J. ENVIRON. SYS., Vol. 4(3), Fall, 1974

ASSESSMENT OF URBAN RESIDENTIAL PROPERTIES: AN EMPIRICAL STUDY OF PITTSBURGH[1]

ROBERT CARBONE
REINHARD S. LAI
School of Urban and Public Affairs
Carnegie-Mellon University
Pittsburgh

ABSTRACT

The purpose of this paper is twofold: to examine whether the current assessment practice in Pittsburgh results in systematic non-uniformity, and to test the feasibility of a computerized assessment procedure based upon multiple regression analysis. Particularly, the predictors of market value used, which include locational and environmental factors, are all objectively quantifiable and justifiable on theoretical ground. A random sample of 245 single and double family dwelling units were drawn from records of approximately 3,000 bona fide transactions in 1970. The result reveals that the current assessment practice systematically underestimate high-value properties. Four groups of factors are postulated important in affecting the market value of a residential property. They are accessibility, site characteristics, environmental features, and building components. The structure presented here is showed feasible in computerizing the assessment of residential properties. Finally, the advantages and disadvantages of implementing the suggested framework are discussed.

The principal function of an assessor is to assess properties in as uniform a relationship to their market values as possible. The annual maintenance of uniform assessment rolls has become an almost impossible task in most urban centers. In the last decade, major criticisms of the current assessment practice have been made vis-à-vis efficiency and equity. The solution of adding more staff to increase frequency and upgrade assessment quality is no longer acceptable as in the past. It would not only create greater internal problems of standardization

[1] The research on which this study is based was given financial support by the Ford Foundation. R. Lai currently is also a consultant to Research Division, Maryland Department of Economic and Community Development. The authors express their thanks to Chang-i Hua, J. B. Kadane and T. McGuire for their helpful comments on an earlier draft.

doi: 10.2190/GVGF-KYBE-AHTP-MPBD
http://baywood.com

but also contribute to rapid increases in government cost which the public would disfavor.

The purpose of this paper is twofold. Using data of the City of Pittsburgh, this study attempts first to present some empirical findings on whether or not the current manual assessment practice in Pittsburgh results in systematic non-uniformity in residential assessment. Second, we propose a computerized assessment procedure based upon multiple regression analysis. The regression formula developed, in comparison to the ones in use in a few urban centers, intends not only to predict the selling prices of residential properties but also to explain the determinants of property values. The predictors of market value used, which include locational and environmental factors, are all objectively quantifiable and justifiable on theoretical grounds.

The principal function of the system presented is not to replace assessors but rather to assist them in their tasks. If attempts to make use of the assessor's time in such a way that it would mostly be consumed in the analytical evaluation of property values.

Bias in Assessment-Sales Ratios

In order to determine whether or not the current assessment practice, as administered in Pittsburgh, produces equalized residential assessment rolls, one must answer questions relative to the extent and nature of the variation of assessment-sales ratios.

The data used here are based upon a random sample of 245 single and double family dwelling units drawn from records of approximately 3,000 bona fide transactions which occurred during 1971 and on information from the 1970 Bureau of Census data on population and housing characteristics in Pittsburgh.

With respect to the extent of the variation, Pittsburgh seems to exhibit wide dispersion about the mean or median ratio in 1971. Whereas the average ratio of the properties sampled was .43, a deviation of only .07 from the ratio set by law, the standard deviation estimated was about 48% of the average.[2] Assuming that assessment-sales ratios are normally distributed, this means that approximately 68% of the observations had ratios ranging from .22 to .64.

In order to describe the nature of the variation a multiple regression analysis is employed. Assessment-sales ratios are regressed (a constant included) on market values and several other variables reflecting the character of the neighborhood where each property is located.[3] The results of the regression based upon our sample are shown in equation (1)

[2] In Pittsburgh the assessed value of a property is set by law at 50% of its market valuation.

[3] David E. Black, "The Nature and Extent of Effective Property Tax Rate Variation within the City of Boston," *National Tax Journal*, June 1972, pp. 202-209. The major difference between our approach and that of Black lies in the dependent variable employed. Whereas Black regressed the average census tract assessment-sales ratios on the corresponding mean value of owner-occupied dwellings and a set of environmental variables, our interest focuses on explaining the variation for individual properties.

$$\begin{aligned} ASR = {} & \underset{(10.62)}{.45} + \underset{(2.24)}{.0065\,LIC} + \underset{(3.59)}{.31\,CRT} - \underset{(-1.18)}{.0037\,PSU} - \underset{(-1.01)}{.0006\,PRC} \\ & - \underset{(-0.47)}{.0006\,DNS} - \underset{(-2.96)}{.000004\,TRV} \end{aligned} \tag{1}$$

$$R^2 = .19$$

$$F = 6.67$$

where

ASR = assessment-sales ratios,

LIC = proportion of low-income families (census tract),

CRT = number of crimes reported by 10,000 population (census tract),

PSU = proportion of housing units lacking plumbing facilities (census tract),

PRC = proportion of non-whites (census tract),

DNS = population density (census tract),

TRV = transaction values.[4]

The coefficient of determination of equation (1) indicates that only 19% of the variation is explained by the above relationship. However, this does not imply that no relationship exists since the value of the F-ratio indicates that we can reject the null hypothesis (no relationship at a .01 significance level.

The direction of the various regression coefficients estimated in equation (1) indicates on the one hand that, *ceteris paribus*, assessment-sales ratios bear a significant positive relationship to the percentage of low-income families and the crime rate in the surrounding neighborhood and a significant negative association to transaction values. On the other hand, the findings show, other things remaining constant, negative but insignificant relationships of the racial composition, the proportion of substandard housing units and the population density with assessment-sales ratios.[5] Therefore, the results tend to suggest that low-value properties which generally are situated in poor neighborhoods exhibiting high crime rates are assessed at a higher proportion of their market valuation than high-value properties in affluent neighborhoods. This implies that the current assessment practice systematically favors individuals living in high-value residential properties.

[4] The parentheses below the estimated coefficients contain the computed t-values. A coefficient is significant at a 0.05 confidence level if its t-value in absolute terms is greater or equal to 1.64.

[5] There is, of course, some multicollinearity in the regression, especially between the five variables reflecting the character of the neighborhood. The supporting simple correlation matrix can be made available upon request. It then becomes very difficult, if not impossible, to disentangle the individual influences of the various explanatory variables. Thus, a more proper way to interpret the results would be to focus on the group effect. Consequently, in our study, we can conclude that the attractiveness of the environment where a property is located has a negative effect on assessment-sales ratios. For a good discussion on the multicollinearity problem see, J. Johnston, Econometric Methods (New York: McGraw-Hill Co., 1972) pp. 159-168.

On the basis of these results, assessors in Pittsburgh appear not to be fulfilling their function. Our intention is not to conjecture on the possible reasons for this non-uniformity in assessments. Instead we will focus our attention on the remedy to this problem, that is, on the development of an objective procedure ensuring that each property be assessed at the lawful proportion of its market value.

A Computerized Residential Property Assessment Program

The California State Board of Equalization has initiated a step towards computerization of property assessments. Multiple regression analysis was used to appraise single-family residential properties in five counties (Orange, Alameda, Riverside, San Mateo and Santa Clara). Success has been claimed in both providing uniform assessment rolls and substantially reducing administrative costs.[6]

The function of the system developed in California is exclusively aimed at predicting selling prices, however. The number of predictive variables used generally exceeds 100. The weights assigned by the regression formula to many variables are therefore more than likely to be of dubious value because of the expected high degree of collinearity between several property characteristics used. More important, a large number of rank variables subjectively determined for each property by assessors are included in the equation. Finally, the impact of locational and environmental factors are not considered. Instead, a different regression equation is run in each sub-area. This would be an unnecessary task if these factors were included. This last drawback also renders any attempt to segregate the land value from improved property sale impossible.

MODEL FORMULATION

If we assume at a given point in time a competitive market situation in equilibrium, the market value of a residential property is in a broad sense a result of the quality and quantity of the housing services it provides. It is influenced by the natural features of a site and the activities of man, i.e., building improvement and provision of public services, which have altered the natural conditions. More specifically, the market value can be assumed to be linearly dependent on various building, site and surrounding environmental characteristics, that is,

$$MV_i = \sum_{k=1}^{n} \alpha_k X_{ik} + \sum_{\ell=1}^{m} \beta_\ell Y_{i\ell} + u_i$$

[6] Robert H. Gustafson, "Can the Computer Assist the County Assessor in the Valuation Process " paper presented at the 11th Annual Meeting of the Institute of Management Sciences Los Angeles, 1970 Ted Gwartney, "A Computerized Assessment Program," in D. M. Holland, *The Assessment of Land Values*, (Madison: University of Wisconsin Press, 1969), pp. 125 142.

where MV_i is the market value of the i-th property; X_{ik}, ($k = 1, \ldots, n$) are accessibility; the site and environmental features; $Y_{i\ell}$, ($\ell = 1, \ldots, m$), the building characteristics; α_k and β_ℓ are the regression coefficients; and u_i, the error term. Given the estimated coefficients and the land and building characteristics of a property, it is possible to predict objectively its market value and, therefore, estimate its assessment.

FACTORS INFLUENCING THE MARKET VALUE OF A RESIDENTIAL PROPERTY

The objective land and building factors considered in our model can be subdivided into four groups:

1. accessibility,
2. site characteristics,
3. environmental features, and
4. building components.

The accessibility of a property to desired urban activities (employment, recreation, social activities, etc.) should influence positively the service level of a property. Since the central business district (CBD) is the primary center of employment and service activities, households should prefer to live close to the central city, other things being equal, in order to minimize their traveling time and expense to their places of employment and leisure. Also, because the study is confined to residential properties, proximity to a public park should also be of concern to households.

Site features also are obvious determinants of value. Two site factors, 1) the foot frontage, and 2) lot size, are considered here despite some overlapping in their characteristics. In the present context, the concept of frontage refers to the length of all sides of a lot bounded by streets. Therefore, implicitly included in the frontage is the corner characteristic of a site. Both factors are expected to have a positive influence on property values. A larger lot size dictates a higher price while a lot with greater frontage is often thought to be more desirable.

Environmental features also play an important role in determining the value of a property. A property located in a pleasant area will usually be more desirable than one located in an unattractive neighborhood. Several indicators can be used to describe the attractiveness of an area. The following four variables are used in our model to reflect the quality of the environment:

1. the maintenance of the structures in the area where a property is located,
2. the parents' perceived quality of the school where their children are expected to attend,

3. the safety of a neighborhood, and
4. the degree of racial segregation.[7]

In the first three cases, the direction of the effect is easily observed. Better maintained surrounding properties, a higher perceived quality of the neighborhood school and a safer district should positively influence property values. In the last case, the relationship is somewhat uncertain. The conventional argument is that an area with a high percentage of non-whites is less attractive to whites. However, the same area could be more desirable to non-whites and hence affect property values positively. In addition, the current availability of housing units to both groups and their relative purchasing power should also influence the direction of the effect.

The above factors are related exclusively to the land component of residential properties. Seven variables are used to describe the physical quantity and quality of the improvements:

1. number of rooms,
2. number of bathrooms,
3. cubical content,
4. age of the building,
5. 0/1 dummy variables for garage,
6. exterior wall finish (wood or brick), and
7. structure (attached or detached).

It is expected that a positive effect will prevail in all cases except for the age of the building. There exist several other attributes reflecting the quantity and quality dimension of a building. However, we consider the above-mentioned attributes the most essential.

An Empirical Study

The model developed in the previous section provides the basis for an empirical study consisting of a cross-sectional regression analysis of the random sample used to examine the extent and nature of the variation in

[7] Two additional environmental factors often used in previous studies have been omitted. They are 1) density and, 2) the median income in an area. Cf. E. F. Brigham, "The Determinants of Land Values," *Land Economics*, November 1965, pp. 325-334; P. B. Downing "Estimating Residential Land Value by Multivariate Analysis," in D. M. Holland, *The Assessment of Land Values* (Madison: University of Wisconsin Press, 1969), pp. 101 123 W. E. Oates, "The Effects of Property Taxes and Local Public Spending on Property Values An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *Journal of Political Economy*, July 1969, pp. 957-970; and D. N. Hyman and E. C. Pasour, 'Real Property Taxes Local Services and Residential Property Values," *Southern Economic Journal*, April 1973 pp. 601 611. The causal relationship for these two variables is unclear. Crowding as well as the median income can be appropriately treated as a function of property values rather than the reverse. Even if the argument is debatable their omission should not affect the final results since they are highly correlated with other environmental variables

assessment-sales ratios. The ultimate aim of the study is to examine the feasibility of implementing the proposed system.

Operational measures of the various explanatory variables were obtained or calculated from several sources. The transaction values were drawn from the 1971 City of Pittsburgh transaction files. As an index of accessibility, the simple linear distance in miles was estimated on a map from the properties which were sampled to the location of the CBD and the closest public park with complete facilities. The lot size and the frontage were measured from the Sanborn maps for Pittsburgh. We have also obtained from this same source and the assessors' files all the building features.

The factors reflecting the attractiveness of a neighborhood are somewhat more difficult to quantify. As a proxy variable of the maintenance of the structure of surrounding properties, we have used the percentage of housing units lacking plumbing facilities in the tract where each property is located (U.S. Bureau of the Census, 1970). To measure the parents' perceived quality of a neighborhood school, the percentage of eligible "Title I" pupils at the nearest public elementary school was chosen.[8] The degree of racial segregation in an area was quantified by the percentage of non-whites as indicated in the 1970 Bureau of the Census data by tract for Pittsburgh. Finally, as a measure of the safety of an area, we have used the 1970 crime rate by census tract as compiled by the Pittsburgh Police Department.

Results

The estimated coefficients of the fifteen explanatory variables obtained through the multiple regression (including a constant) are presented in Table 1. The coefficient for the distance to the CBD is positive and insignificant. The non-influence of this factor on property values indicates that the assumption of one CBD does not hold. In fact, Pittsburgh is characterized by a dispersed employment pattern along its rivers. Also, since the data base is confined to the city rather than its metropolitan area, the CBD is fairly accessible from all points within a half-hour's driving time. Within this range, it has been shown that from the household's viewpoint, neighborhood characteristics outweigh the accessibility feature.[9] The coefficient of distance to the nearest park is negative as expected, but insignificant.

The results reveal that the quality of a neighborhood plays an important role in the market determination of property values. In all cases, the direction of the effect is negative, as expected. Only two out of the four factors tested are

[8] The percentage of eligible "Title I" pupils at a given public elementary school refers to the proportion of pupils whose families have an annual income under $3,000 or are receiving aid for dependent children. The choice of this measure over other indicators such as the ratio of pupils to certified teachers was based on discussions with officials of the Pittsburgh School Board of Education.

[9] J. B. Lansing and N. Barth "Residential Location and Urban Mobility: A Multivariate Analysis," The University of Michigan Institute for Social Research, December 1964.

Table 1. Factors Affecting Residential Property Value in Pittsburgh

Independent variable	*Regression coefficient*	*t statistic*
Constant[a]	8494.11	3.732
Distance to the CBD (miles	13.28	0.086
Distance to the nearest public park (miles)	- 687.41	-0.650
Crimes per 10000 population (tract)	-2291.03	-0.843
% of "Title I" pupils at the nearest elementary public school[a]	- 40.59	-1.954
% of non-whites (tract)[a]	- 34.76	-2.017
% of substandard housing units (tract)	- 13.68	-0.140
Lot size (sq. ft.)[a]	0.34	2.235
Frontage (feet)	- 2.95	-0.245
Exterior wall finish (1 = brick, 0 = wood)[a]	2419.90	2.699
Garage[a] (1 = yes, 0 = no)	1074.14	1.294
Structure (1 = detached, 0 = attached)	924.45	0.875
Age of building[a] (years)	- 103.04	-4.094
Number of rooms[a]	- 636.82	-2.132
Number of bathrooms[a]	5975.21	7.397
Cubical content[a] (cu. ft.)	0.32	5.653

[a] Designates significant variables.
R Square = 0.6923; number of observations = 245; standard error in estimate adjusted = 5873.00.

significant. The percentage of non-whites has a negative and significant effect on property values. Similarly, the percentage of eligible "Title I" pupils is negatively and significantly associated with property values. Both the coefficients of the crime rate and the condition of the structures are negative but not significant. The multicollinearity among the four factors might in part explain the insignificance of the last two results.[10] In the case of the condition of the structures in an area, the indicator used may not reflect the maintenance of the buildings, which would also partly explain the result.

The coefficient of the size of the lot is positive and significant. However, the coefficient of the frontage is negative but insignificant. This probably is due to the overlapping of the two site characteristics as previously mentioned.

Among the seven building components, five have significant effects as we expected. However, the effect of the structure being detached is positive but insignificant. More surprising, the coefficient of the number of rooms is negative and significant. This unexpected result can be explained due to the high correlation with the number of bathrooms (.64) and the cubical content (.80). In fact, the interpretation of some individual variables is somewhat obscure due to their close correlation. It is therefore difficult to interpret the coefficients as the market valuation of the utility provided by an increase in these factors. This is evidenced by the overestimate found for bathrooms ($5,975.00). In other instances the results seem very consistent. For example, given the mean age of

[10] See footnote 4. The matrix of simple correlation coefficients can be made available upon request.

the properties sampled (60 years) and the mean transaction value ($15,000), the average building constructed today would, *ceteris paribus*, be valued at an additional $6,000. This is a very reasonable estimate.

Prediction of Property Values

The above results show the feasibility of using the framework presented to predict the market value of residential properties. Further improvements in the model are needed, however. Based on these results, the accuracy of the model to predict property values at a satisfactory level for direct implementation is somewhat less than would be required. The adjusted standard error in estimate indicated in Table 1 is too large. This is explained by the fact that the model, in its present form, neglects several objective factors for which no data are presently available. For example, it was impossible to differentiate the number of cars per garage in properties which had a garage. Furthermore, many of the present factors are subject to large errors in measurement because of the low-quality information base used. Having corrected for these deficiencies, we are confident, in view of the fairly high R^2 (.70) of this pilot project, that predictions generated by our model would provide assessors with an accurate and objective information base with which to determine uniform and equalized assessment rolls.

The structure presented assumes a market equilibrium situation at the time considered. The prediction would then only be accurate over the period under consideration. For each subsequent period, a new set of improved property sales must be collected in order to generate a set of estimators reflecting the prevailing market situation. The new results would consequently be used to predict residential property values at this new point in time.

Advantages and Disadvantages of Implementation

The steps in the current manual assessment procedure are self-explanatory. They are the assessor's field survey and his appraisal. Needless to say, this type of practice not only exclusively relies upon an assessor's personal judgment but is also wasteful in terms of time and cost involved since a substantial portion of his efforts must be spent on bookkeeping activities.

Under the computerized system, assessors are relieved of all the tedious clerical work and thus are able to devote their time to the judgmental field of determining values. The new administrative process could be described in three steps:

1. the data management,
2. the prediction of property values, and
3. the assessor's analysis of the predictions.

The major responsibility of the data management section would be to both improve and update the records of the quantity and quality measures of land and building components of each property to be assessed. A substantial portion of the data collection could be carried out through a self-reporting system, as for income tax purposes. Each owner could be required by law to fill a form on which he would report the measures of the site and building characteristics of his property. Any major change in the improvements could be recorded thereafter from the building permit office while minor changes would still be reported by the owner.

Based on the data of the bona fide transactions which occurred during the time period under consideration, the multiple regression coefficients would be estimated and the market value of each property then predicted. It is now the assessor's duty to verify the results and to make adjustments if necessary by taking into consideration various conditions which the regression formula had not been able to include.

The multicollinearity problem might appear to limit the procedure since it causes the number of rooms to carry a negative and significant sign in a house-price equation containing only 15 explanatory variables. One may conclude that it would be difficult for an Appeal Board to fall back on the regression equation to defend a change in assessment. However, the difficulty would still prevail even if the multicollinearity problem was solved and all estimators found truly represented the market evaluation of housing attributes.

What presents disadvantages is in fact the cross-sectional nature of the regression analysis on which the procedure is based. It renders impossible the adjustment of property values from one assessment year to another using the weights derived since implicit in this type of analysis is the assumption that the observations represent points in equilibrium. What this means is that no marginal analysis is able to predict the value of a property at a future point in time.

Naturally, it would be preferable if all the weights derived accurately reflected the market valuation of housing attributes: however, this could only be achieved at the expense of the predictive power of the regression formula. What amount of predictive power is willing to be sacrificed in order to have more reliable estimators is an important question for which no predetermined answer is available.