THE EVALUATION OF RESIDENTIAL LIVING SPACE

David W. Rasmussen and Thomas W. Zuehlke*

Introduction

Hedonic regression analysis of a housing market typically includes total floor space and number of rooms in a standard semi-log specification. Although widely employed, this specification does not capture the way in which floor space and the number of rooms interact to determine room size, a prime determinant of the value of residential living space. An intuitive understanding of this problem is straightforward. The estimated coefficients of number of rooms and floor area in the semi-log specification frequently are positive, an expected result since both are generally desirable housing characteristics. However, the implication of this result is that an increase in the number of rooms, *ceteris paribus*, will raise house value. Given a 1,600-square-foot house, persons familiar with real estate markets would not be certain that the addition of another room would in fact have a positive impact. Some researchers have included an average room size variable (floor area divided by number of rooms) to capture the effects of average room size. Although a significant improvement, this approach does not adequately capture the interaction between floor area and number of rooms.

In this paper we compare alternative specifications of the semi-log hedonic regression for a specific market area. We find that the most common specification employing floor area and number of rooms, as well as one that includes average room size, may be misspecified. Further, a straightforward alternative specification is shown to avoid the specification error problem and improve modestly the predictive power of the regression.

Theoretical Considerations

The most frequently employed specification of the hedonic price function is a semi-log regression of selling price (*P*) on a vector of housing attributes that includes floor space (*F*) and the number of rooms (*R*).[1]

$$\ln(P) = B_0 + B_1F + B_2R + B_3X_3 + \ldots + B_kX_k + e \qquad (1)$$

With this specification, the marginal price of each attribute is proportional to its selling price, and the constants of proportionality are invariant to changes in the attribute vector under consideration.

$$\frac{dP}{dX_i} = B_iP, \qquad (2)$$

where

$i = 1$ to k.

Changes in attribute values will affect attribute marginal prices only to the extent that they change selling price. Thus, any two houses with the same selling price will have the same vector of attribute marginal prices, even though their attribute vectors may differ substantially. In addition, the attribute marginal price vectors of any two houses with different selling prices are proportional, and differ by the ratio of the selling prices.

A significant limitation of this typical specification is the constant rate at which any two attributes can be traded without affecting the selling price. This rate of trade is the same for all houses in a given market, regardless of selling price or attribute combination, i.e.,

$$\frac{dX_i}{dX_j} = \frac{-B_j}{B_i}. \qquad (3)$$

Stressing the non-linear nature of the hedonic price function, Murray (1983) shows that the marginal price of an attribute generally will vary in response to attribute variation. The same is true of the ratio of attribute marginal prices. That is, a buyer generally faces a non-linear budget constraint for the restricted problem of choosing the values of an attribute pair, given his choice of the remaining attributes and consumption of other goods. However, the typical semi-log specification requires the market rate of trade for an attribute pair to be constant and the same for all houses.

This limitation seems particularly restrictive for the floor space and number of rooms variables, which interact to determine room size. Relative scarcity of either of these attributes in a house is likely to affect the market rate of trade for these attributes. Specifically, the marginal value of an additional room will be high and the marginal value of additional floor space will be low when floor space is plentiful, relative to the number of rooms. Alternatively, an additional room will have little value and additional floor space will be highly valued when floor space is scarce

*The authors are professor and assistant professor, respectively, in the Department of Economics, Florida State University.

relative to rooms. Thus, we expect the marginal value of floor space will be a decreasing function of floor space and an increasing function of the number of rooms. Correspondingly, the marginal value of a room will be an increasing function of floor space and decreasing function of the number of rooms. Recognizing this interaction, some scholars add average room size to the specification in equation (1).[2]

$$\ln(P) = B_0 + B_1F + B_2R + B_3\left(\frac{F}{R}\right) + \text{other} \tag{4}$$

In this equation, the marginal price of floor space is a function of the number of rooms and selling price, while the marginal price of a room is a function of the number of rooms, floor space, and selling price.

$$\frac{dP}{dF} = \left[B_1 + B_3\left(\frac{1}{R}\right)\right]P \tag{5}$$

$$\frac{dP}{dR} = \left[B_2 - B_3\left(\frac{F}{R}\right)\left(\frac{1}{R}\right)\right]P \tag{6}$$

In order for Equation (6) to be the hypothesized increasing function of floor space and decreasing function of rooms, B_3 must be negative. There are two limitations inherent in this specification. First, the marginal price of floor space does not depend on floor space. Second, variation in the marginal price of either attribute is determined by the single parameter B_3. A more flexible specification can be obtained by adding squares and cross-products for floor space and the number of rooms to the specification of Equation (1).

In the following equation,

$$\ln(P) = B_0 + B_1F + B_2F^2 + B_3FR + B_4R + B_5R^2 + \text{other}, \tag{7}$$

the attribute marginal prices of floor space and rooms are functions of the number of rooms, floor space, and selling price.

$$\frac{dP}{dF} = (B_1 + 2B_2F + B_3R)P \tag{8}$$

$$\frac{dP}{dR} = (B_4 + 2B_5R + B_3F)P \tag{9}$$

If these variables interact as hypothesized, the marginal value of floor space will be decreasing with increasing floor space and increasing with the number of rooms, while the marginal value of a room will be increasing with floor space and decreasing with increases in the number of rooms. That is, we expect $B_2 < 0$, $B_3 > 0$, and $B_5 < 0$.

Choice among the alternative specifications in given equations (1), (4), and (7) is a potentially important empirical issue. If the additional flexibility of Equation (7) is important and the formulations in equations (1) or (4) are employed, then the possibility of specification bias emerges. This possibility is particularly important for studies in which interest centers on specific regression coefficients. Examples of studies with conclusions based upon the significance of specific coefficient estimates include studies of the impact of racial discrimination (King and Meiskowski, 1973), the impact of air and water pollution on housing values (Nelson, 1978), the capitalized value of variations in school quality (Jud and Watts, 1981), the impact of neighborhood quality on housing values (Li and Brown, 1980), and the effects of public policy (Vandell and Zerbst, 1984; French and Lafferty, 1984).

There are two questions of specific interest. First, are the constraints implicit in the usual specification in Equation (1) appropriate? If so, the alternative specifications would not be expected to have significantly higher explanatory power. Because equations (4) and (7) reduce to Equation (1) under linear homogeneous constraints on the coefficients, the usual F test for the exclusion of a subset of regressors may be employed to test the validity of these constraints.[3] Second, if these constraints are not appropriate, are the coefficient estimates from Equation (1) subject to specification bias? The RESET test of Ramsey and Schmidt (1976) was used to test each equation for specification bias. This test is capable of detecting specification bias due to omitted variables, incorrect functional form, or neglected simultaneity. The procedure was to supplement the set of regressors under consideration with powers of the predicted values of the dependent variable, and then test the joint hypothesis that the coefficients of the additional regressors were zero. If the F statistic is significant, then bias due to omitted variables, incorrect functional form, or neglected simultaneity also is significant.[4]

Estimation

The data used in this paper were obtained from the February 1982 Multiple Listing Service *Comparable Book* for Tallahassee, Florida.[5] The sample included observations on selling prices and housing attributes for 138 detached single-family residential properties. The structural attributes included the age of the dwelling in years, the lot size and floor area in thousands of square feet, the number of bedrooms, the presence of a garage or carport, and the presence of a paved road. Variables indicating the availability of an assumable mortgage and the occupancy status of the unit also were available. The neighborhood variables, available by census tract, included median family income in thousands of dollars, the percentage of families headed by a black or Hispanic and the travel time in minutes to the central business district.

Table 1 presents regression estimates for alternative specifications of the floor space and number of rooms variables. The estimates presented use the number of bedrooms rather than the total number of rooms. This is because the regressions using the number of bedrooms performed better than those with total number of rooms, in that they had lower sums of squared errors. This was reflected in higher t statistics for the coefficients of explanatory variables associated with the number of rooms/bedrooms (rooms, rooms squared, rooms times floor space, and floor space divided by rooms). The point estimates and test statistics for the other explanatory variables were not sensitive to the choice of the "rooms" variable. In addition, choice of the competing specifications was not affected by the choice of "rooms" variable. It is not clear why use of the number of bedrooms resulted in slightly higher explanatory power, but the general insensitivity of the point estimates and test statistics of the other variables to the "rooms" variable specification probably isdue to the strong correlation between the number of rooms and number of bedrooms. Most of the variation in the number of rooms result from variation in the number of bedrooms. In all cases, the dependent variable was the log of selling price.

Regression 1 is the typical specification with floor space and the number of bedrooms entering individually. Most of the structural variables have significant coefficients with the expected sign. The coefficient of floor space is strongly significant while the coefficient of bedrooms is insignificant at conventional levels. The coefficient of floor space reveals the marginal price of a unit 91,000 square feet), of floor space as 55 percent of the selling price, measured in thousands of dollars i.e., the marginal price of an extra square foot in a $100,000 house is $55. The coefficient of bedrooms estimate the marginal price of a bedroom to be negative 6 percent of the selling price in thousands of dollars.[6] This confirms the suggestion of French and Lafferty (1984), that more rooms in a house of fixed size may not add value. Taken together, these estimates suggest the improbable circumstance that adding a room of 111 additional square feet would not increase selling price. While the neighborhood variables are of expected sign, none is significant at conventional levels. The coefficient of assumable mortgage is unexpected and significant at the 10 percent level. Given that a buyer can always choose conventional financing, an assumable mortgage should not decrease selling price. Finally, the coefficient of vacancy has an unexpected sign but is insignificant.

This regression shows a good degree of explanatory power; the adjusted R^2 is 0.83 and the sum of squared residuals is 6.05. However, the RESET statistic is 7.23 with a five percent critical value of 2.68. This suggests a specification problem due to omitted variables, functional form, or simultaneity.

Regression 2 adds a measure of average room size, the ratio of floor space to number of bedrooms, as suggested by Vandell and Zerbst (1984).[7] The addition of this variable has little impact on the point estimates, except for those associated with floor space and rooms. The coefficient of average room size is positive and significant, while both floor space and number of bedrooms are insignificant. The marginal prices of floor space and bedrooms are given by the following equations:

$$MP_F = \left[0.09 + 1.53\left(\frac{1}{R}\right)\right]P, \text{ and} \tag{10}$$

$$MP_R = \left[0.24 - 1.53\left(\frac{1}{R}\right)\left(\frac{F}{R}\right)\right]P. \tag{11}$$

Evaluated at mean values of 1,800 square feet of floor space and three bedrooms, the marginal price of 1,000 square feet of floor space is 60 percent of selling price in thousands of dollars and the marginal price of a bedroom is –6.6 percent of selling price in thousands of dollars. Once again, the addition of a fourth bedroom of 110 square feet leaves selling price unchanged. These are, of course, very close to the values given by Regression 1. Finally, the coefficient of median family income is reduced and remains insignificant, while the coefficient of assumable mortgage is still of unexpected sign and marginally significant.

The adjusted R^2 for this regression increases to 0.84 and the sum of squared residuals is reduced to 5.85. This is a statistically significant increase in explanatory power at the five percent level; the F statistic for the exclusion of the measure of average room size is rejected with a test statistic of 4.23 and a five percent critical value of 3.92. However, the RESET test statistic remains significant, with a sample value of 5.92 and a five percent critical value of 2.68.

Regression 3 adds the squares and cross product of floor space and the number of bedrooms to the specification in Regression 1. This specification represents a significant improvement over the previous specifications. The coefficients of the non-linear terms have the expected sign and are significant. The coefficient of floor space squared is significant at the one percent level, a major reason why this specification is an improvement over the one using a measure of average room size. Recall that average room size does not allow the marginal value of floor space to decline with increases in floor space, a hypothesis that is strongly supported by the results reported in Regression 3. The coefficients of number of

Table 1

Alternative Specifications of Residential Living Space

Variable	Regression[a] 1	2	3
Constant	3.05* (12.90)	2.14* (4.27)	2.13* (5.58)
Age	–0.0084* (–3.51)	–0.0088* (–3.73)	–0.0072* (–3.26)
Lot Size	0.0015* (3.72)	0.0017* (4.16)	0.0015* (4.01)
Floor Area	0.55* (12.31)	0.0904 (0.3973)	0.395 (1.54)
Floor Area Squared			–0.24* (–3.30)
Floor x Bedrooms			0.33** (2.39)
Floor + Bedrooms		1.53** (2.05)	
Bedrooms	–0.0610 (–1.43)	0.2421 (1.58)	0.54*** (1.85)
Bedrooms Squared			–0.18** (–2.23)
Garage	0.22* (4.15)	0.22* (4.20)	0.15* (2.90)
Median Family Income	0.0055 (1.42)	0.0027 (0.6698)	0.0089** (2.19)
Minority Population	–0.12 (–0.73)	–0.2371 (–1.36)	0.11 (0.60)
Paved Road	0.27* (3.21)	0.25* (3.01)	0.22* (2.81)
Assumable Mortgage	–0.0662*** (–1.66)	–0.0662*** (–1.68)	–0.040 (–1.07)
Vacancy	0.0011 (0.0266)	0.0047 (0.1195)	–0.025 (–0.66)
Travel Time	–0.0036 (–0.6115)	–0.0030 (–0.52)	–0.0038 (–0.71)
SSE	6.048	5.85	4.97

Adjusted R^2	0.83	0.84	0.86
F^b	——	4.23	8.89
Reset	7.23	5.92	1.41

Note: Numbers in parentheses are corresponding t-values.
[a]The dependent variable is the logarithm of sales price.
[b]This F statistic tests the validity of the exclusion restrictions in equation (1).
*significant at 0.01
**significant at 0.05
***significant at 0.10

bedrooms squared and the interaction term are significant at the five percent level. The marginal value of a bedroom decreases as bedrooms increase and increases with floor space, while the marginal value of floor space decreases with floor space and increases with the number of bedrooms. The coefficient for bedrooms is significant at the 10 percent level, while the coefficient for floor space is significant at about the 15 percent level. The marginal prices of floor space and bedrooms are estimated by the following equations:

$$MP_F = (0.40 - .048F + 0.33R)P, \text{ and} \quad (12)$$
$$MP_R = (0.54 - 0.36R + 0.33F)P. \quad (13)$$

Evaluated at 1,800 square feet and three bedrooms, the marginal price of 1,000 square feet of floor space is 53 percent of the selling price in thousands of dollars, while the marginal price of a room is 5.4 percent of the selling price in thousands of dollars. While the marginal price of floor space at the mean is similar to the estimates obtained under the preceding specifications, the marginal price of a room now is positive. The addition of a fourth bedroom of 110 square feet will increase selling price by 11 percent.

The adjusted R^2 for this regression is 0.86, and the sum of squared residuals is 4.97. While this appears to be a modest increase in R^2 over the specification in Regression 1, it represents an 18 percent reduction in the sum of squared residuals. This is a statistically significant increase in explanatory power at the one percent level. The F statistic for the exclusion of the squared and interaction terms is rejected with a test statistic of 8.89 and a one percent critical value of 3.95. The RESET test statistic for this specification is 1.41, with a five percent critical value of 2.68. Thus, we cannot reject the null hypothesis that there is no specification bias due to omitted variables, incorrect functional form, or neglected simultaneity. A regression also was completed with the floor/bedrooms variable added to Regression 3 to test the hypothesis that this coefficient is not significantly different from zero in the presence of the squared and interaction variables. This regression is not reported, but the hypothesis that the floor/bedrooms coefficient is zero is accepted at the 0.05 level; the F value is 1.49 for a test with a critical value of 3.92.

Finally, there are important differences in the point estimates and test statistics between the third specification and the first two alternatives.[8] First, median family income is insignificant under the first two specifications, but significant with the inclusion of the non-linear terms. The potential importance for hypothesis testing of the inclusion of the non-linear terms easily can be demonstrated. Suppose the problem is to test the Leven and Mark (1977) hypothesis that median family income is a more important determinant of neighborhood quality than racial composition. The significance of median income in Regression 3 provides some support for this view, even though the income variable is not significant in the misspecified Regressions 1 and 2. Other results change significantly. The availability of an assumable mortgage is of unexpected sign and significant in the initial specifications, but is insignificant after the inclusion of the non-linear terms. The addition of the quadratic terms results in a general reduction in the coefficients of variables other than those involving floor space and the number of bedrooms. The point estimates of the coefficients of age, garage, paved road, and assumable mortgage all are reduced in excess of half a standard deviation. Perhaps the significance of many of these variables in the usual specification is due to correlation with the omitted non-linear terms. Finally, the point estimate of the coefficient of vacancy now is of the expected sign, but remains insignificant.

Conclusions

The interactive specification of floor space and number of bedrooms that is proposed here reveals the expected result i.e., the marginal value of floor space is a decreasing function of floor area and an increasing function of the number of bedrooms. The RESET statistic suggests that commonly used hedonic specifications are subject to specification bias, while finding little evidence of bias in the proposed alternative. The predictive power of the interactive model is significantly greater (at the 0.01 level) than the alternatives.

These results are particularly significant for the myriad of studies that employ hedonic regression to analyze problems in which the conclusions depend on the significance of a specific coefficient. Conclusions reported in such studies, including those of neighborhood amenities, capitalization of locational advantages, air and water pollution, and racial discrimination may be inappropriate if the authors fail to consider the interactions between the floor area and room variables. In this study, one important neighborhood variable, median family income, was significant only in the interactive model. Hence, while there can be no presumption that the conclusions of the earlier studies were incorrect, the data in this study suggest that including the interaction between floor area and room size would yield more reliable estimates.

Footnotes

[1]The recent literature does not reveal a preferred specification of structural attributes. Examples of studies employing both floor area and number of rooms (or bedrooms) are Bender and Hwang (1985), Lang and Jones (1979), and Dale-Johnson and Phillips (1984). Census and the Annual Housing Survey data do not include floor area, a serious but inevitable constraint in many studies. Anas and Eum (1984) use tax assessor data for a model that includes floor area but not number of rooms; Jackson, et al. (1984) use real estate data and include floor area and number of bathrooms in their regressions.

[2]Vandell and Zerbst (1984) use average room size, while King and Mieskowski (1973) use floor area squared divided by number of rooms.

[3]Amemiya (1980) has shown that the most frequently used criteria for selection of regressors differ only in their choices of the optimal significance level for an F test of the excluded subset of regressors. Among these criteria are the adjusted R^2 and the "prediction criteria" which minimize the mean square error of prediction. Thus, the test under consideration here is fairly general.

[4]This test is based on the ability to predict the residual means with the higher order moments of the conditional expectation of the dependent variable. Ramsey (1969) shows the residual means are non-zero and non-constant in the case of specification error and can be predicted with the higher order moments of the conditional expectations of the dependent variable unless the residual means are orthogonal to the included variables, in which case there is no specification bias.

[5]The Tallahassee MLS accounts for approximately 75 percent of all residential sales. Since MLS sales are conducted through real estate brokers who receive commissions, it is likely that only "arms length" transactions are included. Hence, each observation in our sample represents a market determined price, a desirable attribute that is not characteristic of all sales outside the MLS system. Thus, the MLS data are highly representative of single family residential sales in this market.

[6]The negative bedroom coefficient is not unusual. Bender and Hwang (1985) make multiple runs of a model similarly specified in which the bedroom coefficient takes significant positive and negative values in different subsample regressions.

[7]While this measure cannot be interpreted as average room size, it is strongly correlated with average room size and performs better as a regressor. Use of average room size did not change any results.

[8]If the additional terms are collinear with each other or any of the original regressors, then the t statistics of these terms are likely to be insignificant. Given the significance of the t statistics of the non-linear terms, the importance of the non-linear terms as a group (reflected in the significant F statistic), and the general increase in the t statistics of the remaining variables, we conclude that multicollinearity is not a problem.

References

Amemiya, Takeshi. "Selection of Regressors." *International Economic Review*. 21 (June 1980). 331–354.

Anas, Alex, and Sung Jick Eum. "Hedonic Analysis of a Housing Market in Disequilibrium." *Journal of Urban Economics*. 15 (January 1984). 87–106.

Bender, Bruce, and Hae-Shin Hwang. "Hedonic Housing Price Indices and Secondary Employment Centers." *Journal of Urban Economics*. 17 (January 1985). 90–107.

Dale-Johnson, David, and G. Michael Phillips. "Housing Attributes Associated with Capital Gains." *AREUEA Journal*. 12 (Summer 1984). 162–175.

French, H. E. III, and Ronald N. Lafferty. "The Effect of the California Coastal Commission on Housing Prices." *Journal of Urban Economics*. 16 (July 1984). 105–123.

Jackson, Jerry R., Ruth C. Johnson, and David L. Kaserman. "The Measurement of Land Prices and the Elasticity of Substitution in Housing Production." *Journal of Urban Economics*. 16 (January 1984). 1–12.

Jud, G. Donald, and James M. Watts. "Schools and Housing Values." *Land Economics*. 57 (August 1981). 459–470.

King, A. Thomas, and Peter Mieskowski. "Racial Discrimination, Segregation, and the Price of Housing." *Journal of Political Economy*. 81 (May/June 1973). 590–606.

Lang, James R., and Wesley H. Jones. "Hedonic Property Valuation Models: Are Subjective Measure of Neighborhood Amenities Needed?" *AREUEA Journal*. 7 (Winter 1979). 451–465.

Leven, Charles L., and Jonathan H. Mark. "Revealed Preferences for Neighborhood Characteristics." *Urban Studies*. 14 (June 1977). 147–159.

Li, Mingche M., and H. James Brown. "Micro-Neighborhood Externalities and Hedonic Housing Prices." *Land*

Economics. 56 (May 1980). 142–154.

Murray, Michael P. "Mythical Demands and Mythical Supplies for Proper Estimation of Rosen's Hedonic Price Model." *Journal of Urban Economics*. 14 (November 1983). 327–337.

Nelson, Jon P. "Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality." *Journal of Urban Economics*. 5 (1978). 357–369.

Ramsey, James B. "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis." *Journal of the Royal Statistical Society*. Series B. 31 (1969). 350–371.

_______, and Peter Schmidt. "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests." *Journal of the American Statistical Association*. 71 (June 1976). 389–390.

Vandell, Kerry, and Robert H. Zerbst. "Estimates of the Effect of School Desegregation Plans on Housing Values Over Time." *AREUEA Journal*. 12 (1984). 109–135.

Appendix A

Mean and Standard Deviation of Variables Used in Regression Analysis

Variable	Mean	Standard Deviation
Sales Price*	79.271	52.030
Age	11.83	9.82
Lot Size**	31.282	59.265
Floor Area**	1.788	0.735
Bedrooms	3.20	0.66
Garage	0.83	0.37
Median Family Income*	23.776	7.146
Minority Population (percent)	0.139	0.146
Paved Road	0.942	0.235
Assumable Mortgage	0.609	0.490
Travel Time (minutes)	17.74	3.731
Vacancy	0.457	0.500

*in thousands of dollars
**in thousands of square feet