Title
Hedonic Index Estimation for Commercial Buildings: Assessors and Economists and the Parallel Search for the Optimal Functional Form

Permalink
https://escholarship.org/uc/item/9682s3d5

Authors
de Silva, Harindra
Gruenstein, John M.L.

Publication Date
1988-12-01

Peer reviewed
CENTER FOR REAL ESTATE 
AND URBAN ECONOMICS 
WORKING PAPER SERIES 

WORKING PAPER NO. 88-154 

HEDONIC INDEX ESTIMATION FOR COMMERCIAL BUILDINGS: ASSESSORS AND ECONOMISTS AND THE PARALLEL SEARCH FOR THE OPTIMAL FUNCTIONAL FORM 

BY 

HARINDRA DE SILVA 
JOHN M.L. GRUENSTEIN 

These papers are preliminary in nature; their purpose is to stimulate discussion and comment. Therefore, they are not to be cited or quoted in any publication without the express permission of the author.
CENTER FOR REAL ESTATE AND URBAN ECONOMICS  
UNIVERSITY OF CALIFORNIA AT BERKELEY

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH

The Institute of Business and Economic Research is a department of the University of California with offices on the Berkeley campus. It exists for the purpose of stimulating and facilitating research into problems of economics and of business with emphasis on problems of particular importance to California and the Pacific Coast, but not to the exclusion of problems of wider import.
HEDONIC INDEX ESTIMATION FOR COMMERCIAL BUILDINGS: ASSESSORS AND ECONOMISTS AND THE PARALLEL SEARCH FOR THE OPTIMAL FUNCTIONAL FORM

by

Harindra de Silva
Analysis Group, Inc.

John M.L. Gruenstein
BPA Economics, Inc. and the University of California at Berkeley

Working Paper No. 88-154

December, 1988
ABSTRACT

This paper investigates the use of several different criteria, including $R^2$, the likelihood function, the sum of squared residuals on the untransformed dependent variable, and a modified version of the coefficient of dispersion, to judge the best functional form for a hedonic index for office buildings in the Los Angeles area. DAMAR data are used to estimate a Box-Cox transformation, for which special cases are linear, log-log, semilog, log linear and other specifications. The results show that there are large differences among the optimal functional form chosen depending on which criterion is used to judge. Substantively, the results show a high percentage of variance explained by the linear form with only the variables building area, age, and the time of the transaction, indicating the usefulness of MRA for evaluating office properties, even in the absence of geographic variables.
HEDONIC INDEX ESTIMATION FOR COMMERCIAL BUILDINGS: ASSESSORS AND ECONOMISTS AND THE PARALLEL SEARCH FOR THE OPTIMAL FUNCTIONAL FORM

by
Harindra de Silva and John M.L. Gruenstein

INTRODUCTION

It does require maturity to realize that models are to be used but not to be believed -- Thiel(1971), p. vi.

Multiple regression analysis (MRA) has been used in a variety of contexts related to the measurement of real estate values. Assessors use MRA along with other types of appraisal techniques for mass appraisal. Economists and policy analysts often use hedonic indexes (the term generally used for MRA as applied to valuation) to assess the impact of environmental amenities and disamenities and particular property characteristics on prices, to test theories of urban structure, and to estimate demand and supply functions for housing and other property. The seminal work of Rosen (1974) provided a foundation for relating hedonic index estimates to underlying demand and supply factors, and therefore to the forces that affect markets.

Other actual and possible uses include appraisal for mortgage and investment decisions, evaluation by financial institution auditors and regulators of collateral underlying residential and commercial mortgages, and as support for litigation involving property damage.

Each of these different uses to which hedonic indexes are put imply somewhat different emphases in the choice of functional
form, goodness of fit criteria, and estimation techniques. Assessors put a high weight on various aspects of the fitted values, including goodness of fit, out of sample prediction, and stability of predicted values across regressions run for the same area, but for different subsets of transactions representing different time periods. Policy analysis, by contrast, emphasizes the size and significance of the particular coefficients -- i.e., the coefficient of the variable like noise level, presence of asbestos containing material, or distance from an electric power transmission line.

This paper applies several functional forms and goodness of fit criteria to the estimation of a hedonic index for a set of office property transactions from Los Angeles. The functional forms and criteria are drawn from both the property tax and the economics literature to stimulate broader thinking and cross-fertilization.

The particular application is also of interest. While MRA has been applied for decades to the evaluation of single family residential properties, its application to income-producing property has been much sparser, to some extent because of data limitations. The few papers that have estimated hedonic indexes for office properties (Hough and Kratz (1983) and Wheaton (1984)) have investigated rents. In the present paper we estimate a hedonic index explaining office building transaction prices, which may be among the first published applications of MRA to this particular topic area.
FUNCTIONAL FORM AND GOODNESS OF FIT

There are many different functional forms used for MRA. Jensen (1986) provides a useful classification of functional forms for regression models as linear, intrinsically linear, or non-linear. Intrinsically linear means that the equation can be transformed into a linear equation, for example by taking logarithms of one or both sides of the equation. True non-linear forms cannot be so transformed. Included in this category would be models with both additive and multiplicative variables and generalized transformations such as the Box-Cox (1964).

A glance through the beginning chapters of most econometrics textbooks might create the belief that there is one correct functional form for any particular hedonic index model -- the one that characterizes the population under study -- and our efforts in estimation should be to find that one true model. As implied in the opening quotation from Thiel, practical experience (and the philosophy of science) show that there are many models to chose from, and which one we ultimately adopt as most appropriate and useful depends on such criteria as goodness of fit (GOF), accord with theory or long experience with the appropriate market, simplicity, ease of estimation, and usefulness for the particular purpose at hand.

Different traditions may have different criteria, not all of which are always explicitly described or even at the conscious level of the researchers estimating and selecting the functional forms. For example, assessors quite naturally prefer functional
forms which are in accord with their own experience with real estate markets, and which are relatively easy to explain to those with whom they must deal, for instance property owners, review boards, and courts. Jensen (1986), for instance, states both the strictly linear and strictly logarithmic models...are inappropriate for handling the concurrent additive and multiplicative effects characteristic of the real-world market (which is demonstrably not inherently linear in nature).

Policy analysts and those involved in litigation support prefer forms which isolate the influence of the specific factor they are concerned with, such as the influence of asbestos or airport noise on property values. Economists like functional forms which allow translation back to relatively simple underlying demand and supply relationships, which is how they understand the real-world market (one of Oscar Wilde's characters once maintained that one could turn a parrot into a good economist simply by teaching it the words "supply" and "demand"!).

All groups have in common that they employ GOF criteria to choose among functional forms. GOF measures based on minimizing sums of squared residuals -- most notably $R^2$ and $R^2$ adjusted for degrees of freedom -- are by far the most commonly used because of their various desirable statistical properties, the large body of literature around least squares, and the consequent easy
availability of computer software designed for least squares estimation.

The use of $R^2$ as a GOF measure when judging different functional forms must be done carefully. In particular, when the dependent variable in a regression is transformed -- for instance, to convert an intrinsically linear form into a linear form for ease of estimation -- the $R^2$ of the transformed regression has a different meaning from that of one with an untransformed dependent variable, and direct comparison across regressions is not appropriate.

To illustrate, suppose we wish to compare the GOF of a linear and an exponential model. The linear model is

$$V = \sum a_i x_i + e$$

(1)

The exponential model can be written

$$V = \exp(\sum a_i x_i) e$$

(2)

We transform the exponential equation into a semi-log equation by taking natural logarithms of both sides

$$\ln V = \sum a_i x_i + \ln e$$

(3)

The $R^2$'s of equation (1) and (3) should not be compared to judge GOF. For equation (1), the $R^2$ measures the percentage of variance explained for $V$; for equation (3), $R^2$ measures the percentage of variance explained for $\ln V$.

Another way to look at this issue is to note that in order to transform equation (2) into a linear model, the error term has to be assumed to be multiplicative. If we are interested in
minimizing the sum of squared deviations around the transaction price and not its logarithm, a more natural specification is

\[ V = \exp(\Sigma a_i x_i) + e \]  \hspace{1cm} (4)

Equation (4) is a true non-linear equation, not an intrinsically linear specification. It must be estimated non-linearly. Offsetting this disadvantage, the results can be compared directly to equation (1).

The more general point is that GOF is not an absolute criterion, but is relative to the particular purpose at hand. Nowhere is this clearer than with regard to the goals of assessment. Assessors typical are directed by state law to use a particular GOF criterion that is quite different from a sum of squared deviations based criterion like \( R^2 \), namely the coefficient of dispersion (COD), defined as

\[ \text{COD} = \frac{\Sigma \{\text{abs}[\hat{V}/V - \text{median}(\hat{V}/V)]\}/[n*\text{median}(\hat{V}/V)]}{\text{COD}} \]  \hspace{1cm} (5)

where \( \hat{V} \) represents the fitted value of the estimation procedure. The coefficient of dispersion has several important differences from \( R^2 \). It is based on the ratio of the fitted value (in essence, the assessment) to the actual value (the transaction price). It involves the median of the distribution of fitted to actual values. Finally, it is a measure base on absolute deviations, rather than squared deviations. The latter property implies that the COD puts less weight on outliers than a least squares based measure.

Because the coefficient of dispersion is the ultimate measure of GOF for assessors, it would be useful to develop a
parameter estimation method that minimized the COD directly. Such a method would stand in contrast to what is in essence an imperfect two-step procedure for choosing among functional forms -- i.e., selecting the "best fit" based on a least squares regression, and using a least squares based GOF measure, such as $R^2$, and then investigating the coefficient of dispersion of the resulting set of fitted versus actual values to see whether it is acceptable (for instance, is below 10%).

Designing a general "least COD" estimation method involves some difficulties, in terms of computational efficiency, treatment of outliers, proof of convergence, and estimating the underlying statistical properties of the resultant estimators. In this paper, we will develop a more limited estimation technique, which will nevertheless be able to shed some light on the shape of more general techniques, and also provide some indications of the differences between the properties of least squares and least COD estimators. We will also discuss some of the further work that would be necessary in order to develop a practical, working technique.

The technique used is as follows. A grid search is carried out for possible parameter values for lambda ($\lambda$) and mu ($\mu$) for the generalized Box-Cox transformation

$$
((Y^\lambda - 1)/\lambda) = a_0 + \sum a_i ((X_i^\mu - 1)/\mu) + e
$$

(6)

For each pair of values of the Box-Cox parameters, lambda and mu, (but not for each set of coefficients), a modified COD (MCOD) is calculated, and a grid search carried out to find the minimum
value of the MCOD. The modified COD has the same formula as the COD, except that the mean of \( \hat{V}/V \) replaces the median of \( \hat{V}/V \):

\[
\text{COD} = \frac{\sum \{ \text{abs}(\hat{V}/V - \text{mean}(\hat{V}/V)) \}}{\text{n} \times \text{mean}(\hat{V}/V)}
\]

Other common measures of GOF, including adjusted \( R^2 \), the value of the likelihood function, and the sum of the squared residuals are also calculated. Because the transformation applies to the dependent variable, yet another GOF measure is calculated, the sum of the squared residuals for the untransformed \( V \), \( SSR(V) \).

Four regressions are then compared: the maximum likelihood estimate (the typical Box-Cox), the maximum adjusted \( R^2 \), the minimum \( SSR(V) \), and the minimum MCOD. An important and useful question is whether these GOF criteria imply the same or different parameters for optimization.

**SPECIFICATION AND DATA**

In theory, the variables that should be useful for estimating transaction prices for office buildings should relate ultimately to net operating income (NOI). On the revenue side, variables that relate to the demand and supply for office space should be present. Other variables that relate to expenses should enter in to complete the NOI equation. Finally, variables that relate to the capitalization rate should also enter.

In practice, data limitations are usually quite severe and dictate the variables that will be used. Very few hedonic indexes for office buildings appear to have been the subject of published articles in the past. We were able to find only two.
Hough and Kratz (1983) used a sample of office buildings in
downtown Chicago to test whether "good" architecture had an
effect on rents that office tenants were willing to pay.
Wheaton (1984), in a paper sponsored by the Lincoln Land
Institute, estimated a similar function for office rents for
properties in the Boston area, in order to look for
capitalization of intrametropolitan property tax differentials.
Both papers used rent, rather than transaction price as the
dependent variable, giving the regressions a somewhat different
meaning. In particular, regressions using rent would relate to
the demand and supply for office space, while regressions using
transactions prices would relate to the demand and supply of real
estate assets. The two markets are related, but not identical
(See, for instance, Corcoran (1987)). Typical variables used in
these regressions included building area, lot area, number of
floors, age of the building, the time of the transaction, as well
as geographic variables.

The data base used for this paper is the DAMAR data base. As
described in Reichert and Kang (1986), the DAMAR data base
includes a large number of residential and commercial property
transactions, and a wide variety of variables. As an on-line
database, the DAMAR data are relatively easy to access and use
and can be downloaded onto a PC, for ease of analysis.
Unfortunately, as Reichert and Kang found for their study of
apartment sales, the number of observations with sufficiently
complete data for hedonic index estimation can be a relatively
small percentage of total sales. For this paper, a sample of 124 sales of office buildings in Los Angeles and Orange County that were sold between 1980 and 1987 was used. Whether the restriction to a subsample with complete information introduces threats to the validity of the regressions -- in particular, selection bias -- is hard to determine without a thorough analysis of the data, which is beyond the scope of this paper (See Cook and Campbell (1979) for a discussion of various threats to validity). However, it would seem reasonable that a higher percentage of the sales lacking complete information on the database, would also be more likely to be non arms-length, and therefore properly to be excluded.

The variables used for the regressions included building area (BLDAREA), age of the building (AGE), time of sale (TIME), lot area (LOTAREA), and the number of buildings (BLDGGS). We would expect that building area, time of sale, and lot area would have a positive effect on transaction price, and that age would have a negative effect. The expected effect of the number of buildings is less clear. Both Wheaton and Hough and Kratz found that the number of floors had a positive influence on rent, probably because of the prestige value of the upper floors. Therefore, since having more buildings in a property with a given amount of square footage is likely (although not necessarily) to imply that each building has fewer floors, we would hypothesize that the number of buildings variable would enter negatively.
Because our resources were concentrated on comparing functional forms, we were unable to devote sufficient time or expense to add geographic variables to the dataset, such as distance from the center of downtown, or distance to highways, mass transit, and so forth. This would appear to be a shortcoming from the point of view of both theory (for example, bid-rent theories of urban location, see Mills (1979)) and experience (note the common saying among real estate brokers that the three most important factors in price are location, location, and location). While we freely admit that the lack of such variables should be corrected in further research, it is nevertheless interesting to note the work of Butler (1982), who investigated the effect of excluding location variables from a hedonic index for single-family houses, for which one would expect at least as great if not a greater effect of location, and found that GOF measures were little affected by the exclusion. Although this result is somewhat surprising, it nevertheless adds support to the use of our data set for testing several different criteria of GOF.

RESULTS

In this section the results of different estimations using different functional forms are examined from three different perspectives. First, a set of regressions using the Box-Cox transformation is examined for GOF on the four criteria listed above. The sample and set of included variables is held constant across regressions in order to maintain comparability. An
important observation is that each criterion is optimized at a different set of values for lambda and mu. Second, the signs and significance levels of each of the coefficients are compared across the four optimal regressions. This can be thought of as a test of robustness for the specification. Third, the linear form is singled out for special attention, both because of its simplicity and because of its good performance on the adjusted $R^2$ GOF measure. Coefficients for several specifications of the linear form are compared to one another and to previous results in order to assess the size of the influences on the value.

**Goodness of Fit**

The basic specification used to test GOF included the variables AGE, TIME, BLDAREA, LOTAREA, and BLDGS. The sample consisted of the same 124 observations for all regressions. Table 1 shows the results for the four GOF measures discussed above. The true coefficient of dispersion (COD) is also included. Although it was not possible to calculate the COD for each value of lambda and mu in order to do the search on that measure, we thought it useful to include its value for each of the optimal regressions, for comparison purposes with the MCOD. Finally, three other forms, the semi-log, log-log, and log-linear were also included, because of their relative simplicity and common application. One striking feature of the results are the quite different values of the two parameters which gave the highest values for the four GOF measures. The likelihood function is maximized at lambda = .25 and mu = .70. The adjusted
<table>
<thead>
<tr>
<th>MEASURE</th>
<th>MEASURE</th>
<th>SSR(V)</th>
<th>MCOD</th>
<th>COD</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAX L</td>
<td>-176.1</td>
<td>.54</td>
<td>366</td>
<td>153</td>
</tr>
<tr>
<td>MAX R2</td>
<td>-271.6</td>
<td>.75</td>
<td>580</td>
<td>174</td>
</tr>
<tr>
<td>MIN SSR(V)</td>
<td>-18.7</td>
<td>.67</td>
<td>342</td>
<td>156</td>
</tr>
<tr>
<td>MIN MCOD</td>
<td>-376.1</td>
<td>.05</td>
<td>40438</td>
<td>140</td>
</tr>
<tr>
<td>SEMILOG</td>
<td>-203.8</td>
<td>.33</td>
<td>10420</td>
<td>149</td>
</tr>
<tr>
<td>LOG-LOG</td>
<td>-202.1</td>
<td>.35</td>
<td>1405</td>
<td>150</td>
</tr>
<tr>
<td>LIN-LOG</td>
<td>-319.5</td>
<td>.45</td>
<td>1255</td>
<td>181</td>
</tr>
</tbody>
</table>

L = VALUE OF LOG LIKELIHOOD FUNCTION
R2 = ADJUSTED R SQUARED OF THE TRANSFORMED VARIABLE
SSR(V) = THE SUM OF THE SQUARES RESIDUALS OF THE UNTRANSFORMED VARIABLE
MCOD = MODIFIED COEFFICIENT OF DISPERSION (MEAN REPLACES MEDIAN)
COD  = COEFFICIENT OF DISPERSION
$R^2$ is highest for the linear regression ($\lambda = \mu = 1.0$). The SSR for the untransformed transaction price is minimized at $\lambda = +.5$ and $\mu = 1.0$. Finally, the MCOD is minimized at $\lambda = -.5$ and $\mu = .4$. These results would certainly seem to indicate that which GOF criterion you pick has a large bearing on how you judge your regression!

It is important to note that we cannot claim that these values actually represent optimal ones, because the grid searched may have been too coarse, and more importantly, because the true optima may lie outside the range searched. In particular, a search of the diagonal elements of the grid for values greater than 1.0 indicates that the highest adjusted $R^2$ appears to be for $\lambda$ and $\mu$ both greater than 1.0, which was outside our range for the two parameter search. The SSR(V) measure may also get smaller for values of $\mu$ greater than 1.0.

A second important aspect of the table is the absolute size of the MCOD measure, essentially an order of magnitude too large to be acceptable for assessment purposes. In principle, one would think that the use of the median rather than the mean might help ameliorate the problem, but as the last column of the table shows, the true CODs are in all cases much larger than the MCOD's. It is true, however, that the relative ranking of the four regressions is the same for the two measures, which might indicate that the computationally more tractable MCOD might be an acceptable proxy for the COD in certain cases. One source of the large values for the COD and MCOD appear to be the large
variance in the dependent variable. The range of variation for transaction prices for office properties is typically much larger than for single-family houses. So relatively small absolute errors for the observations with low transaction prices may still turn out to be very large percentage errors. These large percentage errors have a stronger impact on the MCOD and the COD than on the $R^2$ and SSR because the former two measures are based on percentage deviations. Inspection of the residuals also indicates that the large COD's are partly the result of the presence of a few outliers, a point which it would be instructive to pursue further, given more time and resources.

It is interesting to note that the relatively simple forms involving log transformations on one or the other side of the equation, do not fare particularly well on these measures of GOF, except possibly the semi-log and log-log with respect to the likelihood function and the MCOD (the COD's were not calculated for these functional forms).

**Robustness**

Table 2 shows the results of the four "optimal" regressions and the three other "simple" ones with respect to the sign and significance of the regression coefficients.

Because of the transformations involved there is no simple way to compare the magnitudes of the coefficients themselves. (One way to compare the coefficients is to convert them into prices for attributes by taking partial derivatives of the untransformed left hand side variable with respect to the
### TABLE 2
SIGN AND SIGNIFICANCE OF COEFFICIENTS

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>REGRESSION</th>
<th>AGE</th>
<th>TIME</th>
<th>BLDAREA</th>
<th>LOTAREA</th>
<th>BLDGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX L</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>MAX R2</td>
<td>- **</td>
<td>+ **</td>
<td>+ **</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MIN SSR(V)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>MIN MCOD</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>SEMILOG</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>LOG-LOG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>LIN-LOG</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>**</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

L = VALUE OF LOG LIKELIHOOD FUNCTION  
R2 = ADJUSTED R SQUARED OF THE TRANSFORMED VARIABLE  
SSR(V) = THE SUM OF THE SQARED RESIDUALS OF THE UNTRANSFORMED VARIABLE  
MCOD = MODIFIED COEFFICIENT OF DISPERSION (MEAN REPLACES MEDIAN)

** denotes statistically significant at the .01 level  
* denotes statistically significant at the .05 level

AGE = building age; TIME = year sold; BLDAREA = building area (sq. ft)  
LOTAREA = area of lot (sq. ft.); BLDGS = number of buildings
untransformed right hand side variables, and then evaluating the resulting price functions at a common point in attribute space, for instance the mean. Such a comparison can be misleading. See Cassel and Mendelsohn (1985) for a critique.)

Not surprisingly, the variable BLDAR is the most consistent, always entering positively, and always significantly, except for the minimum MCOD estimate. The value for the coefficient for BLDAR for the log-log form was not significantly different from 1.0, which indicates that price is essentially proportional to building area. Both LOTAREA and BLDGS are also consistent with respect to sign, with the former entering positively, and the latter negatively, although LOTAREA is usually significant and BLDGS is not. These are consistent with expectations and previous results.

The variables AGE and TIME are more problematic. They show variations in signs across regressions; however, they are only significant in the linear functional form, and in this case, they do have the expected signs. This result carries some weight for the linear form, given the clear theoretical justification and intuitive appeal of these variables.

**Linear Regressions**

Because of several reasons cited above, it seemed useful to look more closely at several specifications involving the linear functional form. As table 3 shows, in these comparisons, as the number of variables in the specification increased, the number of
Table 3
Comparison of linear regression coefficients

(Absolute values of t-statistics in parentheses under coefficients)
(Coefficients of CONST, AGE, TIME, and AGETIME are divided by 1000)

<table>
<thead>
<tr>
<th>Regression</th>
<th>n</th>
<th>R2 (F)</th>
<th>CONST</th>
<th>AGE</th>
<th>TIME</th>
<th>BLDAR</th>
<th>LOTAREA</th>
<th>BLDGS</th>
<th>AGETIME</th>
<th>AGEBLDAR</th>
<th>TIMEBLD</th>
<th>AGETMBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>189</td>
<td>.86</td>
<td>-17248</td>
<td>-229</td>
<td>2009</td>
<td>134.61</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(372)</td>
<td>(4.4)</td>
<td>(3.3)</td>
<td>(3.9)</td>
<td>(32.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>174</td>
<td>.91</td>
<td>-17459</td>
<td>-171</td>
<td>1551</td>
<td>143.86</td>
<td>14.22</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(422)</td>
<td>(4.7)</td>
<td>(2.6)</td>
<td>(3.5)</td>
<td>(39.3)</td>
<td>(0.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>124</td>
<td>.75</td>
<td>-12131</td>
<td>-160</td>
<td>1218</td>
<td>129</td>
<td>17.68</td>
<td>-1251</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(73 )</td>
<td>(3.2)</td>
<td>(3.7)</td>
<td>(3.1)</td>
<td>(18.5)</td>
<td>(1.4)</td>
<td>(1.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>189</td>
<td>.93</td>
<td>11605</td>
<td>-355</td>
<td>-2437</td>
<td>-16.72</td>
<td>--</td>
<td>84.5</td>
<td>1.41</td>
<td>24.74</td>
<td>-651</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(368)</td>
<td>(2.4)</td>
<td>(2.4)</td>
<td>(3.5)</td>
<td>(1.1)</td>
<td></td>
<td></td>
<td>(3.5)</td>
<td>(3.2)</td>
<td>(10.9)</td>
<td>(5.9)</td>
<td></td>
</tr>
</tbody>
</table>
observations decreased, because of missing values for the newly included variable.

The first three specifications show the effect of excluding variables which are not significant, while allowing the sample size to increase accordingly. With regard to GOF, the $R^2$'s increase substantially as the variables are dropped, although there is a decrease when the insignificant variable LOTAREA is dropped. The decrease must come from the inclusion of 15 more observations, at least some of which must lie farther from the regression line than the ones included before did, on average. Regressions 1-3 show a reasonably consistent pattern of coefficients, although a fair amount of variation. This could be due to both multicollinearity and differences in the samples.

According to the point estimates from these regressions, office buildings in this area sold for between $129 and $145 per square foot, on average. Each additional year of age decreased the value of the property between $160,000 and $229,000. Over time, properties increased in value between $1,218,000 and $2,009,000 per year. These are all reasonably plausible values. It certainly can be objected that time and age should enter multiplicatively rather than linearly. While in general, this must be true, for the relatively limited sample that we are dealing with, it is remarkable how well the first three linear forms fit the data, judging by $R^2$ and the untransformed sum of squared residuals (although not on the MCOD measure). One way to investigate the non-linearities in time and age is to include
interaction terms. Equation 4 shows the effect of including all possible interaction terms from the specification in equation 1. The regression explains 93% of the variance in transaction price, the highest of any specifications tested. While it is hard to judge the signs and magnitudes of the coefficients directly, because of the need to combine terms to get an overall effect for a variable like age, it is interesting to note that all of the terms are significant at the .01 level except BLDAR, the effect of which can be presumed to be accounted for in the interactions with other variables.

SUMMARY AND CONCLUSIONS

The most important message of this paper is to keep your eye on the ball you want to catch. Different measures may be used to judge goodness of fit of a regression, which are appropriate for different circumstances, and which may give very different results. Among these different measures, the coefficient of dispersion stands out for assessors, because, for better or for worse, in many cases it is required by law or practice.

Because of the special role of the coefficient of dispersion, it seems reasonable to use it as a way of directly estimating parameters. The results of this paper are not terribly encouraging in this regard, given the very high values found for this measure for all regressions tried. These results are of course preliminary in the sense that they apply only to one area and one type of property. More work should be done with other samples, and in deriving direct estimation methods for all
the parameters, not just μ and λ, in which criteria such as COD can be used as the optimizing criterion.

Substantively, the results for offices are interesting. A high percentage of the variation in price could be explained with just a few variables, even without including geographic variables. The estimated parameter values for the linear forms seemed reasonable. These results indicate that MRA could be a useful way of estimating property values for the office sector, as it has shown to be for other property types.
REFERENCES

Journal of the Royal Statistical Society, Series B. 
26:211-252.

Butler, R. 1982. The specification of hedonic indexes for urban 
housing. Land Economics. 58:96-108.

forms for hedonic price equations: comment. Journal of Urban 
Economics. 18:135-142.

Cook, T. and D. Campbell. 1979. Quasi-experimentation: design and 
analysis issues for field settings. Chicago: Rand McNally.

Corcoran, P. 1987. Explaining the commercial real estate market. 


Jensen, D.L. 1986. Comments on Reichert and Kang, alternative 

Reichert, A. and H. Kang. 1986. Alternative approaches to the 
mass appraisal of apartment buildings: guidelines for 

differentiation in pure competition. Journal of Political 
Economy. 82:34-55.