JOURNAL OF URBAN ECONOMICS 1, 127–146 (1974)

Determinants of Real Estate Values

D. M. GRETHER AND PETER MIESZKOWSKI[1]

California Institute of Technology and Queen's University

Received January 17, 1973

INTRODUCTION

The services provided by any particular house clearly depend upon a great variety of physical characteristics of the house and its grounds. Possibly the most important of these characteristics are the size of the house, how that size is divided up for use, and the amount of land that is included with the house. Other features of the structure which presumably are important are the heating, electrical and plumbing systems. There is a wide variety of other characteristics including the parking facilities available, the age and general condition of the house, the presence or absence of a basement, the number of stories, the existence of a family or recreation room, and so on, which may also affect the nature of the services that can be provided. In addition to the physical attributes of the house the characteristics of its location are likely to be of significance. Obvious examples of such features are: distances from place of employment and shopping, population density, the nature of public services (schools, police and fire protection, water supply, sewers, etc.), the amounts of traffic, noise, and air pollution, and so on.

In this paper we present some empirical results on the determinants of real estate values in the New Haven metropolitan area. We attempt to explain housing values using information on the physical characteristics of the house and on some features of the neighborhood in which the house is located. In the latter category we are primarily interested in those features of neighborhoods which can be influenced or controlled by local governments. Determining provisions of zoning codes (or granting variances), routing traffic, allocating services (street sweeping, education, fire protection, etc.) are examples of governmental activities which may affect the nature of neighborhoods. Specific policy changes are frequently advocated or opposed on the grounds of protecting property values so it seems worthwhile to attempt to measure the magnitudes of these effects.

[1] The research reported on in this paper was supported by a grant from the National Science Foundation to the Institute for Social Sciences at Yale University and by grants from the National Science Foundation and the Ford Foundation to the Cowles Foundation for Research in Economics at Yale University.

We have at our disposal information on a large number of single-family homes that were sold during the period 1962–69 in the city of New Haven and its suburbs. We have also collected from public assessors' records information on single-family homes located in residential areas that are close to land used for other purposes, e.g., highways, shopping centers, and high-density residential developments. With these data we hope to determine how proximity to non-residential land use affects the values of single-family dwellings and how those effects vary with distance from the non-residential use. We plan to present the results of this analysis in a subsequent paper. We present here only the results for the city of New Haven (about 830 transactions).

Most previous studies on real-estate values and neighborhood effects utilize aggregate census data where the unit of observation is a census tract. The quality controls on the units are also aggregative and include such variables as percentage of units in a tract classified as dilapidated and/or without private bath, percentage of houses which were more than twenty years old, and median number of rooms per dwelling. Among the studies utilizing census data is the work of Richard Muth (1967, 1970) on urban structure, the relationship between low-quality housing and poverty and between housing prices and race. Also Ridker and Henning (1967) use these data in analyzing the effects of air pollution on property values. Oates (1969) in his study of influence of public services and taxes used average property values for a sample of New Jersey cities.

There have also been several studies using micro (disaggregative) data in which the unit of observation is a single transaction. For example, Bailey (1966) studied the effects of racial composition and population density on housing prices in Chicago. Pendleton (1962) attempted to measure the value of accessibility to job opportunities and to the central business district in Washington, D. C. Kain and Quigley (1970) examined the effects of a variety of neighborhood quality indices on housing prices in St. Louis, and Lapham (1971) used data from Dallas to test the hypothesis of racial discrimination in housing. Two studies specifically concerned with the effects of non-residential land use on property values are Nourse (1962) and Crecine, Davis and Jackson (1967).

The data base for our study is similar to those used in the second group. Our sample is somewhat larger and with the exception of Lapham's, contains more information about the characteristics of the dwellings. Kain and Quigley had a much larger number of variables (39) relating to the quality of the neighborhood and dwelling than we have, though in their regressions they used the first five principal components rather than the entire set of variables. Also, many of these variables were ordinal (building inspectors and interviewers rated various features on a one to five scale) which makes interpretation of the estimated coefficients difficult since the signs of correlations among ordinal variables or between ordinal and cardinal variables are often indeterminate.[2]

THE SAMPLE

The basic data for this study are the multiple listing files of the Greater New Haven Board of Realtors. These files contain a large amount of information about the physical characteristics of the houses sold as well as giving the location of the house, the owner's asking price (and any changes in it over time), the date the house was listed, and the selling price together with the date of the transaction.

The files give information on quite a large number of structural characteristics. Among those covered are: the number of rooms, the total square feet of living space, the type of plumbing, the number of bathrooms and lavatories, the type of heating and the number of independently adjustable heat zones in the house, the building materials used, the kind of roof, the amount of garage facilities (if any), the age of the house, and the kind of floors. In addition, information is provided on insulation, storm windows, the number of small rooms in addition to the primary rooms, the size of the basement, the number of finished rooms in the basement, whether or not there is a laundry hook-up and the electrical wiring. All houses are rated by the realtors as to overall general condition; the categories being: excellent, very good, good, fair, and poor. The files also give information about special or non-standard features of the houses, e.g., what appliances are included, if there is a fireplace or family room, air conditioning information, if there is wall to wall carpeting, if there are stall showers, etc. Finally, there is information which though irrelevant for the city of New Haven will be useful for the suburban sample. The files give the nature of the water supply (city or private well), and of the sewage disposal system. Virtually all the houses in the city of New Haven have city water and are connected to sewers, but in the suburbs this is often not the case.

It is not surprising that for many of the houses in the sample some of the data were missing. One quite important feature, the size of the house, was unreported for about half the sample. For all houses in the sample this variable was obtained from the New Haven assessor's office (the same procedure was followed for the suburban samples as well). The only other major feature that frequently was unreported or reported in a fashion that suggested substantial uncertainty about its accuracy was the age of the house. We were able to obtain this information on all but a small number of houses from the tax assessor and by searching the city directories to see when an address first appeared.

[2] For an example see Battalio, Hulett, and Kagel (1973). Conditions under which the signs of the correlations are identified are given in Grether (1973).

In some cases the missing records were simply filled in with what seemed reasonable guesses. For instance, if no information was given about the voltage, it was assumed that the house had only 110 volts and no 220-volt line. Similar rules were followed if no mention was made of a slate roof, a family room, a fireplace, appliances included, hardwood floors, etc. In general when it seemed likely that the owner would have mentioned a feature if the house had it, we took non-reporting to mean that the feature was lacking. Note that in the vast majority of cases the features in question would be reported as present or absent.

In some instances unreported data could be inferred from other information given on the multiple listing cards. For example, from the style of the house, e.g., bungalow, ranch style, colonial, etc., one can determine the number of stories. While such guesses as to the values of some of the explanatory variables may occasionally be in error we thought it preferable to proceed in this fashion rather than either reduce the sample size substantially or work with only a small number of characteristics. It seemed unlikely that the guesses made would lead to any systematic biases, though it is possible that the reported t statistics somewhat overstate the precision of the estimates.

Suppose that we have a regression model of the form

$$\begin{aligned} Y &= X\beta + \mathcal{E} \\ E(\mathcal{E} \mid X) &= 0 \\ E(\mathcal{E}\mathcal{E}') &= \sigma^2 I. \end{aligned} \tag{1}$$

Suppose that not all the elements of the matrix X are observed, the array of known elements being $\tilde{X}$.[3] If from the array $\tilde{X}$ we delete all the rows which are incomplete we obtain the matrix $\bar{X}$ of the reduced sample. The matrix actually used in estimating β is $\hat{X}$ where $\hat{X}$ contains all the elements of X which are known in addition to our guesses about the unknown elements of X. The estimate of β obtained is

$$\hat{\beta} = \beta + (\hat{X}'\hat{X})^{-1}\hat{X}'(X - \hat{X})\beta + (\hat{X}'\hat{X})^{-1}\hat{X}'\mathcal{E}. \tag{2}$$

If $\hat{X}$ depends only upon the known elements of X, then

$$E(\hat{\beta}) = \beta + (\hat{X}'\hat{X})^{-1}\hat{X}'E(X - \hat{X})\beta. \tag{3}$$

If $\hat{X} = E(X|\tilde{X})$, then $\hat{\beta}$ will be unbiased with variance–covariance matrix given by:

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2(\hat{X}'\hat{X})^{-1} + (\hat{X}'\hat{X})^{-1}\hat{X}'E\{(X - \hat{X})\beta\beta'(X - \hat{X})^1\}\hat{X}(\hat{X}'\hat{X})^{-1}, \tag{4}$$

[3] $\tilde{X}$ is not rectangular Essentially it is a matrix with some holes in it. We assume that the pattern of missing x's is random or at least not systematically related to the disturbance vector $\mathcal{E}$.

where the expectation is taken conditional on the known elements of X. Since $\hat{X}'\hat{X} - \bar{X}'\bar{X}$ is positive definite, the first term in (4) is smaller than the variance–covariance matrix that would have been obtained by working with the reduced sample $\bar{X}$. Because working with $\bar{X}$ alone would have required a substantial reduction in sample size we felt justified in using $\hat{X}$ instead. Further, since in no cases was information on the size of the house (Y) used, we hoped that the condition $\hat{X} = E(X|\bar{X})$ should be at least approximately satisfied.

Houses for which data on several characteristics were missing were excluded as were houses which lacked data on major features such as lot size or number of rooms. The majority of the records were complete, but a large number lacked information on one or two items such as those mentioned earlier. Also excluded from the sample were some houses that were clearly atypical. Examples of the kinds of houses excluded are: a house with a lot reported to be over two acres, a few beach cottages, a shed on an otherwise vacant lot, and one house described as a "hunting lodge."

There were a number of houses that were sold more than once during the period, allowing us to check the general accuracy of the data (at least their consistency). The impression obtained was that the data were in general accurate.

The sample is not only large but also quite varied in the types of houses included. The average house has 1730 square feet of living space (standard deviation—SD of 823 square feet). There are 225 houses with less than 1200 square feet and 98 with more than 3000 square feet. The average lot size was 8500 square feet (SD = 5000). The average age of the houses in the sample was 37 years (SD = 22 years). Approximately one-half of the houses were built during the 1920's or in the period immediately following World War II. The average house has seven rooms (SD = 2), and sold for \$22,000 (SD = \$10,000). Seventy-four percent of the houses had some kind of a garage and another ten percent had a driveway but no garage. Seventy-seven percent of the homes had all hardwood floors with the remainder nearly equally divided between houses with some hardwood floors and houses with all softwood floors (other types of flooring were quite rare in the New Haven sample). Only 5% of the houses sold were rated by the realtors to be in less than good condition, and about half the houses were rated excellent. Approximately 46% of the houses had radiant or water heat, 41% had forced air, and 13% had steam heat. There were no houses with electric heating. In certain dimensions the houses were quite similar. For instance, over 98% of the houses were reported to have basements and over 97% had screens (though only 68% had screens on all windows).

Data for the neighborhood variables came from public sources or were constructed from maps. Reading percentiles for the elementary schools, pupil/teacher ratios, and the racial composition of the schools came from the

New Haven Board of Education. Data on traffic flows (cars per day on each street) were also obtained from public authorities. The density variables (number of neighbors within a 1000 foot square centered at the house and within 500 feet to the right or left of the house) were obtained by counting the number of structures on land use maps. The distance from the New Haven Green (the center of the downtown area) to the house was simply measured off maps. Many public services, e.g., refuse collection, street lighting, water supply, and sewage disposal are uniform throughout the city of New Haven and thus no variables for them are included.

SPECIFICATION OF THE REGRESSION MODEL

We postulate that the value of a house is an additive function of its structural characteristics, the characteristics of the lot, and the characteristics of the neighborhood in which the house is located. Thus

$$v_i = S_i\alpha + L_i\beta + N_i\gamma + \bar{\varepsilon}_i$$
$$E(\varepsilon) = 0,$$

where S_i, L_i, and N_i are vectors of characteristics of the structure, lot and neighborhood and α, β, γ are vectors of unknown coefficients. For the semi-log model the estimating equation is of the form:

$$\log v_i = \bar{S}_i\bar{\alpha} + \bar{L}_i\bar{\beta} + \bar{N}_i\bar{\gamma} + \bar{\varepsilon}_i.$$

We turn now to a discussion of the specification of each of the three components.

The structural characteristics

The basic variables used in this part of the regression are:

v	selling price of the house in dollars
size	square feet of living space
brick–stone	dummy = 1 if the home is made of brick or stone
slate roof	dummy = 1 if the house has a slate roof
steam–radiant	dummy = 1 if the house has steam heat or hot water radiant heat
baseboard	dummy = 1 if the house has baseboard heat
heat zones	the number of individually adjustable heating zones
off boiler	dummy = 1 if the hot water is heated off the furnace
hardwood	dummy = 1 if house has all hardwood floors
softwood	dummy = 1 if house has all softwood floors
1 car	dummy = 1 if there is a one-car garage
2 car	dummy = 1 if there is a two-car garage
2^+ car	dummy = 1 if garage has more than two-car capacity
stories	the number of stories in the house
baths	number of bathrooms
lavs	number of lavatories

tile bathes	number of tiled bathrooms
shower	number of stall showers
brass–copper	dummy = 1 if plumbing is all of brass or copper
laundry	dummy = 1 if there is a laundry hookup
volts	dummy = 1 if there is a 220-V line to the house
amps	number of amperes for which the house is wired
basement	dummy = 1 if seller indicates there is a full basement
insulation	dummy = 1 if there is full insulation
storm-windows	dummy = 1 if house has full storm windows
small rooms	number of small rooms in addition to primary rooms
finished rooms	number of finished rooms in the basement
recroom	dummy = 1 if there is a family room
porch–sunroom	dummy = 1 if there is a porch or sunroom
fireplace	dummy = 1 if there is a fireplace
ceilings	dummy = 1 if there are beamed ceilings
appliances	dollar value (at new prices) of appliances included
improvements	dollar value of recent improvements—painting, etc.
excellent	dummy = 1 if this is realtor's rating of condition
very good	dummy = 1 if this is realtor's rating of condition
good	dummy = 1 if this is realtor's rating of condition
rooms	the number of primary rooms
age	age of the house at time of sale (in years)
carpet	dummy = 1 if wall-to-wall carpeting included

Some of the variables listed above were simply included in the regression in the form listed. In general if it seemed reasonable that the presence or absence of a particular feature should affect the value of all houses equally the characteristic was simply entered among the list of explanatory variables. Included in this category are: the garage variables, bathrooms, lavatories, tile baths, volts, amperes, fireplace, recroom, appliances, and similar variables. The effects of other features may well be expected to vary depending upon the size of the house. These include: the type of flooring, the condition of the house, the number of stories, storm windows, basement, slate roof, building materials, etc. In most cases these variables were entered multiplied by the size of the house (square feet of living space). Included in this category are the condition variables, the flooring variables, and dummies for the number of stories. When some other measure of size seemed more appropriate it was used instead of square feet of living space. For example, the slate roof dummy variable was entered multiplied by the size divided by the number of stories (hopefully a better approximation to the area of the roof). The brass-copper plumbing variable was weighted by the number of bathrooms and lavatories, and the storm-window variable was weighted by the number of rooms (we hoped this would be a better proxy for the number of windows than total size). We assumed that the effect on the value of a house of having brick or stone construction would more likely be proportional to the area of the outside walls than to the floor space. Thus we weighted this variable by (size · number of stories)$^{1/2}$ which if houses were square should be proportional to the desired quantities. The tables giving the results indicate the form in which each characteristic was included.

A large variety of nonlinearities and interactions is possible, but in general we did not include such terms. For instance, diminishing returns to bathrooms may be expected at some point, but we have assumed that this point is outside the range covered by our sample. Two nonlinearities were explicitly included, however. The age of the house and its square are used as separate variables each multiplied by the size of the house, and we also included the square of the size of the house.

In interpreting the coefficients given below it is important to keep in mind the weighting factors used. Thus, the coefficient on hardwood floors of 0.45 indicates that having all hardwood floors increases the value of a house by 45 cents per square foot on the average.

The lot

We postulate that the value of the land is a quadratic function of the size of the lot. Though there could be some interaction between lot size and house size we did not include any such terms in the regressions.

The neighborhood characteristics

It is by no means obvious whether the neighborhood effects will enter in on an additive basis or whether they will be proportional to house size or lot size. It is conventional to argue that neighborhood effects or externalities will be capitalized in the value of land. For example, consider two undeveloped residential areas, one of which has some distinct advantage over the other. If lots are perfectly divisible in both areas a fixed differential per square foot of land should become established between the two areas and this per-square foot differential should be independent of the size of the lot, so that the absolute size of the neighborhood effect will be proportional to the size of the lot.

However if there is originally little variation in the size of lot and once the area is developed it is possible that the absolute size of the neighborhood effect will be more closely related to the size of the house than to the size of the lot. If the absolute amount households are to pay for the more advantageous location is positively correlated with income and more wealthy individuals buy larger houses then it may be more appropriate to "attach" the externality or neighborhood variable with the size of the house rather than with the size of the lot. Of course the distinction will not be very meaningful if there exists a strong positive correlation between lot size and house size. Finally there is the possibility that neighborhood effects can be specified as a fixed charge (premium). The aversion to high traffic, poor schools, etc., might be independent of house size and lot size as the cost of moderating the neighborhood effect is independent of lot size and house size. An example of a fixed charge alternative is the cost of sending a child to a private school.

As we do not have strong *a priori* notions as to how to specify neighborhood effects, we have tried three alternative forms. One assumes these effects to be proportional to lot size, the second to house size and on the third specification the effect enters as a fixed "charge." We also have run some regressions with some features fixed and others proportional to house size or lot size.

It is possible that the relationships between the neighborhood characteristics and housing prices are more complex than assumed here. For example, if occupants of high-priced homes are relatively more successful in obtaining extra resources for their local schools or stop signs, etc. which could alter traffic flows, then there would be a simultaneous equations problem. We felt that this would be a serious problem if we pooled observations from the several towns in the metropolitan area, but hoped that by limiting the investigation to a single city we could avoid this difficulty. To the extent that we have been unsuccessful, the estimated coefficients will not in general be unbiased or consistent.

To a first order approximation it is not unreasonable to characterize the housing market as an auction of different housing units, of varying structural attributes, some of which are located in different "environments." At a given moment of time the price structure of the housing stock is shaped primarily by demand factors, i.e., consumer preferences for various structural characteristics, quiet streets, school quality, and so on. Although this is basically a cross-sectional study our observations span a seven-year period so supply factors will bear on average and possibly relative prices. New and old houses are close substitutes for one another, consequently when construction (replacement) cost changes, the average price of houses and the marginal value of some of the attributes are expected to change in proportion to the changes in replacement costs.

In this paper we assume that the price of the structure increases proportionally with construction costs. Thus each of the structural variables has been multiplied by the national construction price index for new single family homes. We attempt to distinguish between changes in the value of a structure over time (due to the increasing age of the house and to changes in construction costs) and changes in the value of the land that goes with the house. We assume a constant exponential rate of appreciation for land over the period. Thus the estimating equation is of the form

$$v_i = c_t \cdot S_i\alpha + e^{\lambda t} L_i\beta + \mathcal{E}_i,$$

where c_t is the construction price index and the neighborhood variables are included along with either the structural characteristics or with the land. The estimating procedure is a search procedure picking as the estimate of λ that value which gives the smallest sum of squared residuals. If the true disturbances are distributed as $N(0,\sigma^2)$ and independent, then this procedure amounts to the maximum likelihood method. Even if this assumption is not satisfied the estimates obtained will still in general be consistent.

Preliminary analysis revealed that there was evidence of heteroscadasticity with, not surprisingly, more variation associated with larger transactions. To correct for this each observation was divided by the size of the house and the construction price index thus transforming the dependent variable to real price per square foot. The transformation had virtually no effect upon the estimated coefficients and made very little difference in the t statistics.

Having dealt with the specification of the linear model in considerable detail, our discussion of the semilog model will be brief. The basic variables are the same as those mentioned above, though in several cases the form in which a variable appears in the equation is different. In particular, those variables which previously were weighted by house size or lot size are used without weights, e.g., age, age squared, floor type, etc. For this specification it is assumed that having, say, hardwood floors will on average shift the price of a house by a fixed percentage. Variables for most of the other features are used in the same form as before, for instance, the garage dummies. Now it may seem somewhat unreasonable to assume that the values of such things depend upon other characteristics of the dwelling (as they must if the price is shifted by a constant fraction). On the other hand, there really is no such thing as a "two-car garage." The one behind the mansion on the hill and the one beside the house on the corner may be quite different kinds of structures. Thus we have no strong *a priori* opinions as to which form is the better approximation.

In many ways the semilog model is a good deal simpler to deal with. For instance, the problem of associating neighborhood effects with the structure or the land does not arise. Also, by introducing a polynomial in time it is possible to allow for quite general patterns of price change, and we no longer are dependent upon a construction price index which may be inappropriate to the New Haven area. In addition, the logarithmic transformation should eliminate the problem of heteroscedasticity. Since the results from the two models are quite similar, we shall concentrate on the linear model and present only a few results for the semilog form.

THE RESULTS

The estimates obtained from multiple regression are shown in Table 1. On the whole the results are quite encouraging. Nearly all the variables have the expected sign and the majority are statistically significant. In most cases the magnitudes of the coefficients seem reasonable as well. We discussed our results with several realtors, appraisers, and builders and in general the consensus was that the numbers made sense.

TABLE 1

Regression Results[a]

Variable	Coefficient	t	Variable	Coefficient	t
Size (sq. foot of living space)	5.2	6.8	Age squared·size	0.00033	5.6
\$ value of appliances	1.2	2.8	D·size $D = 1$ if excellent	1.8	6.7
$D\cdot(\text{size}\cdot\text{stories})^{1/2}$ $D = 1$ if brick or stone	2.0	2.3	D·size $D = 1$ if very good	1.5	5.2
$D\cdot(\text{size}/\text{stories})$ $D = 1$ if slate roof	1.7	3.1	D·size $D = 1$ if good	0.79	3.0
No. of heat zones	860	2.3	Average room size	−5.2	2.0
D·size $D = 1$ if baseboard heat	−0.9	2.7	Bathrooms	800	4.0
D·size $D = 1$ if all hardwood floors	0.45	2.6	$D = 1$ if laundry hookup	760	3.6
$D = 1$ if 1-car garage	790	4.4	Reading %·Lot size	0.0046	12.6
$D = 1$ if 2-car garage	1270	5.0	(Pupil/teacher)·Lot size	−0.02	3.7
$D = 1$ if fireplace	830	4.4	(No. of neighbors within 500 ft)·Lot size	−0.03	4.1
$D = 1$ if family room	580	2.3	Lot size	0.89	5.6
D·size $D = 1$ if 1 story	0.8	2.0	Lot size squared	−0.0000082	7.6
Age·size	−0.07	8.3	Intercept	36	5.7
			$R = 0.8906$		

[a] Estimated rate of appreciation on land—4% per year. Variables for the following had the expected sign but were not significant: tile bathrooms; softwood floors; more than two-car garage; porch–sunroom; \$ value of recent improvements (painting, etc.); amperes; volts; number of small rooms; lavatories; brass–copper plumbing; insulation; storm windows; traffic flow; distance from center of New Haven; number of neighbors in a 1000 ft square centered at home.

Variables for the following were insignificant and had the wrong sign: number of finished rooms in basement; size square basement.

Variables for the following were also insignificant: one and one-half stories; steam or water heat; hot water heated by furnace.

The dependent variable is price in dollars.

As many of the structural characteristics have effects which vary with the size of the house, we shall discuss the results with reference to a "standard house" which as a two-story frame house with 2000 square feet of living space and 10,000 square feet of land (about $\frac{1}{4}$ of an acre) which was sold in 1960. Increasing all figures by about fifty percent will give the current values. Assume that the house built just prior to World War II has eight rooms, hardwood floors, forced air heating, one heat zone, a fireplace two bathrooms, a laundry hookup, and a two-car garage. If in addition the neighboring houses have lots with 75 feet frontages, and if the local primary school ranked in the 75th percentile with a pupil/teacher ratio near twenty, the results imply that the house would have sold for approximately \$27,000. If the house were of brick or stone construction rather than frame its selling price would have increased by \$1300, and a slate roof would have added another \$1700. A house with all hardwood floors is estimated to sell for \$1700 more than a house with softwood floors. The condition variables which are the realtors' evaluation of the house are all significant and have the expected pattern. Houses rated excellent, very good, and good sell at premiums of \$3600, \$3000, and \$1600, respectively, relative to houses rated fair or poor.

There are three variables which reflect the way the floor space is divided, i.e., the average size of the primary rooms, the number of small rooms, and the number of stories. We expected that the value of a given amount of floor space would be increased if the space were distributed over more rooms and that a one-story house would be, other things being equal, worth more than a two-story house. As it turns out our estimates indicate that the "standard house" is worth only \$200 more with eight rooms than it would have been with seven rooms. A one-story house is worth about \$1600 more than a two-story house (and approximately \$1900 more than a $1\frac{1}{2}$-story house). As two-story houses are more expensive to build than one-story houses the question of land costs will have an important bearing on the type of construction at the margin. Our estimates indicate that an extra 1000 square feet of land is worth about \$800. Consequently it appears that builders should avoid building two-story houses even if this means increasing overall lot size or cutting down on the useable amount of yard space.

A one-car garage is estimated to be worth \$800 on the average and a two-car garage another \$500. The estimates for the electrical variables were subject to high sampling error. The estimates obtained indicated that an extra ampere would add about \$4 to the value of the house (t = 1.6) and having a line for 220 V appliances would add about \$200 (t = 1.2).

The effect of age on the value of a house appears to be clearly nonlinear. Younger houses depreciate more rapidly than older houses and the estimates indicate that houses will continue to lose value until they are around 100 years old. Our "standard house" will depreciate by about \$1300 in the first decade of its life and by only \$800 in the fifth decade assuming constant construction costs. Note that the effect of age is net of the condition of the house and thus reflects obsolescence.

Not all the variables performed as expected. Among the variables that are statistically insignificant and/or have the wrong sign are: basement, brass–copper plumbing, more than two car garages, insulation, stormwindows, the number of lavatories, dollar value of improvements, and the nonlinear term for house size. In some cases, e.g., the basement, stormwindows, and insulation the results may well be due to substantial nonuniformity in the reporting of the data. For example in the case of basements there were generally four reported conditions: no basement (this was rare); a partial basement (usually with no indication of the actual size); full basement, and "yes." We tried various ways of treating the responses, but in general the results obtained were insignificant (t ratios less than 0.5) and the other estimated coefficients were essentially unchanged. Given that the effects are net of condition it is not too surprising that recent improvements (painting, putting on new siding, roof repair, etc.) do not appear to add significantly to the value of a house. In the case of extra large garages inspection of the data gives a possible explanation for the failure of this variable. There are relatively few three- or more car garages in the sample and of these several are very big (five to seven car capacity); these were associated with old houses in poor condition (possibly the structures were old barns or stables). Nevertheless it is a bit unappealing to estimate the value of large garages as being less than the value of a one-car garage (the actual estimate was \$260 with a t ratio of 0.5). We had expected that the nonlinear term in house size would be significant and negative. The results simply did not conform with our expectations, the estimated coefficients being +0.00006 with a t ratio of 0.4.

We expected that the coefficient of the value of appliances would be substantially less than one since the value used was based on new appliance prices. If appliances left behind are likely to be built in then this variable could be picking up the effects of modernizing kitchens. If this hypothesis is correct, then in the suburban samples where the housing stock is younger and more homogeneous the coefficient of this variable should be smaller.

The coefficients of the size of the lot and the square of the lot size are both significant and have the expected signs. From the results in Table 1 we calculate that the marginal value of land falls 49 cents per square foot when comparing a $\frac{1}{4}$-acre lot with a one-acre lot. No term allowing for interaction between house size and lot size is included in any of the regressions reported. We did at one point use yard size instead of lot size (deducting an estimate of the area covered by the house and the garage). The results of that experiment were virtually identical with the results obtained using the size of the lot. The rate of appreciation of land values is estimated to be in the range of 4–6% per year. However, not too much emphasis should be put on this estimate as the residual variation is rather insensitive to changes in this parameter. Thus the data do not allow a precise estimate of this parameter. The estimates of the other parameters are generally quite insensitive to variations in rate of appreciation.

The neighborhood variables were entered under three different assumptions: that they affect the price per square foot of the lot, that they change the value per square foot of the structure, and that the effects are independent of house size and lot size. One regression was run with the school variables and distance having fixed effects with the other variables affecting the price per square foot of the land. The results of that experiment were consistent with the figures shown in Tables 1 and 2. Three of the variables were generally in-

TABLE 2

Results from Different Specifications

Variable	Times lot size	Constant effect	Constant effect	Times· house size	Times· lot size	Times· lot size
Reading %		37.0		0.03	0.0048	0.0059
		(11.7)		(14.7)	(11.3)	(10.0)
Pupil/teachers		−66		−0.14	−0.02	−0.03
		(2.4)		(4.4)	(3.8)	(3.4)
Traffic flow		−0.41		−0.002	−0.000081	−0.00012
		(1.0)		(0.9)	(1.5)	(1.6)
Number of neighbors		−61.0		−0.14	−0.03	−0.04
within 500 ft		(1.2)		(3.3)	(3.9)	(4.0)
Distance from		−48.0		−0.09	−0.006	−0.014
N. H. Green		(0.5)		(1.1)	(0.4)	(0.7)
% white in primary					−0.00015	0.00029
school[a]					(0.5)	(0.7)
Change in % white[a]						−0.00086
						(1.7)
Rate of appreciation	0.04	0.04	0.04	0.06	0.04	0.00

[a] Regressions including this variable were based on a smaller sample than those excluding this variable. The figures in parentheses are t ratios.

significant, viz., the number of neighbors within a 1000 ft square centered at the house, the distance from the New Haven Green, and the number of cars per day driving on the street where the house is located. Since the city of New Haven is fairly compact (houses located on the boundary of the city being about three miles from the central business district), it is understandable that the distance variable is unimportant. The point estimate from the regression in Table 1 is that the "standard house" will sell for about $200 less if it is on the boundary of the city as opposed to being one mile from the central business district. We expect that when the analysis is extended to the suburbs distance will be more important.

A major surprise (and disappointment) was the failure of the traffic variable to be significant. This variable was of considerable interest since zoning changes are often opposed on the grounds that they will result in increased traffic in residential neighborhoods. Nevertheless for the range of traffic flows observed (generally fairly light, i.e., from 0 to 1000 cars per day) this variable is not statistically significant. The estimated impacts of traffic cover a substantial range from a low of 41 cents per car (constant effect) to a high of 4 dollars per car (on house size).

The reading percentile variable is always highly significant, and the estimated magnitude of its effect fairly stable. The pupil/teacher ratio is also significant though the estimates of the effects of this variable are more varied.

Note that these data are all for houses within a single tax district so these effects do not reflect the results of people taxing themselves at different rates to provide the services they demand.

The coefficient of the number of neighbors within 500 feet on either side of the house is negative and statistically significant when the effect is associated either with the size of the house or the lot. As expected houses with the same amount of land packed closely together will sell for less than the same houses located on lots with more frontage.

The regression results predict that if the "standard house" were to be moved from a school district which ranked in the fiftieth percentile to one in the ninetieth percentile its value would increase by approximately $2000 (assuming the effects of school quality are on the price per square foot of land). Treating the school quality as a fixed effect gives an estimate of $1500 and associating this variable with the structure yields an estimate of $2400. It is estimated that an increase in the pupil/teacher ratio from 25 to 30 would reduce the value of the "standard house" by approximately $1000 (on lot), $1400 (on the structure), or $350 (fixed). If the frontage of the neighboring lots were increased from 75 to 100 feet the estimates imply increases in the property value of about $1000 (on the lot) or $900 (on the structure).

Overall the estimates seem satisfactory. The coefficients of the neighborhood variables are by and large significant, have the right signs, and are reasonable magnitudes. In order to attempt to see if these locational variables really were reflecting the effects of the measured traits, we added two other variables to the regressions: the percentage of white children in the primary school and the change in this variable from the previous year. There was little bussing in New Haven during the period so these figures are fairly indicative of the racial compositions of the neighborhoods.

The results of adding these variables are shown in the last two columns of Table 2. Neither variable has a coefficient significantly different from zero. The closest of the two to being significant is the change in percentage white which has a negative coefficient suggesting that property values rise in areas that are becoming relatively more populated by non-whites.[3]

Two experiments were done to check the stability of the estimates of the neighborhood effects. The sample was divided into large houses and small houses, into houses with large lots and houses with small lots, and into those houses sold during the early part of the period and those sold later. For each subsample the regressions were re-estimated and the results are given in

[3] The data for the percentage of white students were only available beginning with the academic year 1963–64 so that the sample size was reduced when this variable was included and still further reduced when the change in the racial composition was used also. There was one primary school which had a policy of open enrollment. The regressions were run both including and excluding this one school district, and the results obtained were not appreciably different from each other.

TABLE 3

Results for some subsamples[a]

Variable	Houses less than 1500 sq. ft.	Houses more than 1500 sq. ft.	Lot less than 9000 sq. ft.	Lot more than 9000 sq. ft.	Sales pre-1966	Sales post-1965
Reading %	0.0048	0.0066	0.0059	0.0060	0.0051	0.0059
	(8.3)	(5.3)	(8.6)	(8.5)	(7.7)	(9.1)
Pupil/teachers	−0.02	−0.04	−0.04	−0.01	−0.01	−0.03
	(3.5)	(1.2)	(4.0)	(1.4)	(1.0)	(3.8)
Traffic flow	−0.00002	−0.00042	−0.00062	−0.00015	−0.00014	−0.000056
	(0.3)	(2.7)	(0.5)	(1.9)	(1.6)	(0.7)
Number of neighbors in 500 ft.	−0.02	−0.03	−0.011	−0.04	−0.02	−0.04
	(2.1)	(1.6)	(1.0)	(3.4)	(1.7)	(3.5)
Number of neighbors in 1000 ft. sq.			−0.00079	−0.00064	0.00026	−0.00028
			(0.8)	(0.7)	(0.3)	(0.3)
Distance from N. H. Green			−0.013	−0.013	−0.011	−0.020
			(0.6)	(0.1)	(0.5)	(0.9)

[a] In all cases the variables are multiplied by lot size; and the rate of application is taken to be zero. The figures in parentheses are t ratios.

Table 3. The overall impression is that the estimates are fairly similar in the different subsamples.

Finally we re-ran the regressions using a highly restricted number of structural variables (Table 4) The variables used were all variables that could be obtained from local tax assessors. While as might be expected the coefficients of the structural variables differ substantially from those shown in Table 1, the estimated coefficients of the neighborhood variables are quite similar to those reported earlier.[4]

We find this result especially encouraging since it seems to indicate that one can obtain estimates of neighborhood effects using publicly available data. If this is indeed the case, then the cost of gathering and processing data for such studies can be greatly reduced. Further, if one does not need all of the detailed information from the realtors' listings, then data could be used on new houses and private party sales, records of which are kept by tax assessors. This could be quite important when studying a small area as it allows one to get a substantially larger (and possibly more representative) sample.

We turn now to a discussion of the results for the semilogarithmic model. Table 5 gives estimated coefficients corresponding to those shown in Table 1, and Table 6 contains the results of some experiments similar to those reported in Table 3. With regard to the structural characteristics of the houses, several

[4] The hypothesis that the omitted structural characteristics all have zero coefficients was rejected.

TABLE 4

Results with Reduced Number of Variables ($\lambda = 0.06$)

Variable	Coefficient	t	Variable	Coefficient	t
Size	8.7	12.9	Average room size	−2.6	0.9
Size squared	0.01	0.9	Reading %·lot size	0.0047	13.6
$D = 1$ if 1-car garage	14.6	7.6	(Pupil/teacher)·lot size	−0.02	3.4
$D = 1$ if 2-car garage	19.6	7.0	(No. of neighbors		
$D = 1$ if more than 2 cars	2.6	0.5	within 500 ft)·lot size	−0.02	3.9
D·size	0.79	3.2	Distance·lot size	−0.02	1.6
$D = 1$ if 1 story			Lot size	0.85	5.4
D·size	−0.35	1.6	Lot size squared	−0.0000078	7.8
$D = 1$ if $1\frac{1}{2}$ stories			Traffic·lot size	−0.00013	2.8
Age	−11	12.7	Constant	34	5.7
Age squared	0.00051	8.1			
No. of bathrooms	11.1	5.3	$R = 0.8451$		

of the coefficients are quite similar to those discussed above. For example, the pattern of depreciation is nearly the same ($1300 in the first decade, $850 in the fifth). A one-story house is estimated to be worth about $800 more than a two-story house and $1300 more than one with $1\frac{1}{2}$ stories, and as before, the latter coefficient is insignificant. Also, the coefficients for slate roof, fireplace, and bathrooms have about the same implications as with the linear

TABLE 5

Regression Results—Semilogarithmic Model

Variable	Coefficient	t	Variable	Coefficient	t
Size	0.04	9.1	$D =$ if good	0.14	4.6
$D = 1$ if brick or stone	0.04	1.4	Average room size	−0.03	1.2
$D = 1$ if slate roof	0.12	3.6	Bathrooms	0.05	4.1
No. of heatzones	−0.18	0.9	$D = 1$ if laundry hookup	0.25	2.4
$D = 1$ if baseboard heat	−0.04	1.1	Reading %	0.002	9.3
$D = 1$ if all hardwood floors	0.03	1.7	Pupil/Teacher	−0.02	5.1
			No. of neighbors within 500 ft.	−0.005	1.0
$D = 1$ if 1 car garage	0.06	3.7			
$D = 1$ if 2 car garage	0.08	4.4	Lotsize	0.002	5.6
$D = 1$ if fireplace	0.05	3.5	Lotsize squared	−0.000003	3.7
$D = 1$ if family room	0.02	0.9	Intercept	4.9	32.9
$D = 1$ if 1 story	0.04	1.7			
Age	−0.006	6.1	$R = 0.94$		
Age squared	0.00003	4.5			
$D = 1$ if excellent	0.24	8.0			
$D = 1$ if very good	0.22	6.9			

TABLE 6

Results from some subsamples
Semilogarithmic Model[a]

	Houses less than 1000 sq. ft.	Houses more than 1000 sq. ft. and less than 1500 sq. ft.	Houses more than 1500 sq. ft.
Reading %	0.002	0.002	0.002
	(1.7)	(4.1)	(7.2)
Pupil/Teachers	−0.02	−0.01	−0.001
	(1.6)	(2.3)	(0.1)
Number of neighbors in 500 ft.	0.03	−0.002	−0.011
	(1.2)	(0.3)	(1.7)
Number of neighbors in 1000 ft. sq.	−0.001	−0.0005	−0.002
	(0.8)	(0.9)	(3.1)
Distance from N. H. Green	0.09	−0.04	−0.04
	(1.6)	(0.9)	(2.5)

[a] The figures in parentheses are t ratios.

form. On the other hand, there are several differences. The quadratic term in house size is significant ($t = 5.8$) with the expected sign. Though both the lot size and its square are significant with the expected sign pattern, the predicted price is less sensitive to variations in lot size than before. The variable for a laundry hookup has a rather large standard error, but the point estimate seems unreasonably high. A number of variables which in the linear form were significant are not in this version (average room size, number of heat zones, family room, etc.).

As to the neighborhood effects, the results are substantially the same as those reported above. Both the reading percentile and the pupil/teacher ratio are significant with the expected signs. It is estimated that moving the "standard" house from the fiftieth to the ninetieth percentile would increase its value approximately $1700 and increasing the pupil/teacher ratio from twenty-five to thirty would lower its price by around $2000. The distance and traffic-flow variables are insignificant in this form also, as are both measures of density. The results of some sample-splitting experiments are shown in Table 6. The estimates are somewhat less stable than for the linear version, and there is a tendency for some of the coefficients to become significant as one goes from smaller to larger houses. The regression equations were re-estimated using a reduced number of variables relating to the dwelling. As noted above this did not change the estimates of the neighborhood effects by very much. All regressions were estimated using a quadratic in time and various patterns of dummy variables. The effects of these alternative specifications upon the other estimated coefficients were generally negilgible.

In summary we have attempted to explain the prices paid for single-family dwellings by the physical characteristics of the buildings and attributes of the location of the house. The statistical results seem to us to be quite good, and the estimates of the effects of the neighborhood variables appear to be robust. In addition, it appears possible to estimate the magnitude of neighborhood effects using a relatively small number of controls for the nature of the dwelling. While it may seem farfetched to think of a housing market as being so precise as to exactly decompose the price of any home into the values of its parts or attributes, our results together with the results of others do suggest that it is definitely possible to measure the contribution to worth of many characteristics of single-family dwellings.

REFERENCES

1. M. J. Bailey, Effects of race and of other demographic factors on the values of single-family homes, *Land Economics*, **42**, May 1966.
2. J. P. Crecine, O. A. Davis and J. E. Jackson, Urban property markets: some empirical results and their implication for municipal zoning, *J. Law Econ.*, 10 October 1967.
3. D. M. Grether, "On the Use of Ordinal Data in Regression Analysis." Social Science Working Paper 27, June 1973.
4. J. K. Kain and J. M. Quigley, Measuring the value of housing quality, *J. Amer. Statistical Assoc.*, **65**, June 1970.
5. A. T. King and P. Mieszkowski, Racial discrimination, segregation, and the price of housing, *J. Political Economy*, **81**, May/June 1973.
6. V. Lapham, Do blacks pay more for housing? *J. Political Economy*, **79**, November/December 1971.
7. R. F. Muth, "Cities and Housing," University of Chicago Press, Chicago (1969).
8. H. O. Nourse, "The Effects of Public Housing on Property Values in St. Louis." Unpublished Ph.D. dissertation, Chicago, 1962.
9. W. E. Oates, The effects of property taxes and property values: an empirical study of tax capitalization and the Tiebout hypothesis, *J. Political Economy*, **77**, November/December 1969.
10. W. C. Pendleton, "The Value of Highway Assessability." Unpublished Ph.D. dissertation, Chicago, 1962.
11. R. G. Ridker and J. A. Henning, The determinants of residential property values with special references to air pollution, *Rev. Econ. Statistics*, **44**, May 1967.