Error Trade-offs in Selection of Comparable Sales for Residential Valuations

By Max Kummerow

Curtin University

kummerom@cbs.curtin.edu.au

Abstract

Regression appraisal methods promised to increase precision, reduce bias and decrease costs of residential valuations. Despite seventy years of academic literature, regression valuation methods remain caught in a dilemma of sample heterogeneity versus sample size. Sample variance, measurement errors and misspecification errors all increase with sample size. This means that in practice, the law of large numbers may not hold-- precision of estimates may decrease as sample size increases. The market outcome whereby valuers use a few comparable sales and adjust for differences probably is mathematically defensible in light of these error trade-offs. The optimal number of comparable sales is usually small, but varies between valuations depending upon a trade-off between random errors that decrease as sample size increases versus biased misspecification and measurement errors that increase with sample size. Because of this problem the best way forward for computer assisted valuation methods is probably to try to develop algorithms that mimic traditional sales comparison methods. A 1983 paper by Colwell, Cannaday and Wu, and work by Robbins, Graaskamp and Dilmore are foundations for attempts to automate sales comparison valuation methods.

Introduction

We begin by reviewing a 1991 *AREUEA Journal* paper by Kerry Vandell on "Optimal Comparable Selection and Weighting."[1] We then discuss reasons why the assumptions of the Vandell paper often are not true in practice. To clarify the issues we refer to Colwell, Cannaday, and Wu (1985) on the distinction between regression methods versus sales comparison methods and papers on comparable sales selection criteria (Tshira, Isakson). We then explore implications of relaxing Vandell's assumptions to allow for omitted variable and measurement errors, developing the idea that there are trade-offs between different kinds of errors as the sample size (number of comparable sales) increases. We cite an empirical paper by Lusht and Pugh (1976) where the data appear to support our view that there are error trade-offs in sales comparison appraisal methods.

Vandell's comparable selection and weighting method

Vandell's 1991 article uses the estimated regression coefficients, variances and covariances of predictor variables from a "submarket" of homogeneous sales to select, adjust, and weight a smaller sub-set of comparable sales. The paper's conclusions section calls it "a methodology…that ranks the desirability of comparables according to an objective criterion, derives optimal weights for each comparable and determines the optimal number and set of comparables to take in the calculation of the final value estimate." (Vandell, 1991:235) The paper therefore claims "To…bring appraisal practices closer to a 'science' with improved reliability and consistency of the value estimates." (Ibid: 216) In contrast, the paper points out, "ad hoc" methods used by practitioners "allow discretion on the part of the analyst, thus permitting bias to enter." (Ibid: 214)

The Vandell paper offers a useful framework for thinking about appraisal methods. First, price is clearly treated as a random variable. Most academic researchers are well aware that observed prices are events from probability distributions of possible prices, but many practitioners still approach appraisal as if they were searching for the one true price, giving too little attention to variation in prices and errors in price estimates. Information

[1] *Real Estate Economics* published three subsequent comments elaborating and debating aspects of Vandell's approach, two by Gau, Lai and Wang and one by Richard Green.

on the price estimate's probability distribution would provide valuable additional information to clients for use in risk management and investment decisionmaking.

A second virtue of the 1991 paper is that in choosing comparable sales Vandell defines “nearness” of comparables in dollars, calculated as adjustment coefficients times hedonic characteristics. This improves on methods that define nearness in terms of physical characteristics without consideration of whether or not the market places any value on those characteristics. Third, Vandell’s paper uses an equation proposed by Colwell, Cannaday and Wu (1985) which distinguishes sales comparison as a method distinct from regression. This could help improve the precision of price estimates (see discussion below). A fourth appealing feature of the paper is that it carries these concepts through as a series of steps: estimating coefficients, comparable selection, then adjustment and weighting of comparables to arrive at a final value estimate using both variances and covariances.

The Vandell paper makes three practical recommendations to practitioners that are “counter to conventional practice.” 1) It may not always be optimal to consider first the “best” comparables (due to covariances among comparables' hedonic predictor variables), 2) It is always desirable to consider more comparables so long as they are optimally weighted and 3) Weights usually selected for “inferior comparables” are typically too small. (Ibid: 213) This paper relaxes the assumptions necessary to Vandell’s results and reaches different conclusions. Vandell and comments on Vandell's paper by Gau, Lai, and Wang and Green all emphasize the importance of empirical testing to see if the assumptions are nearly enough met for practical use of the method.

Vandell’s Empirical Example

Vandell presents an empirical example to illustrate his recommended appraisal method. Using a sample of 360 sales, the following adjustment coefficients were estimated by regressing price on four hedonic characteristics:

Bedrooms	Baths	House area	Lot size
$17,600	$9,400	$56.20/sq. ft.	$2.12/sq. ft.

Variances and covariances of the adjusted house value estimates were calculated from the variation in these hedonic characteristics. A minimum variances/covariances criterion was then used to select and weight a subset of ten "comparable" sales resulting in a value estimate for the subject property, a four bedroom home. (Vandell, 1991: 233)

The ten comparables selected in the example are diverse, ranging from a $93,400 two bedroom, one bath, 1800 sq. ft. home to a $289,500 five bedroom, 4 ½ bath, 3370 sq. ft. home.[2] There are no sales in the 10 comparables sub-set selected with prices between $205,000 and $240,000. In industry practice it would be unusual for an appraiser to use a 2 bedroom, one bath home as a comparable for a 4 bedroom, three bath home that is 1050 sq. ft. larger.

Table 1 Four comparable sales from Vandell, 1991

Sale #	Price	Bedrms	Baths	Area	Lot Area
10	205,500	3	2.5	2500	12100
3	240,000	4	2	2950	11600
Subject	?	4	3	2850	12100
7	249,000	4	3	3200	13400
9	249,300	4	3.5	3100	12000

By casual inspection of the four most similar comparables (Table 1) the subject appears to be inferior to sales 7 and 9 and superior to sale 10. Comparison with sale 3 is ambiguous—the subject has an extra bath and larger lot size, but smaller house area. Using Vandell's estimates of adjustment coefficients, the indicated price for the subject property calculated from sale 3 would be $240,000 + $9400 - 100*56.20 + 500*2.12 = $244,840. But these adjustment coefficients give $22,400 and $18,500 as the calculated difference in prices between sale 7-sale 3 and sale 9-sale 3 respectively rather than the observed value differences of $9000 and $9300. Vandell's estimated adjustment coefficients may be too large.

Vandell concludes that using all 10 comparables provides the best price estimate, because then calculated variance of the weighted average subject property value estimate is smallest. His estimate of value is $233,000 with standard error of $3900.

The value estimate (as well as the standard error) keeps dropping as Vandell adds comparables, from $256,000 with one comparable, to $244,000 with 3 and finally to $233,000 when a weighted average of all ten sales is used. But surely adjusted comparables ought to be equally likely to fall above or below the price estimate instead of always pulling the price estimate downwards. If so, the probability of all 9 sales pulling down the previous estimate is 1/2^9 or .0019. This unlikely pattern is further evidence that the adjustments are biased.

Reasons for bias are easy to propose: a) The "submarket" of sales used to estimate the coefficients may be heterogeneous and the best fitting response surface estimated from this larger sample may not be representative of the pricing process for the subject property. b) The four variable hedonic model used certainly omits important variables (location, neighborhood, design and construction quality, and many other features) and therefore the coefficients are subject to bias.[3] c) There are almost certainly measurement errors in the data. The model predicts poorly because assumptions necessary for consistency are violated.

The Assumptions

Vandell's 1991 assumptions are the usual OLS regression assumptions and that:

1) The population has been adequately delineated by choice of a "submarket" within which the data generating process is reasonably uniform.

---

[2] The 10 sales in the Vandell example were apparently selected for their small errors relative to predicted values. Obviously the method used to select the 10 from the larger sample of 360 resulted in selection of very diverse properties, rather than "near neighbors."

[3] When we attempted to calculate coefficients for Vandell's four predictor model using 22,000 sales from Perth, Australia, and various sub-sets of this population of sales, most samples result in coefficient estimates with "wrong signs", that is negative signs for the bedroom and bathroom variables and coefficients are not stable between sub-samples. This is further evidence of omitted variable bias if a simple four-predictor specification is used.

2) The model is correctly specified.

3) There is negligible measurement error in the predictor variables.

4) Cost of information does not affect the optimal valuation process.[4]

In practical valuation problems, both econometrician and traditional valuer must make decisions in choosing the relevant population of sales, determining sampling strategy to derive a subset of comparable sales and specifying the hedonic model (statistical or ad hoc) to use in adjusting comparable sales. Bias and inefficiency could result from those decisions regardless of method.

Submarkets

There is much more regression valuation literature testing alternative specifications for hedonic pricing models than addressing the questions of how to choose population and sample. Real estate is by nature heterogeneous as we learn in chapter one of real estate texts and from observation. Variances both of price and predictor variables and hence variance of the subject property price estimate tend, therefore, to increase with submarket size. This means defining a smaller, more uniform population will increase precision of price estimates. However, smaller, more uniform populations reduce feasible sample size. Scarcity of sales data is a major issue limiting the precision of regression methods, especially in thinly traded housing markets (Epply, 1996). Lusht and Pugh (1981) tested three alternative geographical strategies for comparable selection to try to expand the set of comparables. They tested single neighborhood, multiple similar neighborhoods, and community-wide samples. The wider choice of comparables from multiple similar neighborhoods gave more adequate sample sizes and smaller prediction errors than single neighborhoods. However, expanding the population further to the whole community reduced precision of price estimates.

Using more homogeneous data sets tends to reduce the range of predictor variables. For example, if all the houses on a street are the same size, house area will disappear as a

[4] Vandell is well aware of these assumptions and the need for empirical testing. For assumption 1, he says “once the analyst has decided on a submarket" referring readers to Lusht and Pugh's article on submarket selection (p. 216). Omitted variable bias, measurement errors, and cost of information are mentioned in footnotes on pages 219, 217, and 227 respectively. This paper will relax all four of these assumptions, based on evidence that all four issues cannot be ignored in applications.

useful price predictor in this sub-sample. Yet a regression model with low R^2 estimated from a very uniform sample could conceivably give a better price prediction (smaller errors, less bias) than a model with more explanatory power estimated from a more heterogeneous (i.e. higher variance) sample. Whipple (1995) maintains that appraisers should refer to market research, that is, they should interview buyers to find out what variables matter, and define submarkets as homogeneous groups of buyers, rather than in terms of property types or neighborhoods.

Correct specification

In recent decades there has been considerable doubt raised about *a priori* models. All parsimonious (i.e. all useful) models leave things out, and so may be subject to omitted variable bias. Kennedy quotes G.P. Box "All models are false, some are useful." Bruce and Sundell's 1977 survey article found that 141 variables had been used in hedonic house price papers surveyed. There is little or no theory to tell us whether in a particular case the market values swimming pools or worries about the stigma of a contaminated site. Interaction effects can be important. In hedonic pricing, there is no logical necessity that coefficients remain constant when the values of other variables change. Most hedonic relationships are non-linear (Brotman, 1990). In fact, the idea that there is a functional relationship may be only a convenient approximation. (See Ekeland, 1988, Giere, 1999).

Without theory, it is necessary to "let the data speak," that is, to test alternative specifications. But data mining results in pre-test biases. If C is the number of candidate variables tested and K is the number included, the corrected alpha level is approximately C/K*alpha, (Charemza and Deadman, 1992). For example, if ten variables are tested and two included the alpha tested is 10/2*.05=.25, not .05. Not surprisingly, if results are due to data mining, out of sample prediction may be poor (Leamer 1983).

Many papers in the regression valuation literature note the prevalence of multicollinearity in housing hedonic predictors. In combination with small sample sizes, this makes it difficult to find reliable models. In practice, it is possible to find better and worse model structures, but it is not possible to say that we have found the "correct" model.

There is little justification for regarding all buyers, even of similar houses, as having identical pricing models. For example, families with children may place a premium on good schools, while empty nesters may not even consider schools relevant. Few hedonic regression models include buyer characteristics, that is, information on homogeneous *buyer* subgroups called “submarkets” by marketing theorists.

The relevant distribution is the possible sale prices of one particular house at a point in time, not the distribution of all house prices or even a subset of houses. This makes the heterogeneity of properties, buyers, and pricing models major concerns. Regression methods produce a predicted value that falls on a response surface, which is in a sense an “average” estimated from the sample. But in the particular region of a particular property’s characteristics, the responses may differ from those that are estimated from a larger sample.[5]

Data sets often omit the qualitative uncountable characteristics that real estate agents tell us are quite important in buyers’ evaluations of properties, such as neighborhood prestige, views, street appeal, or design quality. Vandell's model, for example, omits location variables, neighborhood characteristics, and qualitative variables making its assumption that there are no omitted variables highly questionable.

Misspecification bias is the Achilles heel of hedonic regression methods. Rather than assuming we have the *right* model, *a priori*, we would do better to admit that we always have the *wrong* model and not enough data to find the right one. Because there is little theory, a specification search must test a number of alternative specifications and focus on prediction errors rather than calculated variances to test validity. However, testing more than one model creates pre-test bias increasing the possibility of spurious models. If the model is wrong, the calculated variances and covariances are wrong and therefore so are minimum variance comparable selections and weightings.

Measurement errors

All physical measurements involve some degree of error. At the quantum level, Heisenberg’s principle tells us that measurement error is a property of the universe. With housing data, measurement errors are much larger. Students in appraisal classes reported a distribution of house sizes, even though all of them measured the same building.[6] No doubt practicing valuers also sometimes forget to add a room, disagree about how to account for a finished attic, multiply incorrectly, hold the tape measure slightly crooked or have the batteries run low.

Many of the variables we count and enter on spreadsheets that seem to be measured precisely are measured with error if one looks behind the simple number. A bathroom might be large with marble and a gold plated spa or a tiny closet with leaky forty-year-old toilet and rotting floorboards. Items recorded in a database as a particular number cover a range of property features with widely differing amenity values. The concealed variation can be thought of as measurement errors, since the data does not distinguish these price-affecting differences. These errors might be on the order of several hundred percent in some cases.

Cost of data

Data is costly. In Western Australia tens of millions of dollars have been spent by the public sector to create property databases for use in the land records and property taxation systems and these data sets include only a fraction of the hedonic variables that an appraiser might wish to test in a pricing model. It would cost millions more to clean and verify this data and 100% accuracy could never be attained so long as data is interpreted and entered by fallible humans. For an individual appraiser, inspecting additional comparable sales adds to costs. With fees and profit margins kept low by competition, looking at a few more sales might easily mean a net loss on a valuation job. Cost pressures, therefore, send appraisers towards models and methods requiring a minimum of data.

---

[5] Mike Robbins pointed out this "average response may not describe a particular case" problem of heterogeneous data sets.

[6] In a sample of 53 student valuations of a home known to be 2100 sq. ft. in area, the measurements fell into a roughly normal curve with outliers more than 20% in error, and a majority falling within 10% of the true value. Appraisers have to make numerous decisions in measuring houses about "what counts" and may differ in their conclusions. Comparing the students to appraisers observed measuring houses, the students spent more time and care to "get it right" (the appraisal was a major assignment) but were less expert than the more experienced appraisers.

Sales comparison adjustments versus regression hedonic coefficients

A straight hedonic regression represents sale price as a function of a set of hedonic variables. In equation 1, $\hat{b}_j$ are coefficients and *Xsj* are the measured hedonic characteristics of the subject property. $\hat{V}_s$ is the predicted price of the subject property. The coefficients are estimated from a sample of n sales with k hedonic characteristics.

(1) $$\hat{V}_s = \sum_{j=1}^{k} \hat{b}_j X_{sj} + \hat{e}$$

Colwell, Cannaday, and Wu (CCW, 1985) represent “grid adjustment” sales comparison appraisal methods as a form of regression model, but with one of the key predictors being the unadjusted sale price of a comparable property. The close substitute price proxies for an unknown complex hedonic pricing model. The genius of this technique is that it sidesteps a host of difficult specification and estimation problems, albeit at the cost of using a proxy sold-property price to “indicate” the price of the subject. This is valuation by analogy. Only differences between subject and comparable sales need to be estimated.

(2) $$\hat{V}_s = V_o + \sum_{j=1}^{m} \hat{a}_j (X_{sj} - X_{oj}) + e$$

To distinguish the shorter list of characteristics from the k hedonic variables of equation 1, note that there are m adjustment factors, with m<k. The “s” subscript refers to the subject property, while “o” refers to the observed price of a comparable sale. $\hat{V}_s$ is therefore the “indicated” value of the subject property based on valuing a short list of m differences between the subject property and comparable sale and adding these net adjustments to an observed sale price of a similar property. Characteristics that do not vary between subject and comparable are proxied by the sale price of the comparable. Subtracting the observed comparable sale price from both sides restates this model as a model to estimate the value of hedonic *differences* between subject and comparable sale. The value of these differences should be small relative to the total price, so even a large

percentage error in estimating the differences leads to only a small percentage error in the subject property price estimate provided the comparable's price is a good proxy.[7]

Notice that there is no "hat" over the price of the comparable sale in equation 2—this is an observed price, entered into the equation as a measured quantity, not a random variable. This observed price is, however, a realization of a random variable of possible prices that the comparable might have sold for, and therefore should be thought of as $V_{oi} = V_{im} + e_{oi}$, where V_{im} is the mean of a random variable of possible sale prices for the ith comparable sale. Because we only observe one event from this distribution, it is impossible to know precisely the distribution of Vi, although we can make educated guesses based on circumstantial evidence (Kummerow, 1999).

Note that $a_j \neq b_j$ because the models are different. In equation 2 the coefficient is multiplied by an amenity characteristic's *difference* between two properties, a smaller number, while in equation 1 it is multiplied by the *full* measure of the characteristic for each individual property. Equality of coefficients in each equation is as unlikely as equality of marginal and average prices.[8] And, m<k, there are fewer adjustments in the sales comparison model. The observed comparable price proxies for omitted variables in a hedonic pricing process that may differ from the included variables in a straight regression model. Both the dependent variable and the predictors differ between the two approaches. Sales comparison coefficients therefore differ from straight regression coefficients even where predictor variables are the same.

Comparable Selection and Weighting

In selecting submarkets to comprise a uniform population of sales data and for choosing subsets of comparable sales, some criterion of "nearness" is required. Standardized Mahalanobis distances were proposed by Tchira (1979) and Isakson (1986) where Dij, the "distance" between two sales is calculated from

(3) $$Dij = (Xi - Xj)\,\Sigma^{-1}(Xi - Xj)'$$

---

[7] In the U.S.A., the total absolute value of adjustments must be less than 25% of the comparable's price and the net value of the adjustments not more than 15% in order to satisfy standards imposed by the secondary mortgage markets. These percentages define when a comparable is comparable enough.

where Xi and Xj are vectors of amenity characteristics for properties i and j, and $\sum^{-1}$ is a covariance matrix of these amenities. Tchira, points out, however, that these amenity characteristics are not necessarily translated into price effects. Obviously, if buyers do not change their valuations of properties, it does not matter if physical characteristics differ.

Vandell's solution is to estimate a regression model from submarket data, then use a minimum variance subject property price estimate to find the optimal set of comparables and their optimal weights. He uses the variances and co-variances of the adjustment factors estimated from the submarket regression equation to estimate the variance of the indicated values.

But if the wrong submarket is used in order to estimate the sample, or if the submarket properties are not all priced by the same hedonic process, then Vandell's method suffers from biased estimates and the answers are invalid. The exercise depends on the assumption of uniform hedonic structure throughout the submarket and unbiased choice of submarket in the first place.

But if these assumptions were strictly true a second criticism, pointed out to us by Peter Colwell, becomes operative—the Vandell procedure then becomes mathematically identical to a straight regression with results identical to simply estimating coefficients from the entire submarket sample and using these to calculate the predicted price of the subject. Vandell's finding that more comparable sales reduces standard errors simply means that precision of any regression estimate increases with sample size given a uniform data generating hedonic structure. Confusion in the notation in the 1991 Vandell paper leaves uncertainty about which criticism (ours from the preceding paragraph or Colwell's) is valid. An attempt to sort this out will set the stage for the reformulation presented in the following section of this paper. Let:

Voi Observed sales price of the ith comparable sale, which is one realization of

Vi A random variable representing all possible sales prices on the date of sale for the ith comparable property.

---

[8] Another way of saying this is that relationships are not linear. Empirical support for non-linear

Vm_i The expectation of this distribution, E(Vi), so that $Vi = Vm_i + e$ (a random error)

$\hat{V}si$ Adjusted "indicated" price for the subject property, obtained by adjusting the ith comparable sale. This is an estimate derived from a pricing model, not an observation.

$\hat{V}s$ The final value estimate will be a weighted average derived from q comparable sales, where q is a subset of n, the number of submarket sales.

Because we cannot observe Vm_i, we use Voi in the equation 2 model to obtain an estimate of the subject price:

(4) $$\hat{V}_{si} = V_{oi} + \sum^{m} \hat{a}_j (X_{sj} - X_{ij})$$

If, on the other hand, we used the regression model (like equation 1) to estimate $\hat{V}si$, then we would have

(5) $$\hat{V}_s = \sum^{m} \hat{b}_j (X_{sj})$$

But we suspect that this is not a very good model due to misspecification, etc. And, it was developed for the whole submarket and may not fit the subject property very well. Therefore, the $\hat{b}_j$ coefficients may be biased. In one equation Vandell uses the estimated value, $\hat{V}_{oi}$, rather than the observed value. Our equation 6 is equivalent to equation 2 in Vandell's paper (Ibid: 217).

(6) $$\hat{V}_{si} = \hat{V}_{oi} + \sum^{m} \hat{b}_j (X_{sj} - X_{ij})$$

The reason for preferring to use $\hat{V}_{oi}$ rather than V_{oi}, is presumably that the observed value may be an outlier, that is, not near the mean value, $V\mu_i$. But if the model of equation 5 is inadequate, which one must assume was the rationale for using the sales comparison method (equation 4) in the first place, then the use of the predicted comparable price in equation six also produces a poor prediction. Moreover, if we use the estimate rather than

---

relationships for at least some variables can be demonstrated in most samples of sales data.

the observed price, then the estimate of the subject property price becomes, substituting the right side of equation 5 for $\hat{V}_{oi}$ into equation 6,

(7) $$\hat{V}_{si} = \sum^{m} \hat{b}_j (X_{ij}) + \sum^{m} \hat{b}_j (X_{sj} - X_{ij}) = \sum^{m} \hat{b}_j (X_{sj}) = \hat{V}_s$$

Equation 7 confirms that if the estimated comparable price is used (as in Vandell's equation 2) rather than the observed comparable price and one assumes that the coefficients in both models are the same (i.e. follows Vandell's method of estimating sales comparison adjustment coefficients from the total submarket population of sales), the result is, as Colwell maintains, merely a regression model, not sales comparison. Moreover, all of the estimates of the subject's price produced by different comparables will be equal.

Vandell's ensuing discussion shows (we think) that he really means instead to use the following pricing model

(8) $$\hat{V}_{si} = V_{moi} + e_{oi} + \sum^{m} \hat{b}_j (X_{sj} - X_{ij})$$

Recall that $V_{oi} = V_{omi} + e_{oi}$ is a realization of the comparable sales price random variable V_i. The expectation of e_{oi} is zero, meaning that it will disappear if we estimate V_{si} (indicated subject property price) for a large sample of comparable sales and average the results.

Once Vandell has used this equation to estimate a sample of $\hat{V}_{si}$, he then proceeds to calculate variances and co-variances from the predictor variable data. These are then used to choose the comparable sales and weights for combining $\hat{V}_{si}$ estimates that produce a minimum variance price estimate for the subject property. He suggests a numerical examination of all possible subsets of comparable sales to find the optimum weights. But if the stated assumptions were strictly true, he is taking a long roundabout journey to end up with a conventional regression result. If the assumptions are not true, then the variances and covariances are not right and his results might be biased or inefficient.

Measurement errors and misspecification biases

In sales comparison we use the price of the comparable as a “black box” to avoid problems of efficiency and bias that arise in modeling the price of the subject directly from hedonic characteristics. Misspecification problems are so serious and data so limited and property and people so heterogeneous that the sales comparison approach is probably the best method available for most valuations. A black box observed price provides a more precise value estimate (after we adjust it to reflect differences from the subject) than an estimate from a misspecified model with unstable coefficients.

It may be helpful to list four distinct types of errors in sales comparison methods. The comparable sale is a random variable and does not except rarely take place at the mean value of the unobservable hypothetical distribution of possible prices, instead differing from the mean possible price by random error, e_{oi}, the deviation of the observed price of the ith comparable sale from the mean of the comparable sale's possible price distribution. Next, there is error in the estimates of the a_j coefficients that we might call e_{aj}. Then there are measurement errors in the hedonic characteristic differences between subject and comparable sale: $(X_{sm} - X_{im}) + e_{im}$. Each of the adjusted characteristics will have its own error, so there is a series of m measurement errors to worry about. Finally, there are omitted variables or other misspecification (e.g. wrong functional form, omitted interaction effects) resulting in an error, e_z. If we write the sales comparison equation with all these errors we have equation 9.

(9) $$\hat{V}_{si} = V_{oi} + e_{oi} + \sum^{m} (a_m + e_{am})((X_{sj} - X_{ij}) + e_{mj}) + e_{zi}$$

Only e_{oi} is likely to have an out of sample expectation of zero.[9] With heterogeneous sales, as the number of comparable sales, q, increases errors in the adjustment factors may get larger due to increased likelihood of qualitative differences whose effects are harder to estimate. Measurement errors of the m hedonic characteristics also are an increasing function of the number of comparable sales. As properties become less similar, it becomes harder to measure differences that matter to buyers if only because there are likely to be more of them. For example, in Vandell's example (see table 1) Sale 7 has the same in number of bedrooms and baths as the subject so errors in adjustment coefficients for these variables are irrelevant and only two variables (house area and lot size) must be measured. For sales 3 and 9, however, three factors must be adjusted, and for sale 10, all four variables differ between sale and subject.

Moreover, as the comparable and subject property, become more dissimilar, it is likely that omitted variable biases increase, that is, the "black box" pricing model that produced the observed sale price *Voi* becomes more unlike the process that will price the subject. In addition to the different characteristics of the houses themselves, less similar houses will attract buyers whose preferences and circumstances may differ. The comparable sale analogy for the subject property's hedonic structure becomes less valid. We would expect a $289,000 home to be in a better neighborhood than a $93,000 home, for example, while on the other hand sales with similar measured characteristics are more likely to also be similar in unmeasured variables such as neighborhood quality. As the set of comparable sales used gets larger, and more dissimilar, so do misspecification errors. If this were not true, we could simply use regression.

A preferable definition of "nearness" of comparable sales would be to rank order sales by the mean square error from all sources. Mean square error is defined as error variance+bias squared, where bias is the non-zero expectation of the error. MSE cannot be measured by calculated variances and co-variances because measurement error and omitted variables play important roles in determining total error.

With four types of errors in the equation, error behavior as a function of sample size becomes very complicated unless we assume independence of the various types of errors and simple error functions as q and m increase. We do not think the errors are independent and moreover, at least two of the errors (measurement and omitted variables) are not directly measurable by definition. Since a statistically complete treatment is not feasible, we propose to amalgamate errors in order to simplify analysis. We could write

---

[9] And since prices are generally skewed to the right, the random errors around the observed sale price are probably not normally distributed—which doesn't matter due to the central limit theorem, provided there are sufficient comparable sales to create a normally distributed sampling distribution of the mean.

error for any particular estimate of the subject property price obtained from a sales comparison model (equation 2) to heuristically capture the idea that there are two classes of error. Total errors includes a random component, e_{oi}, and an increasing, possibly biased component, u_i, which represents some amalgamation of measurement and omitted variables errors:

10) $$e_{si} = e_{oi} + u_i$$

The above errors are realizations of random variables with mean square error reflecting random variation and bias:[10]

11) $$MSE(V_{is}) = Var(e_{oi}) + Var(u_i) + E(u_i)^2$$

E(u_i) is the expectation of the non-random errors, that is, the bias resulting from these errors and *Vs* is the subject property price estimate. The expectation of the random error (sampling error) is zero, but the expectation of omitted variable and measurement error biases is not zero. Both variance and bias of these errors tend to increase with sample size.

Multiplying prices by weights when subject price estimates, *Vsi*, indicated by different comparable sales are combined to produce a weighted average final value conclusion (*Vs*) effects errors but does not eliminate the error tradeoff. Since measurement and omitted variables errors are by nature unknown, one cannot use calculated errors to determine proper weights. Therefore, Vandell's approach to determining weights is doubtful on theoretical grounds as well as in his empirical example.

The familiar relationship between the estimated population standard deviation $\hat{s}$ and the standard error of the sampling distribution of the mean, $\frac{\hat{s}}{\sqrt{q}}$ with sample size q, means that the random component of a price estimate standard error derived by averaging a sample of comparable sales decreases inversely to the square root of the number of comparables. If this standard error is 100 with sample size 1, it drops to 50 with sample size 4, to 33 with q=9 and so on. Mean square error from the other sources however, may increase or decrease with q because both bias and variance of these errors increases as the properties become more heterogeneous. MSE = f(q) where f(q) is an increasing function. If the ratio $\frac{f(q)}{\sqrt{q}} > 1$, then mean square error increases with sample size. Therefore, if we order q comparable sales from most similar to least similar to the subject property in terms of measured characteristics then, MSE(u_1) < MSE(u_2)<MSE(u_3) ….< MSE(u_q).

Optimal number of comparable sales depends on the relative size of the various kinds of error and how errors change as a function of increasing sample size q. Measurement errors and omitted variables vary between samples depending upon the characteristics of houses, buyers and data gathering procedures so no general conclusion is possible. It depends on the data.

Although we do not know the size of these errors, nor the function f(q) which will vary between samples, we can use summary statistics of prediction errors to get an idea of the optimum number and weighting of comparables. This is the same basic idea as the Vandell variance minimizing approach, in that it uses value estimate error variances to determine the optimum number of comparable sales. But the treatment differs in recognizing additional sources of error that occur in the real world and in acknowledging that calculated errors are probably misleading. Prediction errors tell us something about the sample distribution of errors from all sources.

Total error is the sum of errors whose variance increases with number of comparables and errors whose sampling distribution variance decreases with number of comparables. Representing a disorderly process of increasing errors by a function f(q) is clearly an oversimplification--the data are not obliged to follow a simple functional form. Nevertheless, error variances calculated in this way give a stylized representation of the essentials of how total error variance varies with the number of comparable sales under various circumstances.

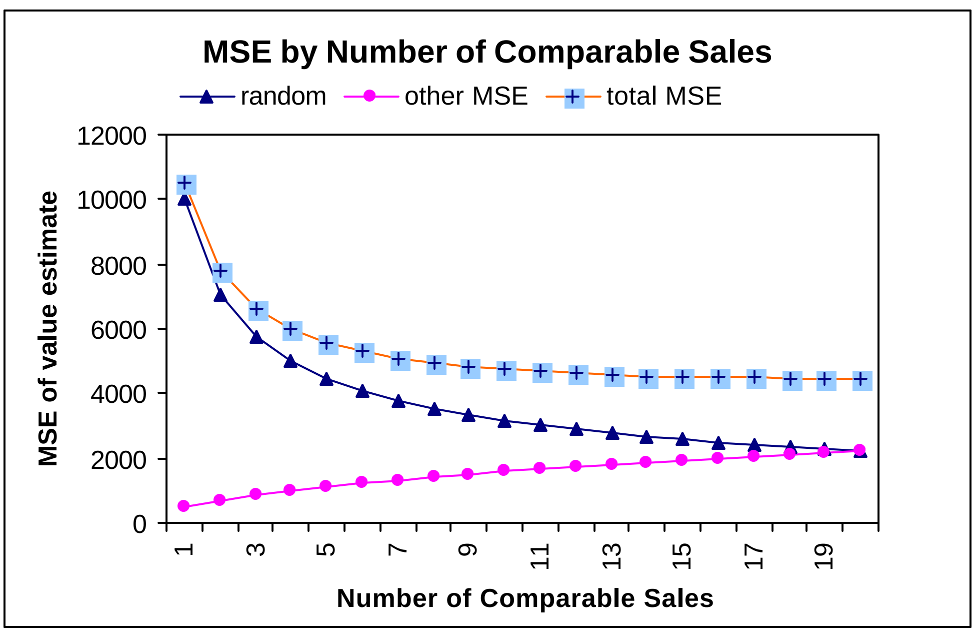
Error trade-offs

If random errors in observed prices are large while mean square adjustment errors are small, that is, $\text{Var}\, e_{oi} >> MSE(u_i)$, and f(q) does not increase $MSE(u_i)$ quickly, then

[10] For simplicity this formulation ignores the variance and covariance of these different types of errors.

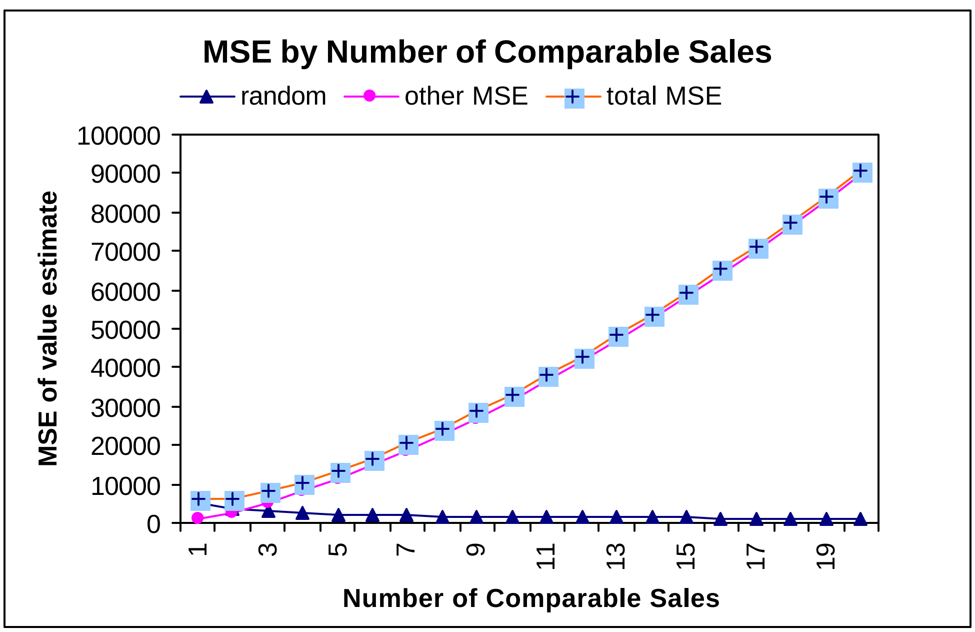
increasing the number of comparable sales will reduce total error variance. For example in figure 1, with $10,000 as the standard deviation of the observed price and a slow linear increase of $500*q as a simplified guess of MSE from other sources, then a pattern like figure 1 results. Under these stylized simple assumptions about the behavior of error variances, use of 20 comparable sales gives smallest total MSE. With fewer than 21 sales, the marginal decrease in random error is larger than the increase in MSE due to other sources of error. Both random variance and "other MSE" are divided by the square root of sample size, so the plotted MSE's represent sampling distribution statistics rather than population statistics, that is, they summarize error variances for the final subject property price estimate derived from a sample of q comparable sales.

Figure 1 Comparable sale price errors large relative to adjustment errors



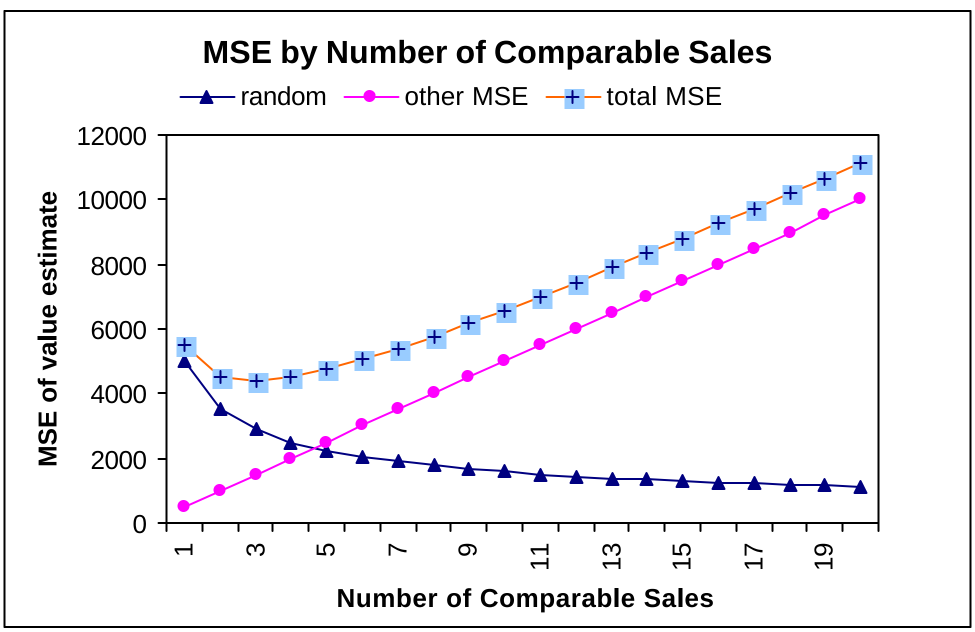
The figure 1 scenario might be called "uninformed buyers, well-informed appraiser" because the former's pricing errors are large, while the latter's price adjustment errors are small. Figure 2 depicts the opposite, "well-informed buyers, uninformed appraiser" in that adjustment errors are big, but observed sales are correctly priced so the possible price distribution has a small variance. In this situation, when the random house price error variance is small relative to adjustment variances and adjustment variances increase rapidly with q, then the optimal number of comparable sales is fewer. For example, in figure 2, standard deviation of observed prices is set at $5000 and adjustment related MSE at an exponentially increasing $1000*q². Under these assumptions, the optimal number of comparables is reduced to one, that is, adjustment error variance increases are larger than random error variance decreases starting with the addition of the second comparable. No doubt this is an unrealistically pessimistic view of the size of adjustment errors in most cases, as shown by the large absolute size of total errors shown on the y-axis. Usually appraisers are not this uncertain about adjustment amounts.

Figure 2 Comparable house price errors small compared to adjustment errors



As noted above, standard errors of the sampling distribution of the mean decrease rapidly with increasing sample size when sample size is small. Therefore, where scarcity of closely matched comparables makes the non-random MSE increase rapidly with q, the optimum number of comparable sales is probably usually in the range of 1-10. There is no single answer to the optimal sample size question since error trade-offs vary between subsets of sales. Figure 3 shows an example where minimum MSE is obtained by using 3 comparable sales. This "*buyers make modest mistakes, appraisers' mistakes start small but get bigger due to increasing comparable diversity*" graph was produced by assuming $5000 standard deviation for the observed sale price, and adjustment errors of $500*q^1.5. Since both error variances are divided by the square root of q, the adjustment error becomes linear.

Figure 3 Optimal number of comparables = 3



Uniform residential appraisal forms widely used in the United States use three comparable sales, implicitly accepting the above error trade-off pattern for typical residential appraisals. The forms, of course, are designed with other factors such as costs of information in mind, so three may not actually minimize MSE in a majority of cases. We believe, however, that this U shaped MSE function with respect to number of sales probably would hold for most sales data.

The reasoning presented leads to the conclusion that the optimal number of comparables varies depending on the relative sizes of various types of errors. If one has an excellent adjustment model, but little faith in the ability of buyers' to price property accurately (Vandell's assumptions), then use of more comparables leads to a more precise value estimate. All the observed sale prices are subject to big random errors due to buyers' and sellers' lack of precision in pricing. On the other hand, if one has little faith in the hedonic model, but more faith that each property has sold for the right price (with little random pricing error), then fewer comparable sales, perhaps only one comparable, will yield the best value estimate. In practice, the situation probably is usually somewhere in between—there are errors in observed prices, in data, and in model specifications—so the optimal number of comparables is generally "a few" with the best number depending upon specific circumstances.

Omitted from the preceding discussion of error variances is the important issue of bias in the adjustments. Expectations of adjustment errors are probably not zero and this bias, like error variance almost certainly increases with sample size. Therefore, the value estimate may tend to "drift" away from the true value as comparable sales are added to the sample, increasing bias. Because of omitted variable biases in a four variable hedonic model, this may very well explain why Vandell's empirical example subject price estimate moves downward from $256,000 to $233,000 as the number of comparable sales goes from one to ten.

Lusht and Pugh's empirical results

Lusht and Pugh (1981) examined the question of how widely one should cast the net in choosing comparable sales. From 325 properties sold in 1978, 150 were randomly selected to serve as subject properties (a withheld sample). Those 150 houses were each appraised three times using comparables drawn from first a "traditional neighborhood," second an expanded group of similar neighborhoods, and third the community at large. Mean absolute percentage prediction error results are shown in table 2.[11]

Table 2 Lusht and Pugh Appraisal % Error versus Number of Comparable Sales

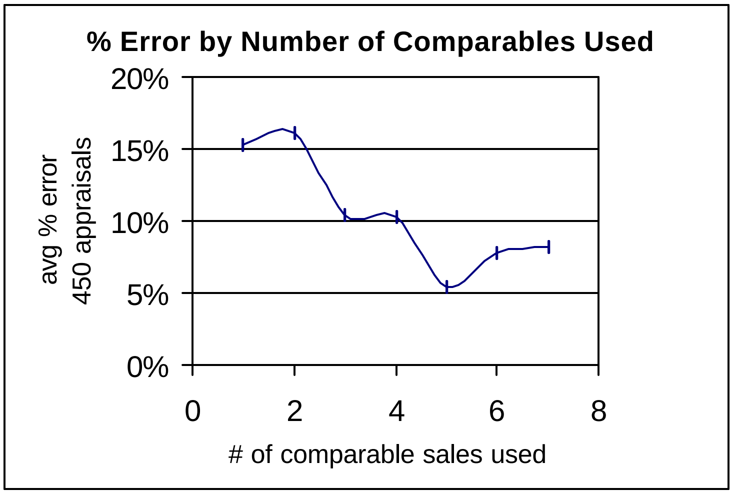
Number of Comparable Sales	One Neighborhood	Similar Neighborhoods	Entire Community
1	15.4	11	19.4
2	8.6	16.2	23.4
3	9.4	8.1	13.9
4	5.3	9.6	16.1
5	6.6	4.7	5
6	11	8	4.3
7	6.3	7.4	10.7

Source: Lusht and Pugh, 1981

As one would expect in real data, the patterns are not absolutely regular, but in all cases, the minimum average percentage prediction error occurs with fewer than the maximum number of comparables. In figure 4 we have summed and averaged the three columns of table 2.

Figure 4 Lusht and Pugh Percent Error in 450 Appraisals by Number of Comparables

[11] Our discussion above has been in terms of MSE (mean square error) while Lusht and Pugh reported MAPE (mean absolute percentage error). Both are summary statistics of error distributions, the difference between the two being that in calculating MSE, individual errors are squared, while the MAPE statistic takes absolute values of errors. Therefore, MSE gives more weight to larger errors. Which is preferable probably depends mostly on whether large errors have bigger consequences.



Source: Data from Lusht and Pugh, 1981

These valuations were performed by a method involving attribute matching for comparable selection, with cost as a basis for adjustment of physical differences, and use of the median comparable's “indicated value” as the final appraised value. With Lusht and Pugh's data set and valuation method, five comparables gave smaller prediction errors (5.4% v 8.1%) than 3 comparables. Optimum number of sales would vary depending on characteristics of the sample. The arguments presented in this paper suggest that while the details would vary, the pattern shown in the Lusht and Pugh results would probably be typical of sales comparison methods in general, regardless of the details of hedonic price model adjustments and comparable selection.

Conclusion

A long tradition of academic research seeks statistical methods to make real estate appraisal more objective, less open to bias, more precise, more replicable, and to increase productivity by use of information technology. There have been sceptical comments on regression methods along the way, notably Lessinger (1969, 1972) and Dilmore (1983), and in practice it has proven difficult to implement hedonic regression models that perform better than traditional appraisal methods. Lentz and Wang (1996) point out the large standard errors of estimate in most regression appraisals. Some processes may be too disorderly to model without large and unstable errors (See Makridakis and Wheelwright, 1989 and Gordon, 1991).

This paper concludes that we will never be able to estimate property prices precisely by hedonic regression methods, or any other method due to the nature of reality. It was not possible for alchemists to transmogrify lead into gold and there is little more reason to expect statistical models to represent complex and evolving markets with parsimonious models. Errors will often be large and ill-behaved. The problem is not in the statistics but in the states of nature.

That said, those making decisions in markets must muddle through somehow, making decisions under uncertainty. Use of statistics and assessment of sources and distributions of errors can lead, it is still sensible to hope, to less bias, more precision and higher productivity for appraisers. Perhaps a realistic hope is that regression methods can provide adequate “approximate” valuations in a majority of cases, but that for unusual properties or thinly traded markets, traditional case study methods will remain the only option.

Although we have disagreed with Vandell’s conclusions, the framework for analysis laid out by Ratcliff, Colwell, Cannaday, and Wu, Vandell and others wherein appraised value is seen as an estimate of the expectation of a random variable offers a theory suitable as a foundation for valuation methods. There is more clarity and precision in representing prices as distributions with outliers, means and variances than in traditional definitions of market value involving vague and confusing concepts like willing and informed buyers. Processes for deriving estimates of subject property prices from comparables are certainly more defensible and transparent if they use statistics rather than educated guesses or at least use statistics to inform educated guesses.

The discussion of errors here is oversimplified, no doubt, but the basic idea that optimal number of comparables involves an error trade-off is probably correct as indicated by our simulations, the Lusht and Pugh data and even, we think, Vandell's empirical example. The best number of comparables depends upon relative sizes of increasing mean square errors of adjustments versus random errors in observed prices. Therefore industry practice--use of 2-5 comparables--is closer to optimum than Vandell’s conclusion that more comparables always decrease total error. It is pleasing to find a theoretical rationale for the market outcome, because one normally expects competitive markets to find a reasonably efficient solution. The fact that competitive markets for appraisals since the 1930s have not found a favorable benefit/cost ratio in requiring more than 3 comparables in residential valuations is probably the strongest empirical support for our analysis of error variances.

This paper points to future research agendas for those interested in econometric approaches to real estate appraisal. First, more diagnostic checking to test for the kinds of errors that have often been assumed away in previous studies and particularly use of prediction errors to establish model validity. We expect calculated errors to not be robust for forecasting out of sample due to measurement and misspecification biases. More care should be taken with respect to data, specification searches and analysis of residuals to ensure that models proposed are adequate, particularly with respect to out of sample stability of the pricing process. Misspecified models will not predict prices well.

Ways of choosing samples and defining populations also merit more empirical research. There should be more effort to combine qualitative expert experience and judgment and qualitative data with quantitative data in the processes of defining submarkets and hedonic specifications. We all know that houses are bought because a buyer liked the kitchen or the overall design, so we need to find ways to get these less countable qualitative hedonic characteristics into quantitative models.

There should be more research on buyers—homogeneous buyer groups are the definition of submarkets. Market survey research techniques could generate new sources of data for hedonic models and insight into how property markets should be classified into homogeneous subsets for more precise price predictions.

There is a bifurcation in methods expressed in equations 4 and 5 between models that are “straight regression” versus those that use sales comparison by adjusting observed sale prices. So far, most academic research has been along the “straight regression” path. More work is needed along the lines of Colwell, et al., Graaskamp, Robbins and Dilmore who have tried to mimic the traditional adjustment grid methods in more replicable, objective fashion, using computer data bases to improve productivity.

Regression and sales comparison may converge as each improves. The “straight regression” avenue can move towards the sales comparison method by improving the richness of data, particularly incorporation of more spatial and neighborhood variables, by improving specifications through more non-linear and interaction terms and by using estimation techniques involving clustering, generalized additive models or spline regressions (Pace, 1998a, 1998b). Sales comparison methods can improve by relying more on statistics in choosing, adjusting and weighting comparables. Data base methods should allow appraisers to automate mass appraisals using sales comparison methods.

Another avenue for research would be how to deal with the risks inherent in imprecise and biased estimates of real estate asset values. Since we can expect to be wrong in valuations, how can we minimize the consequences of errors? Research is also needed on how to reform institutions to reduce principal/agent conflicts. Since we cannot count on technique to make bias impossible (statistical methods also require judgement and decisions), we need to rely on professional ethics and private virtue to keep imprecise estimates near the mean of the distribution of possible sale prices.

Clients may protest that valuations with wide and unknown confidence intervals are of little value, but the opposite is the case. Better understanding of the errors and uncertainty in pricing processes will lead to better understanding of risks, better real estate decisions and increased market efficiency.

References

Bruce, Richard W. and Darrell J. Sundell. 1977. Multiple Regression Analysis: History and Applications in the Appraisal Professon. *Real Estate Appraiser* Jan/Feb: 37-44.

Charemza, W.W. and D.F. Deadman. 1992. *New Directions in Econometric Practice*. Edward Elgar Publishers, London.

Colwell, Peter F., Roger E. Cannaday, and Chunchi Wu. 1983. The Analytical Foundations of Adjustment Grid Methods. *Journal of the American Real Estate and Urban Economics Association* V11(1):11-29.

Des Rosiers, Francois and Marius Theriault. 1996. Rental Amenities and the Stability of Hedonic Prices: A Comparative Analysis of Five Market Segments. *The Journal of Real Estate Research*. V12(1):17-36.

Ekeland, Ivar. 1988. *Mathematics of the Unexpected.* University of Chicago Press, Chicago.

Epley, Donald R. 1997. A Note on the Optimal Selection fnd Weighting of Comparable Properties. *Journal of Real Estate Research.* V14(2):175-181.

Gau, George W., Tsong-Yue Lai and Ko Wang 1992. Optimal Comparable Selection and Weighting in Real Property Valuation: An Extension. *Journal of the American Real Estate and Urban Economics Association* V20(1):107-123.

Gau, George W., Tsong-Yue Lai and Ko Wang 1994. A Further Discussionof Optimal Comparable Selection and Weighting, and a Response to Green. *Journal of the American Real Estate and Urban Economics Association* V22(4):655-663.

Gordon, Scott. 1991. *History and Philosophy of the Social Sciences*. Routledge, London.

Green, Richard K. 1994. Optimal Comparable Weighting and Selection: A Comment. *Journal of the American Real Estate and Urban Economics Association*, V22(4):647-654.

Isakson, Hans R. 1986.The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Methods. *Journal of the American Real Estate and Urban Economics Association*, V14(2):274-286.

Isakson, Hans R. 1998.The Review of Real Estate Appraisals Using Multiple Regression Analysis. *Journal of Real Estate Research.* V15(1/2):177-189.

Kang, Han-bin and Alan K. Reichert. 1991. An Empirical Analysis of Hedonic Regression and Grid-Adjustment Techniques in Real Estate Appraisal. *Journal of the American Real Estate and Urban Economics Association*, V19(1):70-91.

Kennedy, Peter. 1993. *A Guide to Econometrics, 3rd Ed.* MIT Press, Cambridge.

Leamer, Edward E. 1983. Let's Take the Con Out of Econometrics. *American Economic Review.* V7(1):31-43.

Lentz, George H. and Ko Wang 1998. Residential Appraisal and the Lending Process: A Survey of Issues. *Journal of Real Estate Research.* V15(1/2):11-39.

Lessinger, Jack. 1969. Econometrics and Appraisal. *Appraisal Journal*. October: 501-512.

Lessinger, Jack. 1972. A ‘Final’ Word on Multiple Regression and Appraisal. *Appraisal Journal*, July:449-458.

Lusht, Kenneth M. and Frederick Pugh 1981. A Research Note on the Effects of Changing the Search Area for Comparable Sales. *The Real Estate Appraiser and Analyst* Winter:34-36.

Lusht, Kenneth M. 1986. Real Estate Valuation and Appraisal (introduction to special edition). *Journal of the American Real Estate and Urban Economics Association*, V14(2):175-178.

Makridakis, Spyros and Steven C. Wheelwright. 1989. *Forecasting Method for Management*. John Wiley & Sons, New York.

Megbolugbe, Isaac F. and Linda Johnson. 1998. Guest Editors’ Introduction and Summary (to special issue on residential appraisal and lending). *Journal of Real Estate Research*. V15(1/2): 1-9.

Pace, R. Kelley. 1998. Appraisal Using Generalized Additive Models. *Journal of Real Estate Research*. V15(1/2):77-99.

Pace, R. Kelley. 1998. Total Grid Estimation. *Journal of Real Estate Research*. V15(1/2): 101-114.

Schnare, Ann B. and Raymond J. Struyk. 1976. Segmentation in Urban Housing Markets. *Journal of Urban Economics*, V3:146-166.

Tchira, Arnold. 1979. Comparable Sales Selection—A Computer Approach. *Appraisal Journal*, January: 86-98.

Vandell, Kerry D. 1990. Optimal Comparable Selection and Weighting. *Journal of the American Real Estate and Urban Economics Association* V 19(2):213-239, 1991.

Weirick, William N. and F. Jerry Ingram. Functional Form Choice in Applied Real Estate Analysis. *Appraisal Journal*, January: 57-73.

Whipple, R.T.M. 1995. *Property Valuation and Analysis*. Law Book Company, North Ryde, NSW.

Wolverton, Marvin L. 1998. Empirical Investigation into the Limitations of the Normative Paired Sales Adjustment Method. *Journal of Real Estate Research* V15(1/2): 191-203.