Contract H-2882
Task Order 5

A report on research supported by *Department of Housing and Urban Development*
Office of Policy Development and Research

1367-1 December 1980

CHARACTERISTIC PRICES OF HOUSING IN FIFTY-NINE METROPOLITAN AREAS

by

Stephen Malpezzi
Larry Ozanne
and
Thomas Thibodeau

THE URBAN INSTITUTE
WASHINGTON, D.C.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EXHIBITS

LIST OF EXHIBITS (cont'd)

LIST OF EXHIBITS (cont'd)

PREFACE

This report is submitted as fulfillment of subtask 1.3 of task order 5 under Contract H-2882. The objective of task order 5 is to evaluate Fair Market Rents through application of hedonic index methodology to Annual Housing Survey data. Subtask 1.3 calls for the estimation and appraisal of hedonic indexes for all fifty-nine metropolitan areas included in the first three waves of the survey. The report contains this work. Following tasks will apply these estimates in the evaluation of Fair Market Rents.

ACKNOWLEDGEMENTS

This research was supported by funds from the Office of Policy Development and Research, U.S. Department of Housing and Urban Development, under Contract H-2882 to the Urban Institute.

We wish to pay special tribute to James R. Follain who was the moving force behind our original study of hedonic methods using Annual Housing Survey data (see Follain and Malpezzi, 1980a). Readers who are familiar with that study will note a great deal of similarity between it and the present paper; this similarity is a measure of our respect for his work.

The present paper has also benefited from many discussions with John Weicher, who is engaged in similar work at the Institute; Edgar Olsen and David Rasmussen who reviewed the manuscript; and Daniel Weinberg, Terrence Connell and Mark Wynn who contributed to initial design of the study. Steven Gold provided several valuable references.

Christine Hodge prepared the original manuscript. The many pages of discussion and the extensive tables which accompany them attest to the long hours she labored over this paper. They don't reveal, though, the conscientiousness and cooperativeness with which she did her part, and for this we thank her doubly. Charlene Livingston deftly prepared the final version.

CHAPTER I: INTRODUCTION AND SUMMARY

To a large extent, housing market analysis consists of comparing different dwellings. For example, measuring inflation requires comparing the price of housing today to that of some base period; but in the interim the housing stock has changed, through new construction, rehabilitation, conversion and demolition, so that we necessarily compare two different groups of dwellings. Other examples abound: comparing the price of housing in different locations, measuring the effects of racial discrimination in housing, studying the effects of government subsidies and tax policies on how we are sheltered, all require that we compare different dwellings. Such comparisons are made daily, not only by researchers, but also by those interested in more effective government programs, and by bankers, developers, and landlords. In fact, each of us make such comparisons every time we move or consider moving.

Everyone interested in housing markets, then, faces a common problem: how to compare different dwellings. Housing is not a homogeneous good like wheat or oil, but can be thought of as a bundle of diverse characteristics such as a number of rooms, of certain types, in a particular location, of a certain age, and so on. These specific characteristics are more amenable to comparison, so one may compare dwellings by comparing characteristics. Most people agree that comparing the value of, say, two houses with the same number of rooms in nearby locations is easier than comparing two dwellings with unknown characteristics, even though the rooms themselves may differ in size, the proximate location may not reveal that one is next to a freeway, and so forth.

The method of hedonic equations is one way expenditures on housing can be decomposed into measurable prices and quantities so that a market analysis can proceed. A hedonic equation is a regression of expenditures (rents or values) on housing characteristics, and will be explained in detail in the next chapter. Briefly, the independent variables represent the individual characteristics of the dwelling, and the regression coefficients are estimates of the implicit prices of these characteristics. The results provide us with estimated prices for housing characteristics, and we can then compare two dwellings by using these prices as weights. For example, the estimated price for a variable measuring number of rooms indicates the change in value or rent associated with the addition or deletion of one room. It tells us in a dollar and cents way how much "more house" is provided by a dwelling with an extra room.

The method of hedonic equations has been applied many times, and often provides key insights into the workings of housing markets. The results can be used to predict rents and values for standard dwellings in different cities (Follain and Ozanne, 1980), or one can estimate price differentials for housing of constant quality by some variable of interest such as time (inflation), age of structure (depreciation), length of tenure, race or location (Follain and Malpezzi 1980b, 1980c, 1980d). Price and quantity indexes derived from the hedonic estimates can be used to study the supply and demand responses of housing markets (Ozanne and Thibodeau 1980). Past studes were often limited to one or a few markets, for example, St. Louis (Kain and Quigley, 1975) or New Haven (King and Miezkowski, 1973). These studies may give insights into the workings of the particular market studied, but their general usefulness is limited. While markets work in similar ways, varying local conditions such as incomes and changes in population can produce different outcomes including the housing prices we and these earlier studies want to estimate.

The comparability of past studies is further weakened because different estimation procedures and different empirical specifications are usually employed. These differences make comparisons of the many hedonic studies quite difficult (Ball, 1973).

Past hedonic applications have been restricted by data availability to one time period and to one or a few markets. For example, multiple listing data were made available for 1967 in St. Louis (Kain and Quigley) and a mail survey was conducted in 1968 in New Haven (King and Miezkowski) Until recently, there simply has not existed a data set with enough coverage of both markets and dwelling characteristics to permit systematic estimation of a consistent, comparable equation in many markets. Now, however, such a data set exists--the Annual Housing Survey (AHS).[1] The metropolitan Annual Housing Survey presently covers fifty-nine large Standard Metropolitan Statistical Areas (SMSAs) and provides enough information on dwelling and neighborhood characteristics to make hedonic estimation feasible.

In 1978 we estimated an initial set of consistent hedonic equations for thirty-nine SMSAs covered in the first two years of the metropolitan AHS (Follain and Malpezzi, 1980a). The present work represents an extension of that earlier study: first, the indexes are estimated for the full

1. See U.S. Bureau of the Census (1976, 1977, 1978) for a description of the Annual Housing Survey. See, also, chapter 2.

59-area sample of the metropolitan AHS; second, an improved specification is employed. Both reports include renter equations (with rent as the dependent variable) and owner equations (with value as the dependent variable) in all SMSAs. Thus, this paper includes 118 hedonic equations.

The present hedonic equation estimates have been made as part of a larger project. The primary objective of that project is to construct price indexes which we can compare to Fair Market Rents (FMRs) used in the Department of Housing and Urban Development's Section 8 Housing Program. Briefly, FMRs are intended to represent the metropolitan area rent for dwellings that meet Section 8 quality and space requirements. The FMRs serve as rent ceilings and as the determinants of maximum subsidy levels. The FMRs are supposed to vary with market conditions, so they are set by market area and building type. The full hedonic equations reported here will be used to construct indexes of basic housing cost differences among the fifty-nine SMSAs. Both renter and owner cost indexes will be constructed. These indexes will be compared to FMR schedules to determine whether variation in FMRs reflects basic housing cost differences. The cost indexes will also be used in an analysis of the supply and demand factors causing the cost of housing to vary among SMSAs in an attempt to explain why FMRs should be expected to differ. Finally, specific hedonic coefficients estimated in the separate SMSAs will be examined to evaluate Section 8 mark-ups now used for additional bedrooms and elevator buildings as well as potential mark-ups such as for central city location. Comparable models are estimated in all SMSAs to make these analyses more manageable. If the project was instead focused on site-specific estimates of FMRs, then comparability among sites could be sacrificed to obtain site-specific models with lower prediction errors.

Beyond the scope of the present project lie a host of other issues that can be addressed with the hedonic estimates presented in this report. Some of the most obvious are: do racial minorities pay more for housing? Are CPI measures of rent and house price inflation accurate? By how much do current depreciation write-offs exceed actual dwelling decay? Which central cities have higher housing prices than their suburbs and are these differences growing or declining? We draw preliminary implication about each of these questions in reviewing our hedonic estimates to show the directions in which further analysis might proceed.

Because of the multiple uses envisioned for the hedonic estimates, we review both overall equation fit and the estimated prices of individual characteristics in the present report. As part of the review we draw tentative implications concerning some of the issues raised above. Our findings about the quality of the estimates and their policy implications are summarized here.

<u>Findings Relevant to the Quality of the Estimates</u>

(1) The hedonic model succeeds in accounting for much of the observed variation in the log of rent and value. The median multiple correlation coefficient (R^2) is .61 for the 59 owner equations and .67 for the renter equations. Only the Honolulu owner regression performs poorly with an R^2 of .32. Other than this outlier, the R^2 for owners ranges from .49 for Providence to .74 for Memphis. For renters, R^2 ranges from .52 for Newark to .82 for Raleigh.

(2) The standard errors of the models are compared favorably with other hedonic studies on similar data. The standard errors of the owner regressions range from 20 percent

(Paterson) to 37 percent (Birmingham) of average value. Most (36) are less than 30 percent; and the average is about 29 percent. The dollar value of the average standard error evaluated at the average value, is roughly nine thousand dollars. The standard errors of the renter regressions are distributed similarly to the owner estimates, although the renter estimates are more tightly grouped below 30 percent. The range is from 19 percent (Las Vegas) to 35 percent (Honolulu). The average is about 25 percent. The dollar value of the average standard error is, roughly, $40 for a $160 predicted rent. Standard errors in Follain and Malpezzi (1980a) and Ozanne, Andrews and Malpezzi (1980) are slightly larger for comparable models.

(3) Most of the coefficient estimates are significant at any commonly used level of significance. For renters, 76 percent of the estimates have t-statistics greater than 1, 63 percent have t's greater than 1.64, and 48 percent have t's greater than 2.58. For owners the results are similar.

(4) The average estimates of the coefficients are almost always consistent with a priori considerations. For example, the average estimates of the coefficients of the number of baths, the number of rooms, and the number of bedrooms are positive. The average estimates of the coefficients of the house age, and the dummy variable measuring the presence of deteriorated housing are negative. Several variables have the unexpected sign more often than should be the result of chance. For example, for renters the coefficients of structure-type variables such as SFATT and SFDET (single-family attached and detached) have signs which indicate that these variables are probably picking up locational effects rather than structural effects.

(5) The distribution of the residuals conform to model expectations with one exception. The residuals (the difference between each observation's reported log rent or value and that predicted by the estimated equation) are primarily symmetric about zero with half the residuals typically clustered within a range of .263 for renters and .338 for owners. Since these ranges are centered on zero, this means that half the predicted values lie between a plus and minus 14 percent of rent and a plus or minus 17 percent of median value. The one exception to the expected pattern is the finding of several dwellings with extremely low reported rents and values in spite of

fairly common predicted rents and values. We hypothesize that the low reported rents and values do not represent full and current market amounts and if our hypotheses are correct, then the inclusion of these outliers will impart an unknown but probably small downward bias to predictions from the hedonic equations.

Findings Relevant to Policy Issues Include the Following

(1) The average estimated depreciation rate for the flow of rental services is a constant six-tenths of one percent per year. The estimated average depreciation rate for owner-occupied housing starts at nine-tenths of one percent in the first year and falls to three-tenths of a percent in the twentieth year. Estimated rates differ considerably among SMSAs but usually remain well below the 3 to 6 percent depreciation rates permitted on rental property for tax purposes.

(2) Black and Spanish households are estimated to pay less for comparable quality housing than whites. For blacks the average discount is estimated to be 8 percent in rents and 15 percent for house prices. For Spanish households the average discount is estimated at 4 percent in rents and 7 percent for house prices. The rates show considerable variation among SMSAs. However, the black differential is never significantly greater than zero and the Spanish differential is significantly positive in only two cases--both for renters.

(3) Estimated SMSA rent and house price inflation ranges from near zero to almost 15 percent in each of the survey years even though the average stayed between 5 and 8 percent. Therefore, variation among markets appears more important than changes over time or between owner and renters during the 1974-76 years.

(4) House values are estimated to be lower in the central city than the surrounding suburbs in three-fourths of the SMSAs, with the average discount being about 7 percent. Rents are on average about the same in central cities and suburbs, but this average masks large off-setting premiums and discounts in several SMSAs. The largest discounts for values and rents all occur in the older Northeastern cities. In spite of these existing differences most SMSAs are estimated to have similar rates of house price inflation in the central cities and the suburbs. The exceptions are

again concentrated in the older Northeastern cities. Suburban prices are rising relative to the central city in places like Pittsburgh, Rochester and Providence, but central city prices are gaining on the suburbs in Washington and New York.

(5) Rents that include heating costs are estimated to be rising about as rapidly as those where heating expenses are paid separately. This may indicate tenant conservation or landlord absorption of a part of the fuel cost increase.

In conclusion, the results of this work provide analysts with a valuable tool for the study of housing markets. The work is therefore technical, in a sense, but should also be of interest to policy makers because of the light it can shed on racial price differentials, inflation, and other issues. This paper is both documentation for those who use the results as inputs for other studies, and exposition for those who are interested in the implications of these estimates themselves for government policy.

The next chapter describes the method of hedonic equations, and the data, in some detail. The particular specification we employ is discussed, as well as how and why we chose it. The third chapter presents the estimated equations. First, we summarize the overall performance of each equation, and examine the distribution of the individual coefficients. An evaluation of the results includes whether they conform to expectations: whether they are stable, and of reasonable sign and magnitude; and whether they differ by SMSA. We briefly outline how some estimates shed light on current policy issues, but detailed work in this area is left for later papers.

Finally, in Chapter IV we present an analysis of the residuals in our estimated equations. The residuals are examined for additional evidence on how well the data fit the model. Questions of symmetry, clustering, and outlying observations are considered for all equations. Suggestions from this investigation for specification and sample selection are then investigated for five cases.

CHAPTER II: SPECIFICATION OF THE HEDONIC INDEX

This chapter has two purposes. The first section is designed to make clear what a hedonic regression is. We discuss both the intuitive and theoretical arguments which have been put forth to support the use of hedonic regressions. The second section examines practical considerations in estimating the hedonic relation, such as how a market is defined and the choice of functional form for the equation.

SECTION 2.1: THEORETICAL BASIS

In the first part of section 2.1, the goal is to make clear how the hedonic technique works and why the method is valid. The assumptions upon which the method is based are explained with examples. Given these assumptions, the general hedonic relation is discussed.

Intuition Underlying the Hedonic Index

The first, and least controversial, assumption is that a house is a bundle of size, quality, and locational characteristics. An analogy can be made to a bundle of groceries. Some grocery bundles are bigger and better than others, depending upon the number and type of food items in the bundle. So too with housing. A house embodies many features: bedrooms, baths, a heating system, location, etc. The number and type of features embodied in a particular house distinguish it from other houses.

How can housing bundles be compared? It is simple to compare houses which are identical except for one characteristic. For example, a four-bedroom house contains more housing than an otherwise identical three-bedroom unit. Problems occur when units differ in more than one characteristic at a time. Does a three-bedroom unit with two baths represent more housing than a four-bedroom house with one bath? It depends, of course, on the value of a bathroom relative to a bedroom. The problem is easily solved in the grocery bundle example because all individual items have clearly marked prices. The more expensive bundle represents more groceries. This follows because the money used to buy the expensive bundle could be used to buy the less expensive bundle and there would still be money left over to buy more groceries.

Unlike groceries, prices of the individual features which comprise a housing bundle are not directly observable. This is where the second assumption comes in. The second assumption is that the rent or value of a housing unit stems from the quantity and type of characteristics it contains, and that the "prices" of the characteristics can be estimated from the rents or values of many units via multivariate regression analysis. A simple example which demonstrates the reasonableness of this assumption concerns the difference in values between two units which differ only with respect to the type of heating system. If one unit has a central heating system and the other has a fireplace, then the difference in the market value of the two units will equal the market valuation of a central heating system relative to a fireplace. Not all examples are so simple, but by pooling together many dwellings it is possible for multivariate regression to determine the relationship between rents, house values, and dwelling characteristics. The estimated regression coefficients are implicit prices which measure the value of each dwelling and neighborhood characteristic. For example, the regressions might determine that a central heating system adds 10 percent to the value of a house.

More formally, the assumptions suggest the following general hedonic relation:

$\bar{R} = f(S, N)$, where:

R = rent or value,
S = structural charcteristics, and
N = neighborhood characteristics, including location.

If this relationship is true, then a properly specified regression equation applied to appropriate data can provide precise estimates of the relationship. If the relationship is a linear one, then the estimates are interpreted as implicit prices of each of the characteristics which determine rent or value.

Is it likely that the relationship between housing characteristics and rent or value is the same for all types of households in all types of situations? Probably not. Four exceptions seem likely: (1) long-time tenants often receive discounts; (2) large families often pay more than smaller ones for the same unit; and (3) black households pay different amounts for the same unit due to prejudice (the result of prejudice could be higher or lower prices, as we will discuss in Section 3.2). In addition, (4) some renters receive utilities, furniture, and other services in addition to structure and neighborhood. The basic relationship should be modified to reflect these cases. Now,

$R = f(S, N, C)$, where

C = contract conditions, implicit and explicit.

Several of these contract condition variables are tenant characteristics. We emphasize that only those tenant characteristics which affect the prices faced for housing, or the supply prices, are included in the hedonic equation. For example, people with higher incomes can afford better housing, but they should face the same prices for identical dwellings as poor people. Income is not included in the equation because income represents the demand for housing, not the price of housing. Long time renters, on the other hand, are hypothesized to face lower prices for housing because of lower turnover costs. Therefore, length of tenure is included in the hedonic regression as one of the contract conditions.

Is it likely that the hedonic relationship is invariant over time? Probably not. The AHS survey data were collected over a three-year period in which the housing component of the CPI increased at annual rates in excess of 10 percent per year. This means that the differences in the values or rents of two identical units could be substantial if the units are surveyed at two different times. We include a time trend, t, which captures the effect of inflation. Now,

$$R = f(S, N, C, t),$$

which is the key relationship upon which this paper is based.

Search for a Theoretical Basis

Although simple and intuitive explanations serve a purpose, rigorous derivation of the hedonic relation is preferable. A derivation serves three purposes: (1) it places restrictions upon the general relationship which are useful in estimation; (2) it aids in interpretation of the results; and (3) it places checks upon the intuitive logic used above. In fact, several rigorously derived models have been put forth as the basis for hedonic indexes. The models are briefly described here.[1]

1. This discussion is from Follain and Malpezzi (1980a).

One widely cited basis is the household production model of the consumer proposed by Lancaster (1966) and Muth (1969). In this theory, utility is not a function of commodities (units bought in the market, such as a house), but rather of the characteristics embodied in a commodity, such as number of rooms. A specific functional relationship is assumed to exist between the characteristics and the commodities. Maximization of utility, subject to this relation and a normal budget constraint, gives rise to a hedonic function relating the price of a commodity to the characteristics embodied in it. Muellbauer (1974) has shown, however, that the conditions on both the utility and production functions which must hold in order to derive the hedonic relation in this way are quite restrictive.

In Muellbauer's critical review of the theoretical bases for the hedonic approach, he cites two other general models which give rise to the hedonic regression. The first he calls the Houthakker approach. In its simplest form, this model assumes utility is a direct function of the characteristics of a dwelling, and there exists a price schedule for characteristics, which consumers take as given. Maximization of utility, subject to a standard budget constraint and the price schedule for characteristics, implies a hedonic relation with the properties needed for price index construction and comparative analysis of housing quality. Muellbauer argues that the Rosen (1974) approach is basically an extension of the Houthakker approach. According to Muellbauer, all these models require the assumption that consumers face a fixed and known price schedule for characteristics which is based upon production costs. That is, it reflects supply conditions, not demand conditions.

The second family of models discussed by Muellbauer is developed from Fisher and Shell's (1971) "simple repackaging hypothesis." This model asserts that each "market good has a quality index which is a function of a set of physical characteristics" and which is "independent of market variables" (Muellbauer, 1974, p. 988). In other words, the quantity of a particular commodity can be expressed as a function of the characteristics and the relation is not influenced by supply or demand.

The thread which unifies these models is that they all give rise to the desired hedonic relation only under conditions which can be viewed as quite restrictive. In other words, the search for a rigorously derived basis for the hedonic relation is incomplete.

Does it make sense to estimate a model which does not yet have firm basis in theory? Yes. The reason is that time and time again attempts to estimate hedonic relations have produced results highly consistent with simple intuition. This kind of empirical support is hard to ignore. The housing prices we observe in the marketplace are related to structural characteristics of dwellings, location, neighborhood and tenant. In this paper we attempt to obtain precise estimates of the relationship between housing characteristics and housing prices.

The situation is analogous to the aggregation problem which haunts the study of macroeconomics. The frequently estimated macroeconomic relations are only derivable from a microeconomic theory of the consumer under restrictive conditions. Yet such relations continue to be studied and estimated. Why? Because the relationships estimated make sense and appear to be quite strong.

In summary, the search for a widely accepted and rigorously derived basis for the hedonic relation is not yet finished. Ideally, the model should not only demonstrate that a hedonic relation does exist, but should also make clear the conditions under which the relation can be used to construct constant quality price and quantity indexes in different markets. Until such a model is developed, the principal basis for hedonic analysis is relatively simple and intuitive. We shall see that most of our results are consistent with such a basis.

SECTION 2.2: EMPIRICAL IMPLEMENTATION

Overview

This section describes in some detail the actual specification used to estimate the hedonic relation. First, we describe the data, which is gleaned from the metropolitan Annual Housing Surveys. This is followed by our justification for using the SMSA as our definition of a housing market. Then we describe the functional form estimated and the construction of variables. The search which led to this specification is also described. In general, our present specification results from improvements to the model described in Follain and Malpezzi (1980a), so we highlight these changes and discuss their importance. Finally, we present evidence on the effect of some omitted variables. Many previous hedonic studies have included variables such as distance from the central business district, lot and house size, and certain kinds of neighborhood information, which are omitted from our specification because the AHS data do not include the information necessary for their construction. We show that the hedonic estimates we obtain are useful despite the lack of some desirable information.

Data

The data used in this study are from the 1974, 1975 and 1976 metropolitan Annual Housing Survey (AHS). The survey is designed and sponsored by the Department of Housing and Urban Development, and is conducted by the Bureau of the Census. Its purpose is to collect data on certain indicators of housing and neighborhood quality. The survey is designed to be compatible with the decennial Census of Population and Housing, but the AHS includes data on characteristics not included in the Census.

Separate surveys are carried out for the U.S. as a whole and for separate SMSAs. The data used in this study are from 59 SMSA surveys carried out from April 1974 through March 1975 (1974 SMSA surveys), from April 1975 through March 1976 (1975 SMSA surveys), and from April 1976 through March 1977 (1976 SMSA surveys). Although these SMSAs are widely distributed geographically, the SMSA data are not necessarily representative of the country as a whole. An SMSA comprises a central city of at least 50,000 population and one or more contiguous counties. Minneapolis-St. Paul, San Francisco-Oakland and Los Angeles-Long Beach contain two central cities each. Although SMSAs include the rural portions of the contiguous counties rural America is underrepresented. Also, the smaller SMSAs are underrepresented. There are 159 SMSAs with populations of 200,000 or more, but the AHS samples 59 of the largest.

The survey data are from personal interviews with a dwelling's occupants. The enumerators read the questions directly from a copy of the survey, which is reprinted in the printed report available for each SMSA.[1] The survey includes questions about household characteristics such as family size, race, and income; dwelling characteristics such as number of rooms and the presence of various defects; opinions of neighborhood characteristics, and some information on location. There are hundreds of questions in each survey, so no attempt will be made to summarize them here.

---

1. U.S. Bureau of the Census (1976, 1977, 1978).

The survey contains much of the information needed to estimate hedonic equations for the prices of housing characteristics. Structural characteristics are well represented. Lot size and the floor area of the dwelling are not included, although these often make significant contributions to other hedonic studies. There is little objective information on neighborhood characteristics, although there are many questions relating to the occupant's opinion of his surroundings. Finally, there is limited information on the dwelling's location. The effects of the lack of this information will be discussed at the end of this section.

There are two sample sizes. Twelve SMSAs have samples of approximately 15,000 units. In these, about half of those surveyed reside in the central city, and about half outside it. The other 47 SMSA samples are about 5,000 units each, and the number of central city respondents is proportional to the number that actually live there. The sample is selected from three populations: (1) housing units from the 1970 Census of Housing and Population; (2) new construction, sampled from building permits issued since 1970; and (3) new units located in areas not covered by a permit issuing office.

The Census sample is stratified to insure adequate representation of various races, income classes, tenure groups, and family types. The public use copy of the AHS includes a set of weights to make the sample representative of the population as a whole, but the regressions are unweighted. Therefore the means of variables reported in the appendix are not the best estimates of the actual distribution of those characteristics in each SMSA. However, checking these means against weighted population estimates reveals that they are often similar. We report unweighted means because they best characterize the data underlying the estimated regressions.

Several kinds of dwelling units and households are excluded from the estimation sample, either because they lack information needed for a hedonic regression or because respondents do not pay market prices for housing. Most obviously, vacant units and those households not at their usual residence are excluded. No cooperative or condominium owners are included because there is no survey information on the market value of their dwellings. Those who live in public or subsidized housing, or who do not pay cash rent, are excluded because their rent is not determined by the market. Other excluded categories include hotels, rooming houses, trailers, homes on more than ten acres, and owner-occupied dwellings which are part of commercial establishments or medical offices. Of course, any observations with missing data, or with missing responses allocated by Census, are dropped from the regression.[1] Although this seems like a formidable list of excluded categories, the great majority of non-vacant units remain in the sample. For example, in the Pittsburgh file there are about 4,700 occupied housing units, of which we use 947 in the renter estimation and 2,384 in the owner regression. That is, over 70 percent of the total non-vacant units are still included after sample selection.

---

1. There is one minor exception. See the discussion of the variable DFECT, below.

Market Definition[1]

Just as each market for apples produces a market clearing price for apples, each housing market produces a set of hedonic prices. This means that each set of hedonic prices we estimate must be derived from a set of observations from the same housing market. To use too broad a geographical definition of a housing market would produce biased estimates from an improperly aggregated sample. To use too narrow a definition would produce inefficient estimates because the estimates would not be based on all available information.

Much debate has centered on the precise definition of a housing market (Schnare and Struyk, 1976; Murray, 1978). Although most agree it is no larger than an SMSA, finer breakdowns are possible. The principal geographical possibilities are to divide the SMSA into central city and suburban markets, or even further into census tracts or neighborhoods. It is also possible to think of an SMSA market segmented by the kinds of households they serve. For example, separate markets may exist for blacks and whites, due to racial prejudice. Markets could also be defined in terms of housing quality (De Leeuw and Struyk, 1975).

If one believes in the existence of submarkets within an SMSA, there are basically two ways of dealing with them in the estimation of hedonic equations. First, separate regressions could be estimated for each submarket. This implies rather extreme separation because it assumes all the hedonic prices are different in each submarket. The second alternative is to introduce dummy (or indicator) variables for each submarket. This is

---

1. This discussion is from Follain and Malpezzi (1980a).

is more restrictive than the first alternative in the sense that it forces the coefficients to be equal in each submarket. Only the constant term, or the base price, is allowed to differ across submarkets.

This paper adopts the second alternative. The SMSA is defined as the basic housing market, although the rental market is separated from the owner-occupied market. Owner and renter markets may be closely related, but it is not clear how to compare rents and values, which would be necessary if owners and renters were pooled. Two submarket divisions are hypothesized within each SMSA market--central city versus suburbs and black versus white households. The submarkets are assumed to affect the base price or rent of a unit (constant term), with indicator variables for the race and location of a household. The coefficients of these variables represent the base price differential between the submarkets.

There are several reasons for this treatment of submarkets. The first is the strong *a priori* belief that the metropolitan area is the appropriate definition of the housing market. This is based upon the traditional urban economic analysis of Alonso (1964), Mills (1967), and Muth (1969). Second, such a definition is most appropriate for the long run purposes of the research effort, of which this paper is one part. In that larger research effort, the hedonic estimates will be used to analyze variation in Fair Market Rents (ceilings for HUD Section 8 rent subsidies) among metropolitan areas.[1] The impact of intermetropolitan differences can be best highlighted by using the SMSA as the subject of

1. See Ozanne and Thibodeau (1980), and Follain (1979), for examples of the work to be undertaken.

analysis. Third, the data are neither precise enough nor numerous enough to permit finer breakdowns. Census tracts are not identifiable, and breakdowns by race produce very small samples in many SMSAs. Finally, available statistical tests which can be used to study the existence of submarkets are probably not precise enough to warrant a purely empirical approach to defining submarkets (Schnare and Struyk, 1976).

Earlier work described in Follain and Malpezzi (1980a) included tests for segmentation based upon house quality as measured by household incomes. Analysis of metropolitan housing markets using The Urban Institute Housing Model (De Leeuw and Struyk, 1975) has concluded that markets are separated by housing quality. However, until the hedonic model is estimated, no measure of housing quality is available upon which to divide the sample. Rent or value is a possible measure upon which to divide the sample, but it is rejected because estimates obtained using a sample truncated upon the dependent variable are subject to serious bias. Another possibility is to divide the sample based upon household income (adjusted for household size). The idea is that income is positively correlated with housing quality, so splitting the sample upon income effectively splits the sample upon quality.

This was done for the eight SMSAs with which Follain and Malpezzi carried out their specification search. Separate regressions were run to high-income and low-income households for each tenure type. F-tests were computed, and in only four regressions (of a possible sixteen) was the null hypothesis of equality of the coefficients rejected. This suggests that splitting the sample using income as a proxy for quality is not appropriate.

There is one problem of empirical implementation which is unique to this study: estimating the two equations in 59 SMSAs. Most studies are for a particular market, and the final set of estimates which a particular paper reports is often the end product of much experimentation in that particular market. We do not experiment in each SMSA because such a process would be extremely expensive, and because we want to compare the final set of results across markets.

Choice of Functional Form

There is no strong *a priori* notion of the correct functional form. In earlier work, Follain and Malpezzi (1980a) estimated a linear functional form as well as a log-linear (semi-log) specification. In our present work we choose the log-linear for five reasons.

First, the semi-log form allows for the joint determination of expenditures in the regression. That is, the semi-log model allows for variation in the dollar value of a characteristic so that the price of one component depends in part on what else is in the house. For example, with the linear model, the value added by central airconditioning to a six room house is the same as to a ten room home. This seems unlikely. The semi-log model allows the value added to vary proportionally with the size and quality of the home. This fact, all else equal, favors the semi-log model.

Second, the coefficients of a semi-log model have simple and appealing interpretation. That is, the coefficient can be interpreted as the percentage change in the dependent variable given a unit change in the independent variable. For example, if the coefficient of a variable representing central airconditioning is .13, then adding it to a structure adds 13 percent to its value or the rent it commands. (Of course this is really the same as the joint-determination-of-prices advantage described above.)

Third, the semi-log form alleviates the common statistical problem known as heteroskedasticity, or changing variance of the error term. The presence of heteroskedasticity suggests you have not yet derived the best possible estimates. A total lack of heteroskedasticity is rare in applied work, but any simple transformation which reduces it and is otherwise acceptable is useful. Preliminary owner regressions showed that the semi-log model usually exhibited less heteroskedasticity than a linear model. Exhibit 1 presents the results of linear and semi-log regressions for Raleigh owners and renters, including residual plots. (Exhibits are found in numerical order at the end of the text. Tables are displayed within the text.) Heteroskedasticity is indicated if the plots exhibit a widening or narrowing pattern as you move from left to right. Notice that the problem seems greater for owners than for renters, and that there appears to be some improvement in the owner regression when the semi-log form is employed.[1]

Fourth, lacking a strong theoretical reason to prefer one form over the other, another useful criterion is explanatory power. However, the usual measure of explanatory power, the R^2 statistic, cannot meaningfully be directly compared between two regressions with different (linear versus logarithmic) dependent variables. Follain and Malpezzi (1980a, p. 27) reported that, using an appropriate statistical test of

1. For more on heteroskedasticity and the examination of residual plots, see Draper and Smith (1966), Chapter 3.

the hypothesis that the explanatory power of the two regressions is the same, they were only able to reject the semi-log model in favor of the linear model in only a few of their 39 SMSAs.[1] We conclude that explanatory power strongly favors neither model.

Fifth, the model must be computationally feasible. Alternatives to the linear and semi-log forms exist, but they are expensive to implement and difficult to interpret for large data bases such as the AHS.[2] That is why this discussion has revolved around the linear versus semi-log model.

Finally, we note that our independent variables are mostly dummy (or indicator) variables. This allows us maximum flexibility in estimation. Earlier estimates such as those described in Follain and Malpezzi (1980a) often constrained coefficients. For example, if the number of bathrooms is entered as one variable, then the percentage change in rent from adding a second bath is forced to be the same as the change from adding a third. The percentage change may well differ as more are added, so we try to use indicator variables wherever possible. When continuous variables are needed (such as in the age variable) because of the large number of possible values, we try to use higher-order terms (squares and cubes) to allow maximum flexibility. The actual construction of these variables will next be discussed in detail, followed by a description of the specification search that yielded these variables.

---

1. A discussion of the method of comparing explanatory power, known as the Box and Cox test, is in Rao and Miller (1971), pp. 107-11.

2. An example is the general transformation suggested by Box and Cox (1964). Such an estimation would be prohibitively expensive to undertake with our data, and the results are unwieldly for price index construction.

Variable Definitions

The variables which were included in our final specification are defined in Exhibit 2. The means of the variables, the ranges of the means (by SMSA), and some other summary statistics can be found in Exhibit 3. The next few paragraphs set forth our principles of variable construction. Then the variables are discussed group by group.

Variable Construction

Whenever possible, variables are coded as indicator variables.[1] For example, for renters the number of bedrooms is coded into the variables BED0, BED2, BED3, BEDG4. The first 3 take on the value 1 if there are 0, 2, or 3 bedrooms, respectively; otherwise they are 0. BEDG4 takes on the value of the number of rooms if that number is greater than 3, and is 0 otherwise. BEDG4 is not constructed as an indicator variable so we can discern, say, 5-bedroom dwellings from 4-bedroom dwellings without separate indicator variables. This is necessary because until we run the regression we don't know if there are any 5- or 6- or 7- or 8-bedroom dwellings in the sample, so we can't be sure an indicator variable will work. Note that a 1-bedroom dwelling takes on the value 0 for all 4 variables. This is the base case, or the number of bedrooms represented by the constant term. Table 1 gives some examples of how different numbers of bedrooms would be coded using indicator variables.[2]

1. Indicator variables are also commonly known as dummy, binary, dichotomous, or 0-1 variables.

2. See Maddala (1977), pp. 132-41 for a good introduction to indicator variables.

Table 1

Example of Indicator Variable Coding Scheme
for Number of Bedrooms

Actual Number of Bedrooms	Value of Intercept (Constant Term)	Value of BEDO	Value of BED2	Value of BED3	Value of BED4
0	1	1	0	0	0
1	1	0	0	0	0
2	1	0	1	0	0
3	1	0	0	1	0
4	1	0	0	0	4
5	1	0	0	0	5

Several variables (such as POOR, DEFECT, and BADHALL) are linear combinations of indicator variables. This constrains the implicit price of each condition included in one of these variables to be the same. The interpretation of individual coefficients is therefore difficult since we can't separate the effects of the several conditions in a variable. This form is used whenever it is likely that there would be insufficient observations of a condition in some cities to permit inclusion of a separate variable.

A few independent variables are treated as continuous, such as age of the structure. It's not feasible to construct a large enough number of indicator variables in such cases, so we add quadratic (squared) and sometimes cubic terms in addition to the linear to permit flexible estimation. For example, if only the linear variable AGE1 were included in the regression, a negative coefficent estimate would imply that rents or value decline at a constant rate with age. Adding a quadratic term (AGE1SQ) allows us to determine whether rents decline faster in earlier or later years.

Dependent Variables

The variables we wish to explain represent expenditures for housing services. CRENTLN is the dependent variable in the renter regression, and is the natural logarithm of monthly contract rent, as reported by the respondent. The dependent variable in the owner regressions is VALUELN. The enumerator asks homeowners the current market value of their dwelling, but instead of writing down the response, checks a box indicating which of fifteen intervals the response falls in. We recode these intervals to their midpoints, except for the highest interval.

Since the top interval is open-ended, and the distribution of house values varies widely from city to city, we estimate a different value for this interval for each city. The average property tax bill of people in the top category is estimated from a preliminary pass through the survey data. The tax rate for these people is estimated after a careful perusal of the tax rate information in the AHS published reports. The estimated average tax bill divided by the estimated average tax rate yields the estimated average value in this category.

Structural Variables

The first group of variables listed in Exhibit 2 includes a relatively straightforward set of dwelling characteristics such as number of bathrooms (B1, B2, B3) and bedrooms (BED0, BED1, and so on), number of other kinds of rooms (R1, R2, etc.), types of heating and cooling systems (SHEAT, RHEAT, EHEAT, ROOMAC, CENTAC), structure type (SFATT, SFDET, DUPLEX, ELEVP, NGT50), and age of structure (AGE1, AGE1SQ, AGE1CB, DAGE). There are other structural variables which are related to what is loosely termed housing quality, such as the absence of plumbing, presence of holes or cracks in interior surfaces, basement or roof leaks, and the presence of rats, among others.[1] These include NORAD, POOR, NOPRIVCY, NOUT, BADHALL, and DFECT.

Neighborhood Variables

Several variables measure the quality of the respondent's neighborhood. Most of these are based upon the opinion of the household. Such opinion data are not generally used in economic studies, so it is interesting to see if opinions are systematically related to rent and value.

The first three neighborhood variables are constructed from the household's rating of the street upon which the unit is located (EXCELN, GOODN, POORN).[2] A fair rating is the omitted category. The next variable, ABANDON, is constructed from the interviewer's answer to the question: Is there abandoned or dilapidated housing on the street? (Yes = 1, No = 0. The occupant is asked a similar question and the correlation between occupant and interviewer responses is quite high.) Finally, two renter variables are constructed from questions about specific neighborhood conditions (LITTER and NOSHOPS). LITTER takes on the value 1 if there is trash or litter on the respondent's street,

---

1. The presence of rats, which is used in determining the value of the variable DFECT, is the one exception to our rule of deleting observations with missing values for any variable. Since recent movers are not asked this question, we assign them the mean response.

2. The respondent was asked to rate the street in Wave I surveys, the neighborhood in Wave II and Wave III cities.

0 otherwise. NOSHOPS is a similar variable for the absence of convenient shopping.

Locational Variables

Other neighborhood variables represent geographical location. Variables in this set are all indicator variables for the county in which the unit is located or whether the unit is in the central city of the SMSA (CC1).[1] These variables undoubtedly represent many things such as the distance from the center city area and the quality of public services of the county. Forty-two of the fifty-nine surveys identify central city locations; the other seventeen have smaller populations and central cities are not identified because of Census confidentiality requirements. Seventeen SMSAs have at least one additional locational (county) variable. The Allentown SMSA has a county variable but none for central city. New York, with seven variables, has the most locational information.

BLACK and SPAN are indicator variables which equal one if the household head is black or Spanish, respectively. The persistence of residential segregation leads us to interpret these as neighborhood variables, since most minority households live in minority neighborhoods.

---

1. Note there are forty-two central cities identified but only 40 SMSAs with the variable CC1. In the Philadelphia regressions, central city is the omitted category. In New York, the central city is represented by five variables. Locational variables are listed in the separate data appendix, available from the authors.

Contract Conditions

CROWDS is the ratio of the number of persons in the household to the total number of rooms. CLOT, CLOTSQ, and DLOT are constructed from the length of time the tenant has resided in the unit. The first is a linear term, the second quadratic, and the third a dummy for those who moved into their dwellings prior to 1950.

These coefficients are interpreted as price differentials faced by households which live in crowded units, or households who have resided in the unit for a long time. It is expected that long-time renters receive discounts. The coefficient of the crowding variable is expected to be positive for renters and negative for owners, reflecting the costs of faster depreciation. The hypothesis is that crowded dwellings depreciate faster because of harder use. Owners of crowded dwellings would find their value decreasing faster; landlords would require higher rents to recoup the additional costs.

It has been hypothesized that live-in landlords charge lower rents to attract desirable tenants, since they have to face them daily (Merrill, 1977). LLBLG is an indicator variable included in renter regressions for landlord living in the building. If this hypothesis is true we expect a negative coefficient for LLBLG.

We want our rental coefficients to reflect the price of housing structure and location, but some renters pay for additional services such as furniture, parking, and utilities. Indicator variables are used to identify differences in contract rent due to these additional services. FURNINC and PARKINC take on, respectively, the value one if furniture or parking are included in contract rent, and are zero otherwise. HEATINC and NHUINC are similar variables for heat and non-heat utilities included in contract rent.

Measuring Inflation

Housing prices do not remain constant over time, and the Annual Housing Survey is given over the course of a year (April to March). The month of interview is recoded into the variable Q. The first month of the survey, April, is zero, May is one, and so on. The semi-log functional form of the regression allows us to interpret the coefficient of Q as the average monthly percentage change in the price of housing.

Renters often pay for utilities as well as for housing structure and location. It is quite possible that housing utility inflation rates differ from inflation rates for other characteristics. The variable QHEAT is another time trend, similar to Q, except that it is zero whenever heat is not included in rent. The coefficient of Q then measures inflation in rents due to changes in the price of structure and location. QHEAT measures the difference in inflation rates between those who pay extra for heat and those who do not.

Locational differences in demand for housing, as well as differences in supply costs, can result in differing rates of inflation in different locations in the same SMSA. The variable FORAY is an interaction term which measures the difference between inflation in the central city and its suburbs. It is entered in the forty-two owner regressions for which we have the necessary locational information.

The Specification Search

Now that we have described the variables in some detail we will discuss the method used to arrive at this specification. Briefly, the current specification is an extension of that used by Follain and Malpezzi (1980a). The criteria used to choose variables are: (1) consistency with the theory of hedonic indexes outlined in Section 2.1, and (2) the variables yield estimates of the correct sign, and statistically different from zero, in preliminary regressions.

Why the Specification Search is Important

The goal of the search is a model which may be applied to fifty-nine different SMSAs. It is desirable to fit the same specification to each SMSA for the following reasons. First, analysis of the individual coefficients is greatly complicated by estimation of different models in different locations. Secondly, it is very costly and time consuming to fit one hundred eighteen different models. Third, the model chosen does incorporate most relevant information available from the Annual Housing Survey. This model performs well in every SMSA except Honolulu (see Chapter III).

How the Experimentation was Carried Out

As noted, the specification we employ is based on that used by Follain and Malpezzi. Their specification search strategy employed the following four steps:

(1) Intensive experimentation and estimation was carried out for the Los Angeles SMSA--one of the SMSAs in Wave I with a sample of fifteen thousand housing units. Wave II and Wave III SMSAs were not yet available when the research began. The products of this stage were several different specifications and a long list of variables.

(2) Several specifications produced by stage one were estimated for six other SMSAs: Boston, Dallas, Detroit, Minneapolis, Phoenix and Pittsburgh. From this stage, a smaller list and two specifications were selected.

(3) These two specifications were estimated for all SMSAs in Wave I.

(4) After one modification based upon stage three, two specifications were estimated in all thirty-nine SMSAs.

The results of this estimation were carefully perused for several months by members of the Housing Division of The Urban Institute, as well as others,[1] and several modifications were suggested. These improvements were tested in the following four steps:

(1) Several new variables were tested in Pittsburgh and Phoenix as part of an evaluation of the original AHS hedonic indexes (Ozanne, Andrews, and Malpezzi, 1979).

(2) More experimentation was carried out in three Wave III SMSAs: Baltimore, Denver and Raleigh. A preliminary specification was chosen for each tenure group.

(3) These specifications were estimated in fifteen SMSAs.

(4) Examination of the stage three results resulted in several changes, and a final owner and renter model were chosen. These were used to estimate the results presented here. These are the models described above.

Summary of Changes in the Hedonic Specification

For those readers familiar with the Follain and Malpezzi specification we summarize the major changes in the model estimated. This

---

1. Suggestions for specification changes were also made by Edgar Olsen of HUD, and Sally Merrill and Dan Weinberg of Abt Associates.

list does not include every minor change in the way information is recoded into variables, but briefly outlines key differences. There are seven:

(1) New Dependent Variables. The old specification used by Follain and Malpezzi (F&M) used log of gross rent (contract rent plus utility payments). We use contract rent, relying on HEATINC and NHUINC (variables explained above) to account for utilities included in contract rent. Our coefficients are now interpreted as changes in rent for structure and location only, given a change in independent variables (dwelling characteristics). Also for owners, the open-ended value category now varies by SMSA.

(2) More Flexible Variable Construction. More extensive use of indicator variables and higher order (square and cube) terms results in fewer constraints in estimation. For example, the estimated price of a third bedroom is no longer constrained to be the same as that of a second bedroom.

(3) Recent Movers are now Included. Several service breakdown variables which performed poorly (wrong sign, insignificant) have been dropped. Examples are water and sewer breakdowns, and toilet breakdowns. Since recent (less than 90 days) movers were not asked the questions used to construct these variables, they were dropped from the F&M sample. We retain all recent movers. In particular, this assures a more reliable estimate of the inflation rate.

(4) Census-Allocated Responses are Dropped. For several key variables, including rent and value, the Census Bureau coders allocate

responses to respondents who do not answer the questions. When these observations are dropped the predictive power of the model is noticeably improved.

(5) Several Old Neighborhood Variables are Dropped. F&M included seven neighborhood variables constructed from the opinion questions in the AHS. Several performed perversely, and these are no longer included. Those that remain are the general neighborhood rating, now coded in binary form, and LITTER and NOSHOPS. The latter two are included in the renter regressions only.

(6) Property Tax Rates are Dropped from the Owner Model. The tax rate capitalization hypothesis states that the value of otherwise identical houses will vary by the differences in the present value of the future stream of tax payments (negative), and of services (positive). F&M included the log of the property tax rate in their owner model to account for capitalization. However, inclusion of this variable is likely to result in biased and inconsistent estimation. The tax rate is constructed from property taxes divided by value. That is, the dependent variable is used to construct one of the independent variables, so that regressor is correlated with the error term, violating one of the important assumptions of regression analysis.[1] Test regressions indicated that a tax rate variable probably picked up more of the

---

1. When an estimate is unbiased, one expects to estimate the true value of the parameter on average. When an estimate is consistent, adding more observations gives more precise estimates. If a regressor is correlated with the error term, the estimates no longer have these desirable properties. See Wonnacott and Wonnacott (1970), chapter 7.

error term than any capitalization effect, so it was deleted from the final specification.[1]

(7) Several New Variables are Added. ROOMAC indicates the presence of room air-conditioners. SPAN is an indicator variable for Spanish head of household. It measures the premium or discount paid by Spanish households for housing of constant quality (insofar as our other variables account for a unit's quality). Like BLACK, it probably reflects neighborhood characteristics. LLBLG is a variable for the landlord's presence in the building.

City to City Differences in the Model

As noted above, one of the estimation objectives is to use the same model in each SMSA, for two reasons: computational efficiency, and cross-SMSA comparison of coefficients. There are two kinds of exceptions to this rule.

First, if there are no observations of a particular characteristic in the sample for an SMSA, the variable representing that characteristic must, of course, be dropped from the regression. Table 2 presents the modifications made to several SMSA models because of this data problem.

---

1. See Thomas King (1977) for more on tax capitalization.

Table 2

Variables Dropped from Individual Regressions

SMSA	Tenure	Deleted Variable	Variable Description
Miami	Owners	SHEAT	Steam or hot water heating
Honolulu	Owners	SHEAT	
Honolulu	Renters	NORAD	Rooms without heat
Birmingham	Renters	PARKINC	Parking included in rent
Memphis	Renters	PARKINC	
Raleigh	Renters	PARKINC	
San Antonio	Renters	PARKINC	

Second, different SMSAs have different locational variables, because some public use files provide more locational information than others. Details on the interpretation of these variables is presented below in Section 3.3. Finally, some SMSA-specific models were estimated for those cities where our model performed less well than in most SMSAs. These results are discussed in Chapter IV.

Omitted Variables, and their Likely Effects

Now that we have discussed the model in some detail, it is useful to consider what is left out. The Annual Housing Survey does not contain information needed to construct several variables commonly used in hedonic estimation of rents and house values. In particular, several studies have emphasized the importance of distance to the central business district (CBD) or other employment centers (e.g., Muth, 1969), the area of the house and lot size (e.g., Noto, 1976), and objective neighborhood information (e.g., Kain and Quigley, 1970).[1]

To assess the effects of this omitted information on hedonic estimates, Ozanne, Andrews and Malpezzi (1979) estimated hedonic indexes using a data source which had some of the omitted information.[2] They concluded that the absence of this information made little difference in the predictive power of the equation. This means that work which relies on predicted rents and values, such as price index construction, may not be seriously affected by omitted variable bias.

Some problems remain regarding the interpretation of individual coefficients. Ozanne, Andrews and Malpezzi find that the estimates of <u>individual</u> coefficients are biased by lack of square footage, locational, and objective neighborhood information. Studies relying on estimates of individual coefficients are more likely to be affected by omitted variable bias than studies using predicted rents or values. Individual coefficients will be biased if omitted variables are correlated with some included variables. Ozanne, Andrews and Malpezzi find, for example, that the correlation between BLACK and omitted neighborhood characteristics imparts a downward bias to the race coefficient. On the other hand, omission of these neighborhood characteristics does not affect estimates of SMSA-wide inflation from

---

1. Examples of "objective" neighborhood information used in other studies are median census tract income, school expenditures, and crime rates.

2. The data were from the Demand Experiment of the Experimental Housing Allowance Program (EHAP) and were available for low income renters in Pittsburgh and Phoenix.

the variable Q, since the monthly samples are independent of location. Of course, cross-SMSA comparisons of biased estimates still yield useful information if the nature of the bias is known.[1]

---

1. Examples of such studies include Follain and Malpezzi (1980b, 1980c, 1980d, 1980e).

CHAPTER III: ESTIMATION RESULTS

This chapter presents the results of estimating the model described in chapter II to the fifty-nine SMSA Annual Housing Surveys currently available. The actual estimates are available in a separate data appendix (available from the authors). Exhibit 4 presents a typical set of results, for Washington, D.C.

It is difficult to report over five thousand regression coefficients without sacrificing some verve and clarity of exposition. For this reason, we attempt to summarize these results in the following manner. First, we discuss the overall performance of the hedonic model, focusing mainly on its predictive power, and whether individual coefficient estimates are consistent with our prior beliefs. Second, we address the variation of the estimates among SMSAs. It is desirable that the estimates exhibit some stability, yet if the differences in estimates can all be explained away as error, or statistical "noise," there is no point in estimating separate indexes for different markets. A central premise of this work is that housing markets are local and diverse. Estimates of the market clearing hedonic prices should vary because of this diversity.

SECTION 3.1: DOES THE BASIC EQUATION MAKE SENSE?

The answer to this question is based upon three criteria: (1) the explanatory power of the estimated equations; (2) the statistical significance of the coefficient estimates and whether their signs (positive or negative) conform to a priori beliefs, and (3) an examination of the residuals. Unless the estimates of the basic equation explain variations in rents and values for most SMSAs, the estimates are of little use. Likewise, if few of the coefficient estimates are statistically significant, it is unlikely that an analysis of them individually would produce much insight into the operations of housing markets. These criteria are necessary (though not sufficient) to guarantee the validity of the basic equation. This chapter examines the first two criteria; residual analysis is deferred until Chapter IV.

Explanatory Power

Explanatory power is measured by three related statistics: R^2, the F statistic and the standard error of the equation. All give insight into the predictive power of the full equation.

The hedonic model succeeds in accounting for much of the observed variation in rent and value (Exhibit 5). The median multiple correlation coefficient(R^2) is .61 for the 59 owner equations and .67 for the renter equations. The Honolulu owner regression performs poorly relative to other regresions, with an R^2 of only .32. Other than this outlier, the R^2 for owners ranges from .49 for Providence to .74 for Memphis. For renters, R^2 ranges from .52 for Newark to .82 for Raleigh.

Exhibit 6 presents stem-and-leaf plots of the R^2 statistic, by tenure type. A stem-and-leaf plot is a convenient way to summarize the distribution of any set of numbers.[1] To read the plot, find the first digit of the number (here, R^2) on the vertical axis (called the stem). The second digit is read off the leaf which is to the right of the stem, row by row. For example, the bottom row of the renter plot:

.5 22223

tells us that there are four regressions with R^2 of .52, and one with an R^2 of .53.

Forty-one owner regressions and 47 renter regressions have R^2 greater than or equal to .6.

Although all but two R^2s are greater than .50 and most are between .60 and .70, it is interesting to examine their distribution. Two patterns emerge from Exhibits 5 and 6. First, there is a positive correlation between the R^2 for each tenure type, city by city (.66). Second, in 41 SMSAs renter R^2 is greater than owner R^2. The amount of variation explained does depend somewhat on the tenure group, but the differences in R^2 are not overwhelming.

The standard errors of the estimates of the natural log of value and rent are presented in Exhibit 5. The standard error is a measure of the hedonic regression's ability to predict value and rent. Since the equation predicts either the log of value or the log of rent, standard errors are also in logarithms. They can be interpreted as

1. See John W. Tukey (1977) for more on the uses of these plots.

an average percentage error in predicted rent or value, with larger errors receiving larger weights because the deviations are squared before averaging.[1]

The standard errors of the owner regressions range from 20 percent (Paterson) to 37 percent (Birmingham). Most (36) are less than 30 percent; and the average is about 29 percent. The dollar value of the average standard error, evaluated at the average value, is roughly nine thousand dollars.

The standard errors of the renter regressions are distributed similarly to the owner estimates, although the renter estimates are more tightly grouped below 30 percent. The range is from 19 percent (Las Vegas) to 35 percent (Honolulu). The average is about 25 perce cent. The dollar value of the average standard error is, roughly, $40.

Finally, an F-test shows that all 118 regressions are statistically significant. That is, we reject the hypothesis that the observed multiple correlation is due to chance. Both renter and owner equations perform well by this criterion.

Statistical Significance of the Individual Coefficient Estimates

In this section, the statistical significance of individual coefficient estimates is examined. Also, the signs and magnitudes of the average coefficient estimates are examined. Exhibit 7 contains summary information about the individual estimates for renters, includ-

1. The interpretation of the standard error as a percentage is an approximation, because of the logarithmic transformation. That is, the (weighted) average of the logarithmic error term is not the same as the logarithm of the (weighted) average dollar error in predicted rents or values.

ing the mean and median estimates of each coefficient, the range of the estimates and the number of times the estimates are significant.[1] Exhibit 8 contains the same information for owners.

Some general observations should be made. First, most of the coefficient estimates are significant at any commonly used level of significance. For example, for renters, 76 percent of the estimates have t-statistics greater than 1; 63 percent have t's greater than 1.64 (significant level = .1 for 2-tailed test); and 48 percent have t's greater than 2.58 (significant level = .01).

Second, our method of using indicator variables and extremely flexible functional forms (extra quadratic and cubic terms) means that many interesting hypotheses regarding coefficients should be tested using an F-test, not a t-test. An F-test allows us to test

---

1. By statistically significant, we mean estimates which have a t-statistic greater than 1.64 or less than -1.64. This corresponds to a significance level of approximately .1 for a 2-tailed test. A significance level of .1 means that, over many trials, we could expect to <u>reject</u> the hypothesis that the coefficient is zero, when it is in fact <u>is</u> zero, one time out of ten. Our choice of significance level is necessarily somewhat arbitrary. In many econometric contexts, confidence intervals convey more useful information than simply reporting rejection of a null hypothesis. We choose to report the number of times coefficients are significant at an arbitrary level (.1) because it is difficult to summarize 59 confidence intervals for any single coefficient.

These and other hypothesis tests used in this paper assume that the errors in the models are normally distributed. In fact, analysis of the estimated error terms (residuals) show that the error term is probably not normal, but is slightly skewed. Several studies have shown that t-tests are not terribly sensitive to moderate deviations from normality, as long as the error distribution is still bell-shaped (see Theil 1971, pp. 615-16). This means our significance levels are not exact but reasonable approximations.

A second reason hypothesis tests and confidence intervals constructed from these results are not exact is that we estimate a model we have constructed using a specification search (see Leamer, 1978). This means that the probability of an incorrect decision (rejecting or not rejecting a hypothesis) is actually somewhat higher than classical statistical methods indicate, because the <u>data</u> have helped form the <u>model</u>.

the joint effect of several variables. For example, all four variables measuring the age of the structure could yield low t-statistics, while the effect of age on value remains strong.[1] Several F-tests of joint hypotheses have been included and will be discussed below.

The average estimates of the coefficients are almost always consistent with *a priori* considerations. For example, the average estimates of the coefficients of the number of baths, the number of rooms, and the number of bedrooms are positive. The average estimates of the coefficients of the house age, the occupant's rating of the street, and the indicator variable measuring the presence of deteriorated housing are all negative.

Of course, many variables occasionally exhibit coefficients that are not of the expected sign but are "significant," or, more properly, whose t-statistic has an absolute value greater than the chosen cutoff ($|t| > 1.64$). Some anomalies are expected, in a statistical sense, since the coefficient estimates are random variables, and we estimate several thousand.

Several variables have the unexpected sign more often than should be the result of chance. For example, for renters the coefficients of structure-type variables such as SFATT and SFDET (single-family attached and detached) have signs which indicate that these variables are probably picking up locational effects rather than structural effects. That is, we expect that if otherwise identical

---

1. That is, large values of AGE1, AGE1SQ, AGE1CB, and DAGE occur together. The t-statistics measure the effect of each variable, which may be diminished by including several collinear age variables, while their joint effect is strong. The F-test measures this joint effect.

single-family houses and multi-family units were located side by side, the single-family dwellings would command higher rents. The consistently negative coefficients for SFATT and SFDET, and the positive coefficient for NGT50 (greater than 50 units) indicate that these variables are probably proxies for the price of location, since single-family dwellings are concentrated where land prices are lowest.

Only a few variables other than structure-type variables consistently exhibit puzzling signs. NORAD, rooms without heating equipment, is negative and significant in 17 renter and 34 owner regressions, but is the wrong sign and significant in 6 renter regressions and 11 owner regressions. Now, the implicit prices of heating equipment might vary with climate, but the SMSAs where people pay more for dwellings with unheated rooms include northern cities such as Albany and Boston as well as Los Angeles and Miami. POORN, poor neighborhood rating, has the wrong sign and a t-statistic greater than 1.64 in 6 owner regressions, but conforms to expectations in 18 SMSAs.

SECTION 3.2: EXAMINING THE STRUCTURAL COEFFICIENTS

Examining Individual Coefficient Estimates

In the next four sections we discuss the distributions of estimated coefficients for the renter and owner models. The strategy followed is to identify the average estimated coefficient, the spread of the estimated coefficients about the average, and important outliers. The SMSAs with high or low estimated coefficients are examined for obvious similarities in location or growth rate. We indicate the number of statistically significant estimates as well as those that are inconsistent with our prior beliefs. The discussion of the distributions of estimated coefficients basically follows the list of regressors as they appear in Exhibit 2.

The first category includes all coefficients relating to the structural variables. This includes variables which reflect the dwelling's size, age, structure type, heating and cooling equipment, and several quality measures. Neighborhood variables are discussed in the second category of estimated coefficients. Neighborhood variables are constructed from the race of head of household, the respondent's opinion of the neighborhood, and the presence of abandoned housing, trash or litter, and convenient shopping. Central city, county and state variables are also included in this category. The third category covers the conditions of the contract for renters and flow of services for home owners. The estimated coefficients for length of tenure, persons per room, and adjustments for utilities, furniture and parking are discussed in this section. The fourth and final category includes several variables that measure inflation in housing prices.

Within each category the distributions of estimated coefficients for the renter and owner models are compared whenever possible. Finally, we suggest additional analysis of the coefficients to be performed at a later date.

The average coefficient estimates for the renter and the owner equations are presented in Exhibits 7 and 8. These exhibits display the mean and median estimated coefficient for each regressor.

A non-symmetric distribution or the presence of outliers causes the mean and median to take on different values. These exhibits also list the standard deviation and interquartile range for the distribution of each estimated coefficient.

In summary, the important features of a distribution of coefficients are the average value, the statistical significance of the estimated coefficients, the existence of outliers, and the estimated coefficients with unexpected signs. The discussion begins with the estimated coefficients for structural type variables, followed by variables which measure neighborhood effects, contract conditions, and inflation.

Dwelling Size Coefficients

The first group of structural coefficients are those related to the size of the dwelling. These variables are constructed from survey questions about the number of bathrooms, bedrooms, and other rooms; and for owners, whether the dwelling has a garage or basement. The distributions of estimated coefficients in this category are the most well behaved in both the renter and owner models. That is, they are generally symmetric about the average estimated coefficient, with almost all coefficients possessing the correct sign. The dwelling size estimates are almost always statistically significant at the 10 percent level. Recall the renter model constant term represents units having one bathroom, one bedroom, and two other rooms, so the estimated coefficients of the dwelling size variables must be interpreted as differences from this standard dwelling. The standard unit in the owner equation is any unit having one bathroom, three bedrooms, and three other rooms.

By examining the mean, or the median, estimated coefficient for the dwelling size variables, we see the advantage of using dichotomous variables to indicate the number of bathrooms, bedrooms, and other rooms. Consider the bedroom variables in the renter model. There is an average 20 percent discount going from a standard unit with one bedroom to a unit with no bedroom. However, two and three bedroom units command premiums of 12 and 23 percent over comparable one bedroom units. If the 20 percent discount for no bedroom were constant, two and three bedroom units would command premiums of 20 and 40 percent, respectively, compared to the one bedroom dwelling. A single variable for the number of bedrooms would not have captured this nonlinear effect.

The average values of the estimated coefficients for the number of bathroom variables, on the other hand, are approximately linear. In the owner equation, for example, a house with 1.5 bathrooms commands an 11 percent premium over a house with only one bathroom. A two bathroom house is worth, on average, 18 percent more than a one bathroom dwelling. The 18 percent premium is approximately one standard deviation away from a 22 percent premium--twice the premium associated with the 1.5 bathroom house. Similarly, a three (or more) bathroom house requires a 38 percent premium over a one bathroom house. Since the standard deviation for the distribution of estimated coefficients is 8.6 percent, the mean coefficient is within one standard deviation of a 44 percent premium. A 44 percent premium would be required if we used 11 percent for each additional half bath. In spite of the apparent linear relationship among the mean estimated coefficients for the number of bathrooms variables, we recommend using dichotomous variables. The approximately linear relation of the average coefficients does not apply to individual bathroom coefficients in many SMSAs. In addition, there is no need to collapse the categories to one variable to conserve the number of degrees of freedom because there are hundreds of observations in each regression.

Exhibits 7 and 8 list the number of times the estimated coefficients are statistically significant for each distribution. The number of bathrooms coefficients are always statistically significant in the owner model and nearly always significantly different from zero in the renter model. The owner model shows a similar tendency to produce more significant coefficients for the number of rooms other than bedrooms coefficient. This tendency is reversed with the number of bedrooms variables. For example, 45 of the 59 estimated coefficients corresponding to 4 bedroom owner-occupied units are statistically different from zero. All the estimated coefficients for 4 or more bedroom units are statistically significant in the renter model. The indicator variable for the presence of a garage is positive and statistically significant 57 times. The indicator variable for basement is positive and significant 42 times, and negative and statistically significant once.

An estimated coefficient is an outlier if its value is unlike the values of the other coefficients in the distribution. The discussion of outliers, unless otherwise noted, is limited to the coefficients that are statistically different from zero. Few dwelling size coefficient estimates differ radically from the norm, but there are several worth mentioning.

The estimated coefficient for the intercept for the Honolulu renter equation is much greater than the average of the intercepts for the other SMSAs. This implies the standard dwelling is much more expensive in Honolulu than anywhere else. Comparative housing cost data collected by the Bureau of Labor Statistics in 39 cities also find Honolulu to have much more expensive rental housing than other urban areas.[1] The high value of the intercept is accompanied by outlying values for several other coefficients. For example, the estimated coefficient for a unit with 4 rooms other than bedrooms in the Honolulu owner equation is the only estimated coefficient for that variable which is not significantly different from zero. Besides having an unusual concentration of outlying coefficients, the Honolulu equations have the lowest R^2 statistics for both renter and owner models.

Outlying estimates for three other dwelling size coefficients are worth mentioning. The estimated coefficient for no bedrooms in the Baltimore renter equation (-.42) is far below the next lowest estimated coefficient of Cincinnati (-.32). In the owner model, outliers appear in the distributions for the estimated coefficients of the R12 (units with one or two rooms other than bedrooms and bath-

---

1. See the rent component in Urban Family Budgets in U.S. Department of Labor (1977).

rooms) and BEDG4 (4 or more bedrooms) variables. Washington, D.C.'s estimated coefficient for R12 is a positive outlier and the only estimated coefficient not significantly less than zero. Finally, the value of Raleigh's estimated coefficient for a 4 or more bedroom house (0.0498) is more than 3 standard deviations below the mean of the distribution.

To conclude this discussion we indicate those SMSAs with consistently high or low estimated coefficients in several dwelling size categories. The estimated coefficients for the number of bathroom variables in the renter model are consistently low in Anaheim and in Buffalo. The SMSAs with high estimated coefficients in this category are Boston, Fort Worth, and New York. The Baltimore, Albany, Columbus, Detroit, and Rochester SMSAs have consistently low estimated coefficients for the number of bedroom variables in the renter equation. Similarly, the SMSAs with consistently high estimated coefficients in this category are Anaheim and San Antonio. In the owner model, the low estimated coefficients for the number of bathroom variables correspond to the Paterson and Newark SMSAs and the high estimated coefficients are found in Wichita, Dallas, and Oklahoma City. The Honolulu SMSA has a consistently low estimated coefficient for several of the number of rooms less bedrooms variables. Finally, the Raleigh, Dallas, Oklahoma City, and Phoenix SMSAs all have high estimated coefficients for the number of rooms less bedroom variables.

The coefficient estimates for dwelling size characteristics are among the most consistent performers in the hedonic regressions. These estimates are nearly always statistically different from zero, have the expected sign, and exhibit few outliers.

Structure Type

The second category of structural coefficients are the structure type coefficients. The renter model includes the variables SFATT (single-family attached dwellings), SFDET (single-family detached dwellings), DUPLEX (two units), and NGT50 (more than 50 units). The omitted category incorporated in the intercept is a unit in a building that has between 3 and 50 units. The owner equations have only one dwelling type variable, SFATT. Single family detached dwellings is the omitted category contained in the intercept. Multiple unit dwellings are omitted from the owner hedonic because the Annual Housing Survey only reports the value of single family houses.

For renters, the coefficients of structure-type variables have signs which indicate that these variables are probably picking up locational effects rather than structural effects. The more desirable low density structure types, SFDET, SFATT and DUPLEX, have consistently negative coefficients while structures with more than 50 units have largely positive coefficients. Since larger structures are typically located in areas with greater accessibility and higher land costs, the observed pattern could well be reflecting this locational difference. For owners, the distribution of estimated coefficients for single-family attached units is approximately normally distributed around zero. The coefficient of SFATT is negative and significant in

11 SMSAs and positive and significant in 17. The opposing benefits of low density and more accessibility probably account for some of the vacillation of the coefficient between discount and premium in the owner model. Older, slow growing areas often suffer the greatest decline in their central cities while the suburbs grow. On the other hand, rapidly growing areas tend to have vibrant central cities and suburbs. In slow growing areas low density could be dominating accessibility while the reverse occurs in rapidly growing areas. The five lowest estimated coefficients for SFATT are all slower growing, northeastern SMSAs: Albany (-0.38), Baltimore (-0.33), Allentown (-0.33), Philadelphia (-0.28), and Pittsburgh (-0.17). At the other extreme, four of the five largest estimated coefficients appear in western, rapidly growing SMSAs. These SMSAs are Wichita (0.47), Spokane (0.44), Los Angeles (0.28), and Denver (0.26). Milwaukee (0.28) is the fifth SMSA in this latter group.

Age of the Structure

Housing is a durable good which takes many years to be completely consumed. Some of the unit is consumed or used each year as paint and wood age and roofs wear down. If all units are identically constructed, inflation is absent, and the rate of maintenance and repair expenditures is the same for all units, then precise measurement of the rate of depreciation is possible by observing the value or rent of two or more units of different ages. This is not possible, however, because inflation does exist; because units are constructed differently; and because some households spend more on maintenance, repair and alterations than others. In order to estimate accurately the effect of aging on values and rents, it is necessary to control for inflation and quality differences in housing units. The hedonic technique is one way to control for differences in dwelling quality and inflation rates, but it cannot control for most differences in maintenance. The hedonic equations yield estimates of how rents and values depend on the inexorable ravages of time, and maintenance decisions made to combat that decay.

What is the importance of obtaining an accurate estimate of the depreciation rate? First, depreciation is a cost so that the faster dwellings depreciate the higher is the cost of housing. Second, because it is a cost, depreciation is an allowable deduction in computing taxable income from rental housing. Current tax law allows depreciation rates of 3 to 6 percent in early years of a project's life.[1] If true depreciation rates are less than those allowed for tax purposes, this is an important incentive for expanding the supply of rental housing.[2] However, there is little evidence on true depreciation rates so the size of the stimulus, if any, is hard to gauge. The hedonic estimates of depreciation presented here can help in quantifying that stimulus.

Age Variable Specification and Interpretation

The measure of depreciation in the hedonic regressions is derived from several variables measuring the structure's age. AGE1 is the age of the structure, and is constructed from a survey question about

1. Straight line depreciation on 31 years of useful life gives 3.2 percent annual depreciation. Double declining balance depreciation on a 31 year project gives 6.4 percent depreciation the first year, 6.0 percent in the second year, and so on.

2. See deLeeuw and Ozanne (1980), and Wykoff (undated).

when the house was built, and the date of the AHS Interview. The coefficient of AGE1 can be interpreted as the percentage change in rent or value given a one year change in age. However, it is not known that the depreciation rate for a dwelling will be the same when it is new as when it is 30 years old. Certainly automobiles depreciate faster in the first year than in later ones. To avoid constraining the depreciation estimate to be constant we include higher order terms AGE1SQ (age squared) and AGE1CB (age cubed).

Another problem involves measuring the age of very old dwellings. The AHS survey question on the date the structure was built asks the year built for post-1970 dwellings; earlier years are collapsed into six categories (1969-1970, 1965-1968, 1960-1964, 1950-1959, 1940-1949, and pre-1940). The mid-year of these intervals is used to construct the age of five of the six cohorts, but it is difficult to assign a reasonable number to the pre-1940 cohort. This category is open ended, and the average age of structures in this category probably varies greatly from city to city. For example, San Francisco has a fair number of turn-of-the-century dwellings in its current housing stock; Fort Worth has few, if any. The variable DAGE is an indicator variable signifying that a dwelling is in the early cohort. This variable allows estimation of depreciation rates in earlier years to remain unbiased by city-to-city differences in this oldest cohort.

AGE1, AGE1SQ, and DAGE are included in the renter equations; AGE1CB is added to these in the owner regressions (the cubic term was always statistically insignificant in preliminary renter regressions). The discussion which follows revolves around the continuous variables AGE1, AGE1SQ and AGE1CB. The coefficient estimates for DAGE are presented but not interpreted.

The estimated coefficients of the age variables are difficult to interpret individually. To properly explain the effect of age on rent or house value, the effects of all continuous age variables must be considered simultaneously. For example, if the estimated coefficients for AGE1, AGE1SQ, and AGE1CB are -.01, +.001 and -.0001, respectively, the estimated discount of a three year old dwelling relative to a new one is 3(-.01) + 9(.001) + 27(-.0001), or a discount of 2.37 percent. The price of a four year old dwelling is 4(-.01) + 16(.001) + 64(-.0001), or a 3.04 percent discount. The estimated depreciation rate <u>at three years</u> is the percentage decrease in value or rent at that time. Consider the age terms of the semi-logarithmic hedonic model:

$$\text{(A-1)} \qquad \ln V = \hat{\beta}_1 A + \hat{\beta}_2 A^2 + \hat{\beta}_3 A^3 + \hat{\alpha} X + e$$

where V is rent or value,

$\hat{\beta}_i$ and $\hat{\alpha}$ are estimated coefficients,

A is the age of the structure,

X represents all other independent variables, and

e is the residual.

Taking the derivative with respect to age gives the depreciation rate

$$\text{(A-2)} \qquad \frac{dV/dA}{V} = \hat{\beta}_1 + 2\hat{\beta}_2 A + 3\hat{\beta}_3 A^2$$

Using the numbers from our example above, the estimated depreciation rate at three years is:

$(-.01) + 2(.001)\ (3) + 3(-.0001)(9) = -.67$ percent

Note that the percentage decrease from year three to year four, 2.37 percent minus 3.04 percent, is also -.67 percent. This simple difference is a very good approximation to the true depreciation rate for the values of age we deal with. Depreciation rates reported in this section will be calculated from equation (A-2), however.

Estimation Results

Each of the distributions for the age of structure variables have statistically significant negative and positive coefficients. The signs for the coefficients of AGE1, AGE1SQ, AGE1CB, and DAGE typically oscillate for each SMSA. That is, when the coefficient of AGE1 is positive, the coefficients of AGE1SQ and DAGE are negative. Similarly, when the coefficient of AGE1 is negative, the coefficient of AGE1CB is negative and the coefficients of AGE1SQ and DAGE are positive. What do all these numbers mean? One way to effectively summarize the results is to plot the estimated discounts in rent and value by the age of the dwelling. This was done for all 118 models. All the results are not presented here for lack of space, but Exhibit 9 presents several such plots. The first two plots represent typical results. They are the mean discounts in rent and value by age for the 59 SMSAs. The other plots are less representative of the typical city but include features of interest we will discuss below.

Another summary measure of these results is presented in Table 3. These are depreciation rates for renters and owners at selected ages. Columns 1 and 4 give the mean for all SMSAs, the other columns are for the SMSAs named. The averages represent more SMSAs than the selected SMSAs in Table 3, which were chosen to illustrate patterns different from the average.

Do not be confused by the difference between the plots and the table. The plots are of discounts, or the level of the implicit price of age, while the table presents depreciation rates, which are the rates of change in the price of age. The former tells what a dwelling is worth, relative to a new dwelling, while the latter measures how fast its worth is changing with increasing age. The following summarizes our depreciation estimates.

Rents and values consistently decrease with age. This is, of course, intuitively appealing. On the average, a two year old dwelling rents for one percent less than a new dwelling, while a two year

Table 3

Depreciation Rates for Owners and Renters, for Selected Years, Selected SMSAs*

	Renters			Owners		
	Average	Boston	Anaheim	Average	Miami	San Diego
Rate at Year 1	-.0060	.0112	-.0173	-.0092	.0168	-.0033
Rate at Year 2	-.0060	.0104	-.0165	-.0084	.0147	-.0027
Rate at Year 9	-.0059	.0055	-.0108	-.0045	.0031	.0010
Rate at Year 10	-.0059	.0048	-.0100	-.0042	.0019	.0015
Rate at Year 20	-.0059	-.0024	-.0019	-.0029	-.0040	.0052

*SMSAs are selected to show rates significantly different from averages.

old owner-occupied dwelling is worth two percent less than when new. The average discounts for renters and owners can be read from the plots in Exhibit 9.

There are several estimates of appreciation and of essentially zero depreciation. Miami owners and Boston renters pay up to a 10 percent premium for a ten year old dwelling relative to a new one. Detroit renters pay premiums for older dwellings in the first twenty years; past that older dwellings are heavily discounted. Allentown renters pay up to an 11 percent premium for a 16 year old dwelling. The estimates of San Diego and Philadelphia owner depreciation rates, as well as that for Cleveland renters, are statistically zero. That is, age has no measured net effect on rents or values in those SMSAs.

<u>Depreciation rates, and discounts for age, differ among SMSAs</u>. The averages adequately represent a fair proportion of the estimates, but there are several cities which exhibit differences from the average pattern. Future work will explain systematic differences among estimated depreciation rates.

<u>The relative prices of older dwellings differ between tenure groups, even in the same SMSA</u>. This is not surprising. For example, consider that rents are returns to the current flow of housing services, while values are a stock concept. Rents might be expected to change mainly as the flow of housing services decreases as the units deteriorate. Values, on the other hand, reflect the present value of future housing services, as well as current services. Thus present value changes with the expected life of the house as well as with the change in the current flow of services as the unit ages.

Exhibit 9 demonstrates that, on average, owner dwellings depreciate faster in the first five years of a dwelling's life, but that for the next 25 years rents drop faster.

On average, rents decrease at a nearly constant rate, values at a declining rate. The average depreciation rate for renters is remarkably constant, ranging from 0.58 to 0.60 percent. Owner depreciation rates show more variation, from 0.9 percent in year 1 to 0.28 percent in year 20. By year 30, the only observation point after year 20 for measuring depreciation, the rate rises again to 0.6 percent. These depreciation rates are similar to the 0.5 percent estimate of Frank C. Wykoff (undated) for the value of rental properties. He used a hedonic approach similar to ours but lacking as much detail. Depreciation estimates by other methods and from other assets, e.g., commercial buildings, generally find higher depreciation rates, such as 1.0 to 2.5 percent.[1] The difference may be due in part to better controls on physical features of the dwelling. In any event, all hedonic depreciation rates suggest that current tax depreciation schedules considerably overstate true depreciation.

A Caveat. So far in this discussion, the coefficient of our age variables have been interpreted as an accurate indicator of depreciation. It is likely, however, that the model is imperfectly specified. For one example, the AHS lacks some desirable locational information such as distance to the central business district (CBD). Thus we may not be capturing all the influences of location upon dwelling rent or value. If so, and if age is correlated with

1. Bruggeman (1977), Hulton and Wykoff (1978), Taubman and Rasche (1969), Palmquist (forthcoming).

location--as it seems to be--then the coefficients of AGE may be biased. Of course, the direction of the bias depends on whether the older dwellings are located in the more or the less desirable locations.

Another source of bias is that old units which have dropped out of the stock are, of course, excluded from the sample. That is, sixty year old units are included only if they still command a positive rent or value. Sixty year old units which no longer command a value are excluded. This procedure, although it is unavoidable, does produce estimates which understate the average rate of depreciation for all 60 year old dwellings.

Changing construction quality can introduce another bias into the hedonic estimates of depreciation. If in the 1940s and 1950s dwellings were built with higher construction quality than those built in the 1960s and 1970s, then these old units will not have fallen in value relative to new units by as much as they have relative to their replacement cost. Since hedonic equations cannot control for construction quality very well they will understate depreciation in such cases. The reverse bias occurs if older dwellings are of lower construction quality relative to new dwellings.

In which direction is the net bias of our results? It is impossible to know. Our estimates tend toward the lower side of the previous estimates. However, none of the small number of depreciation studies is sufficiently definitive to draw firm conclusions. This must await further study.

Dwelling Equipment

The fourth category of structural coefficients are the estimates for dwelling equipment. These coefficients describe the heating system, cooling system, and for renters of dwellings in multiple unit buildings, whether the building has an elevator. The variables that describe the heating system for renter are RHEAT, an indicator variable for a unit having wall or room heaters with flue, and POOR, a linear combination which includes an indicator variable for primitive heating equipment such as portable heaters. The renter equation includes POOR and RHEAT. The owner equation includes RHEAT, SHEAT, and EHEAT. SHEAT indicates a house heated with steam or hot water heat, and EHEAT indicates a unit heated with electricity. Note that the owner indicator variables are not mutually exclusive because there are both equipment variables and fuel variables. For example, one can have an electric room heater, in which case EHEAT and RHEAT both take the value one. The omitted heating equipment category is central warm air heating for owners and central warm air, steam or hot water for renters. The cooling equipment variables are ROOMAC, an indicator variable for room air conditioners, and CENTAC, an indicator variable for central air conditioning. Finally, ELEVP, an indicator variable for units that are serviced by an elevator, is included in the renter equation.

The average estimated coefficients for the dwelling equipment variables all have the correct sign. For example, the estimated coefficient for RHEAT is negative and statistically significant in 56 SMSAs in the renter model and in 38 SMSAs in the owner model. The estimated coefficient for RHEAT is never positive and statistically significant.

The estimated coefficients for air conditioning enjoy similar success. The estimated coefficient of ROOMAC is positive and statistically significant 48 times in the renter equation and 29 times in the owner equation. This estimated coefficient is negative and statistically significant once in the renter equation (San Francisco) and four times in the owner equation. The estimated coefficient of CENTAC is positive and significant 53 times and 54 times in the renter and owner equations, respectively. The largest average premium or discount associated with dwelling equipment variables in the renter and owner models is the coefficient for CENTAC. The presence of central air-conditioning requires an average premium of 19 percent in the renter equation and 13 percent in the owner equation.

The outlying estimated coefficients for dwelling equipment variables are the coefficients of RHEAT in Honolulu renter (-0.54) and owner (0.47) equations and the coefficient of CENTAC in the Tacoma renter equation (-0.27). The largest estimated coefficients in the owner equation for the airconditioning variables both occurred in Houston (a 22 percent premium for room airconditioning and a 31 percent premium for central airconditioning). In the renter equation, New Orleans has the largest estimated coefficient for ROOMAC (0.19) and the second largest estimated coefficient for CENTAC (0.34). Other SMSAs with high estimated coefficients for the airconditioning variables are Memphis and San Antonio in the renter equation and Dallas and New Orleans in the owner equation. All these large coefficients occur in the South where airconditioning is understandably more important.

Quality Measures

The last category of structural coefficients corresponds to the variables describing the quality of the dwelling. The NORAD variable indicates whether the dwelling has any rooms without heat. The POOR variable provides information on units where water is absent, sewers are absent, there is no bathroom, there is no heat, or whether the unit is shared. The NOPRIVCY variable indicates whether the occupant must pass through a bedroom to reach the bathroom, and NOUT indicates rooms without electrical outlets. BADHALL is a linear combination of several indicators of public hallway condition in rental units only, e.g., hallways with poor lighting. The DFECT variable provides information on a number of nuisances like whether the roof leaks. Finally, the variable COOKE in the owner equation indicates the presence of an electric stove. As expected the coefficients of the POOR, NOPRIVCY, and NOUT variables are rarely positive and statistically significant while these variables are negative and statistically significant 59, 37, and 30 times, respectively, in the renter equation and 45, 39, and 39 times in the owner equation. In addition, COOKE is positive and statistically significant 55 times. The coefficients of BADHALL and DFECT are negative and statistically significant a total of 15 times and positive and statistically significant a total of 7 times.

The reduction in rents and values associated with POOR is much larger than the reductions associated with the other indicators of inferior dwellings. The average estimated coefficients of POOR are -0.25 and -0.17 in the renter and owner equations. The quality coefficients next largest in absolute value are the estimated coefficients of NOUT, -0.07 and -0.10 in the renter and owner equation respectively.

The outlying estimated coefficients in the distribution for structural quality variables in the renter distributions are the estimated coefficients for POOR in Honolulu (-0.50), and NOPRIVCY in Rochester (0.06). The outlying estimated coefficients in the owner distribution are the estimated coefficients for POOR in Paterson (0.40), and NOPRIVCY in Anaheim (0.21).

Summary of the Structural Coefficient Estimates

This concludes our discussion of the various categories of estimated coefficients relating to structural variables. The categories of estimated coefficients in this group are the dwelling size coefficients, the dwelling type coefficients, the dwelling age coefficients, the dwelling equipment coefficients, and the dwelling quality coefficients.

The distributions of coefficients associated with the dwelling size variables were the most stable distributions in the set. The estimated coefficients for the number of bathrooms, number of rooms and bedrooms, garage and basement are almost always statistically significant and have the anticipated sign. The structural type variables are not good proxy variables for lot size or the amenities of low density living. The distributions for dwelling age estimated coefficients show a tendency for the signs of estimated coefficients to oscillate, but taken together, they are consistent with slow depreciation of rents and values. In the dwelling equipment category the airconditioning variables provided the most explanatory power. The coefficients of ROOMAC and CENTAC were nearly always positive and statistically significant. The POOR coefficients in the dwelling quality category were negative and statistically significant more than BADHALL, NOUT, NORAD and NOPRIVACY, the other variables whose coefficients were expected to be negative. In addition, the average discount associated with POOR was greater, in absolute value, than any other average discounts for quality variables.

The discussion of the estimates continues with the neighborhood variables. This is followed by a discussion of the contract condition variables, and the chapter ends with measures of price inflation.

SECTION 3.3: NEIGHBORHOOD COEFFICIENTS

The preceding section emphasized the importance of a dwelling's structural characteristics. However, it's also true that structures have to be somewhere, and housing services are provided by locale as well as structure. For example, location provides access to employment, education, shopping, and recreation. The surrounding locale, or neighborhood, provides such satisfaction as can be derived from its perceived cleanliness, quiet, safety, and feeling of community. Government services, and the prices paid for them, also vary by location.

All these should be related to a dwelling's rent or value. Unfortunately, these neighborhood or location effects are difficult to quantify in a manner suitable for hedonic index construction. In addition, some locational information from the Annual Housing Survey which would be useful is not available because of Census confidentiality requirements.

Still, the AHS data permits us to construct several variables measuring neighborhood or locational effects. The variables describing the neighborhood in which the unit is located are the race variables (BLACK, SPAN), the condition of the neighborhood (ABANDON, LITTER), the residents opinion of the neighborhood (EXCELN, GOODN, POORN), the lack of convenient shopping (NOSHOPS), the central city indicator variable (CC1), and several county location variables.

Neighborhood Conditions

ABANDON, EXCELN, GOODN, POORN, LITTER, and NOSHOPS are included under this rubric. The first four variables are included in both the renter and owner models. The latter two are included only in the renter model, because they were insignificant in preliminary owner regressions.

ABANDON is an indicator variable for the existence of abandoned housing in the neighborhood. The individual coefficients are negative and statistically significant more often in the owner equation (42 times) than in the renter equation (38 times). The estimated coefficients for ABANDON in the Albany renter (-0.20), San Diego renter (-0.20), and Honolulu owner (-0.44) distributions appear as outliers. Philadelphia shows a large discount in neighborhoods with abandoned housing in both the renter (13 percent) and owner (23 percent) models.

The EXCELN, GOODN, and POORN variables represent the respondent's answer to the question "In view of all the things we have talked about, how would you rate this neighborhood as a place to live--would you say it is excellent, good, fair, or poor?"[1] The respondents who described

1. AHS question 104a. (U.S. Bureau of the Census, 1979). The 1974 survey asks the respondent to rate the street rather than the neighborhood.

their neighborhood as fair are the omitted category included in the intercept. The estimated coefficients for these variables usually have the anticipated sign and, with the exception of POORN, are often statistically different from zero.

The variable NOSHOPS, included in the renter model only, is an indicator variable for the absence of convenient shopping. The estimated coefficient for NOSHOPS is negative and significant in 15 regressions, and positive and significant in 3. The indicator variable for LITTER is negative and significant in 12 SMSAs, and positive in one.

The Hedonic Model and Race

The hedonic regressions include two variables constructed from the race or ethnicity of the respondent. BLACK and SPAN are dummy variables for black and Spanish heads of household, respectively. The coefficients for these variables are interpreted as the prices faced by the families, relative to white and oriental families, after adjusting for differences in housing quality. We think of them as neighborhood variables, because we believe the continued existence of housing segregation makes the racial composition of the family highly correlated with the racial composition of the neighborhood. The Annual Housing Survey does not identify neighborhoods, so we cannot test this hypothesis directly, but its reasonableness stems from recent evidence that racial segregation in housing markets continues in the face of increased incomes for minority groups and other social changes.[1]

The persistence of segregated housing has engendered much interest in, and several studies of, racial differences in the price of housing.

1. See Ann Schnare (1978).

That segregation exists is not in doubt, but its effects on housing prices are. Studies by John Kain and John Quigley (1975), John Yinger (1975), and Thomas King and Peter Mieszkowski (1973) have lent support to the idea that blacks pay more for housing. However, much recent work has provided evidence that ghetto housing is actually cheaper. Examples of these studies include Martin Bailey (1966), Brian Berry and Robert Bednarz (1975), Sally Merrill (1977), Ozanne, Andrews and Malpezzi (1979), and Follain and Malpezzi (1980b). Surveys of these and other studies can be found in Follain and Malpezzi (1980d) and Peter Mieszkowski and Richard Syron (1980).

Space precludes a comprehensive discussion of the effects of race on housing prices, but the salient points are these:

1. Either premiums or discounts for blacks can be consistent with segregation.

2. Empirical work has not yet provided conclusive evidence on the existence of premiums or discounts for ghetto housing.

3. Evidence exists that relative prices in the ghetto change significantly over time.[1]

Earlier studies have been limited by data availability. With the exception of Follain and Malpezzi (1980d), previous studies have been limited to one or a few cities. Several key studies used data from a decade or more ago. If diverse housing markets yield different premiums and discounts, and prices change over time, generalization from these market specific studies will be difficult, especially if the data are not recent.

With this as background, we present the estimation results for our racial coefficients BLACK and SPAN.

---

1. See Ann Schnare (1978).

The hedonic approach provides a straightforward way to adjust for quality differences between non-white and white housing, permitting a well-controlled test for racial price differentials. Note that the question is a narrow one, do non-whites and whites face the same prices for housing of similar quality? Broadly, racial discrimination has other deleterious effects; for example, non-whites may be restricted to certain low-quality ghetto dwellings. But so long as whites who live in similar quality dwellings face similar prices for housing with the same attributes (even if few whites live in such low quality dwellings) no price differential will be evident.

The hedonic approach has many advantages in studying racial price differentials. Data on the household level eliminate well-known problems of bias in aggregate studies. The reasonableness of the overall results, as discussed in section 3.1, makes us confident that the hedonic equation does a good job of standardizing for housing quality.

Our model includes two variables which measure ghetto price differentials, BLACK and SPAN. These are indicator variables for black and Spanish head of household, respectively. Earlier studies have focused exclusively on black ghettos, but the growing Spanish population in the southwestern United States faces discrimination as well. Omitting SPAN in SMSAs with large Spanish populations will also bias the results for BLACK. If SPAN is included, it measures the price of Spanish housing, and BLACK the price of black housing, relative to non-Hispanic white housing. Without SPAN, the hedonic yields no information on Spanish ghettos, and understates the black ghetto differential because much Spanish ghetto housing will be included with non-Hispanic white housing.

For renters, the average coefficient of BLACK is -.080, and of SPAN, -.039. But this result, that on average black housing rents for 8 percent less than white, and Spanish 4 percent less, masks the variability of the results. BLACK is negative in 50 regressions, but positive in 9, ranging from -.184 in Atlanta to .082 in Salt Lake City. However, none of the positive BLACK coefficients are statistically significant, while 39 of the negative estimates are. SPAN is positive in 15 cities (2 significant) and negative in 44 (22 significant).

For owners, the average coefficient of BLACK is -.148, ranging from -.006 in Oklahoma City to -.386 in Milwaukee. The Milwaukee estimate is an outlier; no other discount is greater than 29 percent. Note that there are no estimated premiums for black owners, and the discount is significant in 51 SMSAs. The average coefficient of SPAN in the owner equations is -.070, ranging from -.411 in Pittsburgh to .214 in Providence. The estimates were negative and significant in 26 regressions, and never positive and significant. Many SMSAs have few Spanish homeowners, leading to occasional large coefficients with large standard errors. The Pittsburgh owner sample contained only six Spanish owners; the Providence sample, three.

These results are consistent with those previous studies which found that ghetto housing is cheaper, after controlling for housing quality differences. Most estimates imply that blacks and Spanish pay less for housing of comparable quality than non-Hispanic whites, although Spanish renters pay premiums in a few cities. The results support the hypothesis that whites pay a premium for housing in predominantly white neighborhoods.

Present AHS data do not permit distinguishing the race or ethnic background of a respondent from that of his or her neighbors. Such information could be provided by the Annual Housing Survey at a reasonable cost, and without violating confidentiality requirements, by averaging the responses to race questions by neighborhood without explicitly identifying the location. We could then estimate ghetto prices without the bias introduced by having blacks and Spanish who don't live in the ghetto included with those who do.

The range of the results suggests that racial price differentials vary from city to city. Studies which rely on data from one or a few cities may then give conflicting results, but the advantage of this diversity is that we may construct a model which explains the variation in ghetto price differentials. This work can proceed now that we are provided with an extensive set of comparable estimates.

Location

The Annual Housing Survey contains several kinds of locational information. Respondents may live inside or outside the central city of an SMSA; they may have their county identified; and if the SMSA contains more than one central city, or spans more than one state, these locations may be identified. However, because of confidentiality requirements, surveys of smaller SMSAs contain little or no locational information. Sixteen of the fifty-nine SMSA surveys contain no locational information; twenty-five only identified the respondent as living inside or outside the central city (CC1); and seventeen have some additional information on counties, states, or a second central city. The Allentown SMSA has a county variable but none for central city.

All locational variables are indicator variables which take the value one if the respondent resides in that location. They are constructed to be mutually exclusive, i.e., if an identified county contains the central city, the county variable identifies county residents not living in the central city. They are labelled with a mnemonic for the county or state name. All these variables are listed in the data appendix available from the authors. The New York survey contains the most locational information; we were able to construct seven central city and county variables.

For many purposes it would be useful to have some locational variable which could be compared across SMSAs. Exhibit 10 presents the coefficient for inside/outside central city for the forty-three cities for which we have locational information. The coefficient is adjusted so that it reflects the discount or premium for the central city vis-a-vis a population weighted average of prices in the rest of the SMSA. The adjustment is as follows:

$$P^* = P_o - \frac{\sum_{i=1}^{n} P_i X_i}{1-X_o}$$

where P^* = adjusted coefficient,

P_i = coefficient for the i th location,

X_i = proportion of the sample in the i th location,

and $i=0$ = signifies the central city,

$i=1,\ldots,$ n signifies other locations.

The first thing to notice about these estimates is their variation. For renters, estimates range from -18.9 percent (Paterson) to 19.3 percent (Honolulu). Half the estimates are of each sign. The average differential is -0.5 percent for renters; that is, central city rental units are about the same price as comparable suburban units, on average, but the estimates vary widely from city to city.

For owners, the pattern is similar, although the owner differentials show a more persistent negative tendency. The average differential is -6.7 percent and three-fourths of the SMSAs have negative differentials. The variation of the owner differentials is greater than that for renters. Estimates range from -35.5 (Newark) to 11.2 percent (Honolulu) The standard deviation is also greater for owners (11.1 versus 8.7).

Since some SMSAs have several locational variables, and others only one, Exhibit 10 also presents the results of an F-test for the hypothesis that the joint effect of all of the locational variables (central city, county, and state) is zero. The number presented is the probability that an F-statistic as large as the sample value would be observed, given that the null hypothesis is true. A low value for this probability means that it is likely the locational variables do affect rents or values.

Not surprisingly, the F-test usually indicates that location affects rents and values. Miami, San Bernardino, and Oklahoma City are exceptions, with probabilities exceeding .1 for both tenure groups. Four other renter equations, and eight owner regressions, exhibit large probabilities. Of course, this does not mean that location has no effect on rents and values in these cities, but that the central city-suburb distinction is too gross to pick up locational effects.

SECTION 3.4: CONTRACT CONDITIONS

The estimated coefficients that price contract conditions are the length of tenure coefficients, the crowding coefficient, and the coefficient of variables that adjust contract rent for utilities and other services. Each group will be discussed in turn.

Length of Tenure

Both the owner and renter models contain variables constructed from the occupant's length of tenure, CLOT. CLOTSQ is CLOT squared, and DLOT is an indicator variable for the oldest class of tenants, those who moved in prior to 1950. The construction of these variables is very similar to the age of structure variables described in section 3.2.

While the construction of these variables is the same for owners and renters, their interpretation is very different. It is plausible that owners who have not moved recently fail to keep up with changing (and usually increasing) market values. It follows that such errors would be greater for long term owners than relatively recent movers. The owner length of tenure variables are intended to measure the average error in reported value arising from this source. The owner coefficients are not really discounts, but an adjustment to reported values.

Long time renters, on the other hand, have a precise idea of their rent, since it is almost always paid monthly. Long time renters receive discounts for at least three reasons. First, there may be lower supply costs for landlords renting to tenants who are a known quantity, and are often at least perceived as being more stable than many prospective new tenants. Second, it is easier for landlords to raise rents as new tenants move in. Such raises are often customary to recoup the costs associated with the search for a new tenant. Third, tenants have an incentive to remain longer than usual in dwellings which rent for less than market value.

The individual estimated coefficients for CLOT are mostly positive, for CLOTSQ are mostly negative, and are both positive and negative for DLOT in both the renter and owner models. The estimated coefficients for CLOT and CLOTSQ are statistically significant in nearly every case for the renter equations and in about half of the owner equations.

The individual length of tenure coefficients are not as easily interpretable as their combined effects. Exhibit 11 displays plots of the average discounts for length of tenure, for owners and renters, over the 59 SMSAs. The most obvious point is that renter discounts are much larger than the value adjustments. This is true for each individual SMSA as well as the average.

An F-test for the hypothesis that the joint effect of all length of tenure variables is zero is rejected in 58 renter regressions and 38 owner regressions (significance level = .1). On average, renters receive a 3 percent discount per year for the first six years, declining to less than 1 percent per year after the tenth year. The owners' value adjustment is much smaller. The annual adjustment is typically about one-half of a percent per year for the first few years, decreasing to a tenth of a percent or so after a decade. Note that although the value adjustment is small in magnitude, it is statistically significant in most SMSAs.

Crowding

The variable CROWDS is a continuous measure of the number of persons per room. It is included in the hedonic regressions because we hypothesize that crowded dwellings depreciate faster and require more maintenance. If this hypothesis is true, crowded dwellings will command higher rents, as landlords recoup their higher supply costs but lower values, since faster depreciating dwellings will be worth less.

The average renter coefficient for CROWDS is .027, ranging from -.050 in Raleigh to .088 in Newark. Sixteen estimates have the wrong sign (negative) but only two of these are statistically significant. Of the 43 estimates with the correct sign, 16 are significant. That is, most (but not all) results are consistent with the hypothesis that crowded dwellings command higher rents because of increased supply costs.

The average owner coefficient is -.047, ranging from -.117 in San Diego to .042 in Spokane. Only 7 cities exhibit the wrong sign (positive), and all are statistically insignificant. Fifty-two estimates are negative, and 37 of these are significant. The results are again consistent with our hypothesis: crowded dwellings depreciate faster and are therefore worth less.

Adjusting Rents for Utilities and Services

Contract rent includes payment for structure and location, but some renters receive additional services and utilities, while other renters pay separately. In order to properly compare contract rents, the hedonic model must include adjustments for utility payments and other services. We also estimate a rent differential for multi-unit dwellings where the landlord lives on-site. Five variables are in this category There are indicator variables for units that have heat included in rent (HEATINC), parking facilities included (PARKINC), furniture included (FURNINC), a non-heat utility included (NHUINC), and an indicator variable that identifies units in buildings where the landlord also resides (LLBLG).

In general, these estimates usually exhibit the proper signs and are statistically significant. Renters pay an average of 8 percent more for dwellings with heat included in contract rent. The range of the estimates for HEATINC is rather large--from a 27 percent premium in Springfield to a 34 percent discount in Honolulu. The large discount in Honolulu appears to be an outlier; the next largest discount is 15 percent in Sacramento. The coefficient of HEATINC is positive and significant in 44 SMSAs, and negative and significant in 4. The four cities with significant discounts, Miami, San Bernardino, San Francisco, and Sacramento, all have moderate climates.

Tenants whose rent includes utilities other than heat pay 4 percent premiums, on the average. Twenty-four estimates of the coefficient of NHUINC are positive and significant; four are negative and significant. Once again, the four negative and significant estimates are for warmer cities: Miami, San Diego, Birmingham and Honolulu.

The service variables PARKINC and FURNINC also conform to expectations. Renters pay an average premium of 9 percent for parking and 5 percent for furniture. PARKINC is positive and significant in half of the regressions, and never has the wrong sign when significant. FURNINC exhibits the correct sign in 35 of the 39 markets in which it is statistically significant.

Landlords who reside in their buildings have an incentive to exercise more care in the selection and retention of tenants. One way to retain desirable tenants, and attract a greater number of prospective tenants when vacancies occur, is to offer cheaper rents. LLBLG, the indicator variable for landlord living in the building, measures these discounts. The average estimate of this discount is 3 percent, ranging from a 12 percent discount to a 5 percent premium. The coefficient is negative and significant in a third of the regressions; it has the wrong and significant sign in one SMSA.

Summary

By and large, the coefficients for the contract condition variables are well behaved. That is, they are usually of the correct sign, reasonable magnitude, and are often significant. In addition, there seems to be a relationship between utility price estimates and climate.

SECTION 3.5: MEASURING HOUSING PRICE INFLATION

Recent acceleration in housing prices has focussed public attention on housing market inflation. Despite this attention, there are few alternatives to the Consumer Price Index (CPI), which is available for only 25 SMSAs.[1] SMSA specific measurement of housing price inflation is important because these markets are heavily influenced by local conditions, and broad regional aggregates may mask real differences among markets. Further, estimation of many SMSA rates is the first step in explaining inflation rates in terms of local market conditions.

1. The CPI was available for only 23 SMSAs during the 1974-77 AHS survey years used in this report.

Besides its limited availability, the CPI has other potential shortcomings. The method used to compute rent inflation for the CPI is to return to the same unit several times a year and inquire about current rent. Changes in quality are also inquired about so that rent changes are not attributed to inflation if the unit is substantially changed. Researchers have speculated that the index may understate price increases because gradual depreciation is not accounted for and rent increases accompanying substantial rehabilitation but in excess of the rehab costs are omitted. Also, the CPI rent index averages contract rents, some of which include utility costs, others of which do not. It is therefore impossible to identify separate utility and shelter cost increases, which is of increasing interest given the acceleration in utility costs since 1974.

The homeowner component of the CPI measures changes in the outlays necessary to purchase and operate a home. It is a function of movements in interest rates, utilities, and other cost elements faced by homeowners, as well as house values. Consequently, changes in this index do not necessarily match the movements in the value of constant-quality housing--a subcomponent of significant interest. Once the subcomponent is identified, there remain problems with its construction. CPI measurement of house price inflation relies on data from homes purchased with FHA insured mortgages. During the 1974-77 years covered in this report the FHA homes made up a much smaller and less representative sample than that available from the AHS.

Hedonic Estimates of Inflation

The hedonic method allows alternative measures of rent and house price inflation to be calculated. The inclusion of a time trend (variable Q) in the hedonic equations generates estimates of price change after correcting for changes in quality over time, including depreciation. The coefficient of Q in our models measures the average monthly percentage change in rents or values, after standardizing for quality. The average annual rate of inflation, compounded monthly, is:

$$([1 + \hat{\beta}_1]^{12} - 1) \times 100$$

where $\hat{\beta}_1$ is the estimated coefficient of Q. The coefficient of Q is positive and significant in 37 renter regressions and 48 owner regressions.

Two other variables, QHEAT and FORAY, permit the estimation of inflation differentials for rental units with heat included in rent, and owner units in the central city, respectively. QHEAT is an interaction term between Q and HEATINC (heat included in rent); FORAY is an interaction between Q and CC1 (central city dummy). Of course, FORAY is used only in the 42 SMSAs which have their central city identified.

With QHEAT included in the renter model, the coefficient of Q measures rent inflation for dwellings excluding most utilities. We identify this as the shelter component of rent inflation. The inflation rate for dwellings which include heat costs in rent is obtained from the sum of the coefficients of Q and QHEAT. On an annualized rate this gross rent inflation is

$$([1 + \hat{\beta}_1 + \hat{\beta}_2]^{12} - 1) \times 100$$

where $\hat{\beta}_1$, and $\hat{\beta}_2$ represent the estimated coefficients for Q and

QHEAT. The difference between the present and preceding expressions gives the differential inflation rate of the utilities component. The sign of the QHEAT coefficient tells the direction of the differential. If the differential is negative, then the utilities cost component is rising less rapidly than the shelter component; if the differential is positive then the utilities component is outpacing the shelter component.

The coefficients of QHEAT are about evenly divided between positive and negative values, and except for a negative outlier in Honolulu the monthly inflation rate differential for including utilities lies between a plus and a minus 0.2 percent. Five of the negative coefficients are significant (including Honolulu) and 8 of the positive ones are significant. These numbers suggest that inflation rates for the shelter and utilities components have been about the same. If our estimates are accurate they suggest that the large utility price increases of 1974-77 may have been offset in their effects on rents through conservation or through landlords absorbing a share of the increase. Additionally, the shelter component of rent may have been rising faster than suspected.

For owners, Q measures SMSA-wide inflation for cities whose central city is not identified. If the central city <u>is</u> identified and FORAY included in the model, then Q measures suburban inflation, and the sum of the coefficients of Q and FORAY yields our estimate of central city inflation. The SMSA-wide average is a weighted sum, calculated as

$$([1 + \hat{\beta}_1 + \hat{\beta}_3 c] - 1) \times 100$$

where $\hat{\beta}_1$ and $\hat{\beta}_3$ are the coefficients of Q and FORAY, respectively, and c is the proportion of the sample living in the central city.

The estimated coefficients for FORAY are split with about 40 percent being negative and 60 percent positive. Only 3 negative coefficients are significant and only 5 positive ones are significant. Thus, while there is a significant differential in central city inflation in a few markets, most SMSAs show no significant difference between central city and suburban house price changes.

Exhibit 12 contains the hedonic estimates of annual shelter rent and house price inflation for each SMSA. The house price represents an average of central city and suburban rates where they were separately estimated. The exhibit also reports the differential from inclusion of utilities for renters and between central city and suburb for owners. All figures are calculated as described above.

Average inflation rates and average differentials for utilities and central city location are summarized in Exhibit 12. Each wave is presented separately to observe changes in these rates over time. In addition to average rates, the maximum and minimum values are included. The price of rental structure increases by an average of about 5 percent per annum for the first wave, and 7 percent for the second and third. Owner occupied house values go up by an average of 8, 6, and 8 percent in each respective wave.

Several patterns are evident in Exhibit 12. First, the variation in inflation rates among SMSAs within each year is much greater than the average change from year to year. For both renters and owners, the standard deviations of the estimates (by year) are large enough that we cannot reject the hypothesis that the average rates of inflation are the same for all three years. Second, note the wide variance in the estimated differentials for utilities and location. The largest and smallest utility differentials are all in warm SMSAs. The largest positive central city inflation differentials (by wave) are in Washington, New Orleans, and New York; the largest negative differentials are in Pittsburgh, Rochester and Providence. This is suggestive of strong central city demand fueling inflation in the first three, and weak demand tempering inflation in the latter three. Future work with these estimates can provide more systematic explanations.

In summary, our inflation estimates indicate that, although housing price inflation in pervasive, using a single national estimate to measure inflation is misleading. Future work can use these estimates to explain variation in inflation rates.

CHAPTER IV: ANALYSIS OF RESIDUALS

In the previous chapters we describe renter and owner hedonic models and present our price estimates of those models. In this chapter we evaluate the validity of the estimated equations through an analysis of their residuals. The analysis is performed in two parts. In the first section we analyze the general pattern of estimated residuals in all owner and renter equations to see if they are consistent with our model specification. In the second section we analyze a small subset of estimated residuals that lie outside the expected pattern. These outlying residuals are examined in detail for three renter models and two owner models.

This chapter's analyses find that the observed residuals' patterns are generally consistent with the model specifications, although tests for normality of the distributions are rejected in most cases. The residuals are symmetrically clustered about zero in all owner and renter models. Typically, half the estimated residuals are within a range of .263 for renters and .328 for owners. Since these ranges are centered on zero this means that half the predicted values lie between a plus or minus 14 percent of median rent and a plus or minus 17 percent of median value. When observed residuals are plotted against the predicted value of the dependent variable they show roughly a constant variance in the renter model. However, in the owner model they show a definite tendency to cluster more tightly about zero as predicted values increase. More estimated residual get classified as outliers than would be expected from a normal distribution of error terms in both renter and owner models. The most notable feature of these outliers is that three-fourths of them are negative in both the owner and renter models. Still, fewer than one percent of the observed residuals are classified as outliers in almost all models, and the other 99 percent are approximately symmetric. Consequently, we conclude that the validity of the model and of the t-tests and F-tests is adequately supported by the observed pattern of the residuals.

Even though the general validity of the model is supported by the estimated residuals, we think the disproportionate share of negative outliers raises questions that deserve further investigation. Our analysis of outliers in five models finds most negative outliers to have reported rents and house values at the low end of their respective rent and value distributions in spite of fairly typical distributions of those dwelling characteristics included in the model. Attempts to alter model specifications to accomodate these outliers have been largely unsuccessful in bringing the observations back into line with other residuals. As might be expected, deletion of the outliers leads to substantial reductions in equation and coefficient standard errors, and large changes in a few coefficients. The negative outliers indicate that predicted rents and values constructed from our models are likely to have a downward bias. In a forthcoming paper predicting Fair Market Rents using our equations we will be able to make a limited investigation of the severity of this bias on the predictions.

Residuals in the Fifty-Nine Renter and Homeowner Models

In this section we analyze the residuals for fifty-nine renter and fifty-nine homeowner models. For each equation, the distribution of the residuals is examined for symmetry, for clustering and for the existence of outlying observations. When the data fits the model the residuals are symmetric about zero with few outliers. It is important to examine the residuals for symmetry since standard hypothesis testing using the estimated coefficients assumes the residuals are normally distributed. While a random variable that is symmetric about its average value does not imply that the variable is normally distributed, studies of the t-test find that test to be generally robust to the normality assumption as long as the underlying distribution is bell-shaped.[1] Since the hypothesis that our estimated residuals come from a normal distribution can be rejected in most owner and renter models, the questions of symmetry become very important in interpreting our test statistics.

Values of the residuals that are significantly far from their expected value of zero are labeled outliers. An outlier is an indication that for one reason or another the data may not fit the model. There are several reasons for an inadequate fit, however. Respondents may report inaccurate values for some questions or correct responses may be incorrectly transcribed to the data tape. Also, a cluster of outliers with similar underlying data characteristics is an indication that relevant variables are omitted from the equation. The existence of a large number of outliers is contrary to the normality assumption and therefore affects statistical hypothesis testing. In addition,

1. See Theil (1971), pp. 615-16.

the existence of outliers tends to inflate the estimate of the residual variance. Since the estimate of the residual variance is used in hypothesis testing and in the calculation of confidence intervals, outliers reduce the significance level of hypothesis tests and yield wide confidence intervals. The magnitudes of estimated coefficients are likely to be disproportionately affected by outliers as well.

Since we are estimating the same model in fifty-nine SMSAs except for locational variables, it is possible to compare the distributions of the residuals across SMSAs. Two comparisons are undertaken for each tenure type. First, we compare the spread of the residuals for each equation to the average or typical residual spread. We indicate the fitted equations with an unusually large or unusually small residual spread. We also look for similarities among the distributions of residuals by location or by size of the SMSA. Second, we identify outliers for each of the estimated regressions and find a consistent pattern of negative values among the equations.

Before proceeding to the analysis of the residuals we introduce the statistics that will be used. To avoid some of the problems caused by outliers we use order statistics to analyze the residuals. Order statistics are based on the rank of the numerically sorted data. The median, defined as the value for which half of the observations have smaller values, is a familiar order statistic. The closeness of the median value to zero gives one indication of symmetry centered about zero since half will be above and half below the median and since the mean of the residuals is constrained to be zero. To estimate the spread or clustering of the residuals we use the difference between the third and first quartiles, called the interquartile range (IQR), of the residuals. The first and third quartiles are similar to the median of the distribution but have the property that one-fourth and three-fourths of the observations have smaller values, respectively. Unlike the usual estimate of the residual variance, the sum of squared residuals divided by the number of degrees of freedom, an outlier will not inflate the estimate of the interquartile range of the residuals. The similarity of the first and third quartiles in absolute value provides an indication of symmetry of the residuals about zero which is independent of outliers and of the median.

A residual is called an outlier if its value lies much below or above the values of most residuals. We classify a residual as a negative outlier if its value is three or more IQRs less than the first quartile of the residuals. Similarly a positive residual is considered an outlier if its value is three or more IQRs above the third quartile. This definition of an outlier is similar to the idea that a random variable does not conform to a hypothesized distribution if it is several standard deviations away from its expected value. The number obtained by computing the first quartile minus three times the IQR is called the lower fence. The corresponding upper fence is the third quartile plus three times the IQR. The concept of fences using order statistics is similar to the concept of a confidence interval. The probability of an observation appearing outside these fences, assuming a normal distribution for the residuals, is less than .0001. Out of about 250,000 residuals in the 118 equations we have estimated, at most three outliers would be expected to occur. To reiterate, the advantage of using order statistics is that outliers do not inflate the estimate of the residual variance. Similarly, the calculation of the lower and upper fences is not significantly affected by the presence of outliers.[1]

In addition to the above order statistics, the analyses of this section makes use of stem and leaf plots for comparing renter or owner statistics among SMSAs. The method for reading these plots has been discussed in Chapter III.

We begin our residuals analysis with a listing of several statistics for renter and owner equations in every SMSA. Exhibits 13 to 18 list the number of residuals, their median value, the IQR of the residuals, the number of positive outliers and the upper fence, and finally the number of negative outliers and the lower fence. Exhibits 19 to 21 list the total number of outliers and their percentage of all residuals. The first and third quartiles are not listed but can be obtained from the lower and upper fences by adding or subtracting three IQRs as appropriate. The number of residuals equals the number of observations used to estimate the equation and does not reflect the size of the SMSA population.

The IQR, which measures the spread or clustering of the residuals, is given in column 4 of Exhibits 13 to 18. Comparison of renter and owner models in the same SMSA shows the spread to be larger for owners in most cases. This reflects the generally greater spread in reported house value than in reported rent, not the quality of fit in owner versus renter models. The variance of the logarithm of value is, in

---

1. Our use of the IQR in defining outliers follows the work of John W. Tukey (1977).

fact, greater than that for rent in all samples except Boston, Philadelphia, Honolulu and New York.

The distribution of IQRs for residuals in the renter models is displayed through use of a stem and leaf plot in Exhibit 22. The median IQR of the residuals is .263. The distribution looks normal except for large spreads in Pittsburgh (.366), Albany (.360) and Boston (.345). The tightest clustering occurs in Rochester and Las Vegas where both IQRs are .199. A comparable stem and leaf plot in Exhibit 23 shows the IQRs for owners to be generally higher than for renters--as noted above. The median residual spread for owners is .338. Birmingham has the widest spread with an IQR of .419 while the tightest clustering is in Paterson (IQR of .218). The IQRs for Pittsburgh of .366 for renters and .404 for owners rank among the largest for both tenure types.

When comparing IQRs among SMSAs it should be kept in mind that the IQRs reflect both goodness of fit and the underlying variation in the dependent variable. For example, the IQR for residuals in the Paterson owners equation is the lowest for all SMSAs but the Paterson R^2 statistic of .49 is also among the lowest. Paterson's low IQR is more a result of the relatively small variance in reported house values than it is a measure of the model's success. The reliability of predicted values from a model should be viewed as a function of the R^2 or F-statistic, the IQR of the residuals, and the proportion of outlying residuals. The R^2s and F-statistics are discussed in Chapter III; the outlying residuals are addressed next.

Exhibits 19 to 21 list the number of residuals with values that lie beyond the calculated fences. The adjacent number in parentheses is the percentage of observations labeled outliers, which is a better measure of model behavior since the number of total residuals varies widely. The stem and leaf plot in Exhibit 24 shows that the distribution of the percentage of outliers for the renter equations is not symmetric about the median value of .53 percent due to the long tail for high percentages. Honolulu's value of 1.87 percent outliers is clearly larger than the expected value for the distribution. Other SMSAs with large percentages of outliers are Rochester, Anaheim, Denver, Orlando, and Omaha. Four of these six cities are new rapidly growing SMSAs. Rapidly and slowly growing SMSAs are equally prevalent among SMSAs with the lowest fractions of outliers. The distribution of the percentage of outliers for the homeowner equations is given in Exhibit 25. With the exception of Louisville (1.06) and Baltimore (1.00), the distribution is symmetric about the median value of .35 percent. This distribution has a smaller variance than the corresponding distribution for renters. In addition, the owner models have fewer estimated equations with more than one percent outliers compared to the renter equations. These comparisons suggest the owner models provide better fits to the data which is surprising since the renter equations typically have better R^2s and F-statistics. The analysis of the residuals for three specific renter equations, provided later in this section, will suggest an explanation for the apparent paradox.

Columns 5 and 6 of Exhibits 13 to 18 list the number of positive and negative outliers. The number of negative outliers is greater than the number of positive outliers in 53 of 59 renter equations and 47 of 59 owner equations. Out of all outliers in the renter models, 77.7 percent are negative and in the owner models 71.9 percent are negative. Exhibit 26 shows the classification of outliers by sign and tenure group.

The preponderance of negative outliers suggests omitted variables or bad data since one expects the same number of positive as negative outliers to arise by chance. A negative outlier implies either that the unit's rent or value is seriously under reported, over predicted or both. Long time homeowners might under report house value since values have been rising rapidly recently. The length-of-tenure variables in the homeowner model should adjust for the average under reporting of long-time occupants, but there could be wide variability in the amount of such under reporting. Renters are more likely to know their rent precisely. However, if these rents are below market levels because the tenant works for or is related to the landlord, the equations would overpredict rents.

The greater number of negative outliers in most equations also indicates a skewed distribution of residuals. To see whether this skewness also occurs in the other residuals we examine the lower and upper fences and the median. The fences, which are equidistant from the first and third quartiles, have similar numerical values if the inner half of the residuals are symmetric about zero. A larger absolute value for the upper fence indicates a downward skew for these residuals; a smaller value indicates the opposite skew. The upper and lower fences for the fifty-nine renter and owner equations appear in columns 5 and 6 of Exhibits 13 to 18. Albany renters, the first model in Exhibit 13, has upper and lower fences of 1.24 and -1.24 indicating symmetry. In the other equations the fences are generally close in absolute value.

Although the fences are generally close in size, there is a tendency for the upper fence to be greater than the lower one. Out of the 49 times the fences differ in the renter model, the upper fence is larger 47 times. Out of the 41 times the fences differ in the owner model the upper fence is larger 26 times. Thus, the inner half of the residuals are basically symmetric but to a limited extent show the downward skew also found among the distribution of outliers. Median values, reported in column 3 of Exhibits 13 to 18, are all close to zero supporting the basic symmetry of the fences, though the renter medians are disproportionately positive suggesting the same slight downward skew shown by the fences and the outliers.

Our examination of residuals among the fifty-nine renter and owner models finds the inner half of residuals to be basically symmetric about zero and appropriately clustered for most equations. Also, the proportion of outliers is greater than one percent of all residuals in only a handful of cases. For these reasons we believe the models generally fit the data well and that the t-tests and F-tests presented earlier are reliable in spite of the failure of most models to meet strict normality tests for the residuals. We hasten to add that the preponderance of negative valued outliers suggests a specification or data shortcoming needing further analysis. That is the task begun in the following section.

Residuals Analysis and Re-estimation in Five Equations

In an analysis of residuals the choice must be made between attributing an unusual observation to error and deleting it from estimation, or keeping the data point because it contains important information about the model. In the second half of this chapter we estimate revised equations based on an analysis of the outliers, while first keeping, then deleting the remaining outliers. We point out the advantages to deleting outliers but remind the reader that a few will arise naturally from a large number of observations on a normally distributed random variable. Therefore all outliers ideally should not be deleted. The problem of course is that the valid observations are difficult to distinguish from the invalid ones.

Our procedure for analyzing the outliers of an estimated model consists of several steps. First, we examine the plot of the residuals versus the predicted values of the dependent variable. We examine this graph for obvious patterns in the residuals. Characteristics of the outliers are then examined to determine whether there are any similarities among the observations that generated the outliers. We compare the full sample distribution of variables to the distribution of these variables for the outliers.[1] Interaction terms are introduced in an equation whenever the distribution of outliers by regressors is different from the sample distribution. The revised equation is estimated using least squares and differences in the models are noted. Finally, observations which are outliers in the revised model are deleted, and the revised equation is reestimated with the smaller sample. We note changes that occur in the estimated coefficients of

1. Only variables included in the regression are used in the comparison. Other variables available from the AHS were not used because of the cost of merging regression results--the residuals--with the original AHS user tapes. This should be a first step in future analysis of the residuals.

the interaction terms and in the other regression coefficients. In addition, we note the reduction in the standard error and related statistics.

We chose to analyze the Anaheim renter, Baltimore renter and owner, Chicago renter, and Fort Worth owner equations. These models represent a cross-section of the fifty-nine SMSAs by size, location, rate of growth, and sampling period. In addition, these models exhibit interesting patterns in their residuals. Anaheim is a small, rapidly developing SMSA in the Southwest. It has the second largest percentage of outliers for the renter equations but an equal number of positive and negative outliers. The IQR of the residuals for Anaheim is among the smallest in the distribution. Unlike Anaheim, Chicago is a large, already developed SMSA in the Midwest with mostly negative outliers. Anaheim is included in Wave 1 of the Annual Housing Survey (AHS) while Chicago is sampled during Wave 2. Baltimore represents Wave 3 of the AHS and is an older, northeastern American city. In Baltimore, as in Chicago and most of the renter equations, the number of negative outliers is much greater than the number of positive outliers.

The estimated equations for Fort Worth and Baltimore represent the owner models. Fort Worth is a rapidly growing southern city sampled during Wave 1 of the AHS. Unlike the majority of the owner equations, Fort Worth has more positive than negative outliers. Baltimore is representative of the typical owner model since it has a greater number of negative than positive outliers. Both Fort Worth and Baltimore have high IQRs of the residuals and high percentages of outliers. Chicago is the only SMSA with 15,000 observations included in the residuals analysis because of the cost of working with the larger sample.

We begin the analysis of the residuals in individual models with the Baltimore renter equation. The graph of the residuals versus the predicted rents for Baltimore appears in Exhibit 27. The variance of the residuals appears to be constant along the horizontal axis.[1] The noticeable feature of the plot is the large number (9) of negative outliers. The values of the outliers and the data that produced these residuals are listed in Exhibits 28 and 29. We find all negative outliers correspond to low values of reported rent although their predicted rents are spread across the range of other predicted rents. In Exhibit 30 we compare the natural log of rent, for the sample observations to the subsample of outliers. One hundred percent of the outliers are in the lowest two percent of the reported rent distribution! The three lowest reported rents are outliers. This means the model overpredicts rent for a large proportion of all households reporting a low rent. It is possible, but seems unlikely, that the families associated with these outliers are reporting erroneous monthly rent data. These families are nevertheless reporting rents far below their market value as judged by the hedonic equation. These families may be related to their landlord or work in lieu of paying rent. It is not possible to test these hypotheses using the AHS since the relevant questions are asked only if the respondent is paying no cash rent. There is no way to determine whether low rents also reflect such extra considerations.

---

1. It needs to be added that such plots for most SMSA renter equations show no strong pattern for residuals to spread out or become more concentrated as predicted rents rise. Thus, the regression model's assumption of constant variance seems adequately satisfied. The same cannot be said for the owner model residuals which show a strong tendency towards increased clustering as predicted value rises (see Exhibit 33 and the discussion in Chapter II). The owner estimates are consequently less efficient than they could be.

Exhibit 31 also compares the sample frequency to the outlier frequency for three variables included in the regression. The three are 1940, CC1 (an indicator variable for a unit located in the central city), and SFATT (an indicator variable for a single family attached dwelling). The negative outlier frequency is considerably greater than the sample frequency for these characteristics. This outcome may be due to chance or may be caused by significant interaction effects among these variables. The Baltimore renter equation is reestimated with the three combinations of interaction terms included in the regression. Each of the estimated coefficients for the interaction terms is negative and two are statistically significant at the one percent level. Exhibit 32 compares the estimated coefficients in the original equation to the equation including interaction terms. An old unit in the central city offers the tenant a 13 percent reduction in rent. An old, single-family attached unit and a central city, single-family attached unit offer 9 percent and 7 percent discounts, respectively. The inclusion of the interaction terms lowers the value and the statistical significance of the estimated regression coefficients for single-family attached and central city units. The difference in interpreting the estimated coefficients for each of the two models is important. In Model A (the original specification of the model) central city units and single-family attached units offer discounts of 7 percent and 11 percent, respectively. In Model B central city units and single-family attached units are discounted only if these units are also old. The estimated coefficients for the remaining variables in the model and their standard errors do not change.

The final step in the residual analysis is to reestimate the regression coefficients while deleting the outliers observed in Model B. The same observations that produced outliers in Model A produce outliers in Model B indicating that the interaction terms failed to accomodate the original outliers. Model C in Exhibit 32 lists the results of deleting the outliers and reestimating the regression coefficients. The estimated coefficients of CC1DAGE (an indicator variable for an old, central city unit) and of DAGESFAT (an indicator variable for an old, single-family attached unit) remain statistically significant at the one percent level. The discount for an old, central city unit is reduced from 12.9 percent to 9.7 percent, however. This means the statistical significance of the interaction terms included in Model B is not produced by the outliers alone. Two additional changes occur in the estimated coefficients going from Model B to Model C. The estimated coefficient of the four or more bedrooms variable increases by 60 percent and remains statistically significant. The estimated regression coefficient of the indicator variable for black head of household swells by a factor of 5 but still does not become strongly significant. The standard error in the estimated equation goes from 0.2555 in Model A to 0.2300 in Model C, a reduction of 10 percent. Model C shows larger values for levels of statistical significance for most variables as well as higher R^2 and F-statistics because of the smaller residual variance. The confidence interval around predicted rents would be similarly reduced.

The residuals for the Baltimore owner model show a pattern similar to the residuals of most other owner equations. The plot of the residuals versus the predicted values in Exhibit 33 shows a tendency for the variance of the residuals to decline with increasing values of the predicted value of the dependent variable. The plot also shows that 16 of 20 outliers are negative (see also Exhibits 34 and 35). The negative outliers for this model occur in the low range of reported house values even though their predicted values are spread throughout much of the range of all predicted values. A striking 81 percent of the negative outliers appear in the lowest one percent of the distribution for reported house value (see Exhibit 36). All of the negative outliers occur in the lowest 2.1 percent of the distribution for reported house values. It seems likely that many of these homeowners are understating the market value of their homes.

The outliers in the Baltimore owner equation exhibit a greater percentage of old, central city, and single-family attached units as did the outliers in the Baltimore renter equation (see Exhibit 37). The same interaction terms used in the Baltimore renter equation produce statistically significant estimated coefficients in the owner model. The estimated coefficients for the interaction terms in the owner equation are at least 75 percent larger than they were in the Baltimore renter equation. Exhibit 38 lists the estimated coefficients for Models A, B, and C for Baltimore owners. Deleting the outliers from Model B leaves the estimated coefficients for the interaction terms statistically significant at the one percent level while raising the t-statistics for most coefficients. The estimated coefficients for single-family attached units and central city units decrease in value from Model A to Model C but remain statistically significant at the one percent level. The standard error of the equation is reduced 12 percent from Model A to 0.3124 in Model C which causes a 6.4 percent increase in the R^2 statistic, to 0.7339.

The procedure for analyzing the residuals in the Anaheim renter, Chicago renter, and Fort Worth owner models is the same as that used to analyze the residuals in the two Baltimore equations. The estimated coefficients for Models A, B, and C for Anaheim, Chicago, and Fort Worth are listed in Exhibits 39, 40, and 41, respectively.

We summarize the results of the analysis, beginning with Anaheim. Four out of seven of the negative outliers are in the lowest one percent of the reported rent distribution and all the negative outliers are in the lowest 6 percent of the rent distribution. All the postive outliers are in the upper 10 percent of the rent distribution while the four largest rents produce outliers. The inclusion of an indicator variable for a large dwelling interaction is suggested by examining the data on the residuals. The estimated coefficients for a dwelling having three or more bathrooms and four or more bedrooms is .192 and significant at the 10 percent level. The estimated coefficient of the large dwelling interaction variable shows no change after deleting the outliers. The estimated coefficient for single-family detached units (SFDET) is 0.0159 and statistically insignificant in Model B while it is 0.0403 and significant at the one percent level in Model C. The standard error in the Anaheim renter model is reduced by 11 percent after deleting outliers.

The outliers in Chicago are mostly negative with 20 of 21 negative outliers corresponding to the lowest 5 percent of the rent distribution.

An indicator variable for the presence of an elevator in a structure of more than fifty units produces an estimated coefficient of 0.11 and is statistically significant at the one percent level. The inclusion of the interaction terms reduces the magnitude and the statistical significance of the fifty or more units indicator variable. The Chicago standard error decreases 10 percent by including the interaction terms and deleting outliers.

The Fort Worth owner equation exhibits more positive outliers than negative ones. All positive outliers occur in the highest quartile of the reported house value distribution. The four smallest reported house values appear as the four negative outliers! An interaction term for old units with owners having a long length of tenure produces an insignificant estimated coefficient before and after deleting outliers. Deleting the outliers reduces the standard error in the Fort Worth owner model by eight percent to 0.3081.

In this section we have found the disproportionate number of negative outliers to be associated with very low reported rents and values. All negative outliers have reported rents or values in the bottom 6 percent of their respective distributions, and most of them have occurred in the lowest one percent. Examination of characteristics unique to the outliers has suggested interaction terms for inclusion in the models. Addition of these terms mostly fails to bring the outliers back into line, although the variables frequently do well even when the outliers have been deleted. While there appears to be room for improving the hedonic specification by including interaction terms, new information is needed to explain the outliers. Deletion of outliers reduces standard errors and increases sigificance levels substantially as well as altering specific coefficients. Some outliers should be expected even in a complete model, however, so that dropping all outliers as we have done probably overstates the reduction in residual variance that better information could achieve. It is clear though that a better understanding of the negative outliers could lead to important improvements in AHS-based hedonic models.

One implication of these findings is that predictions of rents and values from the equations reported in Chapter III will tend to be biased downward. Inclusion of observations with largely unexplained and very low rents pulls down the average predicted rent in the sample, and is likely to pull down predicted rents for most dwelling specifications. Exclusion of the outliers will not necessarily avoid the downward bias. That depends on the source of the bias. If it is something that effects only the outlying observations then this bias can be avoided by deleting the observations. However, if the source of bias effects other dwellings as well, deletion of the outliers would not eliminate the bias. In a following paper we will examine the effect of eliminating outliers on predictions of rents and values. In the remaining paragraphs of this paper we suggest ways to search for the source of the negative outliers and to measure their impacts on prediction.

The most likely source of the negative outliers in our opinion is that reported rents and values understate actual market prices. Renters could receive reduced rent because they work for the landlord or are related to him. In the Demand Experiment of the Experimental Housing Allowance Program, where these questions were asked, 5.2 percent of respondents in Phoenix worked in lieu of full rent and another 2.2 percent paid reduced rent because they were related to the landlord.[1] The Demand Experiment data could be analyzed to see if hedonic models like those estimated for this project produce a similar majority of negative outliers. If so, those outliers could be examined to determine how many receive subsidized rents for the above reasons. Estimation of the model excluding reduced rent for these reasons could be used to give an idea of the prediction error from this source. Earlier work by Ozanne, Andrews and Malpezzi (1979) found that models for evaluating AHS hedonics can be constructed from the Demand Experiment data and that these models give substantial discounts to tenants that are related to their landlord. Thus, this seems like a promising avenue of analysis.

Long-time homeowners may have widely varying ideas about the current value of their homes. Since values have been rising rapidly in many places, some of these homeowners could substantially under-report the value of their homes, even relative to the average for long time occupants. Perhaps characteristics like age of the survey respondent, when interacted with length of tenure, would characterize some of these outliers.

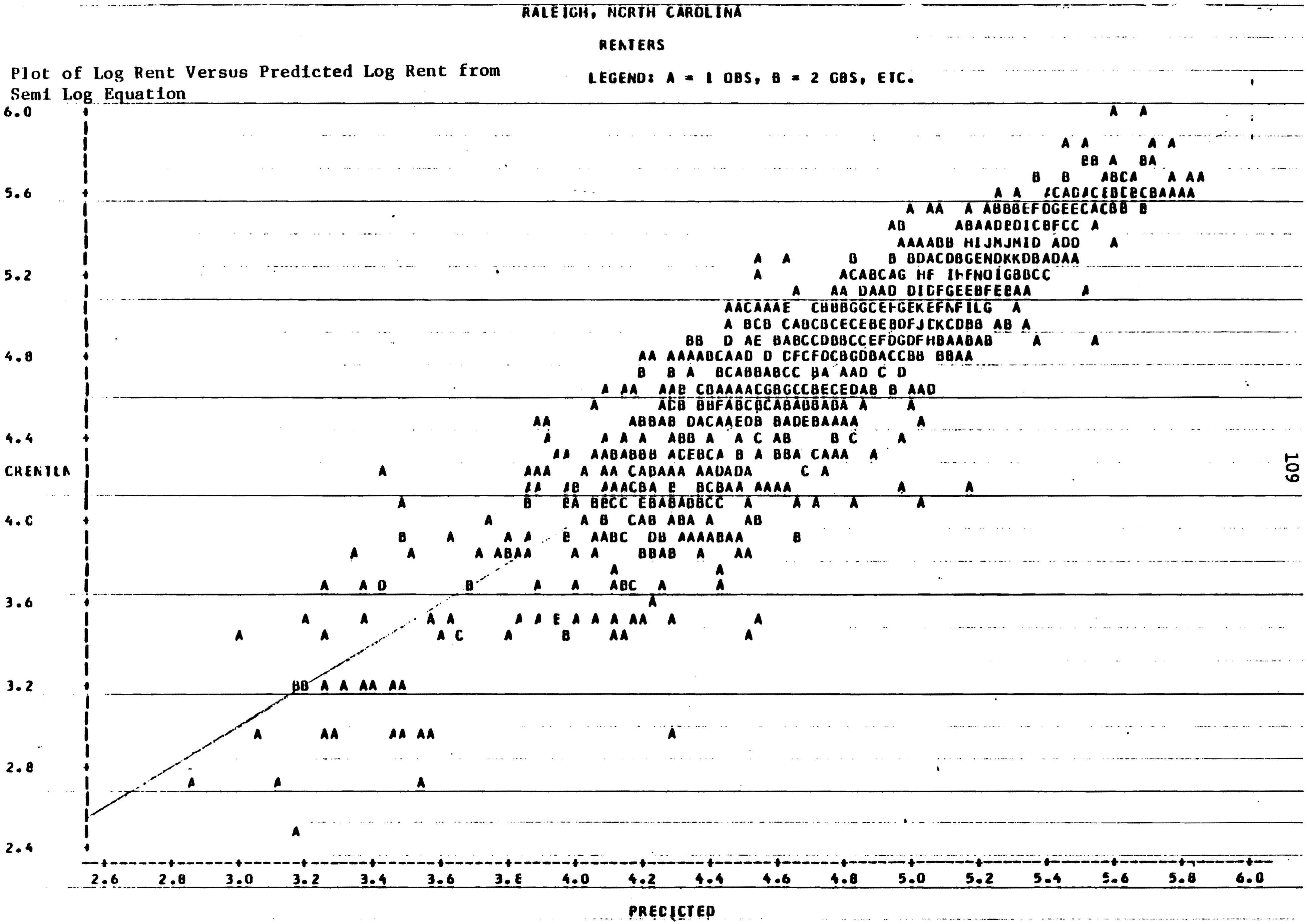
Other sources of under reporting would not be as easy to identify. Renters may receive subsidized housing but not report it, or know it. A few homeowners actually do get very low priced housing from urban homesteading programs.

---

1. Percentages supplied by James Zais from user tapes of Demand Experiment tenants survey. Analyses of market outcomes performed on the experimental data commonly exclude these non-market rents, e.g., Cronin (1979).

Under reporting is not the only potential cause of the negative outliers. Omitted physical, neighborhood or locational information could also be involved. A simple first step in investigating this possibility would be to examine variables omitted from the hedonic but included in the AHS. Future metropolitan AHS user tapes will identify dwellings located in the same sampling cluster. This information could be used to test whether neighborhood location is associated with the low rents and values. In this section we have suggested hypotheses that could account for the observed outliers and ways to test the hypotheses. It remains for future work to investigate them.

Exhibit 1
Comparison of Semilog and Linear Hedonic Regressions



NOTE: 22 CBS HAC MISSING VALUES

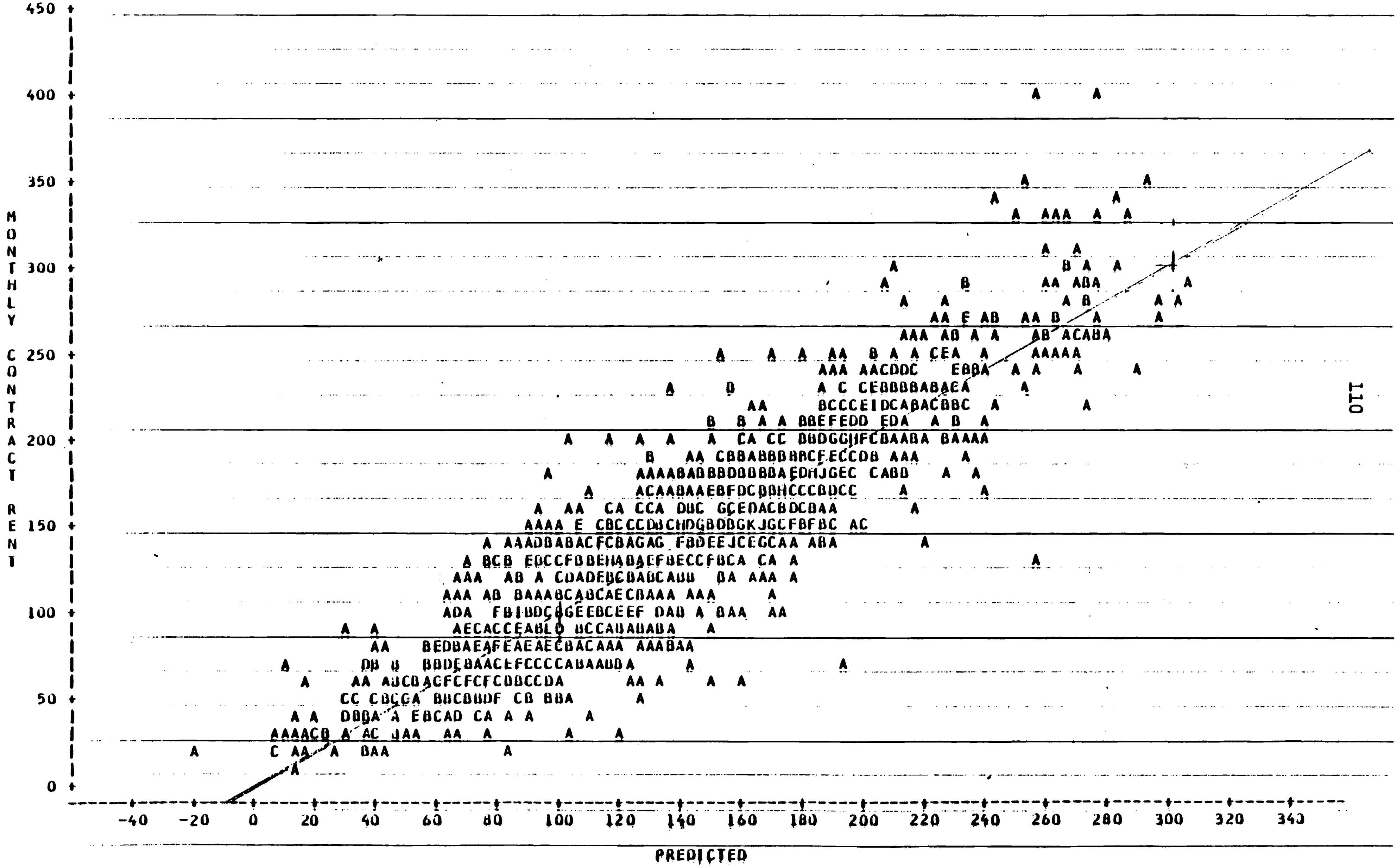
Exhibit 1 (cont'd)
Comparison of Semilog and Linear Hedonic Regressions

RALEIGH, NORTH CAROLINA

RENTERS

Plot of Rent Versus Predicted Rent from Linear Equation

LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 22 OBS HAD MISSING VALUES

Exhibit 1 (cont'd)

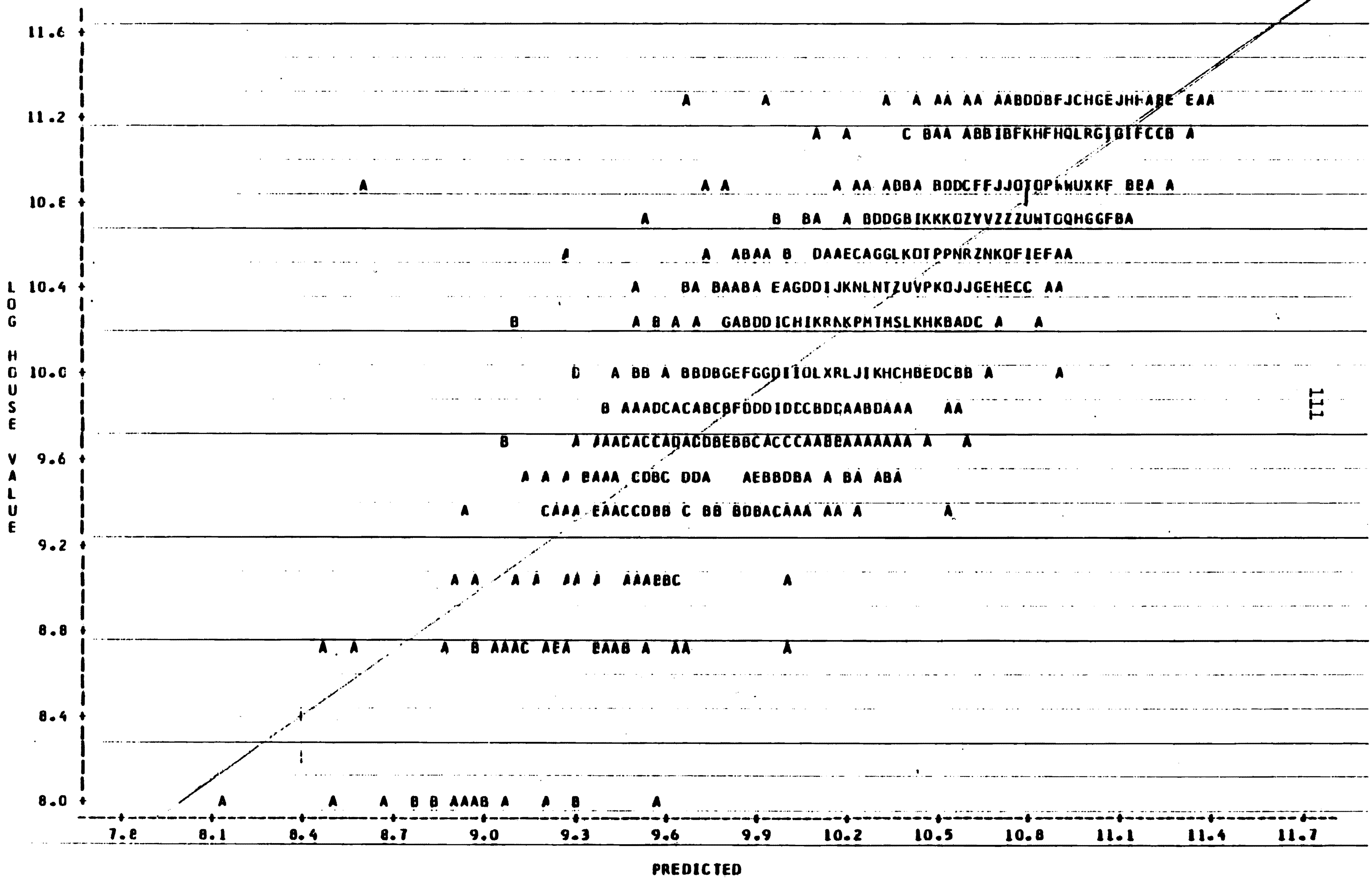
Comparison of Semilog and Linear Hedonic Regressions

RALEIGH, NORTH CAROLINA

OWNERS

Plot of Log Value Versus Predicted Log Value from Semilog Equation

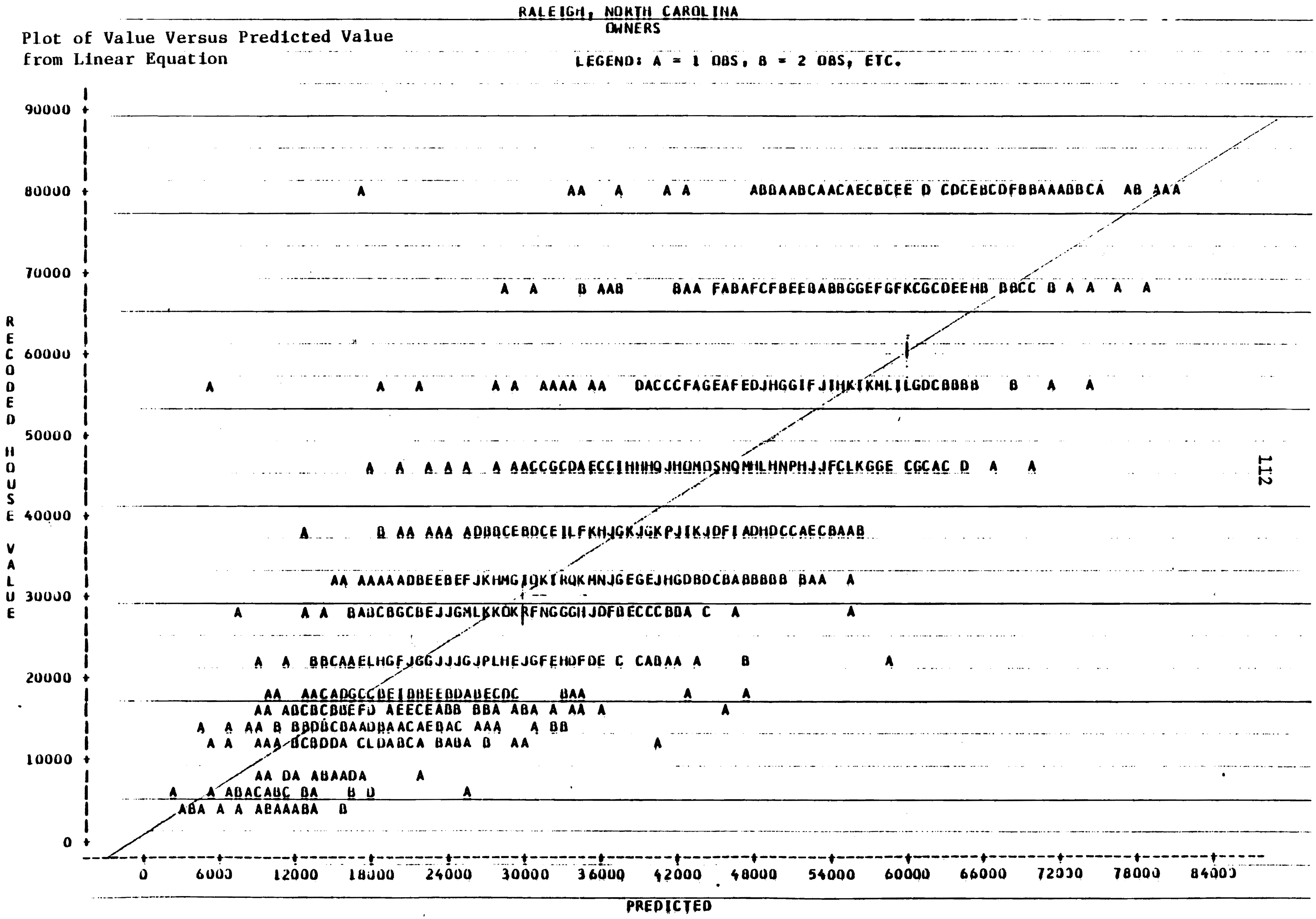
LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 55 OBS HAD MISSING VALUES 19 OBS HIDDEN

Exhibit 1 (cont'd)
Comparison of Semilog and Linear Hedonic Regressions

Plot of Value Versus Predicted Value
from Linear Equation



NOTE: 55 OBS HAD MISSING VALUES

Exhibit 2

Hedonic Variable Definitions

Variable	Tenure*	Definition
I. STRUCTURAL VARIABLES		
Bathrooms		
B1	Both	One and one-half baths = 1, else = 0.
B2	Both	Two baths = 1, else = 0.
B3	Both	More than two baths = 1, else = 0.
Omitted Category**	Both	One bath.
Rooms		
R1	Renters	One room (other than bath and bedrooms) = 1, else = 0.
R3	Renters	Three rooms = 1, else = 0.
RG4	Renters	Number of rooms when number greater than or equal to 4, else = 0.
Omitted Category	Renters	Two rooms.
R12	Owners	Two rooms = 1, one room = 2, else = 0.
R4	Owners	Four rooms = 1, else = 0.
R5	Owners	Five rooms = 1, else = 0.
RG6	Owners	Number of rooms when number greater than or equal to 6, else = 0.
Omitted Category	Owners	Three rooms.
Bedrooms		
BED0	Renters	Zero bedrooms = 1, else = 0.
BED2	Both	Two bedrooms = 1, else = 0.
BED3	Renters	Three bedrooms = 1, else = 0.

Exhibit 2 (cont'd)

Hedonic Variable Definitions

Variable	Tenure*	Definition
BEDG4	Renters	Number of bedrooms when number greater than or equal to 4, else = 0.
Omitted Category	Renters	One bedroom.
BED1	Owners	One bedroom = 1, else = 0.
BED4	Owners	Four bedrooms = 1, else = 0.
BEDG5	Owners	Number of bedrooms when number greater than or equal to 5, else = 0.
Omitted Category	Owners	Three bedrooms.
<u>Structure Type</u>		
SFATT	Both	Attached single-family dwelling = 1, else = 0.
Omitted Category	Owners	Detached single-family dwelling.
SFDET	Renters	Detached single family dwelling = 1, else = 0.
DUPLEX	Renters	Duplex = 1, else = 0.
NGT50	Renters	Structure with 50 or more units = 1, else = 0.
Omitted Category	Renters	Structure with 3 to 49 units.
<u>Dwelling Equipment</u>		
RHEAT	Both	Wall or room heater with flue = 1, else = 0.
SHEAT	Owners	Steam or hot water heat = 1, else = 0.
EHEAT	Owners	Electric heat = 1, other fuels = 0.
ROOMAC	Both	Room airconditioning present = 1, else = 0.

Exhibit 2 (cont'd)

Hedonic Variable Definitions

Variable	Tenure*	Definition
DFECT	Renters	Linear combination of: (1) basement leaks = 1, else = 0, plus (2) roof leaks = 1, else = 0, plus (3) cracks in walls or ceiling = 1, else = 0, (4) holes in floor = 1, else = 0, plus (5) broken plaster or peeling paint = 1, else = 0, plus (6) signs of rats or mice = 1, else = 0.
COOKE	Owners	Cook with electricity = 1, else = 0.
II. NEIGHBORHOOD VARIABLES		
General Neighborhood Rating		
EXCELN	Both	Neighborhood rated excellent = 1, else = 0.
GOODN	Both	Neighborhood rated good = 1, else = 0.
POORN	Both	Neighborhood rated poor = 1, else = 0.
Omitted Category	Both	Neighborhood rated fair.
Neighborhood Location		
These variables vary by SMSA. See the separate data appendix for their definitions.		
Other Neighborhood Variables		
ABANDON	Both	Abandoned housing on street = 1, else = 0. (Enumerator, not respondent, response).
LITTER	Renters	Trash or litter on street = 1, else = 0.
NOSHOPS	Renters	No convenient shopping = 1, else = 0.
BLACK	Both	Black respondent = 1, else = 0.
SPAN	Both	Spanish respondent = 1, else = 0.

Exhibit 2 (cont'd)

Hedonic Variable Definitions

Variable	Tenure*	Definition
CENTAC	Both	Central airconditioning present = 1, else = 0.
ELEVP	Renters	Elevator present = 1, else = 0.
<u>Age of Structure</u>		
AGE1	Both	Age of structure.
AGE1SQ	Both	Age of structure, squared.
AGE1CB	Owners	Age of structure, cubed.
DAGE	Both	Structure built prior to 1940 = 1, else = 0.
<u>Quality Variables</u>		
NORAD	Both	Rooms without heat = 1, else = 0.
POOR	Both	Linear combination of 5 quality variables: (1) water absent = 1, else = 0, plus (2) no sewer or septic = 1, else = 0, plus (3) no bathroom = 1, else = 0, plus (4) bathroom shared with another unit = 1, else = 0, plus (5) no heating equipment, or primitive heating equipment = 1, else = 0.
NOPRIVCY	Both	Pass through bedrooms to bedroom or only bathroom = 1, else = 0.
NOUT	Both	Rooms without electric outlet = 1, else = 0.
BADHALL	Renters	Linear combination of: (1) broken hall lights = 1, else = 0, plus (2) broken steps = 1, else = 0, plus (3) broken railing = 1, else = 0.

Exhibit 2 (cont'd)

Hedonic Variable Definitions

Variable	Tenure*	Definition
III. CONTRACT CONDITIONS		
CROWDS	Both	Persons per room.
CLOT	Both	Length of tenure.
CLOTSQ	Both	Length of tenure, squared.
DLOT	Both	Respondent moved in prior to 1950 = 1, else = 0.
HEATING	Renters	Heat included in contract rent = 1, else = 0.
NHUINC	Renters	Non-heat utility included in rent = 1, else = 0.
PARKINC	Renters	Parking included in rent = 1, else = 0.
FURNINC	Renters	Furniture included in rent = 1, else = 0.
LLBLG	Renters	Landlord lives in building = 1, else = 0.
IV. INFLATION MEASURES		
Q	Both	Time trend constructed from date of interview. First month of survey = 0, second month = 1, ... twelfth month = 11.
QHEAT	Renters	Interaction between Q and HEATINC. Takes on the value of Q if heat included in rent, 0 otherwise.
FORAY	Owners	Interaction between Q and central city dummy. Takes on the value of Q if unit is in central city, 0 otherwise.

Exhibit 2 (cont'd)

Hedonic Variable Definitions

Variable	Tenure*	Definition
V. DEPENDENT VARIABLES		
CRENTLN	Renters	Natural logarithm of monthly contract rent.
VALUELN	Owners	Log value of house, recoded as interval midpoints

* Indicates in which regressions variable appears.

** If dummy variables are mutually exclusive, this defines the omitted category for which no explicit variable appears.

Exhibit 3

MEANS OF RENTER HEDONIC VARIABLES

VARIABLE	LABEL	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE
INTERCEP	VARIABLE	59	1771.068	1259.695	778.000	5532.000
B1	BATH DUMMY 1.5	59	0.063	0.032	0.020	0.180
B2	BATH DUMMY 2	59	0.075	0.049	0.012	0.237
B3	BATH DUMMY GT 2	59	0.011	0.008	0.001	0.037
R1	ONE ROOM	59	0.099	0.034	0.029	0.197
R3	THREE ROOMS	59	0.247	0.077	0.096	0.400
RG4	ROOMS WHEN GE 4	59	0.218	0.084	0.049	0.422
BEDO	NO BEDROOMS	59	0.038	0.023	0.006	0.098
BED2	TWO BEDROOMS	59	0.432	0.052	0.320	0.556
BED3	THREE BEDS	59	0.154	0.040	0.068	0.280
BEDG4	NUMBER BEDS WHEN GE 4	59	0.122	0.054	0.038	0.285
ELEVP	ELEVATOR PRESENT	59	0.059	0.067	0.004	0.339
SFATT	SINGLE FAMILY ATTACHED	59	0.070	0.062	0.012	0.323
SFDET	SINGLE FAMILY DETACHED	59	0.209	0.118	0.038	0.514
DUPLEX	TWO UNITS	59	0.148	0.093	0.038	0.460
NGT50	MORE THAN 50 UNITS	59	0.069	0.056	0.013	0.290
AGE1	AGE OF STRUCTURE	59	29.759	8.172	14.861	44.867
AGE1SQ	SQUARE AGE	59	1356.681	534.930	368.526	2394.576
DAGE	DUMMY OLD STRUCTURE	59	0.392	0.193	0.041	0.784
RHEAT	WALL OR ROOM HEAT W FLUE	59	0.168	0.171	0.001	0.685
ROOMAC	ROOM AIR CONDITIONER	59	0.298	0.136	0.011	0.620
CENTAC	CENTRAL AIR CONDITIONER	59	0.224	0.191	0.003	0.696
NORAD	ROOMS WITHOUT HEAT	58	0.421	0.296	0.063	0.948
POOR	WABSNT+SABSNT+NCBATH+SHARED+NHEAT	59	0.105	0.165	0.011	1.009
NOPRIVCY	PASS THRU BR TO ROOM AND OR BATH	59	0.126	0.070	0.039	0.456
NOUT	NO OUTLETS	59	0.031	0.016	0.007	0.097
BADHALL	BAD HALL LIGHTING	59	0.091	0.047	0.026	0.287
DFECT	BLEAK+RLEAK+CRACKS+HOLES+PLASTER+RATS	59	0.487	0.144	0.197	0.792
CLOT	LENGTH OF TENURE CORRECTED FOR DOI	59	3.596	1.290	1.659	7.394
CLOTSQ	SQUARE LENGTH OF TENURE	59	42.445	24.600	13.125	126.207
DLOT	DUMMY FOR OLD TENANTS	59	0.014	0.013	0.001	0.057
CROWDS	PERSONS PER ROOM	59	0.588	0.046	0.513	0.758
BLACK	BLACK DUMMY	59	0.144	0.109	0.009	0.431
SPAN	SPANISH DUMMY	59	0.052	0.072	0.002	0.379
LLBLG	LANDLORD LIVGS IN BUILDING	59	0.114	0.081	0.021	0.331
NHUINC	NON HEAT UTILITY INCLUDED	59	0.043	0.032	0.009	0.197
HEATINC	HEAT INCLUDED IN RENT	59	0.355	0.188	0.002	0.773
PARKINC	PARKING INCLUDED IN RENT	55	0.027	0.029	0.001	0.117
FURNINC	FURNITURE INCLUDED IN RENT	59	0.148	0.103	0.020	0.481
EXCELN	EXCELLENT NEIGHBORHOOD RATING	59	0.253	0.034	0.159	0.321
GOODN	GOOD NEIGHBORHOOD RATING	59	0.481	0.025	0.422	0.536
POORN	POOR NEIGHBORHOOD RATING	59	0.044	0.014	0.016	0.086
ABANDON	ABANDONED HOUSING, ENUMERATOR	59	0.091	0.041	0.018	0.200
LITTER	LITTER IN NEIGHBORHOOD	59	0.172	0.039	0.104	0.274
NOSHOPS	NO CONVENIENT SHOPPING	59	0.102	0.027	0.062	0.175
Q	TIME TREND	59	5.525	0.135	5.180	5.837
QHEAT	TIME TREND W HEATINC	59	1.973	1.056	0.004	4.421
CC1	PRIMARY CENTRAL CITY	40	0.524	0.129	0.266	0.792
CRENTLN	LOG CONTRACT RENT	59	4.917	0.182	4.528	5.342

Exhibit 3 (cont'd)

MEANS OF OWNER HEDONIC VARIABLES

VARIABLE	LABEL	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE
INTERCEP	VARIABLE	59	2840.678	1502.921	1448.000	7422.000
B1	BATH DUMMY 1.5	59	0.190	0.088	0.049	0.401
B2	BATH DUMMY 2	59	0.258	0.140	0.089	0.630
B3	BATH DUMMY GT 2	59	0.126	0.055	0.055	0.337
R12	ONE OR TWO ROOMS	59	0.286	0.094	0.110	0.492
R4	FOUR ROOMS	59	0.224	0.055	0.116	0.327
R5	FIVE ROOMS	59	0.068	0.027	0.024	0.145
RG6	NUMBER ROOMS GE 6	59	0.190	0.109	0.021	0.535
BED1	ONE BEDROOM	59	0.018	0.009	0.005	0.050
BED2	TWO BEDROOMS	59	0.210	0.056	0.075	0.332
BED4	FOUR BEDROOMS	59	0.194	0.046	0.105	0.325
BEDG5	NUMBER BEDS GE 5	59	0.198	0.104	0.065	0.484
GAR	GARAGE	59	0.793	0.126	0.348	0.972
BASE	BASEMENT	59	0.548	0.381	0.005	0.972
SFATT	SINGLE FAMILY ATTACHED	59	0.037	0.087	0.002	0.490
AGE1	AGE OF STRUCTURE	59	25.281	5.592	12.860	35.268
AGE1SQ	SQUARE AGE	59	976.860	342.308	262.330	1624.310
AGE1CB	CUBE AGE	59	45697.240	19170.187	7317.985	83550.513
DAGE	DUMMY OLD STRUCTURE	59	0.237	0.115	0.016	0.483
SHEAT	STEAM OR HOT WATER	57	0.159	0.229	0.000	0.791
RHEAT	WALL OR ROOM HEAT W FLUE	59	0.101	0.124	0.001	0.522
EHEAT	ELECTRIC UNITS	59	0.033	0.051	0.001	0.224
ROOMAC	ROOM AIR CONDITIONER	59	0.311	0.148	0.016	0.603
CENTAC	CENTRAL AIR CONDITIONER	59	0.272	0.217	0.007	0.889
NORAD	ROOMS WITHOUT HEAT	59	0.400	0.338	0.050	0.999
POOR	WABSNT+SABSNT+NOBATH+SHARED+NHEAT	59	0.055	0.145	0.000	0.998
NOPRIVCY	PASS THRU BR TO ROOM AND OR BATH	59	0.039	0.024	0.005	0.103
NOUT	NO OUTLETS	59	0.016	0.008	0.001	0.038
COOKE	COOK WITH ELECTRICITY	59	0.569	0.207	0.194	0.980
CLOT	LENGTH OF TENURE CORRECTED FOR DOI	59	11.329	1.788	7.111	14.947
CLOTSQ	SQUARE LENGTH OF TENURE	59	214.496	51.224	96.916	319.444
DLOT	DUMMY FOR OLD TENANTS	59	0.097	0.039	0.013	0.207
CROWDS	PERSONS PER ROOM	59	0.551	0.034	0.487	0.660
BLACK	BLACK DUMMY	59	0.077	0.069	0.001	0.280
SPAN	SPANISH DUMMY	59	0.030	0.055	0.001	0.332
EXCELN	EXCELLENT NEIGHBORHOOD RATING	59	0.471	0.054	0.350	0.607
GOODN	GOOD NEIGHBORHOOD RATING	59	0.413	0.028	0.332	0.476
POORN	POOR NEIGHBORHOOD RATING	59	0.015	0.006	0.004	0.031
ABANDON	ABANDONED HOUSING, ENUMERATOR	59	0.045	0.023	0.010	0.115
Q	TIME TREND	59	5.516	0.097	5.171	5.687
CC1	PRIMARY CENTRAL CITY	40	0.313	0.134	0.024	0.633
FORAY	CENTRAL CITY INFLATION DIFFERENTIAL	42	1.712	0.742	0.120	3.408
VALUELN	LOG HOUSE VALUE	59	10.355	0.255	9.872	11.256

Exhibit 4
Sample Regressions

WASHINGTON, D.C. AND ENVIRONS

RENTERS

DESCRIPTIVE STATISTICS

	SUM	MEAN	UNCORRECTED SS	VARIANCE	STD DEVIATION
INTERCEPT	5532.00000000	1.00000000	5532.00000000	0.00000000	0.00000000
CRENTLN	28301.40827049	5.11594510	145705.16703223	0.16574149	0.40711361
B1	306.00000000	0.05531453	306.00000000	0.05226428	0.22861383
B2	356.00000000	0.06435286	356.00000000	0.06022245	0.24540263
B3	167.00000000	0.03018800	167.00000000	0.02928198	0.17111977
R1	567.00000000	0.10249458	567.00000000	0.09200607	0.30332502
R3	1494.00000000	0.27006508	1494.00000000	0.19716557	0.44403330
RG4	1613.00000000	0.29157628	7447.00000000	1.26137904	1.12311132
BED0	543.00000000	0.09815618	543.00000000	0.08853755	0.29755260
BED2	1771.00000000	0.32013738	1771.00000000	0.21768879	0.46657131
BED3	611.00000000	0.11044830	611.00000000	0.09826724	0.31347606
BEDG4	933.00000000	0.16865510	4131.00000000	0.71843153	0.84760340
ELEVP	1456.00000000	0.26319595	1456.00000000	0.19395890	0.44040756
SFATT	480.00000000	0.08676790	480.00000000	0.07925355	0.28152008
SFDET	398.00000000	0.07194505	398.00000000	0.06678103	0.25842026
DUPLEX	209.00000000	0.03778019	209.00000000	0.03635942	0.19068146
NGT50	1390.00000000	0.25126537	1390.00000000	0.18816510	0.43378001
AGE1	156960.00000000	28.37310195	6429980.00000000	357.35634019	18.90387104
AGE1SQ	6429980.00000000	1162.32465654	15593496368.00000000	1468047.74415419	1211.63020107
DAGE	1719.00000000	0.31073753	1719.00000000	0.21421844	0.46283738
RHEAT	73.00000000	0.01319595	73.00000000	0.01302417	0.11412349
ROOMAC	1465.00000000	0.26482285	1465.00000000	0.19472691	0.44127372
CENTAC	2447.00000000	0.44233550	2447.00000000	0.24671940	0.49670857
NORAD	5243.00000000	0.94775850	5243.00000000	0.04952128	0.22253378
POOR	71.00000000	0.01283442	93.00000000	0.01664957	0.12903320

Exhibit 4 (cont'd)
Sample Regressions

WASHINGTON, D.C. AND ENVIRONS

RENTERS

NOPRIVCY	286.00000000	0.05169920	286.00000000	0.04903526	0.22143907
NOUT	105.00000000	0.01898048	105.00000000	0.01862359	0.13646826
BADHALL	684.00000000	0.12364425	954.00000000	0.15719171	0.39647409
DFECT	3385.26800000	0.61194288	7650.16728400	1.00860182	1.00429170
CLOT	25265.75000000	4.56719993	279964.98263887	29.75433937	5.45475383
CLOTSQ	279964.98263887	50.60827597	102079146.22866020	15894.16039174	126.07204445
DLOT	74.00000000	0.01337672	74.00000000	0.01320017	0.11489198
CROWDS	3307.16000000	0.59782357	2567.21260000	0.10669217	0.32663767
BLACK	2383.00000000	0.43076645	2383.00000000	0.24525105	0.49522828
SPAN	125.00000000	0.02259581	125.00000000	0.02208923	0.14862446
LLBLG	136.00000000	0.02458424	136.00000000	0.02398419	0.15486829
NHUINC	500.00000000	0.09038322	500.00000000	0.08222896	0.28675593
HEATINC	3766.00000000	0.68076645	3766.00000000	0.21736278	0.46622182
PARKINC	96.00000000	0.01735358	96.00000000	0.01705552	0.13059577
FURNINC	204.00000000	0.03687636	204.00000000	0.03552291	0.18847523
EXCELN	1411.00000000	0.25506146	1411.00000000	0.19003946	0.43593516
GOODN	2514.00000000	0.45444685	2514.00000000	0.24796974	0.49796560
POORN	261.00000000	0.04718004	261.00000000	0.04496221	0.21204295
ABANDON	510.00000000	0.09219089	510.00000000	0.08370686	0.28932138
LITTER	1082.00000000	0.19558930	1082.00000000	0.15736257	0.39668951
NOSHOPS	968.00000000	0.17498192	968.00000000	0.14438935	0.37998599
Q	31689.00000000	5.72830803	246161.00000000	11.68624968	3.41851571
QHEAT	22140.00000000	4.00216920	172418.00000000	15.15277056	3.89265598
CCI	3202.00000000	0.57881417	3202.00000000	0.24383240	0.49379389
MONTGOM	551.00000000	0.09960231	551.00000000	0.08969791	0.29949509
PRINCEG	717.00000000	0.12960954	717.00000000	0.11283131	0.33590372
ALEXARL	582.00000000	0.10520607	582.00000000	0.09415478	0.30684550

Exhibit 4 (cont'd)
Sample Regressions

WASHINGTON, D.C. AND ENVIRONS

RENTERS

MODEL: A	SSE	261.050658	F RATIO	275.33	
	DFE	5481	PROB>F	0.0001	
DEP VAR: CRENTLN	MSE	0.047628	R-SQUARE	0.7152	

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	4.857290	0.028341	171.3855	0.0001	
B1	1	0.115105	0.014821	7.7663	0.0001	BATH DUMMY 1.5
B2	1	0.199386	0.014091	14.1496	0.0001	BATH DUMMY 2
B3	1	0.402985	0.022054	18.2724	0.0001	BATH DUMMY GT 2
R1	1	-0.081673	0.011712	-6.9736	0.0001	ONE ROOM
R3	1	0.039683	0.007350619	5.3985	0.0001	THREE ROOMS
RG4	1	0.023755	0.003241238	7.3288	0.0001	ROOMS WHEN GE 4
BED0	1	-0.199505	0.012214	-16.3339	0.0001	NO BEDROOMS
BED2	1	0.170256	0.007632282	22.3073	0.0001	TWO BEDROOMS
BED3	1	0.257577	0.013974	18.4323	0.0001	THREE BEDS
BEDG4	1	0.082323	0.005034271	16.3526	0.0001	NUMBER BEDS WHEN GE 4
ELEVP	1	0.130865	0.016456	7.9525	0.0001	ELEVATOR PRESENT
SFATT	1	0.027845	0.013795	2.0185	0.0436	SINGLE FAMILY ATTACHED
SFDET	1	0.054033	0.017126	3.1549	0.0016	SINGLE FAMILY DETACHED
DUPLEX	1	-0.0086424	0.016440	-0.5257	0.5991	TWO UNITS
MGT50	1	0.058240	0.016764	3.4742	0.0005	MORE THAN 50 UNITS
AGE1	1	-0.00333843	0.001849717	-1.8048	0.0712	AGE OF STRUCTURE
AGE1SQ	1	-.0000183895	.00005019064	-0.3664	0.7141	SQUARE AGE
DAGE	1	0.088803	0.063980	1.3880	0.1652	DUMMY OLD STRUCTURE
RHEAT	1	-0.125465	0.026716	-4.6963	0.0001	WALL OR ROOM HEAT W FLUE
ROOMAC	1	0.096907	0.008341561	11.6173	0.0001	ROOM AIR CONDITIONER
CENTAC	1	0.218964	0.011270	19.4283	0.0001	CENTRAL AIR CONDITIONER
NORAD	1	0.020528	0.013733	1.4948	0.1350	ROOMS WITHOUT HEAT
POOR	1	-0.257494	0.023653	-10.8864	0.0001	WABSNT+SABSNT+NOBATH+SHARED+NHEAT
NOPRIVCY	1	-0.071630	0.013986	-5.1217	0.0001	PASS THRU BR TO ROOM AND OR BATH
NOUT	1	-0.058595	0.022071	-2.6549	0.0080	NO OUTLETS
BADHALL	1	-.0000280786	0.007963395	-0.0035	0.9972	BAD HALL LIGHTING
DFECT	1	-0.00529974	0.003352234	-1.5810	0.1139	BLEAK+RLEAK+CRACKS+HOLES+PLASTER+RATS
CLOT	1	-0.025631	0.001777362	-14.4206	0.0001	LENGTH OF TENURE CORRECTED FOR DOI
CLOTSQ	1	0.0007050384	0.0001006498	7.0049	0.0001	SQUARE LENGTH OF TENURE
DLOT	1	-0.139200	0.053225	-2.6153	0.0089	DUMMY FOR OLD TENANTS
CROWDS	1	-0.016738	0.010650	-1.5717	0.1161	PERSONS PER ROOM
BLACK	1	-0.181521	0.008047473	-22.5563	0.0001	BLACK DUMMY
SPAN	1	-0.024766	0.020114	-1.2313	0.2183	SPANISH DUMMY
LLBLG	1	0.054160	0.019414	2.7897	0.0053	LANDLORD LIVGS IN BUILDING
NHUINC	1	0.095523	0.013251	7.2085	0.0001	NON HEAT UTILITY INCLUDED
HEATINC	1	0.103134	0.014204	7.2609	0.0001	HEAT INCLUDED IN RENT
PARKINC	1	0.129680	0.022885	5.6666	0.0001	PARKING INCLUDED IN RENT
FURNINC	1	0.147337	0.016573	8.8903	0.0001	FURNITURE INCLUDED IN RENT
EXCELN	1	0.073148	0.009434108	7.7536	0.0001	EXCELLENT NEIGHBORHOOD RATING
GOODN	1	0.038182	0.007884099	4.8429	0.0001	GOOD NEIGHBORHOOD RATING
POORN	1	0.006458829	0.015200	0.4249	0.6709	POOR NEIGHBORHOOD RATING
ABANDON	1	-0.074153	0.011182	-6.6315	0.0001	ABANDONED HOUSING, ENUMERATOR
LITTER	1	-0.00307057	0.008296565	-0.3701	0.7113	LITTER IN NEIGHBORHOOD

Exhibit 4 (cont'd)
Sample Regressions

WASHINGTON, D.C. AND ENVIRONS

RENTERS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
NOSHOPS	1	-0.020892	0.008262486	-2.5285	0.0115	NO CONVENIENT SHOPPING
Q	1	0.004073701	0.001497718	2.7199	0.0065	TIME TREND
QHEAT	1	0.001623305	0.001839641	0.8824	0.3776	TIME TREND W HEATINC
CC1	1	0.114958	0.013687	8.3990	0.0001	PRIMARY CENTRAL CITY
MONTGOM	1	0.029757	0.013979	2.1286	0.0333	
PRINCEG	1	-0.00101378	0.013356	-0.0759	0.9395	
ALEXARL	1	0.047308	0.014360	3.2944	0.0010	

TEST:TEST001 (Age)	NUMERATOR:	1.52196761	DF: 3	F VALUE:	31.9551
	DENOMINATOR:	0.04762829	DF: 5481	PROB > F:	0.0001
TEST:TEST002 (Length of Tenure)	NUMERATOR:	8.28861423	DF: 3	F VALUE:	174.0271
	DENOMINATOR:	0.04762829	DF: 5481	PROB > F:	0.0001
TEST:TEST003 (Neighborhood Rating)	NUMERATOR:	0.96866163	DF: 3	F VALUE:	20.3379
	DENOMINATOR:	0.04762829	DF: 5481	PROB > F:	0.0001
TEST:TEST004 (Location)	NUMERATOR:	1.52527897	DF: 4	F VALUE:	32.0246
	DENOMINATOR:	0.04762829	DF: 5481	PROB > F:	0.0001

Exhibit 4 (cont'd)
Sample Regressions

WASHINGTON, D.C. AND ENVIRONS

OWNERS

DESCRIPTIVE STATISTICS

	SUM	MEAN	UNCORRECTED SS	VARIANCE	STD DEVIATION
INTERCEPT	4613.00000000	1.00000000	4613.00000000	0.00000000	0.00000000
VALUELN	49427.02944600	10.71472565	530532.14328057	0.20274996	0.45027765
B1	1059.00000000	0.22956861	1059.00000000	0.17690521	0.42060101
B2	889.00000000	0.19271624	889.00000000	0.15561042	0.39447487
B3	1556.00000000	0.33730761	1556.00000000	0.22357965	0.47284210
R12	517.00000000	0.11207457	535.00000000	0.10343830	0.32161825
R4	1424.00000000	0.30869282	1424.00000000	0.21344784	0.46200415
R5	668.00000000	0.14480815	668.00000000	0.12386560	0.35194545
RG6	1766.00000000	0.38283113	11438.00000000	2.33346059	1.52755588
BED1	22.00000000	0.00476913	22.00000000	0.00474742	0.06890149
BED2	572.00000000	0.12399740	572.00000000	0.10864560	0.32961431
BED4	1291.00000000	0.27986126	1291.00000000	0.20158263	0.44897955
BEDG5	1939.00000000	0.42033384	10153.00000000	2.02471220	1.42292382
GAR	2052.00000000	0.44482983	2052.00000000	0.24700980	0.49700380
BASE	3729.00000000	0.80836766	3729.00000000	0.15494298	0.39362797
SFATT	1145.00000000	0.24821158	1145.00000000	0.18664305	0.43202205
AGE1	123522.00000000	26.77693475	4906744.00000000	346.74836640	18.62118059
AGE1SQ	4906744.00000000	1063.67743334	11644248136.00000000	1393116.92801839	1180.30374397
AGE1CB	230278086.00000000	49919.37697811	32373402114664.00000000	4526900158.24662300	67282.24251797
DAGE	1283.00000000	0.27812703	1283.00000000	0.20081592	0.44812489
SHEAT	1485.00000000	0.32191632	1485.00000000	0.21833353	0.46726174
RHEAT	48.00000000	0.01040538	48.00000000	0.01029934	0.10148565
EHEAT	69.00000000	0.01495773	69.00000000	0.01473719	0.12139583
ROOMAC	1664.00000000	0.36071971	1664.00000000	0.23065100	0.48026139
CENTAC	2202.00000000	0.47734663	2202.00000000	0.24954092	0.49954071

Exhibit 4 (cont'd)
Sample Regressions
WASHINGTON, D.C. AND ENVIRONS

OWNERS

NORAD	4224.00000000	0.91567310	4224.00000000	0.07723262	0.27790757
POOR	36.00000000	0.00780403	76.00000000	0.01641783	0.12813210
NOPRIVCY	56.00000000	0.01213961	56.00000000	0.01199484	0.10952094
NOUT	70.00000000	0.01517451	70.00000000	0.01494748	0.12225989
COOKE	1981.00000000	0.42943854	1981.00000000	0.24507421	0.49504970
CLOT	52471.50000000	11.37470193	968533.52083331	80.59103394	8.97725091
CLOTSQ	968533.52083331	209.95740751	539011970.02841040	72779.96160232	269.77761509
DLOT	473.00000000	0.10253631	473.00000000	0.09204257	0.30338518
CROWDS	2336.40000000	0.50648168	1464.30640000	0.06091990	0.24681958
BLACK	1264.00000000	0.27400824	1264.00000000	0.19897086	0.44606149
SPAN	44.00000000	0.00953826	44.00000000	0.00944933	0.09720767
EXCELN	2342.00000000	0.50769564	2342.00000000	0.24999497	0.49999497
GOODN	1783.00000000	0.38651637	1783.00000000	0.23717288	0.48700398
POORN	57.00000000	0.01235638	57.00000000	0.01220635	0.11048235
ABANDON	168.00000000	0.03641882	168.00000000	0.03510010	0.18735019
Q	25577.00000000	5.54454802	197021.00000000	11.97053239	3.45984572
CC1	1507.00000000	0.32668545	1507.00000000	0.22000976	0.46905198
FORAY	8415.00000000	1.82419250	64827.00000000	10.72775805	3.27532564
MONTGOM	944.00000000	0.20463906	944.00000000	0.16279721	0.40348136
PRINCEG	773.00000000	0.16756991	773.00000000	0.13952048	0.37352440
ALEXARL	267.00000000	0.05787990	267.00000000	0.05454164	0.23354153

Exhibit 4 (cont'd)
Sample Regressions
WASHINGTON, D.C. AND ENVIRONS

OWNERS

MODEL:	A	SSE	331.295383	F RATIO	189.21
		DFE	4568	PROB>F	0.0001
DEP VAR:	VALUELN LOG HOUSE VALUE	MSE	0.072525	R-SQUARE	0.6457

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	10.412605	0.035938	289.7380	0.0001	
B1	1	0.105325	0.012904	8.1620	0.0001	BATH DUMMY 1.5
B2	1	0.154126	0.014205	10.8504	0.0001	BATH DUMMY 2
B3	1	0.302760	0.016128	18.7728	0.0001	BATH DUMMY GT 2
R12	1	-0.015417	0.013979	-1.1029	0.2701	ONE OR TWO ROOMS
R4	1	0.045510	0.010333	4.4045	0.0001	FOUR ROOMS
R5	1	0.063049	0.013347	4.7238	0.0001	FIVE ROOMS
RG6	1	0.019313	0.002935609	6.5787	0.0001	NUMBER ROOMS GE 6
BED1	1	-0.070932	0.058142	-1.2200	0.2225	ONE BEDROOM
BED2	1	-0.083489	0.013476	-6.1953	0.0001	TWO BEDROOMS
BED4	1	0.061003	0.010412	5.8589	0.0001	FOUR BEDROOMS
BEDG5	1	0.017588	0.003188006	5.5168	0.0001	NUMBER BEDS GE 5
GAR	1	0.078171	0.008989691	8.6956	0.0001	GARAGE
BASE	1	0.054421	0.011097	4.9041	0.0001	BASEMENT
SFATT	1	-0.122276	0.012081	-10.1217	0.0001	SINGLE FAMILY ATTACHED
AGE1	1	0.003763571	0.005861913	0.6420	0.5209	AGE OF STRUCTURE
AGE1SQ	1	-.0000825003	0.0004235344	-0.1948	0.8456	SQUARE AGE
AGE1CB	1	-.0000010812	.0000854373	-0.1265	0.8993	CUBE AGE
DAGE	1	0.201982	0.411023	0.4914	0.6232	DUMMY OLD STRUCTURE
SHEAT	1	0.003562893	0.011862	0.3004	0.7639	STEAM OR HOT WATER
RHEAT	1	-0.051140	0.041522	-1.2317	0.2181	WALL OR ROOM HEAT W FLUE
EHEAT	1	0.033547	0.033367	1.0054	0.3148	ELECTRIC UNITS
ROOMAC	1	0.040385	0.012226	3.3031	0.0010	ROOM AIR CONDITIONER
CENTAC	1	0.110726	0.014615	7.5762	0.0001	CENTRAL AIR CONDITIONER
NORAD	1	0.006797265	0.014990	0.4535	0.6502	ROOMS WITHOUT HEAT
POOR	1	-0.101586	0.032689	-3.1077	0.0019	WABSNT+SABSNT+NOBATH+SHARED+NHEAT
NOPRIVCY	1	-0.111562	0.037407	-2.9824	0.0029	PASS THRU BR TO ROOM AND OR BATH
NOUT	1	-0.083449	0.032970	-2.5311	0.0114	NO OUTLETS
COOKE	1	0.069360	0.009296161	7.4612	0.0001	COOK WITH ELECTRICITY
CLOT	1	-0.00599095	0.002744868	-2.1826	0.0291	LENGTH OF TENURE CORRECTED FOR DOI
CLOTSQ	1	0.000177684	0.0001230901	1.4435	0.1489	SQUARE LENGTH OF TENURE
DLOT	1	-0.048084	0.042990	-1.1185	0.2634	DUMMY FOR OLD TENANTS
CROWDS	1	-0.110385	0.018208	-6.0624	0.0001	PERSONS PER ROOM
BLACK	1	-0.241638	0.013170	-18.3482	0.0001	BLACK DUMMY
SPAN	1	-0.085693	0.041102	-2.0849	0.0371	SPANISH DUMMY
EXCELN	1	0.158308	0.015550	10.1806	0.0001	EXCELLENT NEIGHBORHOOD RATING
GOODN	1	0.096210	0.014958	6.4318	0.0001	GOOD NEIGHBORHOOD RATING
POORN	1	-0.032355	0.038417	-0.8422	0.3997	POOR NEIGHBORHOOD RATING
ABANDON	1	-0.072323	0.022560	-3.2057	0.0014	ABANDONED HOUSING, ENUMERATOR
Q	1	0.00258428	0.001403317	1.8416	0.0656	TIME TREND
CC1	1	-0.053228	0.022387	-2.3776	0.0175	PRIMARY CENTRAL CITY
FORAY	1	0.006645921	0.002466187	2.6948	0.0071	CENTRAL CITY INFLATION DIFFERENTIAL
MONTGOM	1	-0.029963	0.012507	-2.3958	0.0166	

Exhibit 4 (cont'd)
Sample Regressions
WASHINGTON, D.C. AND ENVIRONS

OWNERS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
PRINCEG	1	-0.132815	0.013263	-10.0137	0.0001	
ALEXARL	1	0.038427	0.019742	1.9464	0.0517	

TEST:TEST001 (Age)	NUMERATOR:	0.14089580	DF:	4	F VALUE:	1.9427
	DENOMINATOR:	0.07252526	DF:	4568	PROB > F:	0.1005
TEST:TEST002 (Length of Tenure)	NUMERATOR:	0.38978594	DF:	3	F VALUE:	5.3745
	DENOMINATOR:	0.07252526	DF:	4568	PROB > F:	0.0012
TEST:TEST003 (Neighborhood Rating)	NUMERATOR:	3.10389342	DF:	3	F VALUE:	42.7974
	DENOMINATOR:	0.07252526	DF:	4568	PROB > F:	0.0001
TEST:TEST004 (Location)	NUMERATOR:	2.35098889	DF:	4	F VALUE:	32.4161
	DENOMINATOR:	0.07252526	DF:	4568	PROB > F:	0.0001

Exhibit 5

Measures of Explanatory Power of the
Hedonic Regressions:
R-Squared and Standard Errors

	Renters		Owners	
	R-Squared	Standard Error	R-Squared	Standard Error
Wave I Cities				
Albany	.65	.32	.54	.34
Anaheim	.59	.21	.59	.23
Boston	.52	.30	.55	.29
Dallas	.77	.24	.71	.33
Detroit	.67	.25	.65	.29
Fort Worth	.78	.25	.71	.33
Los Angeles	.59	.27	.58	.30
Memphis	.81	.24	.74	.31
Minneapolis	.66	.22	.62	.26
Newark	.52	.27	.60	.25
Orlando	.72	.24	.62	.32
Phoenix	.70	.25	.62	.31
Pittsburgh	.69	.30	.62	.36
Salt Lake City	.63	.26	.54	.29
Spokane	.65	.25	.64	.30
Tacoma	.62	.24	.59	.31
Washington, D.C.	.72	.22	.65	.27
Wichita	.73	.24	.68	.31
Wave II Cities				
Atlanta	.73	.26	.68	.30
Chicago	.61	.25	.62	.26
Cincinnati	.71	.25	.66	.30
Colorado Springs	.64	.21	.62	.27
Columbus	.65	.23	.68	.28
Hartford	.59	.25	.58	.22
Kansas City	.73	.25	.70	.33
Madison	.52	.25	.63	.24
Miami	.67	.23	.59	.29
Milwaukee	.64	.23	.65	.26
New Orleans	.72	.28	.66	.30
Newport News	.78	.21	.66	.26
Paterson	.64	.26	.49	.20
Philadelphia	.65	.29	.74	.33
Portland	.59	.23	.60	.28

Exhibit 5 (cont'd)

Measures of Explanatory Power of the
Hedonic Regressions:
R-Squared and Standard Errors

	Renters		Owners	
	R-Squared	Standard Error	R-Squared	Standard Error
Wave II Cities (cont'd)				
Rochester	.59	.23	.65	.29
San Antonio	.82	.26	.73	.33
San Bernadino	.53	.29	.58	.30
San Diego	.57	.25	.56	.27
San Francisco	.56	.26	.59	.28
Springfield	.64	.26	.51	.26
Wave III Cities				
Allentown	.60	.28	.61	.33
Baltimore	.65	.26	.69	.35
Birmingham	.82	.27	.70	.37
Buffalo	.65	.27	.64	.29
Cleveland	.71	.24	.68	.26
Denver	.65	.25	.61	.25
Grand Rapids	.62	.23	.65	.30
Honolulu	.52	.35	.32	.27
Houston	.77	.26	.67	.36
Indianapolis	.74	.24	.71	.31
Las Vegas	.69	.19	.55	.26
Louisville	.72	.25	.71	.31
New York	.58	.32	.57	.21
Oklahoma City	.72	.25	.74	.30
Omaha	.68	.24	.70	.31
Providence	.64	.30	.50	.28
Raleigh	.82	.23	.73	.29
Sacramento	.61	.25	.64	.28
St. Louis	.70	.28	.66	.35
Seattle	.61	.25	.56	.30

Exhibit 6

Stem and Leaf Display for R^2s

Renter

Row Count	Cumulative Count		
4	4	.8	1 2 2 2
4	8	*	7 7 8 8
13	21	.7	0 0 1 1 2 2 2 2 2 3 3 3 4
13	34	*	5 5 5 5 5 5 5 6 7 7 8 9 9
13	47	.6	0 0 1 1 1 2 2 3 4 4 4 4 4
7	54	*	6 7 8 9 9 9 9
5	59	.5	2 2 2 2 3
		*	
		.4	
		*	
		.3	

Owner

Row Count	Cumulative Count		
		.8	
		*	
12	12	.7	0 0 0 1 1 1 1 3 3 4 4 4
15	27	*	5 5 5 5 5 6 6 6 6 7 8 8 8 8 9
14	41	.6	0 0 1 1 2 2 2 2 2 2 3 4 4 4
12	53	*	5 5 6 6 7 8 8 9 9 9 9 9
4	57	.5	0 1 4 4
1	58	*	9
		.4	
		*	
1	59	.3	2

Exhibit 7

MEANS OF RENTER HEDONIC COEFFICIENTS FOR 59 SMSAS

VARIABLE	N	MEAN	MEDIAN	STD DEV	IQR	POS SIG	NEG SIG
INTERCEP	59	4.9376	4.9124	0.224	0.2606	59	0
B1	59	0.1165	0.1153	0.046	0.0583	57	0
B2	59	0.194	0.1938	0.0501	0.0608	57	0
B3	59	0.3722	0.3625	0.112	0.1741	55	0
R1	59	-0.0972	-0.097	0.0466	0.0654	0	49
R3	59	0.0693	0.066	0.0284	0.0352	55	0
RG4	59	0.0291	0.0299	0.0112	0.0161	53	0
BED0	59	-0.1984	-0.1961	0.0686	0.0629	0	57
BED2	59	0.1242	0.1191	0.0328	0.055	59	0
BED3	59	0.2305	0.2249	0.0607	0.0731	59	0
BEDG4	59	0.0645	0.0577	0.0252	0.0376	59	0
ELEVP	59	0.1351	0.115	0.1139	0.1224	37	0
SFATT	59	-0.0365	-0.0298	0.0671	0.0938	3	23
SFDET	59	-0.0381	-0.038	0.0759	0.1011	9	28
DUPLEX	59	-0.0264	-0.022	0.0514	0.0755	2	22
NGT50	59	0.0468	0.057	0.0748	0.0741	27	3
AGE1	59	-0.006	-0.0073	0.0086	0.0119	4	25
AGE1SQ	59	.0000	.0000	0.0002	0.0003	13	9
DAGE	59	0.0546	0.0151	0.2731	0.3099	12	6
RHEAT	59	-0.1668	-0.1614	0.0948	0.1082	0	56
ROOMAC	59	0.0756	0.0751	0.0467	0.0594	48	1
CENTAC	59	0.1868	0.206	0.1068	0.1171	53	1
NORAD	58	-0.0135	-0.0155	0.0451	0.0482	6	17
POOR	59	-0.2535	-0.2562	0.0763	0.0625	0	59
NOPRIVCY	59	-0.0487	-0.0525	0.0328	0.0382	1	37
NOUT	59	-0.0708	-0.0694	0.0709	0.0792	1	30
BADHALL	59	-0.0013	0.0007	0.0283	0.0389	3	6
DFECT	59	-0.0048	-0.0045	0.0114	0.0149	4	9
CLOT	59	-0.0317	-0.0309	0.0094	0.0145	0	59
CLOTSQ	59	0.0008	0.0008	0.0005	0.0005	49	0
DLOT	59	-0.1413	-0.1572	0.2605	0.2519	4	21
CROWDS	59	0.0266	0.0308	0.0356	0.0619	19	2
BLACK	59	-0.08	-0.1032	0.0659	0.0958	0	39
SPAN	59	-0.0386	-0.0485	0.0765	0.0986	2	22
LLBLG	59	-0.028	-0.0269	0.0413	0.0635	1	21
NHUINC	59	0.0447	0.0428	0.0864	0.101	24	4
HEATINC	59	0.0825	0.0904	0.1059	0.1008	44	4
PARKINC	55	0.0869	0.0792	0.0905	0.1117	29	0
FURNINC	59	0.0511	0.0471	0.0648	0.0737	35	4
EXCELN	59	0.0545	0.0524	0.0286	0.0422	40	0
GOODN	59	0.0218	0.0216	0.0241	0.0349	25	1
POORN	59	-0.0033	-0.0097	0.0446	0.0471	2	6
ABANDON	59	-0.0677	-0.067	0.0441	0.0483	0	38
LITTER	59	-0.0158	-0.0142	0.0246	0.0282	1	12
NOSHOPS	59	-0.018	-0.0172	0.0358	0.048	3	15
Q	59	0.0051	0.0051	0.0037	0.0045	37	0
QHEAT	59	-0.001	-0.0005	0.0139	0.0064	8	6
CC1	40	0.0101	0.0043	0.081	0.1054	19	11

NOTE: The last two columns report the number of equations in which the absolute value of the t-statistic exceeds 1.67. This is a test of whether individual coefficient estimates are different from zero. A t-test for the hypothesis that the coefficient estimates are on average different from zero is based on the ratio of the mean of the 59 estimates (col. 3) to their standard deviation (col. 5).

Exhibit 8

MEANS OF OWNER HEDONIC COEFFICIENTS FOR 59 SMSAS

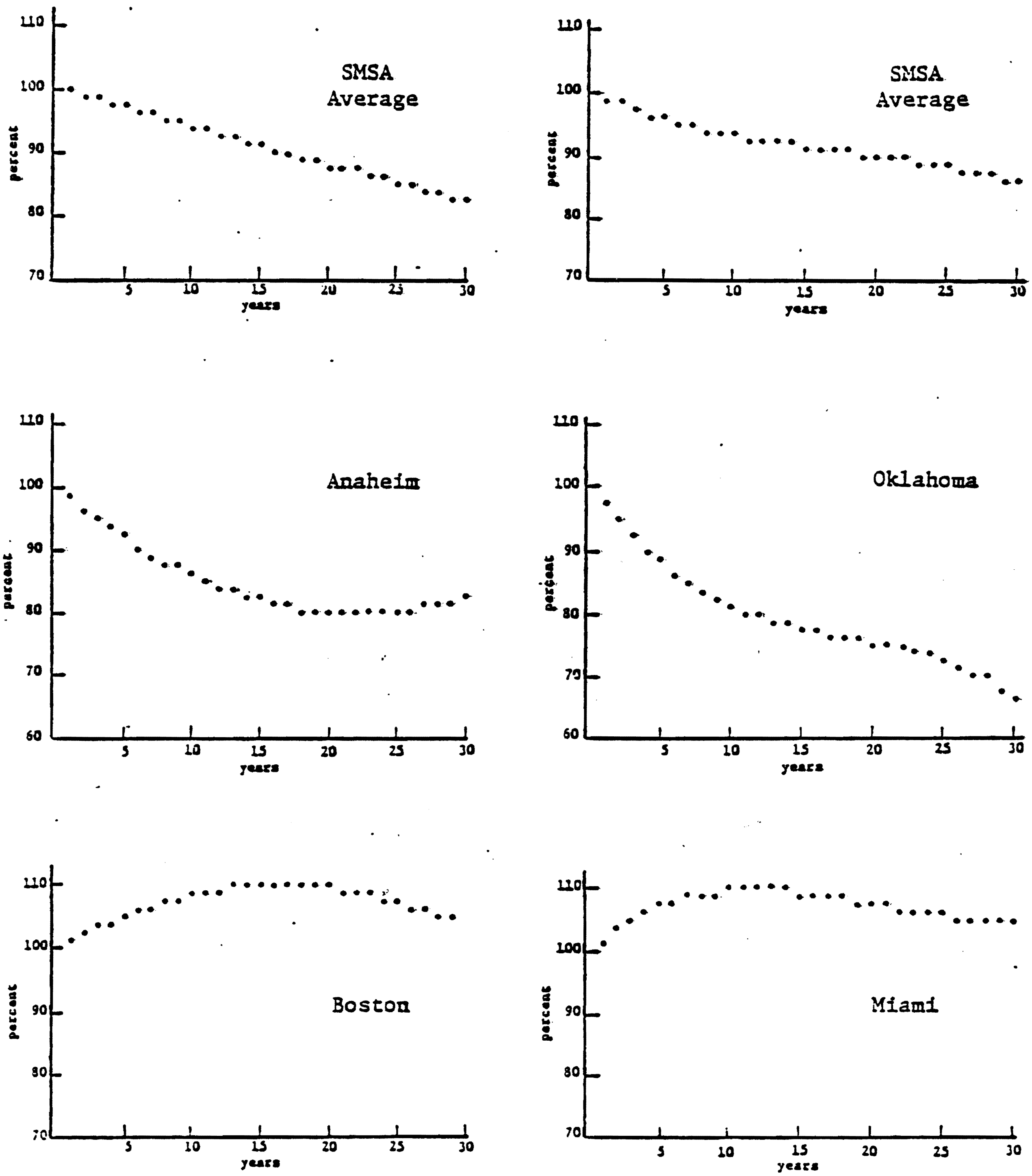
VARIABLE	N	MEAN	MEDIAN	STD DEV	IQR	POS SIG	NEG SIG
INTERCEP	59	10.1051	10.0853	0.2645	0.4045	59	0
B1	59	0.1114	0.108	0.0283	0.0401	59	0
B2	59	0.1765	0.1661	0.0443	0.0703	59	0
B3	59	0.382	0.3774	0.0861	0.1331	59	0
R12	59	-0.0811	-0.0753	0.0277	0.0321	0	58
R4	59	0.0811	0.0775	0.0283	0.0303	58	0
R5	59	0.1404	0.1355	0.0518	0.0775	57	0
RG6	59	0.031	0.0291	0.0128	0.0137	58	0
BED1	59	-0.2564	-0.2579	0.1112	0.1506	0	56
BED2	59	-0.0962	-0.0968	0.0303	0.0419	0	59
BED4	59	0.0516	0.0447	0.0331	0.0464	45	0
BEDG5	59	0.0199	0.0183	0.0093	0.0132	52	0
GAR	59	0.115	0.1098	0.0444	0.0531	57	0
BASE	59	0.0665	0.0633	0.0509	0.0512	42	1
SFATT	59	0.0279	0.0087	0.162	0.1582	17	11
AGE1	59	-0.01	-0.0105	0.0108	0.0139	2	24
AGE1SQ	59	0.0004	0.0004	0.0007	0.0008	14	2
AGE1CB	59	.0000	.0000	.0000	.0000	1	15
DAGE	59	0.3614	0.2561	0.6877	1.0595	16	2
SHEAT	57	0.0951	0.0779	0.1302	0.1041	33	1
RHEAT	59	-0.1054	-0.104	0.1201	0.12	0	38
EHEAT	59	0.0454	0.0433	0.0944	0.115	15	6
ROOMAC	59	0.0376	0.0259	0.0565	0.0445	29	4
CENTAC	59	0.1348	0.1128	0.0799	0.0985	54	1
NORAD	59	-0.0308	-0.0444	0.0661	0.0941	11	34
POOR	59	-0.1688	-0.1811	0.1664	0.2244	2	45
NOPRIVCY	59	-0.0743	-0.0826	0.0691	0.0659	2	39
NOUT	59	-0.0985	-0.0992	0.1089	0.097	1	39
COOKE	59	0.0773	0.0836	0.0334	0.0338	55	0
CLOT	59	-0.0049	-0.0058	0.0055	0.0081	1	26
CLOTSQ	59	0.0002	0.0002	0.0002	0.0004	15	1
DLOT	59	-0.0575	-0.0574	0.0815	0.1008	1	16
CROWDS	59	-0.0469	-0.0597	0.0371	0.0503	0	37
BLACK	59	-0.1481	-0.1509	0.0767	0.1057	0	51
SPAN	59	-0.07	-0.0821	0.1045	0.0905	0	26
EXCELN	59	0.1594	0.1586	0.0452	0.0615	59	0
GOODN	59	0.0676	0.0631	0.0361	0.0488	10	0
POORN	59	-0.0452	-0.0528	0.1019	0.102	6	18
ABANDON	59	-0.0944	-0.0911	0.0788	0.0916	1	42
Q	59	0.0057	0.0056	0.0032	0.0053	48	0
CC1	40	-0.0697	-0.0557	0.1103	0.1236	4	22
FORAY	42	0.0004	0.0005	0.0047	0.005	5	3

NOTE: The last two columns report the number of equations in which the absolute value of the t-statistic exceeds 1.67. This is a test of whether individual coefficient estimates are different from zero. A t-test for the hypothesis that the coefficient estimates are on average different from zero is based on the ratio of the mean of the 59 estimates (col. 3) to their standard deviation (col. 5).

Exhibit 9

Rents and Values as a Function of Dwelling Age

Rents Values



Note: Horizontal axis measures time in years; vertical axis measures rent or value as a percentage of year-zero rent or value.

Exhibit 10

Coefficients for Central City Location, Adjusted for the Presence of other Locational Variables

		Renters			Owners		
	Number of Locational Variables*	Adjusted Coefficient	Proportion of Sample in Central City	Probability Greater than F**	Adjusted Coefficient	Proportion of Sample in Central City	Probability Greater than F
Wave I							
Albany	1	.092	.57	.0001	.040	.20	.2749
Anaheim	1	-.055	.35	.0001	-.123	.25	.0001
Boston	1	-.038	.59	.0002	-.235	.18	.0001
Dallas	1	.061	.69	.0002	.008	.46	.7700
Detroit	3	-.133	.63	.0001	-.257	.38	.0001
Fort Worth	1	.071	.57	.0001	-.006	.47	.8310
Los Angeles	2	.031	.51	.0001	.089	.38	.0001
Memphis	0						
Minneapolis	2	-.071	.34	.0001	-.092	.19	.0005
Newark	3	-.146	.27	.0001	-.355	.02	.0001
Orlando	0						
Phoenix	1	.037	.62	.0411	-.029	.57	.2601
Pittsburgh	2	.139	.33	.0001	.029	.15	.6334
Salt Lake City	0						
Spokane	0						
Tacoma	0						
Washington	4	.096	.58	.0001	-.014	.33	.0001
Wichita	0						
Wave II							
Atlanta	2	.053	.56	.0001	-.016	.35	.0001
Chicago	4	-.022	.65	.0459	-.144	.23	.0001
Cincinnati	2	.008	.53	.0043	.004	.18	.0001
Colorado Springs	0						
Columbus	1	-.064	.76	.0002	-.103	.50	.0001
Hartford	0						
Kansas City	2	.021	.49	.0001	-.075	.35	.0050
Madison	0						
Miami	1	-.013	.36	.4286	-.007	.15	.8571
Milwaukee	2	-.020	.68	.0165	-.183	.36	.0001
New Orleans	1	.062	.64	.0017	.060	.34	.0521
Newport News	0						
Paterson	1	-.189	.33	.0001	-.089	.10	.0013
Philadelphia	6	-.037	.54	.0001	-.097	.42	.0001
Portland	1	.036	.44	.0216	-.108	.32	.0001
Rochester	1	.053	.47	.0081	-.181	.22	.0001
San Antonio	0						
San Bernardino	1	-.021	.29	.3368	.039	.26	.1684
San Diego	1	.072	.59	.0001	.001	.49	.9713
San Francisco	3	.124	.41	.0001	.064	.19	.0001
Springfield	0						

Exhibit 10 (cont'd)

Coefficients for Central City Location,
Adjusted for the Presence of other Locational Variables

	Number of Locational Variables*	Renters: Adjusted Coefficient	Renters: Proportion of Sample in Central City	Renters: Probability Greater than F**	Owners: Adjusted Coefficient	Owners: Proportion of Sample in Central City	Owners: Probability Greater than F
Wave III							
Allentown***	1	-.065	.52	.0023	-.030	.48	.0367
Baltimore	2	-.064	.52	.0086	-.288	.29	.0001
Birmingham	1	.072	.57	.0006	-.053	.33	.0962
Buffalo	1	-.019	.48	.3789	-.305	.15	.0001
Cleveland	2	-.203	.46	.0001	-.231	.22	.0001
Denver	2	-.021	.49	.0001	-.036	.32	.0001
Grand Rapids	0						
Honolulu	1	.193	.65	.0001	.112	.46	.0001
Houston	1	.065	.64	.0001	.068	.36	.0002
Indianapolis	1	-.016	.79	.4196	-.085	.63	.0013
Las Vegas	0						
Louisville	1	-.017	.56	.4046	-.106	.35	.0002
New York	7	-.112	.69	.0001	-.008	.13	.0001
Oklahoma City	1	.002	.62	.9307	.031	.57	.2336
Omaha	0						
Providence	1	-.070	.45	.0003	-.014	.28	.6113
Raleigh	0						
Sacramento	1	-.014	.34	.4459	-.089	.31	.0005
St. Louis	4	-.159	.62	.0005	-.074	.27	.0001
Seattle	1	.067	.54	.0001	-.003	.35	.8480
Average		-.005					

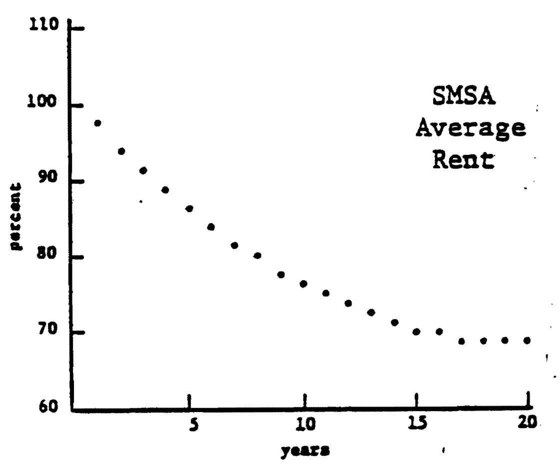
*Includes central city and county variables.

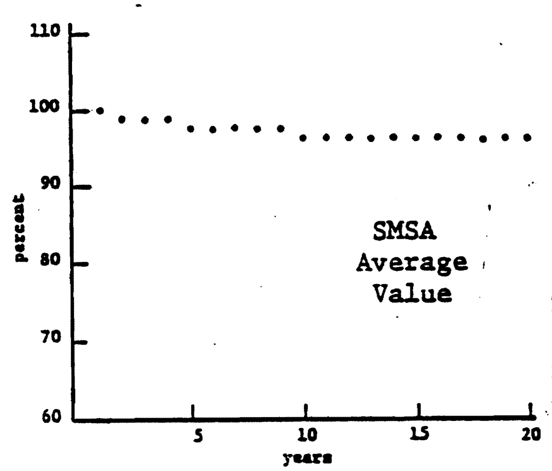
**The probability that the joint effect of the locational variables used to construct the adjusted coefficient is really zero, given the sample.

***Allentown central city is not identified. Coefficient is for Lehigh county.

Exhibit 11

Rents and Values as a Function of Occupants' Length of Tenure





Note: Horizontal axis measures time in years; vertical axis measures rent or value as a percentage of year-zero rent or value

Exhibit 12

Measures of Inflation in Housing Prices

		1974-1975	1975-1976	1976-1977
Rent Inflation				
Structural Rent Inflation	Average	5.2	6.9	6.9
	Maximum	13.9 (Newark)	14.2 (Portland)	17.5 (Buffalo)
	Minimum	-5.0 (Dallas)	2.3 (Columbus)	-2.9 (New York)
	Standard Deviation	5.7	3.5	5.0
Heat Included Differential	Average	-2.2	2.6	-1.1
	Maximum	14.6 (Dallas)	30.6 (San Bernardino)	32.3 (Sacramento)
	Minimum	-15.4 (Memphis)	-12.8 (San Antonio)	-68.4 (Honolulu)[a]
	Standard Deviation	7.8	9.8	18.9
Value Inflation				
SMSA-Wide Inflation	Average	8.1	5.8	8.0
	Maximum	17.6 (Salt Lake)	12.7 (San Francisco)	16.2 (Sacramento)
	Minimum	2.2 (Detroit)	-1.35 (Colorado Springs)	0.6 (Louisville)
	Standard Deviation	3.6	3.4	4.4
Central City Inflation Differential	Average	-0.7	0.1	2.3
	Maximum	8.5 (Washington)	11.4 (New Orleans)	15.2 (New York)
	Minimum	-9.0 (Pittsburgh)	-12.6 (Rochester)	-6.6 (Providence)
	Standard Deviation	6.0	5.8	6.0

Note: All numbers are annual percentages.
[a]Except for Honolulu the lowest rate is -12.3 percent in Buffalo.

Exhibit 13

Analysis of the Residuals for each SMSA
The Renter Equations in Wave 1

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Albany	986	.026	.360	0	(1.28)	6	(-1.24)
Anaheim	1370	-.004	.204	7	(0.71)	7	(-0.72)
Boston	4650	.013	.345	2	(1.21)	11	(-1.20)
Dallas	1283	.005	.250	0	(0.88)	5	(-0.87)
Detroit	2827	.005	.289	2	(1.02)	4	(-1.00)
Fort Worth	1088	.015	.258	0	(0.91)	4	(-0.89)
Los Angeles	5477	.002	.290	8	(1.01)	15	(-1.01)
Memphis	1261	.008	.248	0	(0.87)	8	(-0.86)
Minneapolis	1195	.004	.214	0	(0.76)	5	(-0.74)
Newark	1371	.022	.252	0	(0.90)	12	(-0.87)
Orlando	928	.022	.232	1	(0.82)	9	(-0.80)
Phoenix	924	.004	.225	3	(0.86)	2	(-0.86)
Pittsburg	947	.009	.366	0	(1.29)	2	(-1.28)
Salt Lake City	1085	.013	.271	1	(0.96)	7	(-0.94)
Spokane	1004	.009	.288	0	(1.01)	0	(-1.00)
Tacoma	1138	.009	.245	0	(0.87)	6	(-0.84)
Washington, D.C.	5532	-.001	.232	11	(0.81)	19	(-0.81)
Wichita	1180	.000	.277	0	(0.97)	2	(-0.97)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 14

Analysis of the Residuals for each SMSA
The Owner Equations in Wave 1

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Albany	1952	.011	.352	0	(1.25)	9	(-1.21)
Anaheim	2269	-.012	.282	1	(0.98)	2	(-1.00)
Boston	3862	.016	.315	2	(1.11)	14	(-1.10)
Dallas	2072	-.011	.363	4	(1.26)	4	(-1.28)
Detroit	7422	.007	.313	10	(1.10)	39	(-1.09)
Fort Worth	2453	-.009	.365	11	(1.27)	4	(-1.28)
Los Angeles	5362	-.007	.350	1	(1.22)	6	(-1.23)
Memphis	1912	.007	.338	8	(1.18)	4	(-1.18)
Minneapolis	2458	-.007	.306	3	(1.07)	2	(-1.07)
Newark	1869	.003	.266	1	(0.93)	2	(-0.93)
Orlando	2255	.000	.350	6	(1.22)	4	(-1.23)
Phoenix	2290	-.010	.330	1	(1.17)	11	(-1.16)
Pittsburg	2384	.005	.404	1	(1.42)	7	(-1.40)
Salt Lake City	2579	-.008	.337	2	(1.17)	6	(-1.19)
Spokane	2402	-.005	.342	1	(1.20)	3	(-1.19)
Tacoma	2347	-.017	.340	4	(1.18)	3	(-1.20)
Washington, D.C.	4613	.007	.285	10	(1.00)	13	(-1.00)
Wichita	2286	.000	.367	0	(1.29)	4	(-1.28)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 15

Analysis of the Residuals for each SMSA
The Renter Equations in Wave 2

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Atlanta	3928	.004	.268	4	(0.95)	22	(-0.93)
Chicago	4184	.012	.263	4	(0.93)	19	(-0.91)
Cincinnati	1187	.007	.277	1	(0.98)	4	(-0.96)
Colorado Springs	1226	.002	.214	3	(0.74)	4	(-0.75)
Columbus	1283	.001	.246	1	(0.86)	6	(-0.86)
Hartford	1259	.019	.240	0	(0.85)	9	(-0.83)
Kansas City	1067	.006	.272	2	(0.96)	4	(-0.95)
Madison	1615	.008	.220	0	(0.78)	8	(-0.75)
Miami	1318	.012	.251	1	(0.89)	7	(-0.87)
Milwaukee	1349	.008	.246	0	(0.87)	3	(-0.85)
New Orleans	1382	.013	.285	1	(1.01)	12	(-0.99)
Newport News	1118	.010	.221	0	(0.79)	8	(-0.72)
Paterson	1275	.014	.219	0	(1.03)	4	(-1.01)
Philadelphia	2974	.003	.312	3	(1.10)	8	(-1.08)
Portland	1267	-.003	.238	4	(0.84)	7	(-0.83)
Rochester	1080	.005	.199	2	(0.70)	11	(-0.69)
San Antonio	1084	.007	.259	0	(0.92)	10	(-0.90)
San Bernardino	999	.012	.313	2	(1.10)	1	(-1.10)
San Diego	1398	.016	.263	3	(0.92)	4	(-0.92)
San Francisco	4610	.007	.274	9	(0.97)	24	(-0.95)
Springfield	1311	.018	.293	0	(1.03)	7	(-1.02)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 16

Analysis of the Residuals for each SMSA
The Owner Equations in Wave 2

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Atlanta	5258	.005	.337	16	(1.17)	8	(-1.18)
Chicago	5323	.003	.307	0	(1.08)	10	(-1.07)
Cincinnati	2108	-.001	.343	0	(1.20)	3	(-1.21)
Colorado Springs	2008	-.017	.332	1	(1.16)	2	(-1.17)
Columbus	2163	.004	.339	1	(1.19)	2	(-1.18)
Hartford	2245	-.002	.252	0	(0.89)	7	(-0.88)
Kansas City	2204	.004	.359	4	(1.27)	8	(-1.24)
Madison	1918	.003	.280	1	(0.98)	7	(-0.98)
Miami	1660	-.009	.365	0	(1.28)	1	(-1.28)
Milwaukee	2214	.009	.294	0	(1.04)	5	(-1.02)
New Orleans	1608	-.009	.372	0	(1.30)	0	(-1.30)
Newport News	1872	.005	.275	4	(0.96)	8	(-0.96)
Paterson	2169	.014	.218	0	(0.77)	4	(-0.75)
Philadelphia	7098	.011	.360	7	(1.27)	31	(-1.25)
Portland	2374	-.006	.338	3	(1.18)	3	(-1.19)
Rochester	2308	.004	.314	1	(1.10)	4	(-1.10)
San Antonio	2060	-.005	.367	3	(1.29)	5	(-1.28)
San Bernardino	2052	-.005	.338	3	(1.17)	3	(-1.19)
San Diego	1793	-.018	.338	0	(1.17)	0	(-1.19)
San Francisco	5411	.001	.339	0	(1.19)	12	(-1.19)
Springfield	2148	.006	.305	0	(1.07)	7	(-1.07)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 17

Analysis of the Residuals for each SMSA
The Renter Equations in Wave 3

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Allentown	818	.015	.300	0	(1.07)	3	(-1.03)
Baltimore	1258	.017	.268	1	(0.95)	9	(-0.93)
Birmingham	1020	.006	.289	1	(1.02)	3	(-1.00)
Buffalo	1121	.010	.296	2	(1.04)	3	(-1.03)
Cleveland	1176	-.002	.264	1	(0.93)	2	(-0.93)
Denver	1289	.007	.240	6	(0.85)	9	(-0.83)
Grand Rapids	778	.008	.245	0	(0.87)	4	(-0.85)
Honolulu	1338	.026	.315	0	(1.13)	25	(-1.07)
Houston	4185	.000	.249	23	(0.87)	2	(-0.88)
Indianapolis	1145	-.006	.264	0	(0.93)	4	(-0.92)
Las Vegas	1308	-.001	.199	0	(0.70)	7	(-0.69)
Louisville	1008	-.004	.301	0	(1.05)	2	(-1.05)
New York	4186	.014	.328	3	(1.16)	20	(-1.13)
Oklahoma City	1048	-.002	.244	3	(0.86)	3	(-0.85)
Omaha	1160	.010	.236	4	(0.83)	8	(-0.82)
Providence	1215	.017	.312	2	(1.10)	4	(-1.08)
Raleigh	1416	.005	.258	0	(0.91)	3	(-0.90)
Sacramento	1130	.015	.223	1	(0.79)	6	(-0.77)
St. Louis	3407	.007	.292	4	(1.03)	17	(-1.01)
Seattle	3327	.010	.268	3	(0.94)	10	(-0.93)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 18

Analysis of the Residuals for each SMSA
The Owner Equations in Wave 3

SMSA	Number of Residuals	Median	IQR of the Residuals	Positive Outliers[1]		Negative Outliers[2]	
Allentown	2302	.011	.373	0	(1.31)	13	(-1.30)
Baltimore	2003	.017	.372	4	(1.31)	16	(-1.29)
Birmingham	2292	.018	.419	1	(1.47)	7	(-1.46)
Buffalo	2006	-.001	.316	0	(1.11)	4	(-1.09)
Cleveland	2207	.010	.296	1	(1.03)	9	(-1.03)
Denver	2065	-.003	.288	2	(1.01)	7	(-1.00)
Grand Rapids	2522	-.001	.338	1	(1.19)	9	(-1.18)
Honolulu	1448	.022	.271	0	(0.93)	11	(-0.93)
Houston	6391	.001	.394	10	(1.38)	24	(-1.38)
Indianapolis	2109	.001	.353	1	(1.23)	3	(-1.24)
Las Vegas	2058	-.017	.316	0	(1.09)	1	(-1.12)
Louisville	2356	.003	.335	10	(1.18)	15	(-1.17)
New York	4583	-.001	.254	0	(0.89)	9	(-0.89)
Oklahoma City	2029	-.011	.343	3	(1.19)	7	(-1.20)
Omaha	2166	-.004	.354	4	(1.24)	8	(-1.24)
Providence	1989	-.001	.312	0	(1.10)	8	(-1.08)
Raleigh	2131	-.002	.291	8	(1.02)	6	(-1.01)
Sacramento	2124	-.004	.326	2	(1.14)	5	(-1.14)
St. Louis	5371	.007	.403	5	(1.42)	25	(-1.40)
Seattle	6667	-.015	.351	17	(1.22)	5	(-1.24)

1. The number in parenthesis is the upper fence.
2. The number in parenthesis is the lower fence.

Exhibit 19

Number of Outliers in each SMSA
Wave 1

	Renter Number of Outliers[1]		Owner Number of Outliers[1]	
Albany	6	(.61)	9	(.46)
Anaheim	16	(1.17)	3	(.13)
Boston	13	(.28)	16	(.41)
Dallas	5	(.39)	8	(.39)
Detroit	6	(.21)	49	(.66)
Fort Worth	4	(.37)	15	(.61)
Los Angeles	23	(.42)	7	(.13)
Madison	8	(.50)	8	(.42)
Memphis	8	(.63)	12	(.63)
Minneapolis	5	(.42)	5	(.20)
Newark	12	(.88)	3	(.16)
Orlando	10	(1.08)	10	(.44)
Phoenix	5	(.54)	12	(.52)
Pittsburg	2	(.21)	8	(.34)
Salt Lake City	8	(.74)	8	(.31)
Spokane	0	( )	4	(.17)
Tacoma	6	(.53)	7	(.30)
Washington, D.C.	31	(.56)	23	(.50)
Wichita	2	(.17)	4	(.17)

1. The number in parenthesis is the percentage of outliers.

Exhibit 20

Number of Outliers in each SMSA
Wave 2

	Renter Number of Outliers[1]		Owner Number of Outliers[1]	
Atlanta	26	(.66)	24	(.46)
Chicago	23	(.55)	10	(.19)
Cincinnati	5	(.42)	3	(.14)
Colorado Springs	7	(.57)	3	(.15)
Columbus	7	(.55)	3	(.14)
Hartford	9	(.71)	7	(.31)
Kansas City	6	(.56)	12	(.54)
Miami	8	(.61)	1	(.06)
Milwaukee	3	(.22)	5	(.23)
New Orleans	13	(.94)	0	
Newport News	8	(.72)	12	(.64)
Paterson	4	(.31)	4	(.18)
Philadelphia	11	(.37)	38	(.54)
Portland	11	(.87)	6	(.25)
Rochester	13	(1.20)	5	(.22)
San Antonio	10	(.92)	8	(.39)
San Bernardino	3	(.30)	6	(.29)
San Diego	7	(.50)	0	
San Francisco	33	(.72)	12	(.22)
Springfield	7	(.53)	7	(.33)

1. The number in parenthesis is the percentage of outliers.

Exhibit 21

Number of Outliers in each SMSA
Wave 3

	Renter Number of Outliers[1]		Owner Number of Outliers[1]	
Allentown	3	(.37)	13	(.56)
Baltimore	10	(.79)	20	(1.00)
Birmingham	4	(.39)	8	(.35)
Buffalo	5	(.45)	4	(.20)
Cleveland	3	(.26)	10	(.45)
Denver	15	(1.16)	9	(.44)
Grand Rapids	4	(.51)	10	(.40)
Honolulu	25	(1.87)	11	(.76)
Houston	25	(.60)	34	(.53)
Indianapolis	4	(.35)	4	(.19)
Las Vegas	7	(.54)	1	(.05)
Louisville	2	(.20)	25	(1.06)
New York	23	(.55)	9	(.20)
Oklahoma City	6	(.57)	10	(.49)
Omaha	12	(1.03)	12	(.55)
Providence	6	(.49)	8	(.40)
Raleigh	3	(.21)	14	(.66)
Sacramento	7	(.62)	7	(.33)
St. Louis	11	(.32)	30	(.56)
Seattle	13	(.39)	22	(.33)

1. The number in parenthesis is the percentage of outliers.

Exhibit 22

Stem and Leaf Display for the Interquartile Range of the Residuals in the Renter Equations[1]

.42															
.40															
.38															
.36	0	6													
.34	5														
.32	8														
.30	0	1	$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{5}$									
.28	5	8	9	9	$\bar{0}$	$\bar{2}$	$\bar{3}$	$\bar{6}$							
.26	3	3	4	4	8	8	8	$\bar{1}$	$\bar{2}$	$\bar{4}$	$\bar{7}$	$\bar{7}$			
.24	0	0	4	5	5	6	6	8	9	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{8}$	$\bar{8}$	$\bar{9}$
.22	0	1	3	5	$\bar{2}$	$\bar{2}$	$\bar{6}$	$\bar{8}$							
.20	0	$\bar{4}$	$\bar{4}$	$\bar{9}$											
.18	$\bar{9}$	$\bar{9}$													

1. Add 0.01 to numbers covered by bars. For example, the two entries in the bottom leaf represent 0.199.

Exhibit 23

Stem and Leaf Display for the Interquartile Range of the Residuals in the Owner Equations

.42	
.40	3 4 $\bar{9}$
.38	$\bar{4}$
.36	0 3 5 5 7 7 $\bar{2}$ $\bar{2}$ $\bar{3}$
.34	0 2 3 3 4 $\bar{0}$ $\bar{0}$ $\bar{1}$ $\bar{2}$ $\bar{3}$ $\bar{9}$
.32	6 $\bar{0}$ $\bar{2}$ $\bar{5}$ $\bar{7}$ $\bar{7}$ $\bar{8}$ $\bar{8}$ $\bar{8}$ $\bar{8}$ $\bar{8}$ $\bar{9}$
.30	2 5 6 7 $\bar{2}$ $\bar{4}$ $\bar{5}$ $\bar{6}$ $\bar{6}$
.28	0 2 5 8 $\bar{1}$ $\bar{4}$ $\bar{6}$
.26	6 $\bar{1}$ $\bar{5}$
.24	$\bar{2}$ $\bar{4}$
.22	
.20	$\bar{8}$
.18	

1. Add 0.01 to numbers covered by bars.

Exhibit 24

Stem and Leaf Display for the Percentage of Outliers in the Renter Equation

	1.87
1.2	0
1.1	6 7
1.0	3 8
.9	2 4
.8	7 8
.7	1 2 2 4 9
.6	0 1 1 2 3 6
.5	0 0 1 3 3 4 4 5 5 5 6 6 7 7
.4	2 2 2 5 9
.3	0 1 2 5 7 7 7 9 9 9
.2	0 1 1 1 2 6 8
.1	7
.0	0

Exhibit 25

Stem and Leaf Display for the Percentage of Outliers in the Owner Equation

1.1	
1.0	0 6
.9	
.8	
.7	6
.6	1 3 4 6 6
.5	0 2 3 4 4 5 6 6
.4	0 0 1 2 4 4 5 6 6 9
.3	0 1 1 3 3 3 4 5 9 9
.2	0 0 2 2 2 3 5 9
.1	3 3 4 5 6 7 7 8 9 9
.0	0 5 6

Exhibit 26

Number of Outliers by Tenure Group

	Positive Outliers	Negative Outliers	Total
Renters	129[1]	450	579
Owners	179[1]	459	638
Total	308	909	1,217

1. Chi-square tests of equal proportions of positive and negative outliers reject the null hypothesis for each tenure group. The Chi-square statistic is 176 in the renter equations and 122 in the owner equations. Both tests are statistically significant at the 0.0001 level. For a description of the Chi-square test statistic, see Snedecor and Cochran (1967), pages 211-13.

Exhibit 27

Baltimore, Maryland

Renters

Plot of Residuals Versus Predicted Rents

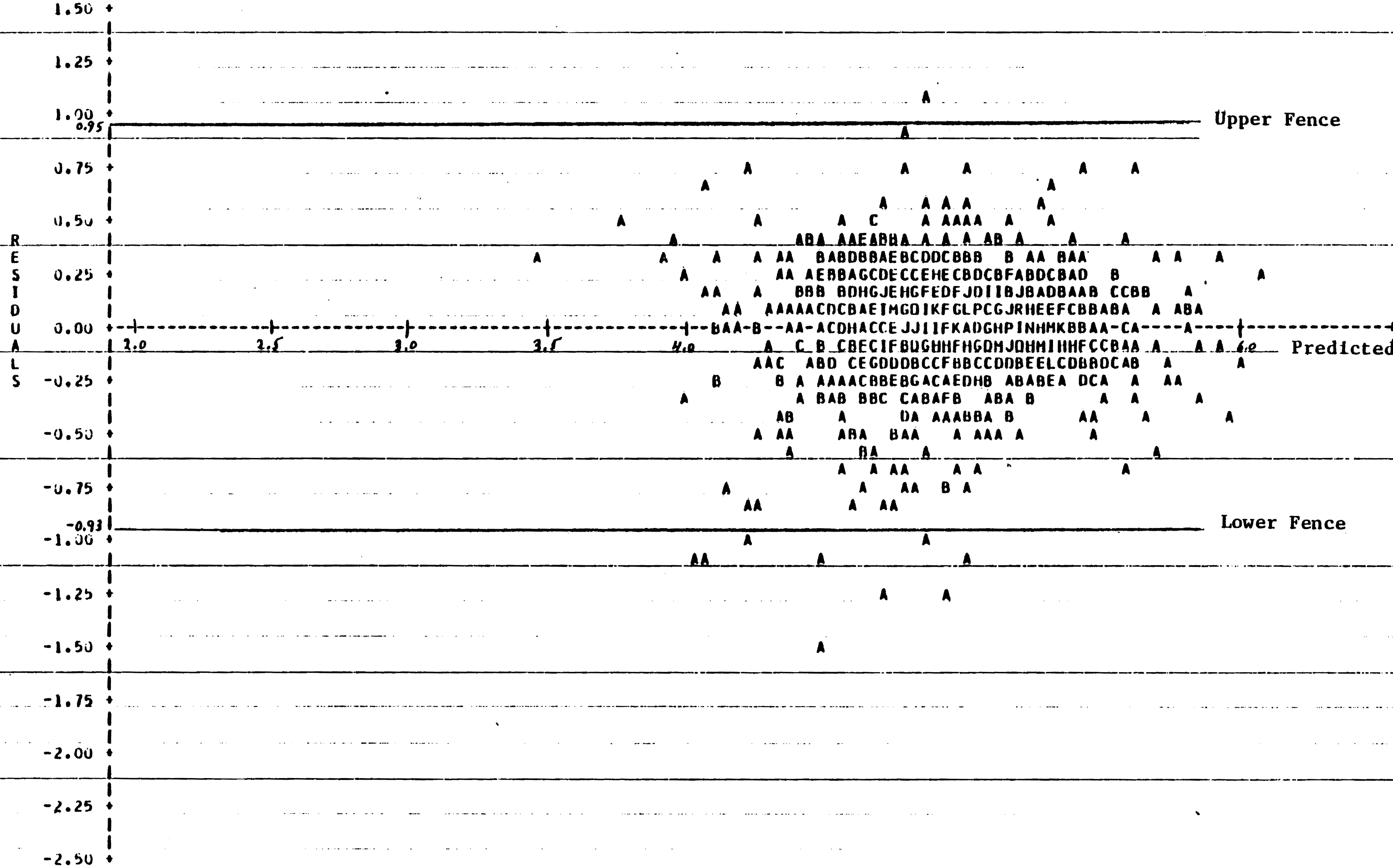


Exhibit 28

OBS	XWEIGHT	XTFI	XCONTROL	Q	AGE1	AGE1SQ	DAGE	CLOT	CLOTSQ	DLOT	CC1	SFDET	SFATT	DUPLEX	NGT50	ELEVP	R1	R3	RG4	BED0	BED2	BED3	BEDG4	RHEAT
1	155.26	15200	97300299100	4	56	3136	1	9.0000	81.00	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1
2	166.12	9000	97300357700	2	56	3136	1	31.5000	992.25	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0
3	166.12	16000	97300436900	3	56	3136	1	0.2500	0.06	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
4	155.26	6516	97300495500	7	56	3136	1	31.9167	1018.67	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0
5	158.80	32754	97300621600	2	56	3136	1	21.5000	462.25	0	1	0	1	0	0	0	0	0	4	0	0	1	0	0
6	153.52	15636	97300712300	7	56	3136	1	4.4167	19.51	0	1	0	1	0	0	0	0	1	0	0	0	0	4	0
7	153.37	4534	97300807100	7	56	3136	1	2.4167	5.84	0	1	0	1	0	0	0	1	0	0	0	0	0	4	0
8	145.19	2300	97302940800	4	22	484	0	0.5833	0.34	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
9	155.50	23605	97303263400	8	22	484	0	22.0000	484.00	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0

OBS	NURAD	ROOMAC	CENTAC	NUUT	NOPRIVCY	B1	B2	B3	POOR	FURNINC	PARKINC	HEATINC	NHUINC	QHEAT	LLBLG	BADHALL	BLACK	SPAN	CRLNTLN	CROHDS	ABANDON	EXCELN	GOODN	POORN	LITTER	NOSHOPS	DEFECT	BCOUNTY	YHAT	RESID
1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3.21888	0.40	0	0	0	0	0	0	0	0	4.21540	-0.9965
2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2.99573	0.17	0	0	1	0	0	0	0	0	4.06912	-1.0734
3	1	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	1	3.49651	0.25	0	0	1	0	0	0	3	0	4.71688	-1.2204
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.99573	0.50	1	0	1	0	0	0	0	0	4.05223	-1.0565
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2.99573	0.29	0	0	0	0	0	0	0	0	4.50297	-1.5072
6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3.91202	0.29	0	0	1	0	1	0	3	0	4.89278	-0.9808
7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.41120	0.60	1	1	0	0	0	0	1	0	4.50330	-1.1021
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.91202	0.75	0	0	1	0	0	1	0	0	5.02080	-1.1088
9	0	1	0	0	0	0	0	0	0	0	0	1	0	8	0	0	0	0	3.66888	0.60	0	0	1	0	0	0	0	0	4.95877	-1.2699

Exhibit 29

OBS	XWEIGHT	XTFI	XCONTROL	Q	AGE1	AGE1SQ	DAGE	CLOT	CLOTSQ	DLOT	CC1	SFDET	SFATT	DUPLEX	NGT50	ELEVP	R1	R3	RG4	BED0	BED2	BED3	BEDG4
1	159.57	26000	97303520700	7	56	3136	1	0.416667	0.173611	0	0	1	0	0	0	0	0	0	0	0	1	0	0

OBS	RHEAT	NORAD	ROOMAC	CENTAC	NOUT	NOPRIVCY	B1	B2	B3	POUR	FURNINC	PARKINC	HEATINC	NHUINC	QHEAT	LLBLG	BADHALL	BLACK	SPAN	CRENTLN	CROWDS	ABANDON	EXCELN	GOODN	POORN	LITTER	NOSHOPS	DEFECT	BCOUNTY	YHAT	RESID
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.94017	0.25	1	0	1	0	0	0	1	0	4.8899	1.05028

Exhibit 30

Distribution of Outliers by Dependent Variable
Baltimore Renters[1]

Log of Rent (CRENTLN)[1]	Sample Frequency	Sample Percent	Cumulative Sample Percent	Negative Outlier Frequency	Negative Outlier Percent	Positive Outlier Frequency	Positive Outlier Percent
2.996	3	0.238	0.238	3	33.33	0	0
3.219	1	0.079	0.318	1	11.11	0	0
3.401	4	0.318	0.636	1	11.11	0	0
3.497	1	0.079	0.715	1	11.11	0	0
3.689	2	0.159	0.874	1	11.11	0	0
3.912	9	0.715	2.067	2	22.22	0	0
5.940	2	0.159	99.205	0	0	1	100.0

1. Not all possible values of the log of rent are represented. We list only the log of rents which produced outliers. For example, over 97 percent of the log of rent distribution takes on values between 3.912 and 5.940.

Exhibit 31

Distribution of Outliers by Regressors
Baltimore Renters

Variable	Value	Sample Frequency	Sample Percent	Negative Outlier Frequency	Negative Outlier Percent	Positive Outlier Frequency	Positive Outlier Percent
DAGE	0	679	53.98	2	22.22	0	0.0
	1	579	46.02	7	77.78	1	100.0
CC1	0	603	47.93	2	22.22	1	100.0
	1	655	52.07	7	77.78	0	0.0
SFATT	0	917	72.89	4	44.44	1	100.0
	1	341	27.11	5	55.56	0	0.0

Exhibit 32

Baltimore Renters

	Model A	Model B	Model C
SSE	79.000	77.0739	63.2838
DFE	1209	1206	1196
MSE	0.0653	0.0639	0.0529
F Ratio	47.15	45.96	51.34
R Square	0.6518	0.6603	0.6865
ESTIMATED COEFFICIENTS[1]			
Intercept	4.9617	4.9526	4.9866
	(93.31)	(91.84)	(101.32)
B1	0.0896	0.0866	0.0853
	(3.31)	(3.23)	(3.48)
B2	0.2123	0.2057	0.1971
	(7.16)	(7.01)	(7.37)
B3	0.3699	0.3257	0.3101
	(4.80)	(4.25)	(4.44)
R1	-0.1308	-0.1347	-0.1138
	(-3.69)	(-3.83)	(-3.52)
R3	0.0677	0.0772	0.0796
	(3.83)	(4.38)	(4.94)
RG4	0.0317	0.0348	0.0350
	(4.24)	(4.67)	(5.14)
BEDO	-0.4173	-0.4106	-0.4171
	(-5.62)	(-5.58)	(-6.23)
BED2	0.0989	0.0958	0.0916
	(5.05)	(4.94)	(5.18)
BED3	0.1932	0.2044	0.2121
	(6.54)	(6.93)	(7.87)
BEDG4	0.0286	0.0363	0.0454
	(2.62)	(3.32)	(4.50)
ELEVP	0.2639	0.2559	0.2496
	(3.45)	(3.38)	(3.62)
SFATT	-0.1147	-0.0392	-0.0492
	(-4.26)	(-1.15)	(-1.58)
SFDET	0.0019	-0.0265	-0.0234
	(0.05)	(-0.71)	(-0.68)
DUPLEX	-0.0457	-0.0630	-0.0589
	(-1.76)	(-2.43)	(-2.49)
NGT50	0.2062	0.1917	0.1851
	(2.48)	(2.33)	(2.47)
AGE1	-0.0024	-0.0047	-0.0021
	(-0.55)	(-1.09)	(-0.55)

1. T-statistics appear in parenthesis.

Exhibit 32 (cont'd)

Baltimore Renters

	Model A	Model B	Model C
AGE1SQ	-2.13×10^{-5} (-0.17)	3.41×10^{-5} (0.28)	-4.47×10^{-5} (-0.42)
DAGE	0.0488 (0.28)	0.0838 (0.49)	0.1797 (1.14)
RHEAT	-0.2196 (-5.11)	-0.2325 (-5.43)	-0.2401 (-6.08)
ROOMAC	0.0583 (2.94)	0.0509 (2.58)	0.0361 (2.00)
CENTAC	0.2090 (7.46)	0.2094 (7.44)	0.1943 (7.57)
NORAD	-0.1042 (-4.12)	-0.0948 (-3.77)	-0.0663 (-2.87)
POOR	-0.2431 (-6.27)	-0.2522 (-6.55)	-0.2711 (-7.72)
NOPRIVCY	-0.1153 (-4.95)	-0.1081 (-4.69)	-0.1035 (-4.88)
NOUT	0.0240 (0.67)	0.0366 (1.02)	0.0042 (0.13)
BADHALL	0.0048 (0.19)	0.0072 (0.28)	0.0107 (0.47)
DFECT	-0.0043 (-0.53)	-0.0011 (-0.13)	-0.0050 (-0.67)
CLOT	-0.0248 (-5.91)	-0.0226 (-5.39)	-0.0259 (-6.75)
CLOTSQ	3.76×10^{-4} (1.71)	3.28×10^{-4} (1.51)	5.78×10^{-4} (2.88)
DLOT	-0.2054 (-1.55)	-0.2148 (-1.64)	-0.1890 (-1.53)
CROWDS	0.0739 (2.45)	0.0706 (2.37)	0.0519 (1.90)
BLACK	-0.0049 (-0.25)	-0.0018 (-0.09)	-0.0242 (-1.36)
SPAN	-0.0941 (-1.08)	-0.0828 (-0.96)	0.0752 (0.89)
LLBLG	-0.0715 (-2.43)	-0.691 (-2.37)	-0.0865 (-3.26)
NHUINC	0.0513 (1.32)	0.0323 (0.84)	0.0297 (0.84)
HEATINC	0.1069 (3.40)	0.0983 (3.16)	0.0867 (3.05)
PARKINC	0.0093 (0.06)	0.0216 (0.14)	7.25×10^{-4} (0.01)

Exhibit 32 (cont'd)

Baltimore Renters

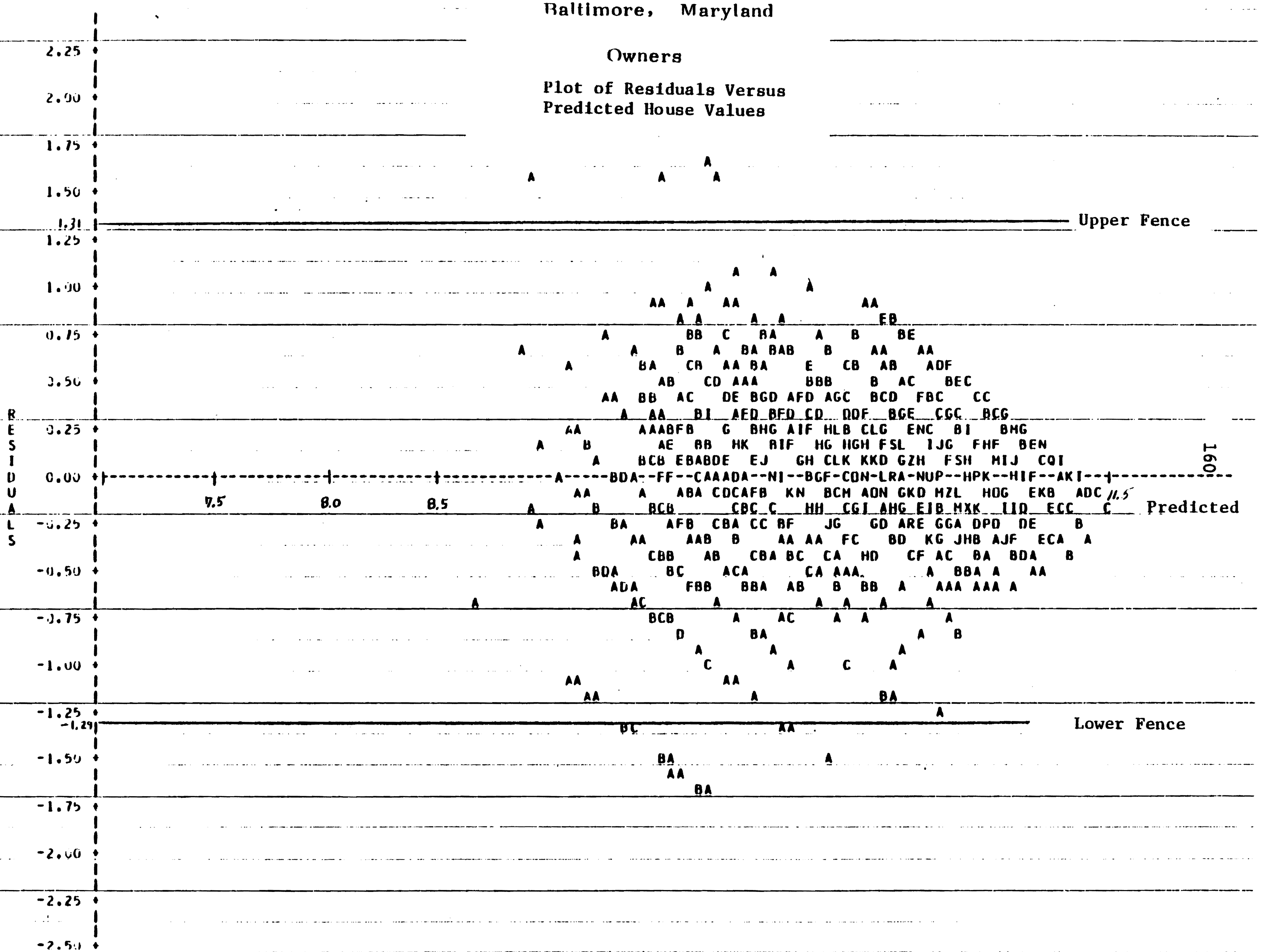
	Model A	Model B	Model C
FURNINC	0.0301	0.0327	0.0061
	(0.84)	(0.91)	(0.19)
EXCELN	0.0363	0.0328	0.0283
	(1.55)	(1.41)	(1.33)
GOODN	-0.0067	-0.0111	-0.0105
	(-0.34)	(-0.57)	(-0.59)
POORN	-0.0187	-0.0132	-0.0176
	(-0.49)	(-0.35)	(-0.51)
ABANDON	-0.0677	-0.0583	-0.0666
	(-2.71)	(-2.35)	(-2.92)
LITTER	-0.0097	-0.0091	-0.0168
	(-0.49)	(-0.46)	(-0.94)
NOSHOPS	-0.0416	-0.0501	-0.0484
	(-1.61)	(-1.98)	(-2.06)
Q	-0.0071	-0.0066	0.0057
	(2.55)	(2.37)	(2.23)
QHEAT	-0.0038	-0.0035	-8.75×10^{-4}
	(-0.85)	(-0.79)	(-0.21)
CC1	-0.0717	-0.0084	-0.0162
	(-2.86)	(-0.28)	(-0.59)
BCOUNTY	-0.0125	-0.0118	-0.0202
	(-0.57)	(-0.54)	(-1.01)
CC1DAGE		-0.1288	-0.0973
		(-3.50)	(-2.93)
CC1SFATT		-0.0662	-0.0453
		(-1.51)	(-1.13)
DAGESFAT		-0.0931	-0.0912
		(-2.20)	(-2.35)

Exhibit 33

Baltimore, Maryland

Owners

Plot of Residuals Versus
Predicted House Values



160

Exhibit 34

BALTIMORE OWNER'S NEGATIVE OUTLIERS

OBS	XWEIGHT	XTFI	XCUNTROL	Q	AGE1	AGESQ	AGE1CB	DAGE	CLOT	CLOTSQ	DLOT	CC1	SEATT	R12	R4	R5	RG6	BED1	BED2
1	155.26	3840	97300150600	0	56	3136	175616	1	31.3333	981.78	1	1	0	0	0	0	6	0	0
2	153.52	13284	97301234200	7	56	3136	175616	1	31.9167	1018.67	1	1	1	0	0	0	6	0	0
3	175.54	2167	97403375700	4	56	3136	175616	1	14.1250	199.52	0	1	1	0	1	0	0	0	0
4	155.07	12079	97300586100	10	56	3136	175616	1	32.1667	1034.69	1	1	1	0	0	0	0	0	0
5	154.80	5200	97300644800	5	56	3136	175616	1	0.2500	0.06	0	1	1	0	0	0	0	0	1
6	156.52	2480	97300776800	6	56	3136	175616	1	21.8333	476.69	0	1	1	0	0	1	0	0	0
7	156.52	6072	97300841000	9	56	3136	175616	1	9.4167	88.67	0	1	1	0	1	0	0	0	1
8	155.52	21162	97300863100	3	56	3136	175616	1	31.5833	997.51	1	1	1	0	0	1	0	0	0
9	157.33	6125	97300882400	11	56	3136	175616	1	32.2500	1040.06	1	1	1	0	0	0	0	0	0
10	155.63	15000	97301393300	0	56	3136	175616	1	21.3333	455.11	0	1	1	0	1	0	0	0	0
11	155.18	6175	97301443400	3	56	3136	175616	1	31.5833	997.51	1	1	1	0	0	0	0	0	0
12	164.56	12000	97301827800	3	56	3136	175616	1	8.9167	79.51	0	1	1	0	1	0	0	0	0
13	154.56	9513	97301823600	3	56	3136	175616	1	8.9167	79.51	0	1	1	1	0	0	0	0	0
14	149.38	3033	97302049900	3	56	3136	175616	1	31.5833	997.51	1	0	0	1	0	0	0	1	0
15	153.80	6196	97302673500	10	56	3136	175616	1	32.1667	1034.69	1	0	0	0	0	0	0	0	1
16	153.57	8340	97303428300	5	32	1024	32768	0	14.2083	201.88	0	0	1	0	0	0	0	0	0

OBS	BED4	BEDG5	SHEAT	EHEAT	KHLAT	NURAD	ROOMAC	CENTAC	NOUT	NOPRIVCY	B1	B2	B3	POOR	COOKE	BLACK	SPAN	CROWDS	GAR	BASE	ABANDON	EXCELN	GOODN	POORN	VALUELN	FORAY	BCOUNTY	YHAT	RESID
1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0.50	0	1	0	0	0	0	8.74034	0	0	10.2405	-1.5002
2	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0.22	0	1	0	0	1	0	8.00637	7	0	9.6899	-1.6835
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0	1	0	0	0	0	8.00637	4	0	9.6709	-1.6645
4	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0.33	1	1	0	0	0	0	8.00637	10	0	9.3748	-1.3685
5	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0.60	0	1	0	0	1	0	8.00637	5	0	9.5065	-1.5001
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0.38	0	1	1	0	1	0	8.00637	6	0	9.5718	-1.5654
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.35	0	1	0	0	0	0	8.00637	9	0	9.5597	-1.5533
8	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0.50	0	1	1	0	0	0	8.00637	3	0	9.3209	-1.3146
9	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.17	0	1	0	0	0	0	8.00637	11	0	9.5098	-1.5034
10	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.29	0	1	0	0	0	1	8.00637	0	0	9.5302	-1.5238
11	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0.33	0	1	0	0	0	0	8.00637	3	0	9.3562	-1.3499
12	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.57	0	1	0	0	0	1	8.00637	3	0	9.3220	-1.3156
13	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0.20	0	1	1	0	0	0	8.00637	3	0	9.3677	-1.3613
14	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.33	1	1	0	0	1	0	8.74034	0	0	10.0827	-1.3424
15	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0.40	1	0	0	0	1	0	8.00637	0	1	9.6594	-1.6530
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.33	0	1	0	0	0	0	8.74034	0	0	10.0534	-1.3130

Exhibit 35

BALTIMORE OWNER'S POSITIVE OUTLIERS

OBS	XWEIGHT	XTEI	XCONTROL	Q	AGE1	AGE1SQ	AGE1CB	DAGE	CLOT	CLOTSQ	DLOT	CC1	SFAT1	R12	R4	R5	RG6	BEO1	BED2
1	156.59	3276	97300021900	2	56	3136	175616	1	31.5000	992.25	1	1	1	0	0	0	0	0	0
2	157.88	11180	97300972300	4	56	3136	175616	1	31.6667	1002.78	1	1	1	1	0	0	0	0	1
3	180.18	10000	97301445700	0	56	3136	175616	1	13.7917	190.21	0	1	1	1	0	0	0	0	0
4	160.13	3802	97304331800	4	22	484	10648	0	21.6667	469.44	0	0	0	0	0	0	0	1	0

OBS	BED4	BEDG5	SHEAT	EHEAT	KHEAT	NURAD	ROOMAC	CENTAC	NOUT	NOPRIVCY	B1	B2	B3	POOR	COOKE	BLACK	SPAN	CROWDS	GAR	BASE	ABANDON	EXCELN	GOODN	POORN	VALUEN	FORAY	BCOUNTY	YHAT	RESID
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	1	0	1	0	0	11.3469	2	0	9.69403	1.65284
2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0.25	0	1	1	0	0	0	10.5321	4	0	8.92591	1.60619
3	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0.50	0	1	1	0	1	0	11.1199	0	0	9.50537	1.61452
4	0	0	0	0	1	1	0	0	0	0	0	0	0	3	1	0	0	0.50	0	0	0	1	0	0	11.3469	0	0	9.76957	1.57730

Exhibit 36

Distribution of Outliers by Dependent Variable
Baltimore Owners[1]

Log of House Value (VALUELN)	Sample Frequency	Sample Percent	Cumulative Sample Percent	Negative Outlier Frequency	Negative Outlier Percent	Positive Outlier Frequency	Positive Outlier Percent
8.006	18	.90	.90	13	81.25	0	-
8.740	42	2.10	3.00	3	18.75	0	-
10.532	210	10.48	55.52	0	-	1	25.00
11.120	156	7.79	91.86	0	-	1	25.00
11.347	163	8.13	100.00	0	-	2	50.00

1. Not all possible values of the log of house value are represented. We list only the log of house values which produce outliers.

Exhibit 37

Distribution of Outliers by Regressors
Baltimore Owners

Variable	Value	Sample Frequency	Sample Percent	Negative Outlier Frequency	Negative Outlier Percent	Positive Outlier Frequency	Positive Outlier Percent
DAGE	0	1446	72.19	1	6.25	1	25.00
	1	577	27.81	15	93.75	3	75.00
CC1	0	1430	71.39	3	18.75	1	25.00
	1	573	28.61	13	81.25	3	75.00
SFATT	0	1332	66.50	3	18.75	1	25.00
	1	671	33.50	13	81.25	3	75.00

Exhibit 38

Baltimore Owners

	Model A	Model B	Model C
SSE	246.6229	236.8306	189.0087
DFE	1960	1957	1936
MSE	0.1258	0.1210	0.0976
F Ratio	103.64	102.38	118.63
R Square	0.6895	0.7019	0.7339
ESTIMATED COEFFICIENTS[1]			
Intercept	10.3657	10.3735	10.4248
	(161.08)	(164.28)	(183.10)
B1	0.0996	0.0928	0.0912
	(4.62)	(4.38)	(4.77)
B2	0.1359	0.1414	0.1431
	(4.95)	(5.25)	(5.89)
B3	0.2652	0.2682	0.2727
	(8.28)	(8.54)	(9.63)
R12	-0.1406	-0.1323	-0.1383
	(-5.60)	(-5.37)	(-6.20)
R4	0.0650	0.0666	0.0723
	(3.19)	(3.32)	(3.99)
R5	0.0765	0.0735	0.0889
	(2.48)	(2.43)	(3.24)
RG6	0.0288	0.0284	0.0341
	(4.32)	(4.35)	(5.76)
BED1	-0.1491	-0.1575	-0.1851
	(-2.68)	(-2.88)	(-3.69)
BED2	-0.1152	-0.1228	-0.1282
	(-5.17)	(-5.61)	(-6.48)
BED4	0.0585	0.0559	0.0482
	(2.35)	(2.29)	(2.20)
BEDG5	0.0204	0.0168	0.0174
	(2.36)	(1.97)	(2.27)
GAR	0.0980	0.0912	0.0905
	(5.26)	(4.99)	(5.49)
BASE	0.0407	0.0340	0.0352
	(1.69)	(1.44)	(1.66)
SFATT	-0.3348	-0.2516	-0.2533
	(-14.70)	(-9.05)	(-10.12)
AGE1	-0.0054	-0.0021	-0.0039
	(-0.47)	(-0.17)	(-0.38)
AGE1SQ	4.80×10^{-4}	1.09×10^{-4}	1.61×10^{-4}
	(0.61)	(0.14)	(0.23)

1. T-statistics appear in parenthesis.

Exhibit 38 (cont'd)

Baltimore Owners

	Model A	Model B	Model C
AGE1CB	-1.01×10^{-5}	-3.26×10^{-6}	-3.11×10^{-6}
	(-0.69)	(-0.23)	(-0.24)
DAGE	0.3573	0.2509	0.1867
	(0.48)	(0.34)	(0.28)
SHEAT	0.0926	0.0965	0.0798
	(4.55)	(4.83)	(4.43)
RHEAT	-0.0935	-0.0818	-0.0892
	(-1.87)	(-1.67)	(-1.97)
EHEAT	-0.0598	-0.0554	-0.0601
	(-1.19)	(-1.12)	(-1.35)
ROOMAC	0.0258	0.0254	0.0352
	(1.30)	(1.30)	(2.00)
CENTAC	0.1130	0.1075	0.1061
	(4.29)	(4.16)	(4.56)
NORAD	-0.0841	-0.0705	-0.0961
	(-3.17)	(-2.70)	(-4.07)
POOR	-0.1798	-0.2085	-0.2872
	(-3.59)	(-4.23)	(-5.94)
NOPRIVCY	-0.1287	-0.0927	-0.0901
	(-2.91)	(-2.13)	(-2.27)
NOUT	-0.2035	-0.1877	-0.1445
	(-3.69)	(-3.48)	(-2.92)
COOKE	0.1078	0.1170	0.1034
	(5.25)	(5.80)	(5.68)
CLOT	-0.0082	-0.0074	-0.0062
	(-1.55)	(-1.43)	(-1.33)
CLOTSQ	1.79×10^{-5}	1.74×10^{-4}	1.26×10^{-4}
	(0.83)	(0.83)	(0.66)
DLOT	-0.0979	-0.1058	-0.0778
	(-1.21)	(-1.33)	(-1.08)
CROWDS	0.0226	0.0063	-0.0013
	(0.62)	(0.17)	(-0.04)
BLACK	-0.1784	-0.1614	-0.1793
	(-6.44)	(-5.90)	(-7.23)
SPAN	-0.0158	0.0024	-0.0142
	(-0.11)	(0.02)	(-0.11)
EXCELN	0.2056	0.1897	0.1485
	(7.49)	(7.02)	(6.05)
GOODN	0.1131	0.0935	0.0616
	(4.37)	(3.67)	(2.65)
POORN	-0.1920	-0.1776	-0.1684
	(-2.64)	(-2.49)	(-2.58)
ABANDON	-0.0578	-0.0246	-0.0370
	(-1.37)	(-0.59)	(-0.97)

Exhibit 38 (cont'd)

Baltimore Owners

	Model A	Model B	Model C
Q	0.0087	0.0085	0.0090
	(3.24)	(3.26)	(3.80)
CC1	-0.3012	-0.1461	-0.1654
	(-7.12)	(-2.60)	(-3.27)
FORAY	8.27×10^{-4}	0.0017	0.0023
	(0.16)	(0.33)	(0.51)
BCOUNTY	-0.026	-0.0301	-0.0355
	(-1.21)	(-1.41)	(-1.84)
CC1DAGE		-0.2209	-0.2218
		(-4.02)	(-4.49)
CC1SFATT		-0.1062	-0.0941
		(-2.04)	(-2.01)
DAGESFAT		-0.1533	-0.1501
		(-2.78)	(-3.03)

Exhibit 39

Anaheim Renters

	Model A	Model B	Model C
SSE	57.999	57.858	44.417
DFE	1322	1321	1308
MSE	0.0438	0.0438	0.0339
F Ratio	40.20	39.50	48.97
R Square	0.5884	0.5894	0.6425
ESTIMATED COEFFICIENTS[1]			
Intercept	5.168	5.1705	5.1628
	(136.89)	(136.99)	(154.93)
B1	0.0396	0.0395	0.0426
	(1.64)	(1.64)	(2.01)
B2	0.0798	0.0814	0.0838
	(4.43)	(4.52)	(5.25)
B3	0.2926	0.2543	0.2822
	(6.35)	(5.02)	(6.06)
R1	-0.0061	-0.0060	-0.0047
	(-0.30)	(-0.29)	(-0.27)
R3	0.0499	0.0492	0.0344
	(2.83)	(2.79)	(2.19)
RG4	0.0211	0.0207	0.0271
	(1.84)	(1.80)	(2.61)
BEDO	-0.1263	-0.1264	-0.1351
	(-3.02)	(-3.02)	(-3.66)
BED2	0.1617	0.1612	0.1577
	(10.05)	(10.03)	(11.12)
BED3	0.3393	0.3423	0.3401
	(12.85)	(12.95)	(14.54)
BEDG4	0.1045	0.0982	0.1026
	(9.48)	(8.49)	(9.92)
ELEVP	-0.1082	-0.1078	-0.1007
	(-1.22)	(-1.21)	(-1.29)
SFATT	-0.0623	-0.0621	-0.0489
	(-2.12)	(-2.11)	(-1.86)
SFDET	0.0173	0.0159	0.0403
	(0.83)	(0.76)	(2.18)
DUPLEX	0.0143	0.0119	0.0218
	(0.58)	(0.48)	(1.00)

1. T-statistics appear in parenthesis.

Exhibit 39 (cont'd)

Anaheim Renters

	Model A	Model B	Model C
NGT50	0.0324	0.0330	0.0291
	(1.19)	(1.22)	(1.22)
AGE1	−0.0181	−0.0180	−0.0172
	(−7.08)	(−7.07)	(−7.59)
AGE1SQ	4.06×10^{-4}	4.08×10^{-4}	3.55×10^{-4}
	(4.78)	(4.81)	(4.74)
DAGE	−0.4726	−0.4764	−0.3958
	(−3.49)	(−3.53)	(−3.31)
RHEAT	−0.0753	−0.0751	−0.0765
	(−3.74)	(−3.74)	(−4.30)
ROOMAC	−0.0227	−0.0221	−0.0147
	(−1.59)	(−1.55)	(−1.17)
CENTAC	0.0311	0.0337	0.0398
	(1.38)	(1.50)	(2.01)
NORAD	0.0129	0.0121	0.0039
	(0.67)	(0.64)	(0.23)
POOR	−0.3261	−0.3254	−0.3151
	(−7.71)	(−7.70)	(−8.46)
NOPRIVCY	−0.0524	−0.0523	−0.0393
	(−2.32)	(−2.32)	(−1.97)
NOUT	−0.0436	−0.0416	−0.0475
	(−0.81)	(−0.77)	(−0.99)
BADHALL	−0.0075	−0.0085	−0.0121
	(−0.31)	(−0.35)	(−0.57)
DFECT	−0.0169	−0.0175	−0.0177
	(−1.36)	(−1.41)	(−1.62)
CLOT	−0.0330	−0.0331	−0.0331
	(−7.11)	(−7.13)	(−8.06)
CLOTSQ	7.97×10^{-4}	7.95×10^{-4}	9.45×10^{-4}
	(2.39)	(2.39)	(3.22)
DLOT	−0.2966	−0.2941	−0.4252
	(−1.25)	(−1.24)	(−2.04)
CROWDS	−0.0083	−0.0079	0.0034
	(−0.39)	(−0.37)	(0.18)
BLACK	0.0053	0.0048	0.0065
	(0.11)	(0.11)	(0.15)
SPAN	−0.0913	−0.0927	−0.0985
	(−4.54)	(−4.61)	(−5.56)
LLBLG	0.0094	0.0083	0.0119
	(0.42)	(0.37)	(0.60)
NHUINC	0.0816	0.0823	0.0353
	(2.53)	(2.56)	(1.21)
HEATINC	0.0496	0.0464	0.0501
	(1.95)	(1.82)	(2.22)

Exhibit 39 (cont'd)

Anaheim Renters

	Model A	Model B	Model C
PARKINC	-0.0682	-0.0658	-0.0758
	(-0.91)	(-0.88)	(-1.15)
FURNINC	0.0614	0.0613	0.0759
	(3.28)	(3.27)	(4.58)
EXCELN	0.0729	0.0721	0.0596
	(4.04)	(3.99)	(3.73)
GOODN	0.0203	0.0195	0.0156
	(1.31)	(1.26)	(1.14)
POORN	0.0044	0.0045	0.0020
	(0.14)	(0.14)	(0.07)
ABANDON	-0.0404	-0.0415	-0.0481
	(-0.92)	(-0.95)	(-1.25)
LITTER	9.97×10^{-4}	0.0019	-2.1×10^{-4}
	(0.05)	(0.10)	(-0.01)
NOSHOPS	-0.0126	-0.0141	-0.0137
	(-0.53)	(-0.59)	(-0.65)
Q	0.0067	0.0065	0.0065
	(3.32)	(3.21)	(3.60)
QHEAT	-0.0055	-0.0051	-0.0040
	(-1.51)	(-1.40)	(-1.26)
CC1	-0.0548	-0.0550	-0.0507
	(-4.29)	(-4.31)	(-4.50)
B3BEDG4		0.1921	0.1984
		(1.80)	(1.82)

Exhibit 40

Chicago Renters

	Model A	Model B	Model C
SSE	264.5049	264.0001	211.099
DFE	4133	4132	4107
MSE	0.06399	0.0638	0.0514
F Ratio	131.16	128.96	158.76
R Square	0.6134	0.6141	0.6635
ESTIMATED COEFFICIENTS[1]			
Intercept	5.1036	5.1086	5.1027
	(192.32)	(192.22)	(213.34)
B1	0.1282	0.1292	0.1304
	(5.637)	(5.68)	(6.35)
B2	0.2354	0.2378	0.2624
	(11.69)	(11.80)	(14.36)
B3	0.4522	0.4519	0.4522
	(8.24)	(8.25)	(9.19)
R1	-0.1332	-0.1330	-0.1292
	(-7.90)	(-7.89)	(-8.53)
R3	0.0976	0.0976	0.0961
	(10.20)	(10.21)	(11.16)
RG4	0.0473	0.0471	0.0422
	(9.73)	(9.70)	(9.62)
BEDO	-0.1945	-0.1941	-0.1890
	(-9.84)	(-9.83)	(-10.65)
BED2	0.0951	0.0945	0.0942
	(9.70)	(9.63)	(10.68)
BED3	0.1622	0.1613	0.1628
	(11.23)	(11.18)	(12.52)
BEDG4	0.0324	0.0319	0.0377
	(4.61)	(4.54)	(5.92)
ELEVP	0.1423	0.1079	0.1019
	(7.50)	(4.78)	(5.03)
SFATT	-0.0565	-0.0584	-0.0669
	(-1.87)	(-1.928)	(-2.46)
SFDET	-0.1346	-0.1361	-0.1479
	(-7.04)	(-7.12)	(-8.60)
DUPLEX	-0.1164	-0.1173	-0.1189
	(-9.79)	(-9.87)	(-11.14)
NGT50	0.1035	0.0352	0.0355
	(5.24)	(1.12)	(1.24)
AGE1	0.0045	0.0041	0.0053
	(1.61)	(1.44)	(2.11)
AGE1SQ	-2.42×10^{-4}	-2.30×10^{-4}	-2.55×10^{-4}
	(-2.78)	(-2.65)	(-3.26)

1. T-statistics appear in parenthesis.

Exhibit 40 (cont'd)

Chicago Renters

	Model A	Model B	Model C
DAGE	0.3367	0.3240	0.3317
	(2.54)	(2.45)	(2.78)
RHEAT	−0.2793	−0.2799	−0.2827
	(−18.18)	(−18.23)	(−20.42)
ROOMAC	0.0721	0.0718	0.0699
	(7.51)	(7.48)	(8.10)
CENTAC	0.2104	0.2099	0.2089
	(11.50)	(11.49)	(12.69)
NORAD	−0.0723	−0.0721	−0.0611
	(−5.92)	(−5.91)	(−5.56)
POOR	−0.2528	−0.2518	−0.2403
	(−11.23)	(−11.19)	(−11.84)
NOPRIVCY	−0.0262	−0.0252	−0.0269
	(−1.49)	(−1.44)	(−1.71)
NOUT	−0.0910	−0.0918	−0.0935
	(−3.99)	(−4.04)	(−4.58)
BADHALL	−0.0237	−0.0229	−0.0172
	(−2.88)	(−2.79)	(−2.32)
DFECT	0.0115	0.0119	0.0081
	(2.53)	(2.61)	(1.96)
CLOT	−0.0244	−0.0244	−0.0235
	(−10.45)	(−10.43)	(−11.15)
CLOTSQ	6.53×10^{-4}	6.50×10^{-4}	6.27×10^{-4}
	(5.32)	(5.29)	(5.67)
DLOT	−0.1595	−0.1578	−0.1647
	(−2.51)	(−2.48)	(−2.88)
CROWDS	0.0406	0.0415	0.0425
	(2.86)	(2.93)	(3.34)
BLACK	−0.0655	−0.0655	−0.0633
	(−5.65)	(−5.66)	(−6.08)
SPAN	−0.0847	−0.0846	−0.0901
	(−5.37)	(−5.37)	(−6.38)
LLBLG	−0.0666	−0.0670	−0.0690
	(−6.35)	(−6.39)	(−7.32)
NHUINC	0.0287	0.0282	0.0295
	(2.02)	(1.98)	(2.31)
HEATINC	0.0263	0.0256	0.0292
	(1.66)	(1.627)	(2.06)
PARKINC	0.0345	0.0346	0.0209
	(1.62)	(1.62)	(1.09)
FURNINC	-5.19×10^{-4}	−0.0013	−0.0070
	(−0.03)	(−0.08)	(−0.44)

Exhibit 40 (cont'd)

Chicago Renters

	Model A	Model B	Model C
EXCELN	0.0872	0.0871	0.0924
	(6.95)	(6.95)	(8.19)
GOODN	0.0221	0.0226	0.0265
	(2.11)	(2.15)	(2.82)
POORN	-0.0449	-0.0430	-0.0266
	(-2.34)	(-2.24)	(-1.53)
ABANDON	-0.0522	-0.0527	-0.0467
	(-3.81)	(-3.84)	(-3.79)
LITTER	0.0104	0.0094	0.0072
	(0.93)	(0.84)	(0.71)
NOSHOPS	0.0066	0.0061	0.0050
	(0.52)	(0.49)	(0.45)
Q	0.0032	0.0032	0.0036
	(2.24)	(2.19)	(2.75)
QHEAT	7.80×10^{-4}	0.0010	-8.50×10^{-4}
	(0.32)	(0.45)	(-0.39)
CC1	-0.0242	-0.0263	-0.0264
	(-2.06)	(-2.23)	(-2.50)
DUPAGE	0.0184	0.0217	0.0356
	(0.89)	(1.06)	(1.93)
KANE	-0.0511	-0.0527	-0.0527
	(-2.19)	(-2.26)	(-2.52)
LAKE	-0.0033	-0.0053	-0.0055
	(-0.13)	(-0.21)	(-0.25)
NGT5ELEV		0.1119	0.1189
		(2.81)	(3.29)

Exhibit 41

Fort Worth Owners

	Model A	Model B	Model C
SSE	268.3847	268.2120	227.1805
DFE	2411	2410	2395
MSE	0.1113	0.1112	0.0949
F Ratio	141.69	138.38	162.76
R Square	0.7067	0.7069	0.7405
ESTIMATED COEFFICIENTS[1]			
Intercept	9.6750	9.6770	9.6286
	(150.90)	(150.90)	(162.25)
B1	0.0886	0.0873	0.0857
	(3.12)	(3.07)	(3.25)
B2	0.1961	0.1954	0.1981
	(8.27)	(8.24)	(9.02)
B3	0.4958	0.4950	0.4971
	(12.43)	(12.40)	(13.47)
R12	-0.1336	-0.1338	-0.1302
	(-9.16)	(-9.18)	(-9.60)
R4	0.1144	0.1147	0.1199
	(5.24)	(5.26)	(5.95)
R5	0.1805	0.1810	0.1905
	(4.41)	(4.41)	(5.04)
RG6	0.0443	0.0438	0.0445
	(4.17)	(4.12)	(4.54)
BED1	-0.2070	-0.2033	-0.1807
	(-4.14)	(-4.06)	(-3.85)
BED2	-0.1055	-0.1054	-0.1101
	(-5.38)	(-5.38)	(-6.04)
BED4	0.1437	0.1440	0.1372
	(6.08)	(6.10)	(6.29)
BEDG5	0.0322	0.0319	0.0305
	(2.66)	(2.64)	(2.73)
GAR	0.0652	0.0647	0.0745
	(3.19)	(3.17)	(3.92)
BASE	0.2090	0.2071	0.2100
	(2.88)	(2.86)	(3.14)
SFATT	0.1603	0.1591	0.1608
	(1.93)	(1.92)	(2.10)
AGE1	-0.0123	-0.0127	-0.0122
	(-1.42)	(-1.46)	(-1.53)
AGE1SQ	1.05×10^{-4}	1.53×10^{-4}	1.11×10^{-4}
	(0.17)	(0.24)	(0.19)

1. T-statistics appear in parenthesis.

Exhibit 41 (cont'd)

Fort Worth Owners

	Model A	Model B	Model C
AGE1CB	2.90×10^{-6}	1.28×10^{-6}	3.06×10^{-6}
	(0.22)	(0.09)	(0.25)
DAGE	-0.3263	-0.1800	-0.3609
	(-0.51)	(-0.28)	(-0.60)
SHEAT	-0.5091	-0.5441	-0.4668
	(-1.49)	(-1.59)	(-1.47)
RHEAT	-0.1489	-0.1480	-0.1312
	(-4.48)	(-4.46)	(-4.26)
EHEAT	0.0102	0.0099	0.0240
	(0.12)	(0.11)	(0.27)
ROOMAC	0.0629	0.0627	0.0852
	(2.49)	(2.49)	(3.64)
CENTAC	0.1694	0.1691	0.2001
	(5.02)	(5.01)	(6.40)
NORAD	0.0360	0.0360	0.0358
	(1.19)	(1.19)	(1.29)
POOR	-0.3269	-0.3271	-0.3100
	(-11.10)	(-11.11)	(-11.23)
NOPRIVCY	-0.0976	-0.0977	-0.1127
	(-3.23)	(-3.23)	(-3.99)
NOUT	-0.0655	-0.0674	-0.1570
	(-1.09)	(-1.12)	(-2.66)
COOKE	0.1059	0.1064	0.1074
	(6.21)	(6.24)	(6.82)
CLOT	-0.0049	-0.0049	-0.0040
	(-1.12)	(-1.11)	(-0.95)
CLOTSQ	2.38×10^{-4}	2.38×10^{-4}	1.92×10^{-4}
	(1.16)	(1.15)	(1.01)
DLOT	-0.0259	0.0151	-0.0450
	(-0.34)	(0.18)	(-0.58)
CROWDS	-0.0485	-0.0490	-0.0496
	(-1.78)	(-1.79)	(-1.96)
BLACK	-0.1703	-0.1711	-0.1621
	(-6.47)	(-6.50)	(-6.60)
SPAN	-0.0995	-0.0996	-0.0961
	(-2.62)	(-2.62)	(-2.70)
EXCELN	0.1863	0.1865	0.1816
	(8.34)	(8.35)	(8.77)
GOODN	0.1288	0.1287	0.1139
	(6.03)	(6.02)	(5.75)
POORN	-0.0267	-0.0264	-0.0219
	(-0.48)	(-0.48)	(-0.43)
ABANDON	-0.0980	-0.0971	-0.0989
	(-2.67)	(-2.64)	(-2.90)

Exhibit 41 (cont'd)

Fort Worth Owners

	Model A	Model B	Model C
Q	0.0089	0.0089	0.0099
	(3.30)	(3.29)	(4.01)
CC1	-0.0056	-0.0049	0.0095
	(-0.21)	(-0.19)	(0.39)
FORAY	-0.0025	-0.0027	-0.0047
	(-0.64)	(-0.67)	(-1.27)
DAGEDLOT		-0.0723	-0.0503
		(-1.25)	(-0.92)

Bibliography

Alonso, William. Location and Land Use. Cambridge, Massachusetts: Harvard University Press, 1964.

Bailey, Martin. "Effects of Race and other Demographic Factors on the Values of Single Family Homes" Land Economics, May 1966.

Ball, Michael. "Recent Empirical Work on the Determinants of Relative House Prices" Urban Studies, June 1973.

Barnett, Vic. "The Study of Outliers: Purpose and Model," Applied Statistics, 27, No. 3, pp. 242-50.

Berry, Brian and Robert Bednarz. "A Hedonic Model of Prices and Assessments for Single Family Homes: Does the Assessor Follow the Market or the Market Follow the Assessor?" Land Economics, February 1975.

Box, G.E.P. and D.R. Cox. "An Analysis of Transformations." Journal of the Royal Statistical Society, Series B, 26, 1964, pp. 211-52.

Bruggeman, William. "Tax Reform, Tax Incentives and Investment Returns on Housing," in Subcommittee on the City, House Banking, Finance and Urban Affairs Committee, Federal Tax Policy and Urban Development, G.P.O. June 1977, pp. 196-293.

Cronin, Francis J. The Housing Demand of Low Income Households. Washington, D.C.: The Urban Institute, 1979.

Cronin, Francis J. "The Efficiency of Demand-Oriented Housing Programs," The Urban Institute, Washington, D.C., 1979.

Court, Andrew. "Hedonic Price Indexes with Automotive Examples," The Dynamics of Automobile Demand. New York: General Motors Corporation, 1939.

de Leeuw, Frank and Larry Ozanne, "Impact of the Federal Income Tax on Investment in Housing," paper presented at Brookings Institution Conference on the Economic Effect of Federal Taxes on October 19, 1979 and revised January 1980.

de Leeuw, Frank and Raymond Struyk. The Web of Urban Housing. Washington, D.C.: The Urban Institute, 1975.

Draper, N.R. and H. Smith. Applied Regression Analysis. New York: John Wiley and Sons, 1966.

Fisher, Franklin and Karl Shell. "Taste and Quality Change in the Pure Theory of the True-Cost-of-Living Index." In Price Indexes and Quality Change. Edited by Zvi Griliches. Cambridge, Massachusetts: Harvard University Press, 1971.

Follain, James R. "How Well Do Section 8 FMRs Match the Cost of Rental Housing? Data From 39 Large Cities." Contract Report 249-15, The Urban Institute, 1979.

Follain, James R. and Stephen Malpezzi. Dissecting Housing Value and Rent: Estimates of Hedonic Indexes for Thirty-Nine Large SMSAs. Washington, D.C.: The Urban Institute, 1980a.

_________. "The Flight to the Suburbs: Insights Gained from an Analysis of Central City Versus Suburban Housing Costs." Journal of Urban Economics, forthcoming, 1980b.

_________. "Estimates of Housing Inflation for 39 SMSAs: An Alternative to the Consumer Price Index." Annals of Regional Science, forthcoming, 1980c.

_________. "Racial Differences in the Price of Housing." Urban Studies, forthcoming, 1980d.

_________. "Are Occupants Accurate Appraisers? " Federal Home Loan Bank of San Francisco Working Paper, forthcoming, 1980e.

Follain, James R. and Larry Ozanne, with Verna M. Alburger. Place to Place Indexes of the Price of Housing: Some New Estimates and a Comparative Analysis. Washington, D.C.: The Urban Institute, May 1979.

Griliches, Zvi. Price Indexes and Quality Change. Cambridge: Harvard University Press, 1971.

Hulten, Charles and Frank Wykoff, "On the Feasibility of Equating Tax to Economic Depreciation," in U.S. Department of Treasury, 1978 Compendium of Tax Research, G.P.O. 1978.

Kain, John and John Quigley. "Measuring the Value of Housing Quality" Journal of the American Statistical Association, June 1970.

Kain, John and John Quigley. Housing Markets and Racial Discrimination. New York: National Bureau of Economic Research, 1975.

King, Thomas and Peter Mieszkowski. "Racial Discrimination, Segregation, and the Price of Housing." Journal of Political Economy, May/June 1973.

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

Leamer, Edward. Specification Searches: Ad Hoc Inferences with Non-Experimental Data. New York. Wiley, 1978.

Maddala, G.S. Econometrics. New York: McGraw Hill, 1977.

Merrill, Sally. Draft Report on Hedonic Indices as a Measure of Housing Quality. Cambridge, Massachusetts, Abt Associates, 1977.

Mieszkowski, Peter and Richard F. Syron. "Economic Explanations for Housing Segregation." New England Economic Review, Federal Reserve Bank of Boston, November/December 1979.

Mills, Edwin. "An Aggregative Model of Resource Allocation in a Metropolitan Area" American Economic Review, May 1967.

Muellbauer, John. "Household Production Theory, Quality, and the 'Hedonic Technique'" American Economic Review, December 1974.

Murray, Michael. "Hedonic Prices and Composite Commodities" Journal of Urban Economics, April 1978.

Muth, Richard. Cities and Housing. Chicago: University of Chicago Press, 1969.

Noto, Nona. The Effect of the Local Public Sector on Residential Property Values in San Mateo County, California. Philadelphia: Federal Reserve Bank, 1976.

Ozanne, Larry, Marcellus Andrews, and Stephen Malpezzi. "The Efficacy of Hedonic Estimation with the Annual Housing Survey: Evidence from the Demand Experiment." HUD Occasional Papers, forthcoming, 1980.

Ozanne, Larry and Thomas Thibodeau. "Market Effects of Housing Allowances: Demand Shifts and the Price of Housing" Washington, D.C.: The Urban Institute, 1980.

Palmquist, Raymond. "Alternative Techniques for Developing Real Estate Price Indexes," Review of Economics and Statistics, (forthcoming).

Rao, Potluri and Roger Miller. Applied Econometrics. Belmont, California: Wadsworth, 1971.

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

Schnare, Ann. The Persistence of Racial Segregation in Housing. Washington, D.C.: The Urban Institute, 1978.

_________ and Raymond Struyk. "Segmentation in Urban Housing Markets" Journal of Urban Economics, April 1976.

_________. "An Analysis of Ghetto Housing Prices Over Time." In Residential Location and Urban Housing Markets. Edited by Gregory Ingram. Cambridge, Massachusetts: National Bureau of Economic Research, 1977.

Selvin, Hanan C., Alan Stuart, and Stephen J. Finch. Introdution to Data Analysis. New York: Random House, 1979.

Snedecor, George W. and William G. Cochran. Statistical Methods. Ames, Iowa: The Iowa State University Press, 1967.

Struyk, Raymond. "The Need for Local Flexibility in U.S. Housing Policy" Policy Analysis, Fall 1977.

_________, Sue Marshall and Larry Ozanne. Housing Policies for the Urban Poor. Washington, D.C.: The Urban Institute, 1978.

Taubman, Paul and Robert Rasche, "Economic and Tax Depreciation of Office Buildings," National Tax Journal, September 1969, 22:3, pp. 334-46.

Theil, Henri. Principles of Econometrics. New York: Wiley, 1971.

Tukey, John W. Exploratory Data Analysis. Reading, Mass.: Addison-Wesley Publishing Company, 1977.

U.S. Bureau of the Census. Current Housing Reports, Annual Housing Survey: Housing Characteristics for Selected Metropolitan Areas (SMSA Name). Washington, D.C., 1976, 1977, or 1978.

U.S. Department of Labor, Bureau of Labor Statistics. "Autumn 1976 Urban Family Budgets and Comparative Indexes for Selected Urban Areas." Washington, D.C.: U.S. Government Printing Office, April 1977.

Wonnacott, Ronald and Thomas Wonnacott. Econometrics. New York: Wiley, 1970.

Wykoff, Frank. "Economic Depreciation of Relatively New Apartments in the U.S." Pomona College (undated, unpublished).

Yinger, John. The Black-White Price Differential in Housing: Some Further Evidence. Madison, Wisconsin: Institute for Research on Poverty, 1975.

THE URBAN INSTITUTE
2100 M Street, N.W., Washington, D.C. 20037