

HOUSING AND RESIDENTIAL DEVELOPMENT: COMMENT

William B. Brueggeman*

Shenkle and Bennett have provided us with the application of a methodology for predicting housing value which utilizes a combination of analysis of variance for grouping sample data into homogeneous subsamples, and multivariate analysis to produce regression equations for each subsample in order to improve the prediction of housing values. Samples of housing transaction data from Montgomery County, Pennsylvania and Bibb County, Georgia are used to demonstrate the technique.

It might be well at the outset to point out the motivation behind the application of this technique in order that its contribution may be judged relative to other attempts to circumvent a knotty and elusive problem in housing market analysis. Basically the problem with estimating value in any housing market stems from the fact that the goods in question are not homogeneous. Any attempt to hypothesize that all housing values can be explained or predicted by one common set of independent variables is destined for failure. The primary problem with an attempt to generalize about virtually any housing characteristic is the failure to recognize that housing submarkets exist. Each submarket is differentiated according to different characteristics (such as location, size, design, etc.) which influence value.

Recognizing the submarket problem researchers have tried various approaches to the study of housing markets. Some have tried to get around the submarket problem by using an abstraction from the physical concept of housing. They have attempted to convert all housing characteristics into the concept of a flow of housing services and from this have tried to study housing market behavior. Other researchers have recognized the submarket problem directly. Attempts to stratify sample data into more homogeneous subgroupings have ranged from limiting the geographical area of analysis to one which exhibits the same physical, locational and economic characteristics such as a "neighborhood" or "the inner city", to employing factor analysis, multiple discriminant analysis, and multiple regression analysis.

In addition to the submarket problem, Shenkle and Bennett recognize an additional problem concerning intervals within characteristic or explanatory variables used in estimation. According to the authors, when a series of characteristics are used for prediction, only certain intervals within a measured characteristic may be relevant for prediction and in developing subsamples for submarket analysis. The procedure used by the authors to identify relevant intervals within explanatory variables is to apply one way analysis of variance on observed values of the dependent variable grouped into intervals within a given explanatory characteristic or independent variable. Successive intervals within each characteristic are combined and the

*Housing and Residential Development Session.

1) "The Selection of Variable Combinations for Predicting Housing Characteristics", by William M. Shenkle and Cathie Bennett.

2) "Vacation Home Location: A Model for Simulating the Residential Development of Rural Recreation Areas", by Raymond J. Burby III, Thomas G. Donnelly and Shirley F. Weiss.

optimal grouping of intervals to create a subsample occurs when the between group sum of squares is highest relative to the total sum of squares of the entire group. Each subsample is in turn categorized into intervals within another explanatory variable and the subsample is again divided, and so on. Ultimately, individual regression equations are developed for each final subsample to predict value for housing units exhibiting characteristics common to all other housing units within a subsample.

The authors should be complimented for their attempt to use this branching technique to housing market analysis. With regard to their specific application, however, some problems do arise. First, an analysis of the explanatory variables used in the regression analysis, and for establishing subsamples, reveal a very heavy concentration on physical characteristics of housing. There are no locational variables which would categorize the data into a rural-urban classification, a distance-travel time classification, or more detailed neighborhood descriptions. Virtually all references to submarkets in the literature refer to location as a prime differentiating factor. Because value is being predicted on a county-wide basis in this study location would seem to be a major characteristic on which a sample might be broken into subgroups. Second, a problem exists with respect to the time of sale variable. In using multiple regression analysis for prediction, estimated coefficients are supposed to represent in some way the weight or importance a variable has in explaining variation in a dependent variable. I fail to see the importance of the year of sale variable as having any predictive power when related to housing value, unless housing sold before some certain data had some inherent quality that housing sold after that data did not possess. I suspect that creation of this variable is really arbitrary and the fact that it is significant in the regression analysis probably results from the fact that other variables, not included in the regression, are "working through" the time of sale variable. Third, a more rigorous constraint must be placed on the number of feasible subdivisions of sample data. The author's criteria of continuing subdivision as long as the number of observations in a subsample justifies multiple regression analysis and reduces error in prediction is inadequate. Continuous subdivision of the sample could theoretically reach the point where for a very large original sample, one could have an indefinite number of subsamples which would explain all variation in the dependent variable. Finally, the problem of choosing intervals arises. While this is a problem common to all research projects, it should be recognized that with the utilization of the grouping technique presented here, there is an infinitely large number of calculations which could be made to insure absolute optimal subgrouping. By breaking the data down into preconstructed intervals there is no assurance that the subgrouping is optimal.

Burby, Donnelly and Weiss provide us with an interesting study in an area which has been largely neglected in regional studies, but which has become an important area of investigation with the increasing concern over environmental quality. Rural recreation areas are a rapidly growing segment of our economy under the impetus of both private and government influence. Residential development of these areas will continue to grow with the increased demand for leisure time by the more affluent segment of our society. This report, therefore, is also a timely contribution to an issue of social importance.

In addition to the simulation model of their study, the authors provide an interesting analysis of the process of residential development of two rural recreation areas: Lake Lanier, Georgia and Lake Norman, North Carolina. The process was described by taking into account the behavior of the relevant decision units involved in development which included landowners, deve-

lopers, and consumers. Using a series of surveys, the motivation of landowners to hold, sell, subdivide or sell to developers was explored. The background and experience of the developers of these properties are presented as well as the results of a survey which concentrated on determining the important locational choice factors cited by development firms when purchasing land. Finally, a socio-economic profile of consumers and their desired locational choice factors are presented to complete the segment of the study dealing with the development process. The descriptive data provided in this segment of the study is important in itself, as well as providing a basis for the development of multiple regression analysis which is a prelude to the simulation model.

The major conclusions reached from the statistical examination of survey data were: 1) seasonal recreational development and permanent residential development in these areas are really segmented markets, the development of each is dependent on different factors, 2) contextual factors peculiar to a given area must be taken into account when analyzing development, and 3) governmental decisions can influence developmental patterns. These conclusions are based on the results of the multiple regression analysis performed on data generated from coding a series of the developer and consumer locational choice attributes deemed important in the surveys.

It should be pointed out that in simulation studies such as this one, the significant variables found in the regression analysis form a basis for the simulation model. Therefore, it is important that variables chosen from the analysis sufficiently explain the phenomena to be simulated. If an important explanatory variable is omitted, future attempts at simulation would fail due to the fact that important influences on the phenomena being simulated are not being considered. It is with the specification of variables used in the regression analysis that present a problem in this study. Regressions relating residential development to various attractiveness factors in the cases considered by the authors, included from 20 to 30 independent variables and yielded coefficients of multiple determination ranging from .192 to .401. The authors indicate their disappointment with these results, and make a point of indicating that a lack of economic data, particularly price on available parcels, severely limited their research. While the lack of data probably hampered the regression analysis, the technique used to quantify the dependent variable should be reexamined. The nine-point scale used to develop grid cells, which form the basis for the dependent variable, may not be a proper quantitative representation of the variable being explained; that is, the extent of residential development. The quantification of the dependent variable is based on an ordinal ranking system and then expressed as a function of some independent variables measured on ordinal scales and some on interval scales. It could be that the quantitative specification of the dependent variable resulted in an improper correspondence between variables in the equation. Perhaps reconsideration of the scaling method used in the study might prove to be of some benefit to the analysis.

One other aspect of the paper that deserves more emphasis is the role of government decision making in the development process. This is not intended as a criticism, but rather as a direction which future research might be undertaken. While these decisions are deemed important in the study, their impact on development is not fully explored. Perhaps this facet of analysis can be incorporated into future studies which I am sure will build on this exploratory research.

The paper provides us with a very clear, well organized, and systematic approach to the decision processes, contextual characteristics, and constraints which must be considered in modeling. The use of the authors' simulation model, previously developed for urban analysis, shows the general applicability of the technique to more than one aspect of regional analysis. The authors should be commended for a very sizeable undertaking in an area of increasing importance. It is hoped that other researchers will profit from their efforts.