Journal of Housing Research • Volume 6, Issue 1 25

Age-Related Heteroskedasticity in Hedonic House Price Equations

Allen C. Goodman and Thomas G. Thibodeau*

Abstract

This article examines the relationship between dwelling age and the market value of owner-occupied housing. The article theoretically establishes and empirically verifies that (1) housing depreciation is nonlinear and (2) dwelling age-induced heteroskedasticity is prevalent in hedonic house price equations. The empirical results are obtained with a semilog hedonic house price equation from data on nearly 8,500 transactions of single-family homes in Dallas. Hedonic parameters are estimated with four alternative dwelling age specifications and two iterative generalized least squares estimation procedures that accommodate heteroskedasticity by explicitly modeling the residual variance. Estimated depreciation rates are sensitive to both the dwelling age specification and the estimation procedure. The article establishes the importance of incorporating second-order effects in obtaining accurate point estimates for housing depreciation.

Keywords: Housing depreciation; Housing econometrics; House prices; Hedonic price theory

Introduction

The hedonic procedure is frequently used to quantify the effect of various housing and neighborhood characteristics on house prices. Empirically, the technique uses regression analysis to explain variation in market values using property characteristics. A hedonic equation for single-family homes relates some market value estimate (the owner's estimate, a real estate appraiser's estimate, a tax assessor's estimate, or, if the property was recently sold, the transaction price) to the property's characteristics (square feet of living space, lot size, dwelling age, whether the property has a swimming pool, variables measuring proximity to transportation arteries, variables measuring the quality of public services, etc.). The relevant housing characteristics to include in the hedonic specification, the most appropriate functional form, the proper estimation procedure, and the correct interpretation of the estimated hedonic parameters are all topics of debate.

This article examines two issues related to hedonic house price equations: (1) how the dwelling age specification influences estimated depreciation rates and (2) whether the error variance is systematically related to dwelling age. The empirical analysis uses a semilogarithmic hedonic house price equation and data on nearly 8,500 sales of single-family detached homes in Dallas. The hedonic parameters are estimated by an iterative generalized least squares (GLS) procedure that models the relationship between dwelling age and the residual variance. Four dwelling age specifications and two variance models are examined.

---

* Allen C. Goodman is Professor of Economics at Wayne State University. Thomas G. Thibodeau is Professor of Real Estate in the E. L. Cox School of Business, Southern Methodist University. The authors thank Ed Coulson, Ron Gallant, Panayiotis Mavros, and a referee for useful comments. Remaining errors are the authors' responsibility.

The article theoretically establishes and empirically verifies that (1) the relationship between house value and dwelling age is nonlinear and possibly nonmonotonic and (2) dwelling age-induced heteroskedasticity is prevalent in hedonic house price equations. We introduce variance modeling procedures that increase efficiency as well as influence hedonic parameter point estimates, in both large and small samples. The article establishes the importance of incorporating second-order effects in obtaining accurate point estimates for parameters.

Hedonic House Price Specification

A hedonic equation for owner-occupied homes relates an estimate of the property's market value to the various characteristics that determine its value. Housing characteristics can be loosely grouped into five categories: (1) characteristics of the lot, (2) characteristics of the improvement, (3) neighborhood amenities, (4) proximity variables, and (5) the period when the housing data are collected. The general specification for a hedonic house price equation is

$$V = f(L, S, N, P, t), \tag{1}$$

where V is the estimated market value of the property (sales price); L denotes a class of variables describing lot characteristics (lot size, shape, topography, site improvements, frontage, etc.); S denotes a class of variables describing structural characteristics (square feet of living space, dwelling age, number of stories, types of equipment and fuels used to provide services, etc.); N denotes a class of neighborhood variables (percentage of improved land area in the neighborhood allocated to owner-occupied homes, percentage nonresidential, percentage undeveloped, employment density, public school achievement scores, police and fire department response times, crime rates, etc.); P denotes a class of proximity variables (distances to the central business district, various nonconforming land uses that may produce externalities, neighborhood recreation facilities, schools, shopping, public transportation, major thruways, etc.); and t denotes the period when property information was collected.

Rosen's (1974) hedonic model (and those derived from it) admits nonlinear price functions. This suggestion has motivated researchers to examine alternative hedonic specifications including the semilog and log-log, as well as empirically search over alternative specifications using Box-Cox transformations (Box and Cox 1964). The Box-Cox formulation provides a useful way of summarizing alternative approaches to estimating hedonic price regressions. Consider the general form:

$$\left(V^{\lambda_0} - 1\right)/\lambda_0 = \beta_0 + \beta_1\left(X^{\lambda_1} - 1\right)/\lambda_1 + \mu . \tag{2}$$

It is easily shown that the hedonic price $\partial V/\partial X$ equals

$$\partial V/\partial X = \beta_1 X^{\lambda_1 - 1} V^{1-\lambda_0} . \tag{3}$$

Thus, the conventional linear form $\left(\lambda_0 = \lambda_1 = 1\right)$ yields constant hedonic prices. Log-log $\left(\lambda_0 = \lambda_1 = 0\right)$ or semilog $\left(\lambda_0 = 0; \lambda_1 = 1\right)$ yields a hedonic price model with a multiplicative structure. Researchers can either test the three forms against each other or estimate the nonlinear parameters in λ directly. Moreover, the second derivative of the hedonic price function $\partial^2 V/\partial X^2 \geq 0$ as $\beta_1 X^{\lambda_1} V^{-\lambda_0}\left(1 - \lambda_0\right) \geq \left(1 - \lambda_1\right)$, providing valuable information on the slope and curvature of the underlying envelope of bid and offer curves.

The main advantage of the nonlinear specifications is that they permit characteristic prices to vary with the quantity of other housing characteristics included in the bundle. Also, some dependent variable transformations correct for heteroskedasticity between house value and the residual (i.e., prediction errors tend to be larger, in absolute value, as property values increase).

Several authors have applied Box-Cox transformations to hedonic house price equations. Goodman (1978) computed house price indices for 15 separate housing submarkets using this technique. He reported that the estimated hedonic coefficients are not constant across submarkets and are not constant over time. In addition, his empirical results support rejection of both linear and semilog specifications. Halvorsen and Pollakowski (1981) also rejected the linear and semilog functional forms using the Box-Cox technique. Linneman (1980) used data from the 1973 national Annual Housing Survey (AHS) to estimate housing hedonic equations for Chicago, Los Angeles, and a pooled sample of the 34 largest metropolitan areas. For owner-occupied properties in Chicago, Los Angeles, and the pooled sample, he reported Box-Cox parameters compatible with the semilog specification.

Effect of Dwelling Age on House Price

Dwelling age influences a property's market value in two ways. First, age is used to quantify economic depreciation. Hulten and Wykoff (1981) define economic depreciation as "the decline in asset price (or shadow price) due to aging." Second, a dwelling's age may also incorporate a vintage effect. The vintage effect occurs when some unmeasured housing characteristic (for example, housing quality) is correlated with the year that a dwelling was built (see Hall 1971 for a discussion of the vintage effect in durable goods).

Consider a standard hedonic house price function of the type

$$P = P_o e^{-\delta\alpha}, \tag{4}$$

where d is the rate of depreciation, and α is the dwelling age. The usual finding with respect to depreciation is $\left(\partial P/\partial\alpha\right)/P = -\delta$.

Now consider a model without renovation, but in which houses with different values of α have different levels of P_o. Hence, $P = P_o\left(\alpha\right)e^{-\delta\alpha}$. In this model, houses built α years ago have the characteristics of vintage α, built according to the decisions of builders α years before. How age affects price in this relationship is jointly determined by the supply decisions of the builders of that housing vintage and the set of consumer preferences in the current period. More generally, we get $(\partial P/\partial a)/P = P_o9 - d$, where P_o9 refers to the percentage change in P_o. This term may be positive or negative, depending on the availabilities and valuations of neighboring vintages of housing.[1]

Finally, consider the renovation decision. In each year, holding depreciation constant, the housing stock is influenced by constructing new units and by renovating the existing stock. Again, consider a house built a years before. Renovations, $R(t)$, at any time t between 0 and α, may either increase or decrease the value of the house, again depending on the supply of housing of that vintage and the set of current consumer preferences. Consider the index A, such that

$$A = 1 + \int_0^{\alpha} R(t)dt, \text{ where } A = 1 \text{ at } \alpha = 0 \text{ and } A > 0 \text{ for all } \alpha. \text{ Then } P = P_o(\alpha)A(\alpha)e^{-\delta\alpha}.$$

Differentiating this with respect to α yields $(\partial P/\partial a)/P = P_o 9 + A' - d$, (with A' equal to the percentage change in A), which again may be positive or negative.

This analysis suggests that over given periods the asset price of housing might vary either positively or negatively with dwelling age and that hedonic house price estimation methods must permit nonlinear variation with dwelling age.

Heteroskedasticity in Hedonic House Price Equations

Since the hedonic price specification is very general and may encompass a large number and variety of variables, it is plausible that any or all of them may contribute to heteroskedasticity of the error term. Indeed, White (1980) provides a heteroskedasticity-consistent ordinary least squares (OLS) variance estimator that includes information from the entire matrix of explanatory variables. We will use this estimator as a benchmark for our estimates.

Although hedonic house price heteroskedasticity may simply be related to errors in specification, economic theory also provides insights into the nature of the error term. Housing is a long-lived durable good, which depreciates, requires maintenance, and is subject to renovation. The prices and attributes of new housing are related to known market conditions at the time it is built. Dwelling age heteroskedasticity is likely because the magnitude of the error in predicting house price probably increases with dwelling age. The older a dwelling, the more likely the property was significantly upgraded or improved at some time during its life. The most common home improvements (upgrading a kitchen or a bathroom, replacing heating or air conditioning systems, installing a new roof) are typically not recorded in publicly available data sets. Consequently, there is no way to incorporate these improvements in the hedonic specification.

Heteroskedasticity with respect to dwelling age would reflect the joint probability of renovation and maintenance, multiplied by the level of renovation or maintenance that took place, integrated over the age of the house. While the functional notation of the relationship may be arbitrary, it is clear that increased age almost certainly leads to heteroskedasticity of the error term.

---

[1] There is a parallel literature on nonmonotonic price-distance functions (e.g., DeVany 1976; Goodman 1979; Polinsky and Shavell 1976). Note that for $\partial P/\partial\alpha$ to be positive, consecutive periods must show the vintage effects posited. Otherwise the vintage effects will be seen as (heteroskedastic) error.

The interaction between dwelling maintenance and vintage also contributes to dwelling age-induced heteroskedasticity. Let the price of a house of vintage o at time t be $P(t) = P_o e^{-dt}$. $P(0)$ is the average price over all houses built at time 0. At any given time, the amount of maintenance up to time T would be $\int_0^T m(t)\,dt$, where $m(t)$ is annual maintenance. This expression has a lower bound of 0 (no maintenance), so as t increased, so would the variance. In addition, even if we assume there is only one renovation during the life of the house, the renovations may differ according to the vintage of the renovation. Presumably, as t increases, the probability that houses have renovations of different vintages increases, further contributing to heteroskedasticity.

There are two compelling reasons to be concerned about age-related heteroskedasticity. The first involves estimating coefficient matrix β (including estimated depreciation rates) relating outcome $\mathbf{y}$ to explanatory matrix $\mathbf{X}$. The standard GLS estimator for β is

$$\tilde{\mathbf{b}} = \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}, \tag{5}$$

where Ω is the symmetric positive definite matrix used to multiply the homoskedastic disturbance matrix. This compares to the OLS estimator of

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{6}$$

Although $\mathbf{b}$ and $\tilde{\mathbf{b}}$ are unbiased (i.e., $E(\mathbf{b}) = E(\tilde{\mathbf{b}})$), they are equal for a given sample only if Ω^{-1} is the identity matrix. Hence, for any given sample, estimated depreciation parameters included in β will be related to the heteroskedasticity of the error term. Hence, in this model, inaccurate characterization of the variance leads to inaccurate estimation of depreciation.

The second reason to be concerned about age-related heteroskedasticity involves the real-world use of hedonic regressions. Various forms of hedonic regressions are used for property tax assessments, in which accuracy is critical, both to ensure equity in property value assessments and to minimize the error variance to limit costly appeals. In this context, the variance truly comprises a "cost function" of inaccurate estimation. Appropriate corrections for heteroskedasticity reduce the variance of the market value estimates, thus making the process more efficient and hence less costly.

We propose an estimation approach using an iterative GLS procedure suggested by Davidian and Carroll (1987). This procedure uses an iterative model in which the regression parameters and the variance function are jointly measured in a semiparametric model. Two variance models are examined here: The first relates the absolute value of the residual to a polynomial in dwelling age, and the second relates the square of the residual to a similar polynomial.

To begin the two-stage iterative GLS estimation procedure, OLS is first used to provide consistent estimates of the hedonic coefficients. Let e_i be the estimated residuals obtained from this stage. The second step uses OLS to estimate parameters for the residual variance model. The absolute-value-of-the-residual model is

$$\left|e_i\right| = \theta_0 + \theta_1\text{AGE} + \theta_2\text{AGE}^2 + \theta_3\text{AGE}^3 + \theta_4\text{AGE}^4. \tag{7}$$

The residual-squared model is

$$e_i^2 = \theta_0 + \theta_1\text{AGE} + \theta_2\text{AGE}^2 + \theta_3\text{AGE}^3 + \theta_4\text{AGE}^4. \tag{8}$$

The first step is then repeated using weighted least squares. The weights are reciprocals of the normalized predicted values obtained from the variance model. We constrain the dwelling age specification in the variance model to be identical to the age specification in the hedonic equation. For example, the weights for the quadratic dwelling age hedonic specification are estimated from variance models that force $\beta_3 = \beta_4 = 0$.

This procedure yields efficient estimates of the regression parameters and of the variance of the estimators. The resulting GLS estimators are more reliable than the OLS estimators. Further, the variance estimates themselves may be important to those who seek to relate possible appraisal errors to observable factors. Finally, the procedures are easily programmed and easily used.

The Empirical Literature

Numerous papers have used hedonic equations to estimate $(\partial P/\partial \alpha)/P$. Kain and Quigley (1970) examined services provided by individual housing characteristics using factor analysis on 39 indices of housing quality. Their study, using a survey of St. Louis households, yielded depreciation rates of 0.7 percent for owner-occupied housing. Grether and Mieszkowski (1974) examined single-family homes sold in New Haven, Connecticut, and its suburbs between 1962 and 1969. They incorporated dwelling age (in addition to neighborhood variables) interactively with lot size on the assumption that land does not depreciate. They reported that the effect of age on house price was nonlinear, with younger dwellings depreciating more rapidly than older ones.

Other authors have examined alternative dwelling age specifications. Palmquist (1979) examined linear, semilog, log-linear, inverse semilog, and Box-Cox transformations to correct a repeat-sales house price index for depreciation. Palmquist ultimately selected the semilog functional form and estimated a 0.8 percent annual depreciation rate. Jones, Ferri, and McGee (1981) empirically examined five alternative dwelling age specifications: (1) geometric depreciation, (2) quadratic depreciation, (3) cubic depreciation, (4) piecewise geometric depreciation (using dummy variables for dwelling age intervals in a semilog hedonic specification), and (5) interactive depreciation that examines the joint effect of lot size and dwelling age. They concluded that the cubic dwelling age specification provided the best fit for their data.

Malpezzi, Ozanne, and Thibodeau (1980, 1987) estimated economic depreciation for renter- and owner-occupied housing located in the 59 metropolitan areas surveyed by the 1974–76 Standard Metropolitan Statistical Area (SMSA) AHS. To avoid constraining the depreciation rate to be constant, the Malpezzi, Ozanne, and Thibodeau semilog hedonic specification included higher order dwelling age terms (dwelling age squared and dwelling age cubed), as well as a dummy variable for the oldest dwellings (those built before 1940). They reported that market values, on average, decrease with dwelling age and that newer units depreciate more rapidly than older ones. The 59-SMSA average depreciation rate for owner-occupied housing ranged from 0.9 percent in year 1 to

0.28 percent in year 20. The rate of economic depreciation for 10-year-old owner-occupied dwellings was half that for new ones. By year 30 the rate rose again to 0.6 percent.

Cannaday and Sunderman (1986) employed a log-linear model to estimate depreciation for single-family homes in the southwest portion of Champaign, Illinois. Their data consisted of 812 transactions for the 1976–84 period. Unlike most of the other empirical literature, they reported a depreciation pattern that was initially less rapid than straight line.

Randolph (1988a) separated economic depreciation from the so-called vintage effect using data from the 1974 and 1977 Detroit AHS. Randolph's example of the vintage effect in housing is that "building technology or material costs may change over time so as to introduce a trend in the initial unmeasured structural qualities of housing units. City growth patterns may have entailed the building of equivalent structures on more valuable land first, consequently introducing a time trend in some unmeasured location and structural characteristics" (p. 164). Randolph demonstrated that longitudinal data do not permit estimation of economic depreciation unless some identifying assumptions are made. The necessary assumptions are that either (1) the long-term vintage effect is negligible or (2) the average market level of unmeasured housing quality is stable. Randolph's "empirical results indicate that the assumption of stable unobserved quality is likely to be superior to ignoring vintage effects" (p. 174). In a subsequent paper designed to correct the shelter component of the Consumer Price Index for the age bias, Randolph (1988b) assumed that the vintage effect was negligible.

Few hedonic house price papers have examined heteroskedasticity, and nearly all that have examined it have focused on heteroskedasticity associated with house value. To correct for heteroskedasticity in their linear hedonic specification, Grether and Mieszkowski (1974) divided each observation by the size of the house and the construction price index, transforming the dependent variable to a real price per square foot. They report that the transformation made very little difference in either the estimated coefficients or the *t* statistics. Palmquist (1979) tested a semilog hedonic specification for heteroskedasticity and concluded that the data did not support rejection of a homoskedastic error variance.

Randolph (1988a) observed that the residual variance increased with dwelling age squared. He estimated regression parameters by the two-step GLS procedure of Hildreth and Houck (1968). Finally, Rachlis and Yezer (1988) illustrated how heteroskedasticity affects the measurement of housing-related real estate risk. They examined heteroskedasticity in hedonic house price equations for five cities, using data obtained from the Federal Housing Administration and from the 1970 Census of Population and Housing. They regress the absolute value of the residual on included housing characteristics and conclude "The variable most consistently related to appraisal risk . . . is age of structure" (p. 294). They correct for heteroskedasticity by a weighted least squares estimation procedure.

Data, Hedonic Specification, and Estimation Procedure

This article examines alternative dwelling age specifications and age-induced heteroskedasticity using data for 8,476 transactions of single-family homes sold during

1984 and 1985 in Dallas. The primary source of information is the Dallas Central Appraisal District (DCAD), which is responsible for estimating property values (for tax purposes) for all real property in Dallas County. DCAD provided a computer file containing each property's address as well as information on each residential property's structural characteristics (square feet of living space, year built, etc.). DCAD obtains sales data from the local multiple listing service and from the North Texas Regional Data Center of the Society of Real Estate Appraisers.

Within Dallas County there are 28 municipalities and 15 school districts. Each municipal government and each school district raises its own revenues and provides its own services. In addition, municipalities frequently overlap with school districts. To avoid tax capitalization issues, only properties that are both in the Dallas Independent School District and in the city of Dallas are included here. Since this limitation eliminates the variation in tax rates across properties, all variables included in traditional tax capitalization studies have been excluded. In addition to dwelling age, only two other housing characteristics are included in the hedonic specification: square feet of living space and month of sale. Square feet of living space is specified as a quadratic, and each month of sale, except for January 1984, is represented by a dummy variable.

The mean sales price for the 8,476 transactions was \$104,297; the median was \$78,250. The average dwelling had 1,703 square feet of living space and was 29 years old. The sales were approximately uniformly distributed over the 24 months beginning in January 1984.

A semilogarithmic specification is employed:

$$\begin{aligned}\ln\left(V_i\right) = \beta_0 &+ \beta_1\text{AGE} + \beta_2\text{AGE}^2 + \beta_3\text{AGE}^3 + \beta_4\text{AGE}^4 \\ &+ \beta_5\text{LIVAREA} + \beta_6\text{LIVAREA}^2 \\ &+ \sum_{j=2}^{24} \delta_j\text{SOLD}_j + \mu_i \end{aligned} \tag{9}$$

where

$\ln(V_i)$ = natural log of the reported selling price of the ith house,
AGE = age of the house in years,
LIVAREA = living area in square feet,
SOLD_j = dummy variable for the jth month of sale.

Four dwelling age specifications are examined. The linear specification forces the depreciation rate to be constant by constraining $\beta_2 = \beta_3 = \beta_4 = 0$; the quadratic forces $\beta_3 = \beta_4 = 0$; the cubic forces $\beta_4 = 0$; and the final alternative estimates the coefficients for a quartic specification. Estimated depreciation rates are computed from

$$\partial V/\partial \text{AGE} = \beta_1 + 2\beta_2\text{AGE} + 3\beta_3\text{AGE}^2 + 4\beta_4\text{AGE}^3 \,. \tag{10}$$

Hedonic parameters are initially estimated by OLS, before the Davidian and Carroll procedure is implemented. The residuals from the OLS regression are then examined for heteroskedasticity. Two separate tests for heteroskedasticity are employed: one developed by Goldfeld and Quandt (1965) and one by White (1980). The Goldfeld-Quandt test partitions the data into two groups according to dwelling age. Separate hedonic parameters are estimated for each sample. The ratio of the estimated error variances follows an F distribution. The White test regresses the squared residuals from the hedonic equation on the squares and cross products of the regressors. The product of the sample size and the squared multiple correlation coefficient follows a χ^2 distribution.

Results

OLS regression statistics for the four alternative dwelling age specifications are presented in table 1. The first two columns list the estimated coefficients and White's (1980) heteroskedasticity-consistent standard errors for the linear dwelling age specification.[2] While theory provides a persuasive case for entering age as a polynomial, there are two reasons to check the linear specification. First, linear coefficients are easily and directly interpreted. Second, by far the largest portion of hedonic price studies have treated age in this manner. Checking this specification provides guidance in interpreting the findings from other studies. The second pair of columns lists regression statistics for the quadratic specification, the third pair for the cubic specification, and the final pair for the quartic specification. In each case, the estimated parameters explain at least 73 percent of the variance in the natural log of sales price. The quartic dwelling age specification explains more than 75 percent of the variance. The estimated coefficients for living area are relatively stable across the specifications and indicate that living area increases valuation at a decreasing rate. Furthermore, the empirical results suggest that Dallas house prices were essentially constant during January 1984 to February 1985 but increased significantly during the remainder of 1985. The estimated coefficients for month of sale are relatively stable across the four specifications.

The constant geometric depreciation rate obtained from the linear dwelling age specification is 0.137 percent. This rate is similar to rates estimated in other studies that constrain depreciation to be a constant rate.

The results for the polynomial dwelling age specifications provide empirical support for nonlinear depreciation rates. Standard F tests on the restriction of polynomial dwelling age terms indicate that increased orders of the polynomials significantly increase the explanatory power of the regressions. The quartic specification yields a 6.7 percent annual depreciation rate for dwellings one year old. The rate decreases to 1.4 percent for dwellings 10 years old.

To examine dwelling age-induced heteroskedasticity, the data were partitioned into 13 dwelling age categories. The categories were constructed to allocate observations equally among categories. The first column in the top part of table 2 defines the dwelling age categories, while the second column lists the number of observations in each category. The mean number of observations in a category is 652; the standard deviation is 45. The

[2] Greene (1993) demonstrates that the OLS estimated standard errors are biased downward.

Table 1. **OLS Regression Statistics**

Variable	Linear in Age		Quadratic in Age		Cubic in Age		Quartic in Age	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Intercept	9.6134	0.0290	9.5558	0.0323	9.6624	0.0309	9.7131	0.0305
(Living Area/100)	0.1207	0.0019	0.1211	0.0019	0.1285	0.0020	0.1337	0.0021
$(\text{Living Area}/100)^2$	–0.0011	0.0000	–0.0011	0.0000	–0.0012	0.0000	–0.0013	0.0000
Age/10	–0.0137	0.0032	0.0255	0.0094	–0.3485	0.0233	–0.7513	0.0356
$(\text{Age}/10)^2$			–0.0061	0.0015	0.1444	0.0095	0.4108	0.0223
$(\text{Age}/10)^3$					–0.0157	0.0010	–0.0760	0.0051
$(\text{Age}/10)^4$							0.0044	0.0004
Feb84	–0.0559	0.0288	–0.0533	0.0288	–0.0564	0.0278	–0.0472	0.0273
Mar84	–0.0418	0.0256	–0.0417	0.0257	–0.0392	0.0249	–0.0313	0.0246
Apr84	–0.0003	0.0253	–0.0013	0.0253	–0.0075	0.0246	–0.0055	0.0244
May84	0.0098	0.0251	0.0120	0.0252	0.0106	0.0245	0.0128	0.0242
Jun84	0.0185	0.0240	0.0203	0.0241	0.0173	0.0232	0.0186	0.0229
Jul84	0.0277	0.0234	0.0279	0.0235	0.0328	0.0228	0.0391	0.0226
Aug84	0.0026	0.0239	0.0049	0.0239	0.0106	0.0236	0.0172	0.0232
Sep84	0.0529	0.0249	0.0547	0.0250	0.0603	0.0244	0.0654	0.0238
Oct84	0.0378	0.0252	0.0415	0.0253	0.0336	0.0244	0.0353	0.0240
Nov84	0.0147	0.0250	0.0164	0.0251	0.0185	0.0242	0.0227	0.0239
Dec84	0.0404	0.0276	0.0433	0.0276	0.0400	0.0265	0.0437	0.0265
Jan85	0.0297	0.0267	0.0315	0.0269	0.0270	0.0260	0.0289	0.0256
Feb85	0.0421	0.0279	0.0443	0.0281	0.0410	0.0272	0.0479	0.0269
Mar85	0.1076	0.0245	0.1094	0.0246	0.0967	0.0238	0.0981	0.0234
Apr85	0.0632	0.0258	0.0645	0.0259	0.0592	0.0252	0.0575	0.0249
May85	0.0825	0.0238	0.0841	0.0239	0.0942	0.0230	0.0949	0.0227
Jun85	0.1473	0.0229	0.1483	0.0230	0.1465	0.0224	0.1486	0.0219
Jul85	0.1345	0.0249	0.1362	0.0250	0.1401	0.0241	0.1444	0.0238
Aug85	0.0985	0.0248	0.1010	0.0249	0.1060	0.0241	0.1087	0.0238
Sep85	0.1421	0.0281	0.1450	0.0282	0.1442	0.0269	0.1495	0.0265
Oct85	0.1841	0.0304	0.1851	0.0305	0.1789	0.0290	0.1801	0.0287
Nov85	0.1070	0.0300	0.1075	0.0300	0.1132	0.0290	0.1112	0.0288
Dec85	0.1278	0.0282	0.1277	0.0283	0.1263	0.0275	0.1283	0.0271
Number of observations	8,476		8,476		8,476		8,476	
Root mean square error	0.349		0.349		0.339		0.335	
Adjusted R^2	0.730		0.731		0.746		0.751	
F value	881.9		852.4		889.5		883.2	

Table 2. **Residual Variances and Heteroskedasticity Tests**

Age Classification	Number of Observations	Residual Variances			
		Linear in Age	Quadratic in Age	Cubic in Age	Quartic in Age
Thirteen age groups					
Age $\leq$ 4	597	0.1041	0.1047	0.1009	0.0991
4 < Age $\leq$ 14	639	0.0604	0.0609	0.0597	0.0602
14 < Age $\leq$ 19	661	0.0568	0.0569	0.0574	0.0583
19 < Age $\leq$ 22	652	0.0612	0.0612	0.0619	0.0628
22 < Age $\leq$ 25	639	0.0757	0.0757	0.0763	0.0772
25 < Age $\leq$ 27	670	0.0803	0.0803	0.0801	0.0804
27 < Age $\leq$ 29	648	0.0776	0.0775	0.0766	0.0764
29 < Age $\leq$ 31	722	0.0984	0.0983	0.0975	0.0974
31 < Age $\leq$ 33	737	0.1360	0.1357	0.1330	0.1311
33 < Age $\leq$ 36	639	0.1561	0.1557	0.1526	0.1508
36 < Age $\leq$ 43	687	0.1791	0.1791	0.1742	0.1743
43 < Age $\leq$ 57	582	0.2002	0.1995	0.1974	0.1970
57 < Age	603	0.2058	0.2032	0.2123	0.1961
Two age groups					
Age $\leq$ 28	4,182	0.0766	0.0717	0.0713	0.0712
Age > 28	4,294	0.1520	0.1487	0.1485	0.1483
Heteroskedasticity tests					
Goldfeld-Quandt test					
F statistic		1.9843	2.0739	2.0827	2.0829
Prob > F (4153,4265)		0.0000	0.0000	0.0000	0.0000
White test					
χ^2 statistic		361.10	635.00	700.06	745.52
Prob > χ^2		0.0000	0.0000	0.0000	0.0000

last four columns in the top part of table 2 list the within-age-category error variances for each specification. Except for the newest homes, error variances increase monotonically with dwelling age. Error variances for the oldest dwellings (more than 57 years old) are consistently more than three times those for dwellings between 14 and 19 years old. The bottom part of table 2 presents results for the Goldfeld-Quandt and the White heteroskedasticity tests. Both tests reject the null hypothesis of homoskedastic error variances at conventional levels for each dwelling age specification.

The estimated parameters for the variance models and the hedonic house price equations are presented in tables 3 and 4. The final variance model parameter estimates (converging after three iterations in all cases) are shown in table 3. The absolute value of the residual variance is less sensitive to outliers than is the square of the residual and consequently provides a slightly better empirical fit for the data. Either specification indicates that the variance increases with dwelling age, suggesting heteroskedasticity in estimation and providing cause for increased attention by home appraisers.

Iterative GLS regression statistics are provided in table 4 for the model based on the absolute value of the residual variance (the results for the residual-squared model are similar and available from the authors on request). As expected, the GLS procedures reduce estimated standard errors. The GLS estimation procedures also improved the overall goodness of fit. The iterative GLS procedure using the absolute value of the variance reduced the standard error of the intercept by 4.5 percent for the linear dwelling age specification and by 3.6 percent for the quartic dwelling age specification. Note that the constant geometric depreciation rate increases from 0.137 percent for the OLS procedure to 0.173 percent for the residual-squared variance model (not shown)—a 26 percent increase in the depreciation rate. Even though the GLS procedure is primarily designed to improve efficiency, the resulting point estimates are sensitive to the estimation technique, even in large samples. In the abstract of their paper, Davidian and Carroll (1987, 1079) noted:

> Standard asymptotic theory implies that how one estimates the variance function, in particular the structural parameters, has no effect on the first-order properties of the regression parameter estimates; there is evidence, however, both in practice and higher-order theory to suggest that how one estimates the variance function does matter.

The within-age-category error variances for the iterative GLS estimation procedure using the absolute value of the residual are listed in table 5. Weighting transactions inversely proportional to the absolute value of the residual reduces estimated standard errors for newer properties but increases standard errors for older ones. The basic pattern in residuals is unchanged, however.

The importance of the dwelling age specification and estimation procedure is also apparent in the estimated depreciation rates. Depreciation rates (results available from the authors) are very sensitive to the estimation procedure when the linear and quadratic forms are employed, but less sensitive in the cubic and quartic specifications. For the linear specification, OLS yields a constant depreciation rate of 0.137 percent annually, the absolute-value-of-the-residual GLS depreciation rate is 0.146 percent, and the residual-squared depreciation rate is 0.173 percent. For the quadratic specification, OLS yields a 0.24 percent annual appreciation rate for dwellings one year old, while the GLS

Table 3. **Final-Stage Variance Model Parameters**

Variable	Linear in Age		Quadratic in Age		Cubic in Age		Quartic in Age	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Model based on absolute value of the residual variance								
Intercept	0.1723	0.0053	0.2306	0.0080	0.2729	0.0094	0.3022	0.0107
Age/10	0.0325	0.0016	–0.0148	0.0052	–0.1246	0.0114	–0.2237	0.0209
$(Age/10)^2$			0.0074	0.0008	0.0533	0.0041	0.1157	0.0123
$(Age/10)^3$					–0.0049	0.0004	–0.0183	0.0027
$(Age/10)^4$							0.0009	0.0002
Number of observations	8,476		8,476		8,476		8,476	
Root mean square error	0.220		0.219		0.211		0.207	
Adjusted R^2	0.044		0.053		0.072		0.074	
F value	390.5		240.1		218.7		171.1	
Model based on residual-squared variance								
Intercept	0.0359	0.0050	0.0906	0.0075	0.1034	0.0090	0.1278	0.0100
Age/10	0.0296	0.0015	–0.0140	0.0048	–0.0644	0.0109	–0.1395	0.0196
$(Age/10)^2$			0.0068	0.0007	0.0284	0.0039	0.0712	0.0115
$(Age/10)^3$					–0.0023	0.0004	–0.0103	0.0025
$(Age/10)^4$							0.0004	0.0002
Number of observations	8,476		8,476		8,476		8,476	
Root mean square error	0.205		0.205		0.201		0.194	
Adjusted R^2	0.042		0.049		0.058		0.058	
F value	371.9		221.3		175.9		132.4	

Table 4. **Iterative Generalized Least Squares Regression Statistics: Model Based on Absolute Value of the Residual Variance**

Variable	Linear in Age		Quadratic in Age		Cubic in Age		Quartic in Age	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Intercept	9.6545	0.0277	9.6083	0.0297	9.6893	0.0292	9.7453	0.0294
(Living Area/100)	0.1175	0.0017	0.1188	0.0018	0.1257	0.0017	0.1300	0.0017
(Living Area/100)^2	–0.0010	0.0000	–0.0010	0.0000	–0.0012	0.0000	–0.0012	0.0000
Age/10	–0.0146	0.0027	–0.0010	0.0083	–0.3549	0.0180	–0.7309	0.0339
(Age/10)^2			–0.0012	0.0013	0.1469	0.0067	0.3981	0.0203
(Age/10)^3					–0.0158	0.0007	–0.0735	0.0045
(Age/10)^4							0.0042	0.0003
Feb84	–0.0616	0.0293	–0.0549	0.0293	–0.0532	0.0280	–0.0453	0.0276
Mar84	–0.0432	0.0267	–0.0372	0.0265	–0.0302	0.0253	–0.0224	0.0250
Apr84	–0.0121	0.0268	–0.0064	0.0266	–0.0086	0.0255	–0.0060	0.0252
May84	0.0048	0.0263	0.0105	0.0261	0.0130	0.0251	0.0172	0.0247
Jun84	0.0163	0.0246	0.0219	0.0245	0.0228	0.0235	0.0245	0.0232
Jul84	0.0160	0.0253	0.0242	0.0252	0.0339	0.0240	0.0420	0.0237
Aug84	–0.0041	0.0248	0.0046	0.0247	0.0183	0.0237	0.0266	0.0233
Sep84	0.0482	0.0260	0.0543	0.0259	0.0638	0.0248	0.0676	0.0244
Oct84	0.0355	0.0262	0.0405	0.0262	0.0364	0.0251	0.0397	0.0248
Nov84	0.0107	0.0262	0.0173	0.0261	0.0228	0.0250	0.0271	0.0247
Dec84	0.0342	0.0276	0.0428	0.0275	0.0420	0.0264	0.0468	0.0260
Jan85	0.0330	0.0270	0.0334	0.0269	0.0323	0.0258	0.0350	0.0255
Feb85	0.0313	0.0274	0.0363	0.0273	0.0359	0.0261	0.0428	0.0257
Mar85	0.1067	0.0246	0.1107	0.0245	0.1014	0.0236	0.1018	0.0233
Apr85	0.0658	0.0255	0.0675	0.0254	0.0656	0.0243	0.0651	0.0240
May85	0.0816	0.0239	0.0859	0.0238	0.0955	0.0228	0.0972	0.0224
Jun85	0.1411	0.0234	0.1450	0.0233	0.1469	0.0223	0.1477	0.0220
Jul85	0.1311	0.0250	0.1354	0.0249	0.1409	0.0238	0.1458	0.0235
Aug85	0.0924	0.0254	0.0961	0.0252	0.1014	0.0242	0.1046	0.0238
Sep85	0.1287	0.0270	0.1368	0.0270	0.1361	0.0258	0.1404	0.0255
Oct85	0.1766	0.0295	0.1792	0.0294	0.1736	0.0282	0.1750	0.0278
Nov85	0.1126	0.0275	0.1153	0.0273	0.1235	0.0261	0.1241	0.0258
Dec85	0.1209	0.0268	0.1236	0.0267	0.1245	0.0255	0.1264	0.0252
Number of observations	8,476		8,476		8,476		8,476	
Root mean square error	0.346		0.346		0.335		0.331	
Adjusted R^2	0.746		0.743		0.761		0.767	
F value	955.7		906.1		963.2		963.6	

Table 5. **Residual Variances for the Model Based on the Absolute Value of the Residual Variance**

Age Classification	Number of Observations	Residual Variances			
		Linear in Age	Quadratic in Age	Cubic in Age	Quartic in Age
Age ≤ 4	597	0.1033	0.1039	0.1004	0.0986
4 < Age ≤ 14	639	0.0598	0.0601	0.0590	0.0591
14 < Age ≤ 19	661	0.0566	0.0566	0.0570	0.0577
19 < Age ≤ 22	652	0.0609	0.0609	0.0614	0.0620
22 < Age ≤ 25	639	0.0754	0.0755	0.0759	0.0766
25 < Age ≤ 27	670	0.0805	0.0803	0.0799	0.0800
27 < Age ≤ 29	648	0.0782	0.0779	0.0769	0.0765
29 < Age ≤ 31	722	0.0988	0.0985	0.0975	0.0973
31 < Age ≤ 33	737	0.1375	0.1368	0.1340	0.1322
33 < Age ≤ 36	639	0.1575	0.1568	0.1537	0.1521
36 < Age ≤ 43	687	0.1802	0.1795	0.1748	0.1752
43 < Age ≤ 57	582	0.2011	0.2007	0.1981	0.1976
57 < Age	603	0.2069	0.2064	0.2127	0.1970

rates are essentially zero. As dwellings age, the OLS appreciation rates decline and eventually (for 25-year-old dwellings) become negative. The GLS depreciation rates in the quadratic specification are essentially zero throughout the dwelling age distribution.

The depreciation patterns produced by the cubic and quartic dwelling age specifications provide evidence that depreciation rates vary considerably with dwelling age. When dwelling age is specified as a cubic, the estimated annual depreciation rates are between 3.2 percent and 3.5 percent for new dwellings. The rates decline rapidly (in absolute value) with dwelling age: The rate for 8-year-old dwellings is half the rate for new homes. The depreciation rate declines to zero for dwellings 15 to 20 years old. Dwellings 20 to 40 years old appreciate slightly, while older dwellings depreciate. The depreciation pattern is similar, but more pronounced, for the quartic dwelling age specification. The annual depreciation rate is between 6.5 and 6.7 percent for 1-year-old dwellings. Within 6 years, the depreciation rate is reduced by half. Dwellings between 15 and 40 years old appreciate, while older units depreciate.[3]

Conclusion

A closer look at the economics of housing prices suggests several reasons why the age of a home has a complicated effect on the price. Both depreciation and vintage effects suggest that the effect is almost certainly nonlinear and quite likely nonmonotonic. Moreover, the decisions made about construction and renovation imply a heteroskedasticity that is almost certainly related to age alone.

This heteroskedasticity is pervasive in many types of house price modeling. Consider the analyses based on "sale-resale" methods, concentrating on houses with multiple sales. Consider a house n, built at time i, and sold at time t:

$$P_{int} = \alpha + \beta X + \gamma \text{AGE} + \sum_j \delta_j D_j + \varepsilon_{int}, \tag{11}$$

where D_j is a set of year dummies for when houses were built. Let the error term ε be

$$\varepsilon_{int} = \rho_n + \kappa_t + \lambda_{\text{AGE}} + \mu_{int}, \tag{12}$$

where ρ_n is house specific, κ_t is year specific, λ_{AGE} is related to the house age, and μ is uncorrelated. The resale method, with a one-year lag, nets out house-specific and vintage effects, yielding

$$\Delta P = P_{in,t+1} - P_{int} = \gamma + \left(\kappa_{t+1} - \kappa_t\right) + \left(\lambda_{\text{AGE}+1} - \lambda_{\text{AGE}}\right). \tag{13}$$

Here, coefficient γ gives the effect of another year of age (the joint effect of depreciation and house price inflation).

---

[3] The small-sample properties of the iterative GLS estimator were examined using a 10 percent random sample of the 8,476 sales. The parameter estimates in the smaller sample were similar to their large-sample counterparts; however, the small-sample estimated parameter standard errors are more than three times their large-sample counterparts. The absolute-value-of-the-residual iterative GLS procedure reduced the standard error of the intercept by 6.4 percent in the linear dwelling age specification and by 1.0 percent in the quartic specification.

What can be said about the error term? There is no reason to believe that $(\kappa_{t+1} - \kappa_t)$ is heteroskedastic. However, unless λ is a linear function of age, $\lambda_{\mathrm{AGE}+1} - \lambda_{\mathrm{AGE}}$ is likely to be heteroskedastic. Thus, resale indices suffer from the problem of heteroskedasticity, even if there is no variation in the duration between sales. Heteroskedasticity is even more likely if the coefficients vary with time, as is found in most studies.

Our results support these conjectures. Using a large sample of housing transactions, we observe that the empirical relationship between transaction price and dwelling age is most sensitive to the dwelling age specification. The depreciation rates are nonlinear—higher for new homes and then declining with dwelling age—as well as nonmonotonic.

This study introduces two iterative GLS procedures that have not been applied to hedonic or appraisal analyses. These procedures accommodate heteroskedasticity by explicitly modeling the residual variance. Introduced by Davidian and Carroll, these models increase efficiency by reducing estimated standard errors, and they also influence point estimates of dwelling age coefficients.

The results suggest at least two real-world applications. The more efficient estimators are easily programmed for any regression-based appraisal package. Property tax assessors may reduce the costs that may accompany unnecessarily large variance in point estimates of property values.

The method may also be useful in the valuation of mortgage-backed securities. An explicit, and efficient, estimate of house price variance among bundled properties, using the method of Davidian and Carroll, may be quite useful in evaluating the overall risks of securities that are backed by home mortgages.

References

Box, G. E. P., and D. R. Cox. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society*, ser. B, 26(2):211–43.

Cannaday, Roger E., and Mark A. Sunderman. 1986. Estimation of Depreciation for Single-Family Appraisals. *AREUEA Journal* 14(2):255–73.

Davidian, M., and R. J. Carroll. 1987. Variance Function Estimation. *Journal of the American Statistical Association* 82:1079–91.

DeVany, Arthur S. 1976. An Economic Model of Airport Noise Pollution in an Urban Environment. In *Theory and Measurement of Economic Externalities*, ed. Steven A. Y. Lin, 205–14. New York: Academic Press.

Goldfeld, Stephen M., and Richard E. Quandt. 1965. Some Tests for Homoskedasticity. *Journal of the American Statistical Association* 60:539–47.

Goodman, Allen C. 1978. Hedonic Prices, Price Indices, and Housing Markets. *Journal of Urban Economics* 5(4):471–84.

Goodman, Allen C. 1979. Externalities and Nonmonotonic Price-Distance Functions. *Urban Studies* 16(October):321–28.

Greene, William. 1993. *Econometric Analysis.* 2nd ed. New York: Macmillan.

Grether, David M., and Peter Mieszkowski. 1974. Determinants of Real Values. *Journal of Urban Economics* 1(2):127–45.

Hall, Robert E. 1971. The Measurement of Quality Change from Vintage Price Data. In *Price Indexes and Quality Change*, ed. Zvi Griliches, 240–71. Cambridge, MA: Harvard University Press.

Halvorsen, Robert, and Henry O. Pollakowski. 1981. Choice of Functional Form for Hedonic Price Equations. *Journal of Urban Economics* 10(1):37–49.

Hildreth, Clifford, and James P. Houck. 1968. Some Estimators for a Linear Model with Random Coefficients. *Journal of the American Statistical Association* 68:584–95.

Hulten, Charles, and Fredrick Wykoff. 1981. The Measurement of Economic Depreciation. In *Depreciation, Inflation, and the Taxation of Income from Capital*, ed. Charles R. Hulten, 81–125. Washington, DC: The Urban Institute Press.

Jones, Wesley H., Michael G. Ferri, and L. Randolph McGee. 1981. A Competitive Testing Approach to Models of Depreciation in Housing. *Journal of Economics and Business* 33(3):202–11.

Kain, John F., and John M. Quigley. 1970. Measuring the Value of Housing Quality. *Journal of the American Statistical Association* 65:532–48.

Linneman, Peter. 1980. Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market. *Journal of Urban Economics* 8(1):47–68.

Malpezzi, Stephen, Larry Ozanne, and Thomas Thibodeau. 1980. *Characteristic Prices of Housing in Fifty-nine Metropolitan Areas.* Washington, DC: The Urban Institute Press.

Malpezzi, Stephen, Larry Ozanne, and Thomas Thibodeau. 1987. Microeconomic Estimates of Housing Depreciation. *Land Economics* 63(4):372–85.

Palmquist, Raymond B. 1979. Hedonic Price and Depreciation Indexes for Residential Housing: A Comment. *Journal of Urban Economics* 6(2):267–71.

Polinsky, A. Mitchell, and Steven Shavell. 1976. Amenities and Property Values in a Model of an Urban Area. *Journal of Public Economics* 5:119–29.

Rachlis, Mitchell B., and Anthony M. J. Yezer. 1988. Evaluating Risk in Real Estate Investment and Mortgage Lending Using Evidence from an Appraisal Equation. In *Real Estate Market Analysis: Methods and Applications*, ed. J. M. Clapp and S. D. Messner, 275–98. New York: Praeger.

Randolph, William C. 1988a. Estimation of Housing Depreciation: Short-Term Quality Change and Long-Term Vintage Effects. *Journal of Urban Economics* 23(2):162–78.

Randolph, William C. 1988b. Housing Depreciation and Aging Bias in the Consumer Price Index. *Journal of Business and Economic Statistics* 6(3):359–72.

Rosen, Sherwin. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82(1):34–55.

White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48(4):817–38.