that is, incoherent responses to conjunctive and disjunctive events, is shown in Table 5. For each type of event, there are six opportunities to provide an incoherent probability.

Table 5. Average Number of Coherence Violations

Type of Event	Violations by May Subjects	Violations by July Subjects
$(p \wedge q)$	3.94 (1.66)	4 (1.76)
$(p \wedge \neg q)$	4.17 (1.29)	4.52 (1.47)
$(p \vee q)$	3.94 (1.26)	3.7 (1.33)
$(p \vee \neg q)$	4.33 (1.46)	4.61 (1.44)

Correcting for Statistical Incoherence

We have determined that incoherence is present in the set of responses given by the subjects, as would be expected in any such study—even experts are not immune to assigning incoherent estimates to complex and conditional events, according to Deines et al. (2002). Using an algorithm developed by Deines et al. (2002), we can adjust the set of probability estimates to ensure that statistical coherence is achieved, with the divergence from the original estimates minimized. The idea is that the original judgments may contain genuine insights of the subjects which could be lost if we deviate too much from the values that people provide. An abbreviated discussion of the optimization problem and technique is provided in current paper, and is heavily borrowed from Deines et al. (2002), pages six through eleven. A more rigorous and detailed explanation is available in Batsell et al. (2002).

We will consider the function *Plaus* (meaning "plausibility") to represent the probability that the subject assigns to a certain event. For example, $Plaus(\varphi) = x$ is equivalent to the subject saying "the probability of φ is *x*," while $Plaus(\chi \mid \psi) = y$ is equivalent to the subject saying "the probability of χ assuming ψ is *y*." Note that $Plaus(\varphi) = x$ refers to all types of events except for conditional events (those of the "assuming if" form). We consider *Plaus* to be coherent if the numbers generated by the function agree with those generated by a probability distribution. Since this is generally not the case, it is necessary to "reconstruct" the function through the following optimization problem:

> Let *Plaus* map formulas $\varphi_1 \cdots \varphi_k$ and pairs of formulas $(\chi_1, \psi_1) \cdots (\chi_j, \psi_j)$ into [0,1]. Find a (coherent) distribution *Pr* such that
>
> $$\sum_{i \le k} \left| Plaus(\varphi_i) - \Pr(\varphi_i) \right| + \sum_{i \le j} \left| Plaus(\chi_i \mid \psi_i) - \Pr(\chi_i \mid \psi_i) \right|$$
>
> is minimized.

In our case, k=34, since φ_i includes elementary, conjunctive, and disjunctive statements, while j=12 since $(\chi_i \mid \psi_i)$ refers to the twelve "assuming that" statements. The proximity measure in the optimization problem is absolute deviation. We could also use squared deviation, although Deines et al. (2002) choose to use the absolute deviation approach for algorithmic reasons.

In order to determine a probability distribution that is as close to the original estimates as possible, yet coherent, we use a simulated annealing search technique (van Laarhoven, 1988) applied to probability arrays. That is, the technique searches for a probability array whose probabilities for a certain body of estimates provide a minimum absolute deviation, as outlined in the optimization problem above. We abbreviate the technique as SAPA. Table 6 includes the average probability estimates provided by the subjects (shown in Table 4), and the corresponding corrected and coherent estimates provided by the SAPA method.

Table 6. Average Probability Estimates and Average Corrected Probability Estimates

Type of Event	May Subjects	May Subjects Corrected	July Subjects	July Subjects Corrected
p	.499 (.3)	.491 (.226)	.584 (.321)	.56 (.25)
$(p \mid q)$	.455 (.309)	.456 (.284)	.548 (.31)	.538 (.277)
$(p \mid \neg q)$	.443 (.314)	.457 (.279)	.466 (.307)	.472 (.288)
$(p \wedge q)$	.42 (.308)	.282 (.208)	.49 (.293)	.36 (.221)
$(p \wedge \neg q)$	.439 (.325)	.287 (.186)	.399 (.267)	.286 (.195)
$(p \vee q)$	.558 (.3)	.702 (.192)	.64 (.294)	.748 (.208)
$(p \vee \neg q)$	.501 (.303)	.68 (.185)	.544 (.291)	.7 (.192)

T-tests show, as we would expect for *coherent* corrected estimates, that for both sets of subjects the average probability assigned to simple events is higher than that assigned to conjunctive events and lower than that assigned to disjunctive events.

It is also interesting to determine whether the coherent estimates are statistically different from the corresponding subjective (non-corrected) estimates. For both the May and July subjects, all of the coherent conjunctive and disjunctive estimates are significantly different than the subjective estimates. In addition, for the July subjects, the average coherent estimate for the simple statements is significantly smaller than the average subjective estimates for the simple statements ($t(230)=2.16$, $p<0.05$).

For our subsequent analysis of housing values we are most interested in the elementary probabilities of the ozone events. The average probability estimates and the average corrected estimates for simple events involving ozone are listed in Table 7

Table 7. Average Probabilities and Average Corrected Probabilities for Simple Ozone Events

Event to which probability is assigned	May Subjects	May Subjects Corrected	July Subjects	July Subjects Corrected
There will be an ozone watch in Meyerland.	.31 (.241)	.387 (.218)	.432 (.314)	.486 (.278)
There will be an ozone watch at Hobby Airport.	.424 (.293)	.476 (.212)	.517 (.288)	.522 (.243)
There will be an ozone watch at the Medical Center.	.366 (.236)	.333 (.138)	.488 (.276)	.518 (.248)
There will be an ozone watch at the ship channel.	.434 (.284)	.468 (.184)	.579 (.275)	.533 (.226)

Again, we would like to compare the average probabilities assigned to simple ozone events in the various parts of the city, but for the coherent estimates rather than the originals. For the May subjects, we find significant differences between the average probabilities assigned to the Hobby Airport and Medical Center events (t(34)=2.38, p<0.05), and also between the Medical Center and ship channel events (t(34)=2.47). There were no significant differences between the combinations of other events in May, and no significant differences at all in July.

Also, we want to determine whether the coherent estimates are significantly different from the subjective estimates. The coherent estimate for an ozone watch in Meyerland is significantly larger than the subjective estimate in both the May and July samples (t(18)=3.02, t(18)=1.87), while there is no significant difference for any other simple event in the May sample. For July, the average coherent estimate for the ship channel is barely significantly lower than the subjective (t(18)=1.71, p<.1).

Do Homeowners Know Anything?

Of course, as mentioned in the introduction, after all judgments are made and then corrected, the accuracy of the estimates when compared to real events should be considered. Since we intend to use the subjective and coherent estimates in a regression model of housing values, it is crucial to determine whether "people know anything." We consider two methods (both outlined by Deines et al., 2002) for determining accuracy before and after coherent approximation by SAPA. The first is the quadratic penalty method (von Winterfield and Edwards, 1986), which can be defined for both absolute (non-conditional) and conditional events as follows:

a) Suppose that p is the estimated probability of an absolute event E. The *quadratic penalty* for the estimate is $(1-p)^2$ if E is true. It is p^2 if E is false.

b) Suppose that p is the estimated probability of the conditional event "*E assuming that F*." The *quadratic penalty* for the estimate is $(1-p)^2$ if both F and E are true. It is p^2 if F is true but E is false. It is undefined if F is false.

Thus, the lower the quadratic penalty, the more accurate is the estimate.

Yates (1990) outlines another method for evaluating the predictive accuracy of probability statements. It involves calculating the average probability assigned to events that come true minus the average probability assigned to events that do not come true. Thus, higher values of this "slope" measure correspond to more accurate estimates.

The quadratic penalties and slopes for both May and July subjects, both before and after SAPA correction, are listed in Table 8. For our purposes it is most meaningful to consider the quadratic penalty for all estimates (conditional and non-conditional).

Table 8. Average Quadratic Penalties and Slopes for Subjective and Coherent Probabilities

Measure of Accuracy	May Subjective Estimates	May Coherent Estimates	July Subjective Estimates	July Coherent Estimates
Quadratic Penalty for All Estimates	.27 (.304)	.205 (.214)	.303 (.311)	.247 (.26)
Slope	.161 (.024)	.249 (.018)	.154 (.023)	.253 (.019)

First of all, we notice that the July estimates result in larger differences in absolute value between coherent and subjective quadratic penalties and slopes. This means that more correction for incoherence was necessary in July.[23] We can use a t-test to evaluate whether the correction for coherence leads to significantly smaller quadratic penalty in simple or conditional events. For both months, t-values indicate that the correction for coherence significantly reduces the quadratic penalty.

We can also test the "ignorance" of the subject by testing whether the quadratic penalty is significantly different from .25. We use this number as the threshold for ignorance because a subject who answers .5 to all questions, although incoherent, can guarantee a quadratic penalty of .25. For subjective estimates in May, the quadratic

[23] It may be the case that conjunctive and disjunctive events have become more important in July.

penalty is not statistically different from .25 at the one percent or five percent levels, but is significantly different from .25 at the ten percent level (t(728)=1.79). For subjective estimates in July, the quadratic penalty is significantly different from .25 at the one percent level (t(919)=5.18). Therefore we can say with some confidence that subjects appear to have some knowledge of the ozone pattern in July, but the evidence that they have knowledge of ozone patterns in May is much weaker.

For the slope of subjective estimates in May, the t-value indicates that the slope is significantly greater than zero (t(727)=6.8), which suggests that the participants are not ignorant regarding the truth. In fact, the same result holds for the slope of subjective estimates in July (t(918)=6.69), slope of coherent estimates in May (t(727)=13.38), and slope of coherent estimates in July (t(918)=12.99). It is also important to test whether the coherent estimates lead to a significantly higher slope. For both May and July subjects, the coherent estimates result in significantly higher slopes than the subjective estimates (t(727)=3.12, t(918)=3.93).[24]

The most striking result from the slope measure is that homeowners indeed possess at least some knowledge about the levels of ozone pollution in various neighborhoods around the city in *both* months, while the calculation of quadratic penalty provided weak evidence that subjects have knowledge of events in May. Thus our original hypothesis that people are unaware of ozone patterns across the city may be faulty. It is equally striking that coherent reconstruction makes homeowners better judges of ozone and related weather conditions. The improvement is reflected in both quadratic penalty and slope.

It may, however, be possible to explain our results while retaining the idea that people are truly unaware of ozone. Subjects could confuse ozone with other types of pollution. For instance, it is possible that when subjects are asked about ozone pollution, they actually think of the more odorous sulfur compounds. Sulfur compounds are typically a byproduct of the petroleum refining process, while some of the VOCs required for the creation of ozone are produced by petrochemical plants. Since the petrochemical plants and petroleum refining plants are located near each other, winds that carry the sulfur compounds to other neighborhoods may also lead to elevated ozone levels. Ozone and odorous compounds may therefore be positively correlated temporally, and in relation to particular weather conditions. Spatially, however, the odorous compounds are more prevalent in the southeast where the petrochemical facilities are located. It may therefore be the case that subjects are equating sulfur compound pollution (which they can smell) with ozone pollution, and are therefore making accurate predictions about ozone levels and weather even though their expectations about the geographical prevalence of ozone are inaccurate.[25] In any case, we will proceed with using the subjective and coherent estimates of ozone in the hedonic model of housing prices outlined earlier in the paper.

Incorporating the Subjective Measure of Ozone Level

In our previous analysis of housing prices, only objective levels of ozone were used. We will now consider the average subjective and coherent estimates of probability

[24] This finding is consistent with results reported by Batsell et al. A possible explanation is that the frequency of different events have to satisfy the laws of probability. Hence, it is plausible that adjusting stated probabilities to make them coherent could also induce more accurate forecasts.

[25] Alternatively, we could imagine a situation in which residents are aware of weather conditions that are conducive to sulfur compound pollution. However, there is no known correlation between weather conditions and odors produced by petroleum refining plants. (Dr. Matt Fraser, Department of Civil and Environmental Engineering, Rice University, personal communication, January 30, 2003)

of an ozone alert in both May and July for three areas of the city[26] as the measure of ozone level. We hope that this will allow us to test one explanation for the unexpected positive sign on ozone found in earlier—namely the idea that subjective expectations differ from the objective evidence on ozone pollution.

Recall that in the model using objective ozone values, the highest log likelihood was achieved using average daily ozone measures. For analysis of the model with non-objective data, we have four measures available: 1) subjective estimates of the probability of an ozone watch for May, 2) coherent (corrected) estimates for May, 3) subjective estimates for July and 4) coherent estimates for July. However, in order to provide a meaningful comparison between the models using the objective data and the models using subjective and coherent data, we create two other measures of ozone – the average of all subjective estimates (May and July) and all coherent estimates.

Similar to Table 1, Table 9 presents the results of the Box-Cox model in which the statistically insignificant variables have been dropped, and only the results for neighborhood variables are shown. The measure of ozone is the average of the *coherent* estimates across May and July.

Table 9. Box-Cox Specification with Insignificant Variables Excluded

Number of Observations	3230
LR chi-squared	10917.46
Prob>chi-squared	0.000
Log likelihood	-32642.47

Transform Value	Coefficient	Standard Error	z	P>|z|
θ	.2009	.0205	9.82	0.000
λ	.4823	.0489	9.87	0.000

[26] Earlier we considered four geographical areas of the city, but in the current paper the north section is dropped because it was not an area for which participants were asked to assign a probability of an ozone watch. The result is a loss of 1005 observations.

Nontransformed variables	Coefficient	chi-squared	P>chi-squared
hspc1	1.48	48.46	0.000
hspc2	1.04	56.96	0.000
mspc1	.4171	9.08	0.003
mspc2	.5463	5.59	0.018
espc1	-.5579	81.7	0.000
park dummy	-.3129	31.96	0.000
constant	29.02		

Transformed Variable	Coefficient	chi-squared	P>chi-squared
downtown	-4.56	44.41	0.000
crime ranking	-1.29	19.55	0.000
average coherent ozone	-47.05	88.56	0.000

Hypothesis Tests	chi-squared	p>chi-squared
$\theta = \lambda = -1$	2341.35	0.000
$\theta = \lambda = 0$	136.06	0.000
$\theta = \lambda = 1$	2269.47	0.000

Among the non-transformed variables, the most notable difference between the objective model and the model using the corrected subjective (coherent) estimates of ozone level is the change in the coefficients on school variables. This is consistent with the notion that pollution measures and school variables are essentially acting as neighborhood dummies. Thus, changing the pollution measure altered the coefficients of the other neighborhood effects.

The changes in the school variables imply different effects of school quality on housing prices than we found in the model with objective ozone data. First, the high school variables are even more positive and significant than in the objective model, implying that when subjective measures of ozone are used, high schools lend more to house values. In comparison to the model presented earlier in the paper, we still find that

McArthur high school decreases house value, while all other high schools add to the house value. Bellaire High School still has the largest effect, but now adding $19,412 (measured at mean values for the variables) to the value of a home in the neighborhood. Secondly, we see that the middle school variables have gone from being negative and significant in the objective model, to being positive and significant in the subjective model. The consequence is that all middle schools, except Fondren Middle School in southwest Houston, add to mean house prices. Fondren Middle School subtracts $1,672 from the value of a home in the neighborhood. It is difficult to know a priori whether the measured school effects in the current model are more plausible than the effects found in earlier. In our view, however, it seems unusual that the high school and elementary school for our southwestern neighborhood should add so prominently to the house value, while the middle school subtracts from house value. It would then seem that perhaps the objective values are better than the subjective values for ozone level. However, it is important to note that Bellaire High School is a magnet school at which good students from around the city attend, and it is therefore possible that it may lend positively to house values while Fondren Middle School subtracts. At least anecdotally, Fondren is not highly favored by parents in southwest Houston. Also, while we have considered only one elementary school in the southwest, students from several schools attend Fondren, which may explain why the selected elementary school lends positively to house values while Fondren does not. Therefore, it still may be that subjective measures for ozone value are better to use than objective measures.

Among the transformed variables, crime ranking remains significant in the subjective model, whereas it was dropped in the objective model. The coefficient on crime ranking has the expected negative sign, as a higher value for crime ranking implies a higher crime area. This perhaps adds support to the notion that the subjective measures are preferable, since we surely would expect high crime rates to negatively affect housing values.

However, the most notable change from the objective model is the negative and significant coefficient on ozone level, which is what we expected at the outset of our study. While the ozone measure in Table 3.5 is the average coherent estimate of ozone alerts in both May and July, average subjective estimates, as well as subjective and coherent estimates for May and July separately also result in negative and significant coefficients. The average coherent estimates lead to slightly higher log likelihoods, and therefore are used in the model.

As for interpretation, the elasticity of market value with respect to ozone level is –3.302. That is, a 0.005 increase in the average coherent estimate of the likelihood of an ozone watch is associated with a $3,012 decrease in house value. Recall that in the first model ozone was measured in parts per billion, while in the now we characterize ozone level as the average probability of the occurrence of an ozone alert, as estimated by subjects in the survey. The elasticities are, however, unit free and therefore are comparable. In the earlier model the corresponding elasticity was 1.903. One also can compare the elasticities of house prices with respect to ozone with the elasticities with respect to the other factors – crime, schools, etc. The elasticity of house price with respect to crime ranking is -.309, the elasticity of house price with respect to distance from downtown is –1.676, and the elasticity of house value with respect to the first high school principal component is .1281.

As another test of the relative explanatory value of the objective and subjective expectations, we substitute the objective values into the subjective model, and the subjective values into the objective model. In the survey, participants were asked about ozone levels in the southwestern, southeastern, eastern, and central parts of the city. However, the neighborhoods for which we collected the housing data used in the hedonic model are in the southwestern, southeastern, eastern, and *northern* parts of the city. Therefore, in order to have a meaningful comparison between the models using objective and subjective ozone data, it is necessary to drop the 1005 homes in the northern neighborhoods from the objective sample.

Table 3.6 reports the log likelihoods and coefficients and chi-squared values for the ozone variable in four different regressions: 1) the objective model from earlier in the paper using the objective ozone values (with the aforementioned 1005 observations excluded), 2) the objective model using a subjective ozone measure (average coherent estimate), 3) the subjective model (reported in Table 3.5) using the objective ozone measure and 4) the subjective model using the subjective ozone measure.

Table 10. Various Models Using Objective and Subjective Ozone Measures

Type of Model and Ozone Measure	Log Likelihood	Coefficient on Ozone Measure	chi-squared	P>chi-squared
Objective Model Using Objective Ozone	-32655.97	4.49	81.0	0.000
Objective Model Using Subjective Ozone	-32655.42	-34.84	82.11	0.000
Subjective Model Using	-32643.33	6.31	86.83	0.000

Objective Ozone				
Subjective Model Using Subjective Ozone	-32642.47	-47.05	88.56	0.000

The most striking result from Table 3.6 is that the objective values of ozone always result in positive and significant coefficients, while the subjective values for ozone lead to negative and significant coefficients, as we initially expected. For this exercise we have used average coherent estimates of ozone as the measure of ozone level. In fact, any of the subjective or coherent measures available to us produces a negative and significant coefficient. The reason for using the average coherent value was so that a comparison could be drawn to the average ozone level used in the subjective model.

However, recall that we were interested in not only the accuracy of the estimates provided by the survey participant (the subjective estimate), but also the effect of modifying the subjective estimate to correct for statistical incoherence (the coherent estimate). We have determined that some sort of estimate based on homeowner expectations is more appropriate than an objective measure for our hedonic model. Now we would like to more closely examine the effect of using one of the four estimates provided directly and indirectly by the participants: 1) subjective estimates for May, 2) subjective estimates for July, 3) coherent estimates for May and 4) coherent estimates for July. In essence, we would like to observe how much, if at all, the correcting of statistical incoherence affects the results of the hedonic regression model. All four of the potential subjective ozone measures result in the same specification in the Box-Cox framework; that is, the final set of significant variables is the same for both subjective and coherent estimates in May and July. Thus, it is convenient to compare the log likelihoods and coefficients of the ozone variables. Table 3.7 contains the results.

Table 11. Log Likelihoods and Ozone Coefficients Using Subjective and Coherent Estimates for May and July

Ozone Measure	Log Likelihood	Coefficient	chi-squared	P>chi-squared
May Subjective Estimates	-32642.49	-21.81	88.52	0.000
May Coherent Estimates	-32642.44	-34.25	88.62	0.000
July Subjective Estimates	-32642.66	-23.93	88.18	0.000
July Coherent Estimates	-32642.55	-71.29	88.39	0.000

Although the variation among models is slight, the coherent estimates generally offer larger (less negative) log likelihoods and more significant ozone coefficients than the subjective estimates. Thus, correcting the estimates of individuals for coherence appears to produce at least as good a predictor of housing prices. As for the distinction between the May and July figures, the May log likelihoods are slightly larger than those for July, and the coefficients for May are slightly more significant – indicating perhaps that the May estimates contain more variation than those for July. This is not an unreasonable result, since residents may feel that ozone levels for the extremely hot Houston July must be high everywhere in the city, whereas the levels in May could vary by location. However, comparison with the model using objective values of ozone indicates that both May and July (subjective and coherent estimates) provide higher log likelihoods and more significant coefficients.

In summary, the coherent probability values are better than subjective values, and both coherent and subjective are better than objective values. Also, the May values tend to be better than July, but both May and July are better than objective values.

However, it is interesting to return to the negative and significant coefficient on the non-objective ozone levels. Even though the sign is as we would have expected at the outset of the project, we have found that objective levels of pollution are virtually the same in the southeastern and western parts of the city. We have also discovered that residents indeed have some knowledge about the probability of ozone watches around the city, given weather conditions. However, we have hypothesized that residents tend to have more knowledge about the odorous sulfur compounds than they do about ozone levels. This is a reasonable assumption given that precursors to ozone (nitrous oxides and VOCs) will occur in the same geographic areas as sulfur compounds, because petrochemical facilities and petroleum refining plants are located in the same part of the city. However, what residents may not realize is that nitrous oxide and VOCs, which are also produced by vehicles, can occur independently of the sulfur compounds, resulting in high ozone concentrations but no odorous signs of pollution. If the sulfur compounds are indeed what subjects are responding to when asked about ozone, the result in the hedonic model would be to assign a penalty to those houses in the southeastern and eastern parts of the city, which appears to be what has happened in the exercise.