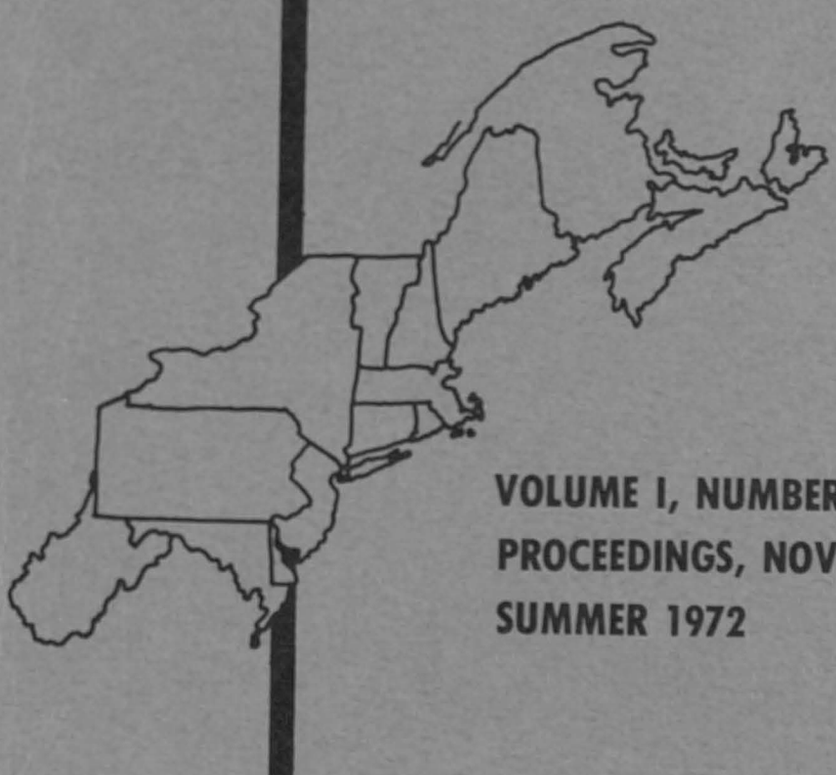
JOURNAL OF

Northeastern Agricultural Economics Council



VOLUME I, NUMBER I
PROCEEDINGS, NOVA SCOTIA
SUMMER 1972

USING MULTIPLE LISTING SERVICE DATA TO ANALYZE DETERMINANTS OF RESIDENTIAL PROPERTY VALUES

J. B. Penn and George D. Irwin*
Agricultural Economists
Economic Research Service
U.S. Dept. of Agriculture
Purdue University

Housing needs in the U.S. since World War II have experienced tremendous expansion and this expansion has been accompanied by a myriad of problems--transportation lags, urban blight and sprawl, decay of central cities, poverty, pollution, racial discrimination, rural slums, and the like.

The recent focus of national attention on domestic social issues has indicated that residential areas are often of central concern to these problems. Attempts to find practicable solutions have accentuated the need for a better understanding of the complex facets of housing structures. The relation of property value to use and to property tax base are central components of such analyses. Recent studies in this area have predominately concentrated on property value as related to one or more social problems such as race, pollution, location, other amenities, developmental patterns, and population densities [1,2,4,5,6,7,9].[1/] However, an information gap is keenly acute in the identification and measurement of the overall determinants of residential property values. A principal problem has been to obtain data. The objective of this paper is to suggest the use of a previously untried and more comprehensive data source.

An Overview. Multiple listing service (MLS) information on residential transactions was used for an empirical analysis of the determinants of urban residential property values for a sample area at the edge of Lafayette, Indiana. A simple economic model is postulated[2/] to give context in which to examine the empirical data. Statistical models are then developed, estimated, and results presented. Finally, implications are drawn and conclusions made.

---

* Helpful comments were received during preparation of this paper from George S. Tolley while visiting professor at Purdue.

1/ Bracketed numbers refer to references at the end of this article.

2/ Economists have long been concerned with real property values and costs. While much of the early work, such as the theorizing of Ricardo, concentrated on agricultural land, more recent studies have treated urban property.

Data Sources. The basic data for this analysis are residential unit transfers which occurred during the period January, 1965 through June, 1970. The area chosen for analysis is a housing development on the edge of a small city of approximately seven square blocks in size and relatively homogeneous in terms of housing stock, lots, age of development, and occupants income. Average values of selected characteristics are shown in the appendix.

Information was obtained on 80 units transferred after being listed for sale with Lafayette real estate agencies. This information included location, attributes of house and lot, information on financing, and other data relevant to each property. A sample form illustrating these data is shown below.

SAMPLE FORM ILLUSTRATING DATA AVAILABLE ON RESIDENTIAL TRANSFERS*

8060	2400 Potawatomi Dr.	Laf. Ind	6	3	1	Residence	$18900.00
Listing No.	Address	Location	Rooms	Bdrms	Baths	Type Property	Price

Lafayette Board of Realtors - Multiple Listing Service - Lafayette, Indiana

GENERAL INFORMATION		CONSTRUCTION DETAILS			ACCESSORIES, EXTRAS	
			Size-1st floor	Size-2nd floor		
Construction	Frame Alum.	L.R.	12 x 18		Insulation	2 - 2
Builder	Price & Price	D.R.			Storms	alum. comb
Age	6 years	Kit.	14'6 x 15'2		Screens	
Condition	Excellent	Bath	5 x 7 w/ shower and tile		Fireplace	
Possession	30 DAFC	B.R.	14'3 x 9'4		Carpeting	LR Hall
Encumbrance	$13895.52 Int. 5½+½	B.R.	12'2 x 9'2		Draperies	all
Favor of	NHAC #103290-13	B.R.	9'0 x 8'10		Built-ins	
Mo.payments	$ 122.00 PITI	B.R.	8'10		TV antenna	cable
Commitment	(78.67PI)	Fnd.size	28'10 x Mat. conc		Disposal	
Terms	cash equity-MGIC	Rec.room	12 x 12		Dishwasher	
Trade for		Basement	no		Water htr.	30 gas
Lot Size	94x119x113x93	Garage	12'6 x 20'		Air Cond.	
Lot Number	55 Tec. II	Porches	front		Driveway	conc
Asd. Value	5090-ME= 4090	No.closets	5 2 double		Fenced yard	
Taxes	$377.92	Heat	gas perimeter		City water	yes
Public school	Miami Tec.	Floors	tile and carpet		Well depth	
Assessment	5020-ME= 4020	Int. walls	DW painted		San.sewer	yes
Transportation	bus Jeff Square	Roof	asph seal tab		Septic tank	
Legal Description		Windows	alum		Gas	yes
					Elec.- volts	220 v

REMARKS -- All information herein, while believed correct, is not guaranteed.

TV antenna not included

Owner Harbolt Address same Occupant owner Phone 474 5303

Listed by Price & Price Phone 447 5051 Date 2-9-70 Rec'd. 2-9-70 Xp. 5-9-70

Printed by NATIONAL REAL ESTATE ADVERTISING COMPANY, Box 1007, Dayton, Ohio 45401

8060 2400 Potawatomi Dr. 6-3-1 $18,900

* These forms usually include a picture of the property and the final sales price is appended.

Most studies of this type conducted in the past have relied on transaction records filed in the various offices of governmental units. These records consist of deed or mortgage recordings or of assessment information compiled for tax purposes. The disadvantages of these data sources are numerous and a major purpose of this paper is to evaluate the data available on urban transfers from Board of Realtor multiple listing services.

The use of MLS data appears to overcome some of the deficiencies inherent in public data sources. MLS data are very detailed for each transaction including physical, financial, and location information which should afford more explanatory power (higher R^2's) than usually available from large samples. In addition, the data are usually easily accessible (although not public records), and easy to prepare for analysis. The remainder of this paper is a report of an analysis utilizing this data source. A major limitation is that transactions accomplished by private sale are omitted. To the unknown extent that such sales may be of the more desirable properties, the results here would be unrepresentative. Too, private sales may be relatively more common in rural housing areas. More likely though, the rural market may be a mixture of rentals on surplus farm houses by low income persons and private purchase of rural residences for more affluent persons seeking amenities of rural living.

Economic Model. The determination of residential property price occurs in an imperfect market. Housing stock and accompanying land (lot) are greatly differentiated, trading is relatively infrequent, varying degrees of bargaining occur, transaction costs are high, and often a large information gap is present. In addition, prices are affected by outside forces such as "tight" and "easy" money, interest rates, the current situation in the construction industry, and the general state of the national economy.

For the analysis reported here, the value of a residential unit (v) is hypothesized as being functionally related to two broad categories of variables, physical attributes (PA) and amenities (A). The economic model may thus be written as

$$V = V(PA,A)$$

Numerous explanatory variables are subsumed under each category. These subsumed variables will obviously differ among urban areas within and among geographic areas. The specific variables chosen for this analysis are detailed in the section describing the results of estimation. Since the estimates are made for a fairly small homogeneous area some local market considerations required in a broadly applicable model could be omitted in this study. However, the 1965-70 period did include a good deal of diversity of external economic conditions.

Statistical Model. A single equation statistical model of the multiplicative type was specified for this analysis. This model may be written as

$$V = \alpha Z_1^{\beta_1} Z_2^{\beta_2} \ldots Z_k^{\beta_k} D_1^{\gamma_1} D_2^{\gamma_2} \ldots D_j^{\gamma_j} \varepsilon$$

where V is an endogenous variable.
$Z_1 \ldots Z_k$ are exogenous variables,
$D_1 \ldots D_j$ are zero-one variables,
β's, α's, and γ's are parameters,
ε is the disturbance.
Since this model is intrinsically linear, it may be expressed, by suitable transformation of the variables, in the conventional linear model form and estimated by standard matrix linear regression procedures. Taking logarithms to the base ε in the above equation converts the model into the linear form

$$\ln V = \ln\alpha + \beta_1 \ln Z_1 + \beta_2 \ln Z_2 + \ldots + \beta_k \ln Z_k + \gamma_1 \ln D_1 + \ldots \gamma_i \ln D_j + \ln\varepsilon ,$$

It should be noted that the requirements for valid tests of significance and confidence internal tests are now $\ln\varepsilon \sim (0, I\sigma^2)$ (rather than ε). An examination of the residuals from the equation estimated here provides no evidence to refute this assumption. Such a model implies declining marginal value of additional units of each variable (when coefficients are less than one).

Results - Logarithmic Model. Ordinary least squares estimates of the parameters are shown here, along with the standard error of estimates and t values in parenthesis directly below

P =	-3.287	$\underline{YR}^{2.447}$	$AGE^{0.004}$	$ENC^{0.128}$	$LS^{-0.080}$	$\underline{RM}^{0.297}$	$\underline{BRM}^{0.214}$
	(2.94)	(0.70)	(0.05)	(0.07)	(0.09)	(0.10)	(0.11)
	(-1.12)	(3.49)	(0.07)	(1.87)	(-0.88)	(2.86)	(1.98)
	$BTH^{0.057}$	$\underline{HS}^{0.198}$	$\underline{GR}^{0.055}$	$TS^{-0.012}$	$TP^{0.010}$	$\underline{OT}^{-0.067}$	$\underline{LOC}^{0.095}$
	(0.04)	(0.06)	(0.02)	(0.03)	(0.007)	(0.03)	(0.03)
	(1.48)	(3.43)	(2.62)	(-0.38)	(1.42)	(-2.21)	(-2.89)

$R^2 = 0.857$ 80 observations

Definition of variables:
P sales price (000 dollars),
YR year of sale,
AGE age of housing stock,
ENC amount of existing encumberance (000 dollars),
LS lot size (square feet),
RM total number of rooms,
BRM number of bedrooms,

BTH number of bathrooms,
HS house size (square feet),
GR existence of garage (0 if no garage, 1 if garage existed),
TS type of structure (0 if conventional, 1 if bi-level),
TP time until possession (days),
OT occupancy by owner or tenant (0 if owner, 1 if tenant),
LOC location (0 if not on street bordering industrial area, 1 if on street adjacent to industrial area.

Overall, the equation is quite satisfactory in explanatory power and indicates that approximately 86 percent of the variation in sales prices is associated with the exogenous variables specified in this equation. Seven of the 13 exogenous variables, underlined in the equation, are significant at the 0.05 significance level.

The coefficients of the variables can be interpreted directly as elasticities indicating the percentage change in the sales price of the property for a one percent change in the exogenous variable, ceteris paribus.

An examination of the coefficients (which indicate the relative effects of the variables) reveals that year of sale has the strongest influence on sales price, number of rooms and number of bedrooms are of intermediate importance, and the other variables exhibit lesser influence. Price inflation and other associated variables apparently averaged 2.447 percent per year.

The variable representing existing encumberance was included to determine if the possibility for assumption of an existing mortgage, at an interest rate usually less than prevailing rates on new mortgages, was valued by buyers (an asset to sellers) and subsequently bid into price. The non-significance of the variable does not support the contention for this sample. However, the positive sign on the coefficient may be indicating that buyers do not possess the relatively larger amounts of cash necessary for assumption and that they are willing to pay more in total if a smaller down payment will gain possession of the unit. If so, size of downpayment would be more important than size of subsequent monthly payment in the eyes of the buyer. Too, newer properties may have higher mortgages, and thus some of the age effect might be picked up by this variable.

The negative sign on the coefficient of the lot size variable was not as expected. However, the sign and non-significance of the variable can be explained by an examination of the sample data. The subdivision was developed by one firm and lot sizes show little variability. All observations on the size variable are clustered closely around the mean (7,333 square feet), thus the variable has little or no explanatory power in this sample.

Variables representing number of rooms and number of bedrooms are significant and relatively strong while the variable representing number of bathrooms is non-significant. Both the variable representing house size and garage are significant and possess the expected positive sign.

The zero-one variable representing type of structure is not significant and has a negative sign. Bi-level houses tend to have a larger number of square feet and possibly some of the influence of this variable was assumed by the house size variable.

The variable representing time until possession after closing has a positive sign apparently indicating (for the range of data in this sample) that immediate possession after purchase is generally not desirable to buyers, or that owners anxious to sell were willing to accept lower prices.

Whether a residential unit was occupied at time of sale by the owner or tenant was included to test the preconception that houses occupied by tenants are associated with lower sales prices. The negative coefficient and significance of the variable does not refute this assertion. The lower sales price for tenant occupied units could possibly be explained by owners living outside the area being unfamiliar with current prices and selling for less. Or this variable could be measuring quality, indicating that rented houses are less well-kept than owner-occupied ones.

An examination of the residuals after estimating the equation with only the variables discussed so far indicated that an important variable had been omitted. Inspection of the data indicated that units located on a street which is adjacent to an industrial area had large negative residuals. Hence a zero-one variable to represent location was added, and it indicates that some discounting of price occurs for units located on this street. All coefficients discussed above are for the model including a location variable.

Results - Alternative Formulations. Several alternative formulations of the above model were constructed and estimated, concentrating primarily upon amenities. Several location variables were included in an attempt to measure location effects of area schools and a nearby park. Also, location variables for units on streets with heavy traffic flows were included at one stage. These alternative formulations were made after the one presented above because it was felt the area chosen for study was too small (approximately 7 square blocks), and had too few observations for distance and location variables to be meaningful. The results of the reformulations tend to bear this out.

The model discussed above was also respecified by changing the functional form to linear in the variates. The hypothesis for doing so is that the data range is so narrow in our sample that the intrinsic

value of having a declining marginal term, which is provided by the log form, is not of great importance with current data. The results of this estimation are presented below in the same format as above.

$$
\begin{aligned}
P = {} & -42516.77 + 660.20\,\underline{YR} + 30.89\,AGE + 0.187\,ENC - 0.112\,LS \\
& (13284.98) \quad (203.00) \quad (181.77) \quad (0.097) \quad (0.222) \\
& \quad (3.20) \quad (3.25) \quad (0.17) \quad (1.92) \quad (-0.503) \\
& + 816.01\,\underline{RM} + 1156.5\,\underline{BRM} + 808.17\,BTH + 2.73\,\underline{HS} + 1045.18\,\underline{GR} \\
& (296.53) \quad (588.5) \quad (470.57) \quad (0.868) \quad (400.93) \\
& \quad (2.75) \quad (1.97) \quad (1.72) \quad (3.14) \quad (2.61) \\
& - 240.64\,TS + 10.03\,TP - 839.98\,OT - 1425.79\,\underline{LOC} \\
& (596.36) \quad (8.80) \quad (590.70) \quad (619.29) \\
& (-0.404) \quad (1.14) \quad (-1.42) \quad (-2.30)
\end{aligned}
$$

$R^2 = 0.837$ 80 observations

Comparing the results of the equation with the log form reveals one less variable significant at the 0.05 level (occupancy by owner or tenant) and no divergences in signs of coefficients. Explanatory power remains high (84 vs. 86 percent), and the same variables are important elsewhere.

The coefficients of this equation can be read directly as the marginal value of an additional unit of the variable. The coefficients appear realistic in magnitude and provide direct dollar valuation estimates not generally available.

The coefficient of the variable representing year of sale indicates an annual price appreciation of $660. On the average price ($18,118) of the sample data, this is 3.6 percent. The variable representing age of the housing stock is also positive. This is not unexpected as houses in the area range in age from 3 to 9 years and additional development such as landscaping, fencing, etc., has occurred. This tends to increase value or at least overcome depreciation in the range of ages studied here.

The remaining variable noteworthy of comment here is the location variable. The coefficient of this variable indicates the magnitude of price discounting ($1,426.00) occurring on units located on the street adjacent to the industrial area. Residential units on this street sell on the average for about 8 percent less than units located elsewhere in the area. The remainder of the coefficients have a straightforward interpretation and will not be discussed to conserve space.

It appears that the range of data analyzed here is so limited that the choice of one functional form over the other could not be easily justified. But it is equally clear that the linear form would

not be useful in an extreme case. For example, it is doubtful that one could get an additional $1,156.50 for an eighth bedroom, though he might do so for a 3rd or 4th.

Conclusion. This analysis was directed toward identifying and measuring the determinants of residential property values using Board of Realtor MLS data. The use of transfer data from the multiple listing service of real estate boards shows considerable merit for analyses of this type. It might be especially useful also, to make a study of comparative characteristics of properties moved privately or by sole listing. Such an approach might be particularly appropriate in open areas. Of course, detailed data of the multiple listing type would have to be gathered, the sample would be smaller, and the specification of variables would probably differ. But our results do suggest that such data might be expected to use useful for analysis.

Both physical attributes and amenities were found to be major components of residential property values as expected. A weakness of this test of the use of MLS data was consideration of an area too small to ascertain distance and location influences. Subsequent studies of the residential real estate sector should consider much larger areas of a less homogeneous nature to fully explore amenity influences, and will ultimately need to relate to the rental housing market. However, this exploratory study does provide estimates for one section of the city not heretofore available.

Selected References

[1] Adams, F. G., G. Milgram, E. W. Green, and C. Mansfield, "Undeveloped Land Prices During Urganization: A Micro-Empirical Study Over Time," Rev. of Econ. & Stat., Vol. L, No. 2, May 1968.

[2] Bailey, Martin J., "Effects of Race and Other Demographic Factors on the Values of Single-Family Homes," Land Economics, Vol. XLII, No 2, May 1966.

[3] Bailey, M. J., R. F. Muth, and H. O. Nourse, "A Regression Method for Real Estate Price Index Construction," J. of the Amer. Stat. Assn., Vol. 58, No. 304, December 1963.

[4] Brigham, Eugene F., "The Determinants of Residential Land Values," Land Economics, Vol. XLI, No. 4, November 1965.

[5] Harris, R. N. S., G. S. Tolley, and C. Harrel, "The Residence Site Choice," Rev. of Econ. & Stat., Vol. L, No. 2, May 1968.

[6] Kitchen, J. W., and W. S. Hendon, "Land Values Adjacent to an Urban Neighborhood Park," Land Economics, Vol. XLII, No. 3, August 1967

[7] Nourse, Hugh O., "The Effect of Air Pollution on House Values," Land Economics, Vol. XLIII, No. 2, May 1967.

[8] Olsen, Edgar O., "A Competitive Theory of the Housing Market," The Amer. Econ. Rev., Vol. LIX, No. 4, Pt. 1, September 1969.

[9] Ridker, R. G., and J. A. Hemming, "The Determinants of Residential Property Values with Special Reference to Pollution," Rev. of Econ. & Stat., Vol. XLIX, No. 2, May 1967.

Appendix Table 1
Mean Values of Selected Characteristics, Eighty Observations, Sample Area, Lafayette, Indiana.

Characteristic	Unit	Mean
Sales Price	$	18,118.04
Age of housing stock	Years	6.63
Encumberance	$	13,062.39
Lot Size	Square Feet	7,333.60
Rooms	Number	6.09
Bedrooms	Number	3.16
Bathrooms	Number	1.48
House Size	Square Feet	1,213.53
Garage	Number	0.61
Type of Structure	Number	0.28
Time to Possession	Days	10.03
Occupancy (Owner-Tenant)	Number	0.13
Location	Number	0.13