Working Paper 77:2

INFLUENCE OF MULTICOLLINEAR RELATIONSHIPS ON EMPIRICAL RESEARCH IN REAL ESTATE AND URBAN ECONOMICS

by

George W. Gau

and

Daniel B. Kohlhepp

Center for Economic & Management Research
College of Business Administration
UNIVERSITY OF OKLAHOMA, NORMAN, OKLAHOMA 73019

Working Paper 77:2

INFLUENCE OF MULTICOLLINEAR RELATIONSHIPS ON EMPIRICAL RESEARCH IN REAL ESTATE AND URBAN ECONOMICS

by

George W. Gau

and

Daniel B. Kohlhepp

Assistant Professors of Business Administration
Division of Finance
University of Oklahoma
Norman, Oklahoma

February 1977

INFLUENCE OF MULTICOLLINEAR RELATIONSHIPS
ON EMPIRICAL RESEARCH IN REAL ESTATE AND URBAN ECONOMICS

I. Introduction

Recently a substantial amount of research by urban economists, real estate appraisers, and assessors has been directed at the estimation of reduced-form price equations for housing. The statistical technique most commonly employed by these researchers has been ordinary least-squares regression analysis. The methodology utilized with this technique is based on the assumption that the predetermined (explanatory) variables are linearly independent. Unfortunately, this assumption is violated to some extent in most estimations, that is, multicollinearity is present in the data set. The purpose of this paper is to investigate and to evaluate alternative methods for dealing with the multicollinearity problem in reduced-form price equations (RFPE) for housing. Specifically, this paper examines stepwise regression and principal components analysis as alternative ways to reduce the undesirable effects of multicollinearity. These methods are evaluated on the basis of the explanatory and predictive power of the RFPE estimated by each technique.

The scope of this paper is limited to a discussion and evaluation of the statistical techniques as they relate to the reduction of multicollinearity. The paper does not address the question of the appropriate specifications

---

* The authors benefited in this study from the helpful comments of a colleague, Professor R. Dacey, and the data preparation and computational assistance of J. Lawson, M. Jones, T. Strickland, R. Palmer and G. Bray.

of the theoretical model. The paper also does not consider the problem of the correct measures as proxies for the theoretical variables. Since the same theoretical variables and measures are used by each technique to estimate the RFPE, the effect of the theoretical specification and measurement is held constant in the study. Also, the theoretical model and data used in this study reflect the models and data most commonly used by economists and real estate appraisers.

The next section of the paper reviews the procedures and problems in estimating RFPE using ordinary least-squares multiple regression analysis. Following this section the research design is outlined, and then in Section IV the statistical methods used in the analysis are discussed. Section V reports the statistical estimates of the RFPE using each method and the predictive results obtained from these estimates as well as measures of multicollinearity. Finally, the conclusions and implications of the study are reported in Section VI.

II. Overview of the Procedures and Problems in Estimating Reduced-Form Price Equations for Housing

In this section the procedures usually employed to estimate a RFPE for housing are reviewed as well as the problems which multicollinearity causes in the process. First, there is a brief discussion concerning the derivation of the theoretical model, and then the necessary steps to estimate the parameters of the model are considered. Multicollinearity is then defined, and an examination is made of the problems which it causes in estimations, specifications, and forecasting. Finally, there is a review of the general procedures that have been recommended for reducing the effect of multicollinearity.

Theoretical Model

The first step in the estimation of the reduced-form price equation is the statement of a general theory concerning the behavior patterns in the marketplace as follows:

$$f\ (\theta) = \mu \tag{1}$$

where θ = a matrix of real variables which are used to represent and describe the interaction of the participants in the marketplace.

μ = a vector of random disturbances in the marketplace, $E\ (\mu) = 0$

A linear theory is given by

$$A\ \theta = 0 \tag{2}$$

Let θ be partitioned by the subvectors α $(m \times 1)$ and λ $(n - m \times 1)$, and partitioning A in form (Λ,ξ) that

$$\Lambda\ \alpha + \xi\ \lambda = 0 \tag{3A}$$

is the structural form of the theory. The reduced-form of the theory can be obtained by premultiplying by Λ^{-1} and rearranging terms,[1]

$$\alpha = -\Lambda^{-1}\xi\gamma \tag{4A}$$

The particular behavior patterns in the housing market are stated simply that the demand for housing (D_H) is given by

$$D_H = f\ (p,\ x,\ldots x_n)$$

where

p = vector of housing prices

x_j = vector of other housing characteristics

and that the supply of housing (S_H) is given by

$$S_H = g\ (p,\ y_1\ldots y_n)$$

where

$$p = \text{vector of housing prices}$$

$$y_j = \text{vector of other housing characteristics}$$

The price of housing is implicitly defined in the structural equation

$$D_H - S_H = 0 \tag{3B}$$

or

$$f(p,x_1 \ldots x_n) - g(p,y,\ldots y_n) = 0$$

and explicitly defined in the reduced-form

$$p = h(x_1 \ldots x_n,\ y_1 \ldots y_n) \tag{4B}$$

The theoretical determinants of the price of housing, $h(x_1 \ldots x_n,\ y_1 \ldots y_n)$, have been theorized by several writers and can be summarized in the following broad categories:[2]

(1) Accessibility

(2) Quantity

(3) Quality

(4) Physical and Social Environment

(5) Fiscal Environment

(6) Financing terms and other conditions of the sale

Statistical Model

Given the theoretical variables of the RFPE, the next step is to establish operational or observable measures or proxies for these variables. The researcher's imagination comes into play at this time, as the variables can be measured or proxied in an infinite number of ways. Table 1 suggests some of the measures which have been used to proxy the different variables.

TABLE 1

Theoretical Variables	Possible Measures
Accessibility	Distance, time, or cost to and from another economic activity
Quantity	size of heated living area, number of bedrooms, lot size, etc.
Quality	age of the home, type of construction, construction materials, maintenance levels, number and types of built-ins, etc.
Physical and Social Environment	types and conditions of surrounding properties, income or education levels of neighbors, etc.
Fiscal Environment	property tax rate, quantity and quality of municipal services, quality of public schools, etc.
Financing Terms	FHA, VA, or conventional mortgages, assumable low interest mortgage, junior mortgage availability.

When the theoretical variables are converted to operational measures, the matrix $[-\Lambda^{-1}\xi]$ (exogenous variables) and the vector α (endogenous variables) are observable in the marketplace. Consequently, the parameters of the theoretical model, γ, can be estimated by consistent point estimates using the least squares criterion as follows:

$$\alpha = -\Lambda^{-1}\xi\ \gamma \tag{4A}$$

or

$$p = Xb + d \tag{4C}$$

where

p = n x 1 vector $<p_1 \ldots p_i \ldots p_n>$ of observed prices, α (endogenous variable)

b = m x 1 vector $<b_1 \ldots b_j \ldots b_m>$ of estimated parameters, γ

X = n x m matrix $<x_{11} \ldots x_{ij} \ldots x_{nm}>$ of the observed measures of the determinants of housing prices, $-\Lambda^{-1}\xi$ (exogenous variables)

d = n x 1 vector $<d_1 \ldots d_i \ldots d_n>$ of the estimated disturbance terms, with the assumptions that $E(d) = 0$ and $E(d'd) = \delta^2 I$.

The true relationship, b, between the exogenous variables, X, and the endogenous variable, p, can be estimated with the ordinary least-squares (OLS) procedure

$$\hat{b} = (X'X)^{-1} X' p \tag{5}$$

which has a variance-covariance matrix

$$V(\hat{b}) = \delta^2_d (X'X)^{-1}$$

where δ_d is the underlying variance of the disturbance term.

Once the estimates of the parameters are obtained, they are predominantly employed in two ways: (1) hypothesis testing; and (2) forecasting. The theorized relationships between the housing prices and housing characteristics are stated as

$$H_o \ : \ b_j = 0$$

$$H_a \ : \ b_j \neq 0$$

and are empirically tested by

$$\text{pr } (\hat{b}_j - t_{\alpha/2}\, s_{\hat{b}_j} < b_j < \hat{b}_j + t_{\alpha/2}\, s_{\hat{b}_j}) = 1-\alpha$$

where

$$1-\alpha = \text{confidence level}$$

$$t = \hat{b}_j / s_{\hat{b}_j}$$

$$s_{\hat{b}_j} = (\text{var } \hat{b}_j)^{1/2}$$

Besides hypothesis testing, the economists also use the estimated parameters of the RFPE of housing, $\hat{b}$, to forecast or predict a selling price, $\hat{p}_i$, given certain values for the exogenous variables.

$$\hat{p}_i = (x_{i1} \ldots x_{im})\ \hat{b}$$

where

$(x_{i1} \ldots x_{im})$ is a vector of given values.

Real estate appraisers and assessors use the estimated parameters almost exclusively to forecast or predict the market value (most probable selling price) of a house. Assessors commonly use the RFPE for housing to estimate the assessed values of the housing stock in a community for taxation purposes, while appraisers employ the estimated parameters to aid them in predicting the most probable selling price of a house in the "Direct Sales Comparison Approach" to value estimation.

Definition of Multicollinearity

An important assumption of the statistical model is that the data matrix has rank = m, the explanatory variables are linearly independent. The vectors $< x_1 \ldots\ldots x_j \ldots\ldots x_m >$ are linearly independent if there does not exist a set of non-zero constants $a_1, \ldots\ldots, a_m$ such that

$$\sum_{j=1}^{m} a_j x_j = 0 \qquad (6)$$

Since linear independence requires that the exogenous variables are mutually orthogonal, multicollinearity can be defined as "departures from orthogonality" [10, p. 100].

At the extreme if the variables are perfectly dependent, the matrix (X'X) approaches singularity and cannot be inverted as required in the estimating equation (5). At this point the parameter estimates are indeterminate. Yet, multicollinearity also exists in cases where the variables are not perfectly dependent, but only exhibit some level of interdependence--equation (6) is approximated for some subset of the column vectors. The problem of multicollinearity is thus not just one of existence, but also of degree.

Problems with Multicollinearity

If the level of multicollinearity is significant in a data set, the statistical model will be plagued by problems in the areas of estimation, specification, and prediction.[3] If the orthogonality assumption (along with others relating to unbiasedness and minimum variance of the estimators) is fulfilled, the influence of each of the explanatory variables on the dependent variable (price) can be accurately measured by the coefficients $\hat{b}_1, \ldots\ldots, \hat{b}_m$. With the presence of multicollinear relationships, the resultant estimates of the population parameters tend to have very large variances. In these situations hypothesis testing is quite difficult since the contribution of the explained variance of one variable cannot be distinguished from the contribution of another variable with which it covaries. Furthermore, the parameter estimates are extremely sensitive to changes in the sample as well as changes in the model specification.

The model specification begins with theoretical variables and moves to operational measures. When these measures are observed, the hypothesized relationships can be empirically tested. The presence of multicollinearity in the data set makes the correct specification of the model very difficult. Although the theoretical model can be misspecified, the operational proxies are the cause of most specification errors as the theoretical richness of the model is reduced and compromised to accommodate the available data. Typically with housing data, a large number of measures are used to proxy a theoretical variable. Unfortunately, as the number of variable measures taken from a sample increases, each measure tends to reflect different dimensions of the same few theoretical variables. This results in multicollinear relationships in the data which cause increases in the standard errors of the regression coefficients and the possibility of incorrectly discarding important variables based on the tests of significance.

Quite often economists and real estate appraisers contend that they are not concerned about multicollinearity in the data set because the estimated parameters are used to predict and not to explain. Although it may provide a sense of well-being for its believers, this contention is subject to question. For the parameters that are estimated with multicollinear data to be used successfully for prediction or forecasting, two necessary conditions must be met:

1. The estimated dependency relationship between the endogenous and exogenous variables must remain stable.

2. The interdependency relationships among the exogenous variables must remain stable.

The first condition is usually assumed on the basis of the hypothesized theoretical relationships. However, there is generally no theoretical reason for assuming that the second condition holds and in most studies little effort is made to empirically test for this condition. Implicitly assuming that the interdependency relationships among the explanatory variables stays constant, when in fact they do not, can lead to substantial prediction errors.

Corrections for Multicollinearity

When multicollinearity is present in a sample, there are three broad types of remedies which are generally suggested for reducing its influence [26, pp. 383-386]:

(1) change the specification of the model. The specification can be modified by simply deleting one or more of the variables which have the highest degree of multicollinearity. While this remedy may reduce the level of interrelationships in the model, it also introduces the possibility of a serious specification error.

(2) redefine the explanatory variables. The redefinition of the variables is usually in the form of transformations (ordinary first differences or log first differences) or aggregations (linear or non-linear combinations of the original variables to form new composite variables). Of course in the case of the first difference transformation, multicollinearity is reduced at the cost of possibly increasing the autocorrelation in the disturbance term.

(3) acquire additional data. This remedy is usually accomplished by increasing the sample size. The acquisition of additional observations can improve the estimation of the parameters with multicollinear data providing the new information increases the range of the original variable values.

III. Research Design

This study evaluates two quantitative methods that have been proposed for reducing the influence of multicollinearity in a multiple regression model--stepwise regression and principal components regression.[4] The statistical methodology utilized with these techniques, especially the procedures employed for multicollinearity reduction, are briefly described and contrasted in the next section of this paper. Then, using a sample of residential housing observations, four reduced-form price equations are derived with a ordinary least-squares procedure and with the two alternative techniques. Subsequently, tests are performed to determine the level of multicollinearity affecting the equations, and the predictive ability of each model is examined with a holdout sample from a different time period.

The residential sales data considered in this study is drawn from single-family home sales in the Norman, Oklahoma area during the period of January, 1973 to December, 1975.[5] On the basis of the date of sale, the total sample is segmented into two groups: (1) an analysis sample of 504 observations having a sale's date on or before March 1975; and (2) a holdout (prediction) sample of 228 houses that were sold after March, 1975. The analysis sample is applied to the development of the price-equation models and corresponding multicollinearity tests, while the holdout sample is retained for the prediction tests.[6]

In this study two statistical measures are applied to the evaluation of the degree of multicollinearity affecting the regression coefficients in the models. One of the measures is the value $|X'X|$, the determinant of the mxm correlation matrix--X being a nxm matrix of n observations on m explanatory variables with the observations normalized on the basis of the sample size and the standard deviations of m.[7] If pairwise independence (orthogonality) exists between all of the m variables, then X'X reduces to an identity matrix and $|X'X|=1$. Alternatively, if perfect collinearity is present, $|X'X|=0$. Thus, $0 \leq |X'X| \leq 1$ serves as an ordinal measure of the level of pairwise correlation or dependence influencing the regression estimates.

The other measures that are derived for assessing multicollinearity (beyond just simple pairwise correlation) are the coefficients of multiple determination, R_j^2, which are obtained when each individual variable, x_j, is regressed on the remaining m-1 explanatory variables.[8] If any one of these mR_j^2's is close to 1, then there is a linear relationship among two or more of the explanatory variables and the degree of multicollinearity for these interrelated variables is high. Conversely, if all of the R_j^2's are near zero, the level of interrelationship among the variables is low. Both the R_j^2's and $|X'X|$ are calculated with the observations in the analysis sample and the explanatory variables <u>in the form</u> that they are applied to the construction of the reduced-form price equations.

To test the relative <u>ex post</u> forecasting ability of the models, the regression coefficients estimated in the analysis sample with each technique are multiplied times the values of the explanatory variables in the prediction sample and summed across the observation to obtain a forecasted sale price for each house. A comparison is then made of the difference between the forecasted price (P_{fi}) and the actual sale price (P_{ai}) and two measures are calculated for each model:

1. mean absolute error (MAE)

$$\frac{\sum_{i=1}^{n} \left(\left| P_{fi} - P_{ai} \right| \right)}{n}$$

2. root mean square error (RMSE)

$$\left(\frac{\sum_{i=1}^{n} \left(P_{fi} - P_{ai} \right)^2}{n} \right)^{1/2}$$

While the mean error would be influenced by the sign of the deviation of the forecasted from the actual price, the MAE takes into account only the magnitude of the prediction error. Similar to the standard deviation as a measure of dispersion, the RMSE--by squaring the difference between the prices--gives greater weight than the other measures to those predictions which are further away from the actual price.

IV. Methodology

Initially in this study, an OLS model is estimated with the housing observations in the analysis sample. The data matrix X in this sample has dimensions of 504 x 47. However, since three of the housing characteristics are being measured through three sets of dummy variables that for each characteristic are mutually exclusive and exhaustive in the sample (e.g., the type of construction), one dummy variable in each set must be dropped from X to allow the inversion of (X'X). The influence of the three variables, of course, is then measured in the constant b_o. Therefore, the estimation of the regression coefficients for only 44 of the 47 variables is attempted with the OLS model.

Stepwise Regression

With the presence of multicollinearity (X having rank < m), one of the techniques which has been proposed for reducing the level of multicollinearity is stepwise regression (SWR). Stepwise procedures have been employed for variable selection and reduction in previous appraisal studies.[9] The usual criterion in most SWR algorithms is to select that subset of explanatory variables x_j^* ($j=1,\ldots, s$; $s < m$) which is most highly correlated with the dependent variable. SWR methods can be combined with a F test to reduce the dimensions of X without significantly influencing the explanatory power of the model.[10] The resulting SWR model will then be

$$p = X^*b + d \qquad (7)$$

with X* being a nxs matrix and the b's being estimated with a least-squares procedure.

Of the three general types of remedies for multicollinearity reduction described in Section II, the SWR technique is attempting to empirically weaken the influence of multicollinearity by changing the specification of the model. Yet, it should be noted that besides the general concern over the possible sample dependence of the stepwise selection and the need to replicate the results with other samples to insure that the selected variables are the optimum in the population, this type of procedure may still leave a subset of explanatory variables that are interrelated. While the SWR technique will allow for the reduction of the number of variables and thus the possible multicollinearity, the criterion and methodology do not necessarily guarantee that the selected variables will be independent. The actual rank of X^* may be $< s$ and the $\hat{b}$ still subject to the influence of interrelationships among the x_j^*'s.

Principal Components Regression

An empirical method that has been utilized for deriving a reduced-form price equation with the minimal influence of multicollinearity is principal components regression (PCR).[11] Principal components analysis is a multivariate statistical technique that groups the variation in a data set into distinct dimensions while trying to account for as much of the total information in the sample as possible. This procedure extracts the maximum amount of variance from the explanatory variables while simplifying a large number of variables into a smaller set of independent dimensions. The principal components model takes the form:[12]

$$x = Af' \tag{8}$$

with x = $m \times 1$ vector of explanatory variables, A = $m \times c$ matrix of factor loadings, and f = $1 \times c$ vector of factors.[13] Each factor loading, a_{ji}, represents a product-moment correlation between x_j and f_i. In a components model the initial number of factors is equal to the number of variables, $c = m$. However, in many research situations, the number of factors retained for future analysis and interpretation (r) will be less than m, with the residual variance of the remaining $c-r$ factors assumed to be attributable to random error.

To facilitate the interpretation of the substantive meaning of each f_i, this study rotates the loadings matrix of the retained factors using a varimax procedure. This technique trys to describe a factor in terms of as few variables as possible while still preserving the independence (orthogonality) of the factors. This transformation results in a matrix of rotated factor loadings, $\bar{a}_{ji}$, which tend to have values close to either one or zero. Those x_j's with high $\bar{a}_{ji}$ on a f_i can be interpreted as being highly related to the dimension associated with f_i.

Also, the rotated loadings matrix is utilized in this study in the formation of factor scores to represent each of the factors. The factor scores are calculated through the equation

$$Z' = (\bar{A}'\bar{A})^{-1}\bar{A}'X' \qquad (9)$$

with Z = $n \times r$ matrix of factor scores representing the values on the factors for each observation, $\bar{A}$ = $m \times r$ rotated factor loadings matrix, and X = $n \times m$ matrix of values on the variables for each observation.

The matrix Z is then employed in the PCR model of the form

$$p = Zb + d \tag{10}$$

to estimate a $\hat{b}$ for each factor using the OLS estimation procedure. The factor scores matrix is also utilized in the tests of the level of multicollinearity affecting the estimates of the coefficients for the PCR model.

Unlike the SWR technique, PCR tries to reduce the influence of multicollinearity by redefining the explanatory variables in a reduced-form price equation. The redefined variables are in the form of linear transformations of the original explanatory variable. On a theoretical level in an appraisal study, a researcher in attempting to fully identify those characteristics that explain the price of a house may employ a number of explanatory variables that, in essence, measure the same determinant in the population. For example, in examining the effect of the neighborhood on the value of a house, the appraiser may include in his reduced-form price equation observable variables such as average income, school district, etc. as measures of the neighborhood quality. Or, perhaps the problem is that no single variable alone is capable of measuring the effect of the neighborhood characteristic on the observations and only through the consideration of a set of multiple-interrelated measures will the characteristic be adequately represented in the model. PCR enables the researcher in these situations to statistically combine the observable interrelated variables and attempt to develop constructs that serve as better measures of the characteristic than the explanatory variables applied on an individual basis.

The controversies surrounding this technique have generally been in the areas of the interpretation of the statistically-formed factors and the proper criterion for factor retention in a PCR model. Farrar and Glauber [11,p.97] have taken the position that the PCR technique is fruitful only if the factors have a recognizable substantive meaning as economic variables. On the other side McCallum [25] argues that the lack of a clear economic interpretation of the factors is not critical, so long as the factors allow for a better estimation of the dependent variable. Two alternative criteria have been proposed for determining the optimal number of factors to be retained in a PCR: (1) delete the factors which are relatively unimportant as predictors of the original explanatory variables--those factors with the smallest eigenvalues; and, (2) delete the factors which are relatively unimportant as predictors of the dependent variable--those factors with the smallest t ratios. Coxe [7] and Greenberg [13] have presented findings which indicate that the choice of the retention criterion should be dependent on the purpose of the study. If the primary objective of the study is structural analysis and thus the researcher would like to minimize the standard errors of the coefficients, retaining only those factors with eigenvalues greater than one is the best selection procedure. However, in forecasting applications of PCR where minimization of the prediction error is the primary goal, the optimal criterion is choosing those factors that are important predictors of the dependent variable.

V. Analysis of Results

Reduced-Form Price Equations

The equations estimated with the OLS and SWR technique are presented in Table 2. An examination of the summary statistics for the OLS equation finds that the OLS model derived a RFPE that explained a major percentage of the variation of the sales price for the housing observations in the analysis sample. The R^2 in the OLS regression is .9038, and the standard error of estimate (SEE) is $3,737 or approximately 12% of the mean sales price. Of the forty-four explanatory variables included in the OLS model, twelve coefficients are significant at the .05 level. On the basis of the standardized coefficients, the square feet variable is the most important determinant of the sales price, followed by the year built (YRBUILT) and central air conditioning (CLC).

However, for some of the explanatory variables the signs of the regression coefficients are the opposite of the direction that would be expected from the theoretical relationships. For example, with each additional bedroom or bathroom, or an increment in the total number of rooms in a house, it would be hypothesized that there would be an increase in the sales price of the dwelling. Yet, for all three variables (BEDRMS, BATHRMS, and NBRMS) the coefficients have negative signs, indicating an inverse relationship between the sales price and each additional room unit.[14] Obviously, something is wrong, and the problem rests in the area of multicollinearity. While the results of the multicollinearity measures for the OLS model will be presented in the next section, even a cursory examination of the pairwise correlation among the explanatory variable in the analysis sample (see

TABLE 2

OLS AND SWR MODELS

Variable[1]	OLS Regression Coefficient	OLS Standardized Coefficient	SWR Regression Coefficient	SWR Standardized Coefficient
SQFT	17.179 (27.275)**	.799	17.150 (31.386)**	.818
LOTSIZE	.175 (3.008)**	.052	.212 (3.988)**	.178
BEDRMS	-877.782 (-1.810)	-.042	-1019.325 (-2.453)*	-.048
BATHRMS	-1058.141 (-1.922)	-.047	-1096.783 (-2.139)*	-.049
NBRMS	-190.090 (-.743)	-.021		
YRBUILT	195.571 (4.702)**	.155	171.501 (6.466)**	.136
DATESOLD	97.808 (2.497)*	.041	102.912 (2.908)**	.043
TCF	1548.681 (.870)	.025		
TCB	520.606 (.378)	.014		
TCR[2]	----	----		
TCO	-196.470 (-.113)	-.004		
FLS	173.290 (.198)	.005		
FLC[2]	----	----		
COE	-1004.126 (-1.076)	-.042		
COG	-1899.385 (-2.231)*	-.082	-945.290 (-2.577)**	-.041
COF[2]	----	----		
ORC	-851.517 (-1.522)	-.028	-1036.664 (-2.151)*	-.034

TABLE 2 (Continued)

OLS AND SWR MODELS

Variable[1]	OLS Regression Coefficient	OLS Standardized Coefficient	SWR Regression Coefficient	SWR Standardized Coefficient
ORU	515.879 (1.126)	.020		
ORD	227.488 (.354)	.006		
ORO	544.962 (.197)	.003		
PTC	2406.102 (2.830)**	.052	2483.976 (3.316)**	.053
PTO	775.077 (1.467)	.033	1010.654 (2.464)*	.043
CDC	1034.427 (1.545)	.025		
CDD	-210.036 (-.312)	-.006		
FNS	2663.639 (1.620)	.025	3412.029 (2.183)*	.032
FNC	-321.093 (-.202)	-.003		
FNO	-673.437 (-1.169)	-.019		
OF	1595.096 (2.468)*	.043	1313.440 (2.196)*	.036
HTC	-322.503 (-.180)	-.007		
HTF	-1241.328 (-.698)	-.026		
CLC	4243.918 (5.229)**	.128	4411.551 (7.431)**	.133
CLW	509.317 (.488)	.010		
TYLNC	198.684 (.305)	.008		

TABLE 2 (Continued)

OLS AND SWR MODELS

Variable[1]	OLS Regression Coefficient	OLS Standardized Coefficient	SWR Regression Coefficient	SWR Standardized Coefficient
TYLNV	291.674 (.366)	.011		
TYLNF	769.641 (1.004)	.031		
AVEINC	.022 (.149)	.005		
TOCAMPUS	982.649 (2.867)**	.074	612.612 (2.483)*	.046
TOI35	900.394 (1.653)	.102		
JACKSON	162.658 (.120)	.004		
CLEVELAND	-418.455 (-.326)	-.015		
WILSON	-1398.679 (-1.222)	-.029		
JEFFERSON	-4517.949 (-3.122)**	-.103	-2242.201 (-3.368)**	-.051
EISENHOWER	-3936.733 (-2.390)*	-.147		
LINCOLN	712.475 (.464)	.010		
MCKINLEY	1130.874 (.769)	.024		
MONROE	413.184 (.769)	.011		
KENNEDY	-2456.877 (-2.213)*	-.056	-1401.252 (-2.094)*	-.032

TABLE 2 (Continued)

OLS AND SWR MODELS

	OLS	SWR
Constant	-18650.864	-16215.720
R^2	.9038	.8995
$\bar{R}^2$	.8946	.8962
SEE	3737.090	3709.3525
F	98.039	272.345

Figures in parentheses are t values. * = significant at .05 level. ** = significant at .01 level.

[1]An interpretation of the variable symbols can be found in Table 1A in Appendix.

[2]Variables withheld from OLS to allow for inversion of (X'X).

Table 3A in the Appendix) shows a number of high correlation coefficients for the variables with the wrong regression signs. The regression coefficients for these variables are therefore being influenced by other explanatory variables, and the interrelationships are causing the OLS model to clearly provide inaccurate information concerning the direction of the marginal influence of each of these variables on the sales price.

The SWR equation in Table 2 is constructed with a stepwise regression procedure using a partial $F(1,\infty, .95) = 3.840$ for selection and retention of the explanatory variables [10, pp.171-172]. A total of sixteen variables are ultimately selected in the stepwise procedure for inclusion in the model, and even though the R^2 for the SWR equation is slightly lower than in the OLS model, the adjusted coefficient of multiple determination ($\bar{R}^2$, adjusted for degrees of freedom) is higher than in the OLS equation. These results imply that the reduced subset of variables in the SWR model can explain <u>in the population</u> approximately the same level of variation of the sales price as the total set of variables in the OLS equation.

Unlike the OLS model all sixteen variables in the SWR equation have coefficients which are significant at the .05 level. Also, note that four variables (BEDRMS, BATHRMS, PTO, and FNS) which had insignificant t values in the OLS equation are now exhibiting significant coefficients, due to a reduction in their standard errors with the model specification derived by the stepwise procedure. While the OLS model indicated that these four variables did not have a significant impact on the sales price and should possibly be discarded in forming the RFPE, the SWR shows that these variables do have a significant relationship with the dependent variable (based on the partial F criterion) and therefore should be retained in the RFPE. However, both the bedrooms and bathrooms variables--along with others in the SWR equation--continue to exhibit theoretically-incorrect regression signs with the likelihood that multicollinear relationships are still affecting the SWR equation.

In the development of the PCR equations, the analysis sample is subjected to a principal components analysis of the explanatory variables, and the two alternative criteria are utilized in the determination of the number of factors retained for rotation. On the basis of the eigenvalue-one criterion (retained factors limited to those with eigenvalues greater than one), a total of sixteen factors--accounting for 72 percent of the original variance in the analysis sample--are selected for rotation and further analysis. Table 3 gives a summary of the resulting factor loading matrix after a rotation of the sixteen factors. Additionally, a second PCR model is derived through the retention of all forty-four factors, and Table 4 summarizes the rotated factor loading matrix with the forty-four factors.

From an examination of Table 3 for the sixteen factor model, most of the factors seem to have a substantive meaning as representations of the theoretical determinants of property value. For example, Factor 1 forms a "quality" dimension through the combination of variables such as the type of heating and cooling system (HTC,HTF,CLC, and CLW), the type of construction (TCB and FLS) and the age of the structure (YRBUILT). An observation with a high positive factor score on this dimension would be expected

TABLE 3

SUMMARY OF ROTATED FACTOR LOADINGS MATRIX
(16 FACTOR MODEL)

Factor	Variable[1]	Loading[2]	Factor	Variable[1]	Loading[2]
1	HTC	.879	6	WILSON	.729
	HTF	-.846		TCO	.551
	CLC	.784			
	CLW	-.660	7	FNS	.708
	TCB	.644		LINCOLN	.619
	FLS	.608			
	YR BUILT	.606	8	FNO	.710
2	NBRMS	.838	9	CDD	.803
	SQFT	.793			
	BEDRMS	.749	10	ORD	.776
	BATHRMS	.737			
			11	TYLNV	.798
3	TOI35	.893		TYLNF	-.697
	EISENHWR	.823			
	AVEINC	-.755	12	JEFFRSON	.838
	MONROE	-.579			
			13	KENNEDY	.865
4	COG	.892			
	COE	-.856	14	JACKSON	.865
5	CLEVLAND	.847	15	PTC	.773
	TOCAMPUS	.589			
			16	ORO	.778

[1] An interpretation of the symbols for the variables can be found in Table 1A in the Appendix.

[2] Loadings greater than the .550 level.

to be of better quality and have a higher sales price. Factor 4 would also represent another "quality" dimension, except in this case (given the signs of the loadings) there would be the theoretical expectation of an inverse relationship between these factor scores and the dependent variable. Similarly, Factor 2 can be interpreted as representing a "quantity" dimension and Factor 11 as a "financing" dimension.[15] Other factors (i.e., Factors 3,5,6,7,12,13, and 14) are formed primarily by school district variables and would seem to be representing "neighborhood" dimensions which would tend to be unique to the Norman housing population.

An analysis of the rotated loadings in Table 4 with the full forty-four factor model indicates that, when compared to the sixteen factor model, these forty-four factors are not as easily interpretable as theoretical determinants of property values. Even though the first sixteen factors listed in the table have similarities with the factors derived in the sixteen factor model, there are some obvious differences in the size of the loadings for the explanatory variables. Also, some of the sixteen factors in the first model split in the second model into a multiple number of dimensions. For example, three of the variables that had high loadings in Factor 2 in the sixteen factor model (i.e., NBRMS, SQFT, and BATHRMS) have less influence in Factor 2 of the forty-four factor model and these variables are dominant in defining other dimensions (Factors 32, 35, and 36) in the second model. These additional twenty-eight factors in the second model seem to represent other dimensions in the sample that are not included in the sixteen factor model. However, the low level of the loadings in many cases make the identification of the economic meanings of these factors difficult.

TABLE 4

SUMMARY OF ROTATED FACTOR LOADINGS MATRIX
(44 FACTOR MODEL)

Factor	Variable[1]	Loading[2]
	HTC	.902
	HTF	-.925
	CLC	.517
1	CLW	-.282
	TCB	.278
	FLS	.373
	YRBUILT	.292
	NBRMS	.539
2	SQFT	.448
	BEDRMS	.934
	BATHRMS	.352
	TOI35	.919
3	EISENHWR	.918
	AVEINC	-.516
	MONROE	-.217
4	COG	.957
	COE	-.919
5	CLEVLAND	.337
	TOCAMPUS	.834
6	WILSON	.962
	TCO	.099
17	TCO	.919
	TCB	-.701
18	TCF	.959
19	TYLNC	.911
20	MONROE	.947
21	LINCOLN	.966
22	MCKINLEY	.949
23	OF	.955
24	LOTSIZE	.954
25	CLW	.914
26	CDO	.967
27	FNC	.978
28	ORU	.962
29	PTO	.883
30	ORC	.954

TABLE 4 (cont.)

Factor	Variable[1]	Loading[2]
7	FNS	.985
	LINCOLN	.100
8	FNO	.976
9	CDD	.983
10	ORD	.972
11	TYLNV	.939
	TYLNF	-.583
12	JEFFRSON	.983
13	KENNEDY	.972
14	JACKSON	.960
15	PTC	.970
16	ORO	.995
31	DATESOLD	.978
32	BATHRMS	.804
33	FCS	.761
34	CLC	.726
35	SQFT	.593
36	NBRMS	.592
37	AVEINC	.586
38	YRBUILT	.527
39	TYLNF	.469
40	CLEVLAND	.338
41	TCB	.306
42	COE	.211
	COG	.183
43	HTC	.203
	HTF	.185

TABLE 4 (cont.)

Factor	Variable[1]	Loading[2]	Factor	Variable[1]	Loading[2]
			44	TOI35	-.125
				EISENHWR	.125

[1] An interpretation of the symbols for the variables can be found in Table 1A in the Appendix.

[2] For Factors 1-16, loadings given for variables of similar factors in Table 3 with 16 factor model. For Factors 17-44, loadings given for dominant variables.

Applying the two sets of factor scores formed with the two rotated loadings matricies for the sixteen and forty-four factor models and the standardized variable values for the observations in the analysis sample (see equation (9)), two PCR equations are constructed with the factors as explanatory variables and the sales price (in standardized form) as the dependent variable. Tables 5 and 6 show the two equations formed with the PCR technique. PCR16 is derived by employing the ordinary least-squares procedure with the sixteen factors formed on the basis of the eigenvalue-one criterion. PCR44 is developed with the least-squares procedure using the full forty-four factor model.

For the PCR16 equation, the sixteen factors exhibit less explanatory power in the analysis sample than the original variables in the OLS and SWR models. The R^2 for PCR16 is .7668, compared to .9038 and .8995 in the OLS and SWR equations. Also, the standard error of the estimate is higher in this equation than in either the OLS or SWR models.[16] For the PCR44 equation, the forty-four factors have exactly the same explanatory power as the original forty-four variables in the OLS equation. The R^2 and SEE for PCR44 are equivalent to the measures in the OLS equation. This result is not surprising since, unlike PCR16, the PCR44 model is employing the same total information (variance) of the explanatory variables as used in the OLS model. Only in the case of PCR44, the information is utilized in the form of statistical dimensions constructed on the basis of the interrelationships present among the explanatory variables rather than the actual variables.

Unlike the OLS and SWR models, the coefficients in the PCR16 equation seem to have the theoretically-correct signs. While COE, HTC, and CDD all

TABLE 5

PCR16 EQUATION

Factor	Regression Coefficient	Factor	Regression Coefficient
1	.3700	9	.1223
	(16.910)**		(5.587)**
2	.6861	10	.0206
	(31.354)**		(.941)
3	-.1043	11	-.0145
	(-4.765)**		(-.662)
4	-.2149	12	.0702
	(-9.821)**		(3.210)**
5	.2387	13	-.0711
	(10.910)**		(-3.248)**
6	-.0025	14	-.0060
	(-.115)		(-.274)
7	.0451	15	.0266
	(2.060)*		(1.215)
8	-.1278	16	-.0217
	(-5.842)**		(-.993)

Constant	.00000202
R^2	.7668
$\bar{R}^2$	.7592
SEE	.49075
F	100.107

Figures in parentheses are t values. * = significant at .05 level. ** = significant at .01 level.

TABLE 6
PCR44 EQUATION

Factor	Regression Coefficient	Factor	Regression Coefficient	Factor	Regression Coefficient
1	.2384	17	-.0715	33	.1011
	(16.470)**		(-4.938)**		(6.987)**
2	.3426	18	-.0503	34	.1969
	(23.667)**		(-3.475)**		(13.603)**
3	-.1408	19	.2397	35	.4872
	(-9,726)**		(16.557)**		(33.658)**
4	-.2473	20	-.0023	36	.0997
	(-17.082)**		(.162)		(6.885)**
5	.2081	21	-.0145	37	-.0005
	(14.378)**		(1.001)		(.037)
6	-.0722	22	-.0204	38	.0945
	(-4.990)**		(1.409)		(6.531)**
7	.0677	23	.2137	39	-.0103
	(4.675)**		(14.763)**		(-.712)
8	-.0680	24	.1517	40	.0012
	(-4.698)**		(10.479)**		(.080)
9	.0997	25	-.0962	41	.0026
	(6.889)**		(-6.644)**		(.181)
10	.0554	26	.0224	42	-.0208
	(3.826)**		(1.548)		(-1.435)
11	-.0466	27	-.0285	43	-.0073
	(-3.217)**		(-1.967)*		(-.501)
12	-.0450	28	.0424	44	-.0303
	(-3.112)**		(2.928)**		(-2.095)*
13	-.0488	29	.1625	Constant	.00000039
	(-3.374)**		(11.225)**	R^2	.9038
				$\bar{R}^2$	.8946
14	-.0285	30	.1028	SEE	.32464
	(-1.969)*		(7.104)**	F	98.039
15	.1474	31	.0626		
	(10.182)**		(4,322)**		
16	.0052	32	.2330		
	(.362)		(16.099)**		

Figures in parentheses are t values

* = significant at .05 level.

** = significant at .01 level.

exhibit negative coefficients in the OLS model, the signs for Factors 4, 1, and 9 indicate that these variables are having a positive influence on sales price, as would be expected given the hypothesized theoretical relationships. Similarly, even though the SWR derives negative coefficients for the BEDRMS and BATHRMS variables, the positive sign of the regression coefficient for Factor 2 correctly measures the direction of the effect of these variables on the dependent variables.

The relative magnitudes of the regression coefficients in the PCR16 equation shows that Factor 2 (the "quantity" dimension) is the single most important factor affecting the sales price of a house in the analysis sample--followed by Factor 1 (a "quality" dimension), Factor 5 (a "neighborhood" dimension), and Factor 4 (another "quality" dimension). Through the combination of the PCR coefficients and the factor loadings, an analysis can also be made of the comparative impact of some of the explanatory variables on the dependent variable. As an example, the large loadings for the NBRMS variable in Factor 2 means that this variable is important in defining this "quality" dimension and (given the large coefficient for this factor in the PCR16 equation) determining the sales price.[17] Conversely, the insignificant coefficient for Factor 11 indicates that the "financing" dimension and the variables forming this dimension (primarily, TYLNV and TYLNF) are not important determinants of the sales price. While this type of analysis does not provide the cardinal measures of marginal impact for the explanatory variables that result from the OLS and SWR equations, the choice of which technique to employ in a structural analysis of a RFPE should be dependent upon the comparative influence of multicollinearity in the estimation of each equation.

Multicollinearity Measurements

Table 7 presents the values for the two statistical measures of the degree of multicollinearity affecting the regression coefficients in the four equations. These measures suggest that there are decidedly different levels of multicollinearity influencing each of the models. While the variables in the OLS model have an extremely high degree of interrelationship, the factors in the PCR model are independent (the $|X'X|$ is equal to 1 and all the R_j^2's are zero to five decimal places), and thus the coefficients in the PCR equations have no multicollinear relationships affecting their directions or magnitudes. While the SWR model is between the two extremes, the R_j^2's in the last column of the table show that certain coefficients in the SWR equation are being plagued by multicollinearity.

These results have important implications for any attempt at structural analysis of a RFPE. The equations themselves, of course, gave indications of the direct effects of multicollinearity--the high standard errors and the wrong signs in the OLS and SWR models. Yet, the indirect influences of multicollinearity on the coefficients are often more subtle and not obvious in the examination of a RFPE. For example, the square feet variable has the correct sign and a significant t value in both the OLS and the SWR equations, leading an observer to the possible conclusion that the marginal value of each square foot on the sales price is being correctly measured at approximately \$17 per square foot. But, is this the correct value?

With the multicollinearity tests, the R_2^2's are .75573 (OLS) and .68028 (SWR). These measures indicate that the SQFT variable is highly related to other explanatory variables in the models (e.g., BEDRMS,BATHRMS,and NBRMS). The regression coefficients for the SQFT variable are therefore also representing the influence of other variables, and the values of the regression

TABLE 7

MULTICOLLINEARITY MEASURES

Model	$\lvert X'X \rvert$	R_j^2 Average	R_j^2 Five Highest[1]
OLS	.00010	.60886	R_{39}^2 = .94446 R_{44}^2 = .94438 R_{47}^2 = .88830 R_{30}^2 = .87580 R_{15}^2 = .86507
SWR	.02445	.28664	R_{2}^2 = .68208 R_{5}^2 = .59974 R_{7}^2 = .53139 R_{4}^2 = .46673 R_{38}^2 = .39966
PCR16	1.00000	.00000	
PCR44	1.00000	.00000	

[1] For OLS and SWR models, j's are variable numbers in Table 2A.

coefficients in the OLS and SWR equations will thus be affected by the model specification. The PCR model is the only technique that allowed for a definitive assessment of the relative influence on sales price of the inputs in a regression equation, although in this case the inputs are statistical constructs in the form of linear combinations of the explanatory variables rather than the actual variables.

Prediction Tests

The results of the prediction tests with the four equations and the holdout sample are presented in Table 8. The two measures of forecasting ability generally find that the four models do not perform equally in the prediction of the actual sales price of an observation in the holdout sample. The absolute sizes of the forecasting errors (MAE) indicates that the PCR44 equation is the best predictor of the value of the dependent variable in the holdout sample, followed closely by the OLS and SWR equations. The forecasting errors for the PCR16 equation are significantly higher when compared to the results with the other equations. Given the lower R^2's for PCR16 in the analysis sample, the poorer predictive performance of this equation in the holdout sample could be expected, and these results indicate that some of the twenty-eight factors that were not retained in the sixteen-factor model (with the eigenvalue-one criterion) could have been useful in the prediction of the sales prices in both the analysis and holdout samples.

A comparison of the root mean square errors (RMSE) in the holdout sample with the standard errors of the estimate (SEE) in the analysis sample also shows that all four equations had a reduction in their forecasting ability between the two samples.[18] The RMSE's for the four equations are higher

TABLE 8

PREDICTION RESULTS

Model	MAE	RMSE
OLS	3368.51	4499.58
SWR	3409.24	4504.17
PCR16	5085.32	7151.23
PCR44	3338.32	4496.50

than the SEE's: OLS by \$762; SWR by \$795; PCR16 by \$1,052; and, PCR44 by \$759. Generally, there are two possible explanations for this type of drop in the predictive capability of the models.[19] First, the true relationships in the population between the dependent variable and the explanatory variables (or factors in the PCR models) has changed over the time periods. Second, the interrelationships among the explanatory variables has not remained stable. In this second situation the OLS, SWR, and the two PCR models will experience--but, perhaps in differing degrees--an increase in prediction errors, since all four equqtions are constructed either implicitly (OLS and SWR) or explicitly (PCR16 and PCR44) on the basis of certain multicollinear relationships present in the analysis sample.

To test the stability of the interrelationships among the explanatory variables in the two samples, a comparison can be made of the correlation matricies for the variables across the observations in each sample. An analysis of the comparative pairwise correlations (see Tables 3A and 4A in the Appendix) suggests that the interrelationships remained fairly constant. As a descriptive measure of the equivalence of the entries in the two matricies, a correlation coefficient is calculated for the non-diagonal elements of the lower triangles for the forty-six explanatory variables in the two samples--variable 21 (ORO) is not considered since no observations in the holdout sample exhibited this characteristic. The correlation between the entries in the two matricies is .8754, indicating a high degree of stability of the interrelationships between the two samples. Thus, the reduction in the forecasting ability of the four equations in the holdout sample does not seem to be primarily a function of changes in the multicollinearity among the explanatory variables.

However, it is interesting to note that, with this small change in the multicollinearity structure, the PCR44 model did slightly better in forecasting than the OLS model. Since the OLS and PCR44 equations had equal R^2's and SEE's in the analysis sample, any differences in the predictive power of the two equations in the holdout sample must be attributable to: (1) differences in the stability of the sales price-explanatory variable dependency relationships in the OLS equation versus the sales price-factor relationships in the PCR44 equation; and, (2) any differences in the effect on the two equations of the minor instability in the multicollinear relationships. In both cases, these results suggest that the full-factor PCR model retains a higher level of forecasting power over time than models developed with the alternative procedures.

VI. Summary and Conclusions

This study has examined three quantitative models that can be applied in the estimation of reduced-form price equations for housing under conditions of multicollinearity. With a sample of residential housing observations, four equations were formed and tested for the degree of multicollinearity influencing the equations and the forecasting power of each model in a holdout sample. This study then attempted to evaluate the suitability of the OLS, SWR, and PCR techniques for structural analysis and prediction.

The results of this study show that the application of an OLS or SWR model in the structural analysis of a RFPE for housing is a very tenuous proposition. Under conditions of multicollinearity, this research demonstrated how the regression coefficients estimated by the OLS and SWR techniques are often troubled by theoretically-incorrect signs, high standard errors, and coefficients that generally are not accurate measures of the marginal impact of the explanatory variables on the sales price. Even though smaller models could be heuristically-designed with possibly lower levels of multicollinearity than exhibited in the statistical models employed in this study, such attempts would explicitly require certain variable choices that may directly lead to model specification errors. The simultaneous consideration of the large number of explanatory variables that seem to measure different aspects of the theoretical determinants of property value inherently involves models with a high degree of multicollinear relationships. While the PCR model provided accurate coefficients for the factors, this technique did result in the loss of direct absolute measures for the explanatory variables. Yet, the findings of this study suggest that this level of aggregation in the form of statistically-derived factors may be the only suitable method in most cases for an assessment of the impact on the sales price of the hypothesized determinants of property value.

The results of this study provide additional evidence concerning the question of the optimal criterion for factor retention in a PCR model applied to the estimation of a RFPE for a residential market. The sixteen factors retained on the basis of the eigenvalue-one criterion exhibit in this study a more discernible economic meaning as representations of determinants of property values than the forty-four factors retained with the full-factor model. Conversely, the PCR model with the greater number of factors is able to explain a larger percentage of the variability of the dependent variable and show a better prediction performance in the holdout sample than the PCR model with only those factors having eigenvalues greater than one. From the findings of this study it seems clear that there is a tradeoff between interpretation of the factors and predictive accuracy for a researcher estimating a RFPE with a PCR model, and that the decision concerning the number of retained factors should be based on the primary objective of the study. If the purpose of the study is structural analysis, then the optimal factor selection procedure is the eigenvalue-one criterion. However, if the goal is prediction, then the researcher should retain the total factor set for initially estimating the RFPE.

Another tentative conclusion of this study is that a researcher may find that over time the full-factor PCR model provides better forecasting results of the sales price of a property than either the OLS or SWR models. While the differences in the predictive ability in the holdout sample of the PCR and OLS equations is not dramatic, the dispersion is sufficient to at least suggest the possibility that dependency relationships in the OLS model are less stable than in the PCR model.

Finally, the interesting question remains of whether there would be any further differences in the relative forecasting capabilities of the alternative techniques when there is a substantial change in the multicollinear relationships among the explanatory variables. Since an appraiser or assessor employing a RFPE for predictive purposes probably has little *ex ante* information concerning the stability of the interrelationships over time, his choice of techniques would be affected by the likely increase of forecasting errors in the future with each model if there are fluctuations in the inherent multicollinearity. While the samples employed in this study do not allow for a comprehensive evaluation of this question, this topic is an area for future research.

FOOTNOTES

[1]In order to generate the reduced form from the structural form, three conditions must be satisfied:

a. Λ^{-1} must exist (i.e., Λ must be a nonsingular matrix).

b. The underlying individual supply and demand functions must be continuous and possess continuous first partial derivatives in all functions in order to satisfy the Implicit Function Theorem.

c. The entrees in the matrix A must be explicitly definable in terms of α (the endogenous variables) and the probability distribution, $H(\mu)$, of the disturbance term. [32, p. 102]

Furthermore, if A is not explicitly definable, then the researcher cannot obtain consistent point estimators of the regression parameter. The general linear model of housing markets is based upon the Lancaster model of consumer behavior [21, 22]. If the underlying concepts within that model are not explicitly definable (e.g., as in c.), then the Lancaster model is, at least in part, non-testable [8].

[2]Accessability, measured by distance to market place, has been theorized as a major determinant of housing prices by all urban economists using Von Thüenen-like models. Among these writers include Alonzo [1] and Muth [27]. Also, see Muth for analysis of quality and quantity determinants. Examples of the analysis of the physical and social environment

include Brigham [3], Ridher and Henning [29], and Wercher and Zerbst [33]. Barr [2] examines the effect of the fiscal environment, and Doherty [9] considers the influence of financing terms on housing prices.

[3]Unfortunately, there is no appropriate statistical procedure to test whether a given level of multicollinearity in a data set is statistically significant. While Farrar and Glauber [11, p. 101] propose such a measure, it has been shown to have little value in a sample test [28]. Thus, even though multicollinearity can be measured, the level at which it is statistically significant cannot be directly tested in the formulation of a model.

[4]Another technique, ridge regression, was also tested in the study, but the results are not included in the paper because of the inability of the technique to derive a stable model. First proposed by Hoerl and Kennard [15, 16], ridge regression forms an estimate of b on the basis of

$$\hat{b}^* = (X'X + kI)^{-1}X'p \tag{5*}$$

where $k > 0$ is a scalar constant and I is a m x m identity matrix. The idea behind the ridge regression technique is that, in a situation of non-independent explanatory variables, through a careful selection of k a biased estimator of b can be found, $\hat{b}^*$, that would be expected to be closer to the actual population b than $\hat{b}$. In essence, the methodology attempts to reduce the effect of multicollinearity on the estimated coefficients by respecifying the model--in this case by placing a priori restriction on the value of b, increasing the residual sum of squares, and potentially reducing the predictive power of the model.

The only previous application of ridge regression known to the authors in the area of residential appraisal is by Churchill [5]. However, unlike the results reported in the Churchill article, this study was not able to derive reasonable estimators of the coefficients. In the fourteen runs ($.02 \leq k \leq 1$) with equation (5*) performed in this research, the $\hat{b}*$'s did not stabilize at specific values and no clear determination could be made of the appropriate level for k. The probable explanation for our conflicting results is the much larger number of explanatory variables considered in this study (compared to Churchill's thirteen variables) with a multicollinearity structure more complex than the relatively simple correlation matrices previously tested. Our tentative findings tend to question the value of ridge regression in applied reduced-form price equations which include the estimation of a large number of parameters. For a critical discussion of other problems associated with ridge regression, see Conniffe and Stone [6].

[5]A more detailed description of the sample and the variables included in the analysis is provided in the Appendix.

[6]Various programs in the BMDP package are utilized in this study to estimate the equations and test for multicollinearity. The primary programs are BMDP1R (multiple regression), BMDP2R (stepwise regression), and BMDP4M (factor analysis).

[7]The application of this statistic as a measure of multicollinearity has been proposed in a number of sources, including Farrar and Glauber [11] and more recently in Mason, Gunst, and Webster [23].

[8] See Johnston [17, p. 163], Kmenta [20, p. 390], and Mason, Gunst, and Webster [23, p. 285].

[9] Examples include Smith [31] and Gloudemans and Miller [12].

[10] Draper and Smith [10, Chapter 6] provide a thorough description of the stepwise regression technique.

[11] This technique was first proposed by Kendall [19] and explored in depth by Massy [24]. Among the applications of principal components and factor analysis in reduced-form price models include Church [4] and Kain and Quigley [18].

[12] Instead of estimating the communality of the variables as done in the common factor model, principal components analysis retains the total variance and the correlation matrix is factored with 1's in the diagonal. For a comprehensive description of principal components and factor analysis, consult Harman [14] and Rummel [30].

[13] In recognition of the nomenclature generally utilized in applied multivariate analysis, the term "factors" will be employed in this study instead of "components" to identify the dimensions derived in the principal components analysis. Technically, this distinction results from the subsequent rotation of the loadings matrix.

[14]Other examples of regression coefficients in the OLS equation with theoretically-incorrect signs include the coefficients for variables COE (condition of structure--excellent), ORC (club room), CDD (both carpets and drapes), HTC (central heat), TOCAMPUS (distance to campus), and TOI35 (distance to major interstate).

[15]At this point a caveat must be given concerning the interpretation of the results of this study. While the statistical constructs found in this application may be the correct measures of the theoretical determinants of property values in the <u>general housing population,</u> before such a conclusion is made, further testing should be performed to examine the replication of these factors with alternative geographical samples, variable sets, and factoring levels.

[16]Since all inputs in the PCR equations are in standardized form, the regression coefficients and the standard error of the estimate will also be in terms of standardized values. To solve for the unstandardized SEE (for comparison with the values in the OLS and SWR models), the standardized SEE is multiplied times the standard deviation of the dependent variable. For PCR16, the (unstandardized) SEE = $5,649 and for PCR44, SEE = $3,737.

[17]Neither the OLS nor SWR equations identify the NBRMS variable as significant.

[18]In comparing these two statistics, it should be noted that there is a small difference in the method of calculation of these two measures. The SEE is formed by dividing the squared errors by n-2, while the RMSE divides by n.

[19]There are also two statistical effects which would tend to cause the RMSE to be higher than the SEE: (1) sample bias; and, (2) time trend. Since in the derivation of the regression equation the least-squares procedure attempts to form a linear function that best fits the particular characteristics of the analysis sample (characteristics which may not be common in the population), the predictive ability of the equation would be expected to have an upward bias in the analysis sample. The forecasting power of the models in the holdout sample would also be affected by rising sales prices of houses over time. With the holdout sample being drawn from the sales of residences in a period after the sampling for the analysis group, the holdout observations tend to have higher sales prices--that are further from the mean value--than would be found in the analysis sample. Thus, the time trend for the dependent variable would have the effect of increasing the forecasting errors in the holdout sample.

APPENDIX
SAMPLE DESCRIPTION

The sample used in the statistical analysis of this paper consisted of 732 single-family residences which were offered for sale and sold in Norman, Oklahoma during the period January 1, 1973 to December 31, 1975. The information collected for each observation was determined by two factors: (1) the appropriate measures of theoretical variable (Section II), and, (2) the availability of information. The data were gathered from several sources which included Realtor Multi-List information, 1970 Census data, zoning maps, and when necessary personal inspection.

The selling price for each house as well as most measures for the quantity, quality, and financing condition variables were obtained from the Multi-List Sold Books provided by the Norman Association of Realtors. The physical and social environmental variable measures were obtained from the 1970 Census of Housing and Population while the measures of accessibility and fiscal environment were obtained from local zoning maps. A complete listing of the measures used for each variable and the source of the data is in Table 1A of this Appendix.

The data set was initially analyzed by calculating the descriptive statistics for each measure, i.e., means, standard deviations, ranges, frequency plots, and skewness and kurtosis measures. From this analysis, it was determined that it would not be necessary to perform any transformation on the measures (logs, etc.), since they would meet the criteria necessary for the statistical analysis in their original form. Of course, if the objective of the study were to estimate the best regression equation, rather than to evaluate alternative methods, then several transformations could have been performed to attempt to improve the fit of the model. Table 2A contains selected descriptive statistics for the variable measures in the total sample and Tables 3A and 4A give the pairwise correlation coefficients for the observations in the analysis and prediction samples.

TABLE 1A
VARIABLE MEASURES

Theoretical Variable	Measure Used	Source of Data
I. Accessibility	Straight line distance in miles	
	TO CAMPUS-to intersection of Boyd and Asp Streets. (approximate CBC)	zoning maps
	TO I35-to nearest inter-state form	zoning maps
II. Quantity	SQFT-square feet of heated living area	Multi-List
	LOT SIZE-square feet of lot	County Assessor Office
	BDRMS-number of bedrooms	Multi-List
	BATHRMS-number of bathrooms	Multi-List
	NBRMS-total number of rooms	Multi-List
III. Quality - Age	YRBLT-year dwelling was built	Multi-List
Type of Construction		
	TCF-Frame	Multi-List
	TCB-Brick veneer	Multi-List
	TCR-Rock	Multi-List
	TCO-Other	Multi-List
Floor Construction		
	FLS-Slab	Multi-List
	FLC-Conventional	Multi-List
Overall Condition		
	COE-Excellent	Multi-List
	COG-Good	Multi-List
	COF-Fair	Multi-List
Specialty Rooms		
	ORC-Club Room	Multi-List
	ORU-Utility Room	Multi-List
	ORD-Separate Dining Room	Multi-List
	ORO-Other	Multi-List

TABLE 1A (Con't.)

Theoretical Variable	Measure Used	Source of Data
Patio		
	PTC-Covered	Multi-List
	PTO-Open	Multi-List
Carpets and Drapes		
	CDC-Carpet only	Multi-List
	CDD-Both Carpets and Drapes	Multi-List
Fence		
	FNS-Stockade	Multi-List
	FNC-Chain Link	Multi-List
	FNO-Other	Multi-List
	OF-Special Features	Multi-List
Heating System		
	HTC-Central Heat	Multi-List
	HTF-Floor Heat	Multi-List
Cooling System		
	CLC-Central Air Conditioning	Multi-List
	CLW-Window Air Conditioning	Multi-List
IV. Physical and Social Environment	AVEINC-Mean family income in census tract	U.S. Census
V. Fiscal Environment		
	Elementary School District	
	JACKSON-Jackson Elementary School District	zoning maps
	CLEVELAND-Cleveland Elementary School District	zoning maps
	WILSON-Wilson Elementary School District	zoning maps

TABLE 1A (Con't.)

Theoretical Variable	Measure Used	Source of Data
	JEFFERSON-Jefferson Elementary School District	zoning maps
	EISENHOWER-Eisenhower Elementary School District	zoning maps
	LINCOLN-Lincoln Elementary School District	zoning maps
	McKINLEY-McKinley Elementary School District	zoning maps
	MONROE-Monroe Elementary School District	zoning maps
	KENNEDY-Kennedy Elementary School District	zoning maps
VI. Financing Terms and Other Conditions of the Sale		
	DATESOLD-Date of Sale, number of months since January, 1970	Multi-List
Types of Mortgage Loan		
	TYLNC-Conventional	Multi-List
	TYLNV-VA Guaranteed	Multi-List
	TYLNF-FHA Insured	Multi-List

TABLE 2A

VARIABLE DESCRIPTIVE STATISTICS

	Variable	Mean	Standard Deviation	Range
1.	PRICE	31204.543	12362.773	96,000 - 9,000
2.	SQFT	1641.802	552.637	4,433 - 676
3.	LOTSIZE	10176.789	3514.620	37,440 - 1,097
4.	BEDRMS	3.254	.588	8 - 2
5.	BATHRMS	1.863	.514	3.5 - 1
6.	NBRMS	6.202	1.307	12 - 4
7.	YRBUILT	67.980	9.729	75 - 20
8.	DATESOLD	58.693	6.392	70 - 41
9.	TCF	.034	.182	1 - 0
10.	TCB	.898	.303	1 - 0
11.	TCR	.019	.137	1 - 0
12.	TCO	.049	.216	1 - 0
13.	FLS	.852	.355	1 - 0
14.	FLC	.148	.355	1 - 0
15.	COE	.359	.480	1 - 0
16.	COG	.583	.493	1 - 0
17.	COF	.057	.233	1 - 0
18.	ORC	.214	.411	1 - 0
19.	ORU	.236	.425	1 - 0
20.	ORD	.131	.338	1 - 0
21.	ORO	.003	.052	1 - 0
22.	PTC	.078	.268	1 - 0
23.	PTO	.366	.482	1 - 0
24.	CDC	.915	.279	1 - 0
25.	CDD	.115	.319	1 - 0
26.	FNS	.018	.132	1 - 0
27.	FNC	.026	.159	1 - 0
28.	FNO	.143	.351	1 - 0
29.	OF	.116	.321	1 - 0
30.	HTC	.921	.270	1 - 0
31.	HTF	.070	.255	1 - 0
32.	CLC	.855	.352	1 - 0
33.	CLW	.055	.227	1 - 0
34.	TYLNC	.362	.481	1 - 0
35.	TYLNV	.205	.404	1 - 0
36.	TYLNF	.313	.464	1 - 0
37.	AVEINC	11095.008	2700.623	15,919 - 5,330
38.	TOCAMPUS	2.004	.899	5.08 - .23
39.	TOI35	1.885	1.298	5.84 - .23
40.	JACKSON	.096	.294	1 - 0
41.	CLEVELAND	.223	.416	1 - 0
42.	WILSON	.072	.259	1 - 0
43.	JEFFERSON	.070	.255	1 - 0
44.	EISENHOWER	.235	.424	1 - 0
45.	LINCOLN	.038	.192	1 - 0
46.	MCKINLEY	.061	.240	1 - 0
47.	MONROE	.105	.307	1 - 0
48.	KENNEDY	.068	.252	1 - 0

TABLE 3A

CORRELATION MATRIX
ANALYSIS SAMPLE

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1.000																							
2	.913	1.000																						
3	.344	.326	1.000																					
4	.565	.641	.248	1.000																				
5	.633	.666	.311	.592	1.000																			
6	.657	.747	.386	.681	.574	1.000																		
7	.440	.299	.050	.308	.500	.170	1.000																	
8	.084	.032	.046	.014	.002	.003	.045	1.000																
9	-.147	-.084	-.149	-.084	-.160	-.037	-.442	-.094	1.000															
10	.212	.119	.062	.206	.247	.061	.535	-.004	-.580	1.000														
11	.096	.101	.040	.032	.055	.086	.064	.049	-.024	-.383	1.000													
12	-.226	-.153	-.066	-.235	-.240	-.104	-.403	.059	-.043	-.674	-.028	1.000												
13	.395	.289	.082	.324	.484	.217	.768	.029	-.383	.514	.051	-.418	1.000											
14	-.395	-.289	-.082	-.324	-.484	-.217	-.768	-.029	.333	-.514	-.051	.418	-1.000	1.000										
15	.439	.383	.201	.244	.338	.326	.366	-.059	.084	.109	.097	-.136	.281	-.281	1.000									
16	-.326	-.293	-.147	-.166	-.158	-.261	-.161	.053	.022	.064	-.078	-.063	-.109	.109	-.873	1.000								
17	-.207	-.163	-.099	-.145	-.345	-.116	-.393	.008	.122	-.342	-.034	.393	-.331	.331	-.207	-.296	1.000							
18	.231	.299	.165	.231	.262	.356	.076	.159	-.058	.008	.028	.024	.061	-.061	.031	.008	-.076	1.000						
19	.047	-.022	.050	-.039	.007	-.096	.155	.047	-.038	.027	.035	.018	.107	-.107	.198	-.171	-.045	-.264	1.000					
20	.101	.102	.108	.019	.014	.164	-.073	.103	.075	-.062	.061	-.015	-.069	.069	.125	-.090	-.063	-.153	-.199	1.000				
21	.003	.007	-.060	.030	-.033	.068	-.009	.008	-.012	.021	-.008	-.014	.026	-.026	-.049	.057	-.017	-.028	-.037	-.021	1.000			
22	.259	.244	.094	.135	.184	.199	.023	-.011	-.008	.034	-.034	-.022	.061	-.061	.025	-.005	-.037	.074	-.063	.042	-.017	1.000		
23	.407	.370	.156	.245	.242	.306	.259	.009	-.092	.177	-.039	-.145	.203	-.203	.313	-.271	-.069	-.012	.123	.095	-.052	-.217	1.000	
24	.182	.149	.039	.144	.175	.082	.196	.034	-.099	.144	.038	-.138	.109	-.109	.083	.041	-.244	.037	.008	.006	.019	-.039	.155	1.000
25	.066	.086	.052	.007	-.008	.026	-.167	-.135	.111	.004	-.043	-.076	-.088	.088	.098	-.090	-.010	-.013	.055	.078	-.021	.068	.109	.101
26	.084	.074	.005	-.014	-.007	.017	-.102	.124	.176	-.086	-.014	-.025	-.061	.061	-.048	.061	-.029	.048	-.023	.023	-.007	.045	-.015	-.034
27	-.065	-.045	.037	-.052	-.146	-.026	-.229	.141	.160	-.074	-.015	-.027	-.244	.244	-.058	.038	.037	.037	-.069	.127	-.008	-.031	.076	.035
28	-.159	-.126	.052	-.075	-.128	-.063	-.237	.139	.026	-.138	.001	.172	-.161	.161	-.143	.043	.194	.009	.016	-.048	-.024	-.075	-.134	-.043
29	.399	.404	.198	.244	.202	.348	.019	.086	.001	-.012	.007	.011	.014	-.014	.186	-.161	-.041	.148	-.045	.111	-.022	.190	.168	.034
30	.395	.285	.072	.304	.385	.171	.534	.024	-.309	.509	.036	-.466	.576	-.576	.208	-.029	-.350	.048	.047	-.027	.018	.076	.203	.162
31	-.371	-.267	-.101	-.293	-.341	-.155	-.477	-.013	.257	-.485	-.033	.477	-.550	.550	-.187	.036	.293	-.030	-.041	.045	-.016	-.069	-.180	-.101
32	.517	.385	.104	.292	.407	.227	.522	.066	-.263	.366	.051	-.314	.497	-.497	.278	-.105	-.334	.043	.078	.004	.025	.106	.270	.216
33	-.286	-.202	-.052	-.201	-.228	-.096	-.407	-.084	.301	-.291	-.029	.164	-.275	.275	-.159	.131	.050	-.029	-.008	.043	-.014	-.061	-.150	-.133
34	.457	.496	.210	.246	.244	.465	.059	-.040	.033	.029	.036	-.091	.104	-.104	.227	-.230	.017	.089	-.093	.069	.018	.167	.435	-.017
35	-.189	-.189	-.087	-.061	-.052	-.234	.018	-.017	-.031	.011	.005	.008	.022	-.022	-.075	.079	-.014	-.024	-.066	-.034	.039	-.146	-.318	.079
36	-.352	-.369	-.145	-.219	-.211	-.299	-.096	.059	-.008	-.090	-.013	.141	-.138	.138	-.192	.168	.039	-.024	.098	-.091	-.041	-.154	-.317	-.109
37	.213	.205	.292	.146	.131	.271	-.042	-.004	-.061	.125	-.023	-.109	.046	-.046	.036	-.024	-.023	.118	-.058	.056	.082	.137	.168	-.050
38	.473	.400	.135	.286	.404	.293	.553	.005	-.191	.222	.089	-.198	.421	-.421	.313	-.199	-.215	.081	.082	-.034	-.014	.093	.235	.089
39	-.225	-.272	-.224	-.162	-.141	-.332	.165	-.042	-.006	.039	-.041	-.026	.095	-.095	.051	-.024	-.052	-.199	.088	-.097	-.064	-.130	-.110	.062
40	-.053	-.047	.036	-.043	-.036	-.018	-.081	.063	.010	-.028	-.041	.054	-.199	.199	-.031	.018	.023	.107	-.004	-.021	-.021	.023	-.059	-.027
41	.345	.336	.108	.195	.176	.279	.108	.019	-.047	.074	.051	-.093	.111	-.111	.169	-.146	-.038	.119	-.002	.001	.045	.198	.134	-.060
42	-.187	-.138	-.108	-.112	-.235	-.098	-.289	-.009	.084	-.219	-.033	.253	-.276	.276	-.115	-.020	.266	-.049	.039	-.005	-.016	-.034	-.058	-.075
43	.026	.034	.041	.059	.049	.042	.119	.003	-.054	.068	-.036	-.027	.092	-.092	.157	-.131	-.044	-.086	.131	.029	-.018	-.013	.158	.056
44	-.132	-.199	-.200	-.086	-.029	-.281	.296	-.043	-.058	.126	-.035	-.105	.216	-.216	.046	.011	-.112	-.143	.019	-.100	-.036	-.112	-.122	.083
45	-.075	-.046	.066	-.140	-.165	-.037	-.372	.046	.163	-.146	-.022	.076	-.279	.279	-.057	.030	.053	-.012	-.044	.101	-.011	.053	-.114	.006
46	-.031	-.005	.125	.032	-.021	.130	-.231	.043	.122	-.073	-.034	.016	-.123	.123	-.075	.059	.027	.009	-.063	.174	.111	-.038	.111	.020
47	.052	.050	.115	.034	.071	.068	-.014	-.050	-.033	.052	.007	-.048	.087	-.087	-.077	.083	-.016	.046	-.089	.007	-.022	.010	.025	-.012
48	-.061	-.027	-.090	.031	.100	-.029	.099	.014	-.054	.043	.086	-.063	.092	-.092	-.157	.191	-.075	.056	.027	-.071	-.018	0.075	-.122	.001

TABLE 3A (cont.)

	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
5	1.000																							
6	.023	1.000																						
7	.127	-.013	1.000																					
8	.072	-.041	-.044	1.000																				
9	.258	.079	.013	-.015	1.000																			
0	-.027	.031	-.095	-.122	.028	1.000																		
1	.045	-.029	.108	.125	-.039	-.912	1.000																	
2	-.015	-.009	-.099	-.112	.086	.689	-.625	1.000																
3	.103	.059	.129	.054	-.051	-.454	.428	-.568	1.000															
4	.029	.031	.016	-.207	.159	.044	-.078	.124	-.078	1.000														
5	-.080	.026	-.066	.107	-.133	.069	-.048	.046	-.105	-.418	1.000													
6	-.163	-.071	-.003	.128	-.213	-.144	.152	-.243	.192	-.489	-.358	1.000												
7	.106	.048	.011	.069	.157	.066	-.085	.049	.014	.263	-.208	-.150	1.000											
8	-.162	-.031	-.123	-.167	.163	.313	-.297	.326	-.235	.229	-.074	-.156	.002	1.000										
9	-.089	-.056	-.051	-.041	-.167	.073	-.058	.003	-.070	-.276	.232	.109	-.634	.154	1.000									
0	-.021	.027	-.039	.043	.017	-.138	.137	-.065	.019	-.020	-.004	.012	.249	-.280	-.227	1.000								
1	.033	.033	-.019	-.015	.225	.092	-.094	.052	-.051	.218	-.159	-.099	.274	.459	-.376	-.167	1.000							
2	-.087	-.028	.040	-.021	-.063	-.239	.238	-.208	.056	-.039	-.005	.088	-.303	-.224	.031	-.083	-.132	1.000						
3	-.071	-.031	-.033	-.013	-.049	.052	-.042	.091	-.064	.104	-.030	-.132	.017	.218	.204	-.091	-.145	-.072	1.000					
4	-.116	-.062	-.067	-.085	-.124	.161	-.147	.066	-.108	-.282	.224	.141	-.470	.190	.848	-.183	-.291	-.145	-.159	1.000				
5	.101	.204	.083	.047	.018	-.043	.055	-.107	.128	-.052	.049	-.030	-.040	-.308	.013	-.055	-.087	-.043	-.048	-.096	1.000			
6	.174	-.029	.174	.096	.010	-.137	.161	-.102	.198	.084	-.109	-.031	.217	-.266	-.195	-.086	-.137	-.068	-.075	-.149	-.045	1.000		
7	.049	.020	.013	.024	.020	.052	-.091	.067	.008	.093	-.088	-.032	.526	-.113	-.401	-.114	-.181	-.090	-.099	-.198	-.059	-.093	1.000	
8	-.020	-.031	-.033	-.036	-.049	.052	-.042	.069	-.029	-.086	.114	.034	-.276	-.015	-.087	-.091	-.145	-.072	-.079	-.159	-.048	-.075	-.099	1.000

TABLE 4A

CORRELATION MATRIX
PREDICTION SAMPLE

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	22*	23	24	25
1	1.000																							
2	.918	1.000																						
3	.425	.407	1.000																					
4	.413	.567	.154	1.000																				
5	.718	.695	.369	.501	1.000																			
6	.633	.744	.297	.670	.600	1.000																		
7	.350	.121	.074	-.146	.380	-.072	1.000																	
8	.063	.077	.024	-.053	-.056	.026	-.132	1.000																
9	-.145	-.023	-.041	.113	-.086	.103	-.439	.078	1.000															
10	.029	-.107	.076	-.099	.059	-.146	.392	.075	-.507	1.000														
11	.299	.308	.081	.174	.277	.239	.094	-.142	-.029	-.468	1.000													
12	-.143	-.054	-.133	-.072	-.215	-.047	-.277	-.064	-.042	-.672	-.039	1.000												
13	.392	.218	.204	.013	.451	.079	.706	-.044	-.335	.340	.072	-.269	1.000											
14	-.392	-.218	-.204	-.013	-.451	-.079	-.706	.044	.335	-.340	-.072	.269	-1.000	1.000										
15	.326	.201	.156	.016	.261	.135	.364	-.130	-.066	.027	.124	-.076	.299	-.299	1.000									
16	-.264	-.182	-.123	-.089	-.211	-.156	-.243	.114	.027	.033	-.107	.051	-.177	.177	-.915	1.000								
17	-.129	-.033	-.070	.181	-.106	.063	-.271	.031	.225	-.145	-.033	.053	-.277	.277	-.138	-.273	1.000							
18	.328	.337	.134	.171	.268	.419	.029	.075	-.066	-.033	.183	-.033	.145	-.145	-.096	.109	-.041	1.000						
19	-.147	-.169	-.122	-.062	-.089	-.201	.009	-.140	.042	-.113	-.080	.183	-.026	.026	.026	-.008	-.042	-.332	1.000					
20	.097	.044	.038	-.078	.030	.077	.058	.056	.106	-.006	-.011	-.065	.004	-.004	.289	-.287	.015	-.332	-.239	1.000				
22*	.186	.125	.118	.086	.171	.142	.040	-.035	.022	-.017	.122	-.081	.073	-.073	.135	-.104	-.069	.012	-.095	.122	1.000			
23	.285	.211	.088	.042	.085	.089	.131	-.015	-.057	.069	.076	-.107	.149	-.149	.253	-.213	-.079	.003	-.018	.178	-.218	1.000		
24	.196	.101	.107	-.065	.145	.061	.291	.029	.056	.131	.052	-.265	.271	-.271	.151	-.078	-.169	.086	-.036	.079	.059	.069	1.000	
25	.096	.125	.058	.089	.020	.145	-.194	.179	.148	-.060	.012	-.039	-.062	.062	-.057	.088	-.082	-.030	-.005	-.006	.026	.159	.128	1.000
26	.144	.120	.142	.112	.079	.103	.018	.033	-.032	.062	-.029	-.042	.009	-.009	-.012	-.027	.094	.097	-.022	-.022	.021	.055	.057	.074
27	-.126	-.113	-.030	-.043	-.095	-.076	-.180	-.041	.072	-.105	-.039	.120	-.109	.109	-.033	.051	-.047	-.075	.183	-.065	-.016	.066	.075	.301
28	-.402	-.365	-.126	-.191	-.232	-.224	-.242	.009	.109	-.118	-.079	.137	-.274	.274	-.134	.100	.075	-.062	.077	-.065	-.092	-.283	-.079	-.163
29	.418	.395	.201	.141	.334	.327	.141	-.091	-.069	.012	.098	-.034	.171	-.171	.266	-.266	-.079	.126	-.092	.106	.162	.209	.079	-.007
30	.331	.201	.196	.133	.386	.137	.486	-.042	-.214	.387	.051	-.412	.494	-.494	.210	-.165	-.096	.111	-.084	-.005	.056	.164	.277	-.053
31	-.320	-.186	-.193	-.129	-.370	-.125	-.487	.043	.222	-.402	-.049	.426	-.512	.512	-.204	.157	.102	-.102	.053	.013	-.052	-.157	-.288	.061
32	.418	.291	.161	.132	.427	.228	.396	.026	-.201	.348	-.003	-.329	.429	-.429	.190	-.145	-.097	.109	-.032	-.001	.030	.249	.236	.036
33	-.249	-.121	-.104	-.009	-.307	-.082	-.374	.005	.260	-.359	.067	.254	-.315	.315	-.104	.049	.127	-.066	.049	.004	-.033	-.130	-.160	.097
34	.494	.521	.138	.302	.277	.436	-.043	.155	-.027	-.117	.219	.028	.081	-.081	.139	-.112	-.058	.139	-.135	.096	.159	.206	.080	.223
35	-.193	-.176	.021	-.098	-.088	-.207	.037	.015	.074	-.019	-.066	.018	-.027	.027	-.057	.062	-.017	-.057	.026	-.102	-.139	-.174	.129	-.091
36	-.429	-.449	-.270	-.234	-.224	-.381	.004	-.131	-.024	.081	-.120	-.008	-.100	.100	-.202	.181	.039	-.104	.153	-.080	-.132	-.348	-.147	-.244
37	.276	.230	.274	-.092	.109	.138	.076	.185	-.145	.111	.129	-.135	.142	-.142	.040	-.030	-.021	.218	-.194	-.006	.053	.049	.153	.182
38	.402	.276	.111	-.016	.319	.165	.523	.022	-.182	.155	.152	-.186	.349	-.349	.369	-.291	-.164	.089	-.074	.184	.109	.164	.180	-.074
39	-.221	-.258	-.155	-.098	-.148	-.250	.120	-.191	-.019	.061	-.092	-.004	.015	-.015	.151	-.117	-.073	-.253	.097	.113	.016	.074	-.015	-.163
40	-.049	-.053	-.091	-.030	-.026	-.063	.033	.137	-.058	.114	-.053	-.077	-.058	.058	-.062	.087	-.066	.065	.028	-.009	-.064	-.078	.053	-.049
41	.377	.295	.225	.037	.208	.221	.166	.143	-.102	.008	.095	.000	.172	-.172	.218	-.227	.039	.196	-.128	.051	.066	.123	.044	.087
42	-.230	-.167	-.132	.036	-.134	-.074	-.138	-.098	.200	-.171	-.053	.123	-.218	.218	-.158	.149	.010	-.062	.066	-.084	-.112	-.110	.001	-.047
43	-.128	-.127	-.129	-.110	-.130	-.147	-.063	.026	-.045	-.086	.072	.103	-.184	.184	-.056	.075	-.052	-.134	.060	.060	-.088	.078	-.108	.002
44	-.085	-.165	-.066	-.102	-.022	-.169	.277	-.157	-.094	.186	-.087	-.124	.204	-.204	.210	-.160	-.107	-.154	.036	.117	.095	.036	.095	-.123
45	-.070	.079	-.035	.243	.009	.148	-.395	-.081	.272	-.319	-.042	.266	-.234	.234	-.017	-.077	.229	-.017	.013	.013	.090	-.042	-.171	.002
46	.033	.064	.165	-.014	-.079	.039	-.273	.155	-.042	.082	-.038	-.056	-.109	.109	-.118	.134	-.048	-.033	-.065	.034	-.016	.023	-.128	.074
47	.042	.060	.076	-.008	.035	.067	-.039	-.036	.027	.027	.131	-.077	.063	-.063	-.126	.118	.010	.129	-.084	-.122	.033	-.045	.104	.124
48	-.072	-.057	-.099	.006	.022	-.064	.092	-.062	-.043	.086	-.040	-.058	.108	-.108	-.126	.103	.047	-.086	.167	-.120	-.084	-.032	-.053	-.099

TABLE 4A (cont.)

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
26	1.000																						
27	-.042	1.000																					
28	-.086	-.114	1.000																				
29	.156	.083	-.154	1.000																			
30	.055	-.066	-.207	.075	1.000																		
31	-.054	.071	.179	-.070	-.972	1.000																	
32	.077	-.005	-.221	.097	.674	-.653	1.000																
33	-.047	.096	.053	-.103	-.543	.496	-.612	1.000															
34	.026	-.095	-.315	.060	.038	-.028	.099	.022	1.000														
35	.074	.188	.063	-.045	.036	-.030	.036	-.005	-.303	1.000													
36	-.131	-.173	.421	-.231	-.097	.078	-.185	-.009	-.551	-.297	1.000												
37	.097	.046	-.045	.107	.172	-.171	.178	-.084	.245	.037	-.286	1.000											
38	.043	-.083	-.174	.290	.251	-.235	.194	-.182	.119	-.003	-.149	.069	1.000										
39	-.076	-.080	-.058	.101	-.054	-.063	-.103	-.011	-.319	.028	.146	-.580	.119	1.000									
40	-.028	-.077	.032	-.127	.049	-.045	.101	-.087	-.090	.039	.071	.252	-.249	-.244	1.000								
41	.132	.000	-.123	.135	.179	-.174	.194	-.112	.264	.000	-.212	.289	.504	-.437	-.189	1.000							
42	-.058	-.189	.070	-.127	-.161	.170	-.225	.093	-.152	-.003	.195	-.416	-.223	.059	-.107	-.189	1.000						
43	-.046	-.060	.017	-.099	-.144	.121	-.189	.005	-.001	-.051	.042	-.063	.145	.196	-.084	-.148	-.084	1.000					
44	-.033	-.077	.043	.169	.089	-.083	.055	-.055	-.287	.060	.099	-.334	.354	.850	-.173	-.305	-.173	-.136	1.000				
45	-.045	-.060	.063	.008	-.243	.253	-.189	.226	.189	-.103	-.073	-.136	-.410	.021	-.083	-.148	-.084	-.065	-.136	1.000			
46	-.042	-.056	.037	-.034	-.135	.071	-.113	.175	.069	-.095	-.049	.099	-.285	-.146	-.077	-.136	-.077	-.060	-.125	-.060	1.000		
47	.028	.189	-.005	.005	.048	-.045	.060	-.027	.157	.039	-.147	.558	-.131	-.353	-.107	-.189	-.107	-.084	-.173	-.084	-.077	1.000	
48	-.044	-.058	-.070	-.096	.076	-.074	.106	-.065	-.105	.009	.136	-.277	-.028	-.079	-.080	-.142	-.080	-.063	-.130	-.063	-.058	-.080	1.000

*Variable 21, other speciality rooms, had all zero values in the prediction sample.

REFERENCES

1. Alonzo, William. Location and Land Use. Cambridge, Massachusetts: Harvard University Press, 1964.

2. Barr, James L. "City Size, Land Rent, and the Supply of Public Goods." Regional and Urban Economics, 2 (January 1972), 67-103.

3. Brigham, Eugene F. "The Determinants of Residential Land Values." Land Economics, 41 (November 1965), 324-334.

4. Church, Albert M. "Utilizing Principal Components in Multiple Regression Valuation." Automated Mass Appraisal of Real Property. Chicago: International Association of Assessing Officers, 1974.

5. Churchill, Gilbert A., Jr. "A Regression Estimation Method for Collinear Predictors." Decision Sciences, 6 (Ocrober 1975), 670-687.

6. Conniffe, D. and Stone, Joan. "A Critical View of Ridge Regression." The Statistician, 22 (September 1973), 181-187.

7. Coxe, Keren. "Do Principal Components Solve Multicollinearity? The Longley Data Revisited." Proceedings of the Statistical Computing Section, American Statistical Association, (1975), 90-95.

8. Dacey, Raymond. "Theory Absorption and the Testability of Economic Theory." Zietschreft Für Nationalökonomie, forthcoming.

9. Doherty, James L. "Favorable Financing Effects on Values of Single-Family Residences." Appraisal Journal, 39 (July 1970), 398-405.

10. Draper, N. R. and Smith, H. Applied Regression Analysis. New York: John Wiley and Sons, Inc., 1966.

11. Farrar, Donald E. and Glauber, Robert R. "Multicollinearity in Regression Analysis: The Problem Revisited." Review of Economics and Statistics, 48 (February 1967), 92-107.

12. Gloudemans, Robert J. and Miller, Dennis W. "Multiple Regression Analysis Applied to Residential Properties: A Study of Structural Relationships Over Time." Decision Sciences, 7 (April 1976), 294-304.

13. Greenberg, Edward. "Minimum Variance Properties of Principal Components Regression." Journal of the American Statistical Association, 70 (March 1975), 194-197.

14. Harman, Harry H. Modern Factor Analysis. Second Edition. Chicago: University of Chicago Press, 1967.

15. Hoerl, Arthur E. and Kennard, Robert W. "Ridge Regression: Biased Estimation For Nonorthogonal Problems." Technometrics, 12 (February 1970), 55-67.

16. ____________________. "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics, 12 (February 1970), 69-82.

17. Johnston, J. Econometric Methods. Second Edition. New York: McGraw-Hill, Inc., 1972.

18. Kain, John F. and Quigley, John M. "Measuring the Value of Housing Quality." Journal of the American Statistical Association, 65 (June 1970), 532-548.

19. Kendall, Maurice G. A Course in Multivariate Analysis. London: Griffin Press, 1957.

20. Kmenta, Jan. Elements of Econometrics. New York: The Macmillan Company, 1971.

21. Lancaster, Kelvin. "A New Approach to Consumer Theory." Journal of Political Economy, 74 (March/April 1966), 132-157.

22. ____________________. Consumer Demand: A New Approach. New York: Columbia University Press, 1971.

23. Mason, Robert L., Gunst, R.F., and Webster, J.T. "Regression Analysis and Problems of Multicollinearity." Communications in Statistics, 4 (1975), 277-292.

24. Massy, William F. "Principal Components Regression in Exploratory Statistical Research." Journal of the American Statistical Association, 60 (March 1965), 234-256.

25. McCallum, B.T. "Artifical Orthogonalization in Regression Analysis." Review of Economics and Statistics, 52 (February 1970), 110-113.

26. Murphy, James L. Introductory Econometrics. Homewood, Illinois: Richard D. Irwin, Inc., 1973.

27. Muth, Richard. Cities and Housing. Chicago: University of Chicago Press, 1969.

28. O'Hagan, John and McCabe, Brendan. "Tests For the Severity of Multicollinearity in Regression Analysis: A Comment." Review of Economics and Statistics, 56 (February 1975), 368-370.

29. Ridher, R. and Henning, J.A. "The Determinants of Residential Property Values with Special Reference to Air Pollution." Review of Economics and Statistics, 48 (May 1967), 246-257.

30. Rummel, R.J. Applied Factor Analysis. Evanston: Northwestern University Press, 1970.

31. Smith, Theodore R. "New Dimensions in Appraisal Theory." Appraisal Journal, 42 (January 1974), 47-61.

32. Toumela, Raimo. Theoretical Concepts. New York: Springer-Verlag, 1973.

33. Wercher, John C. and Zerbst, Robert. "The Externalities of Neighborhood Parks: An Empirical Investigation." Land Economics, 49 (February 1973), 99-105.