An Iterative Approach to Minimizing Valuation Errors Using an Automated Comparable Sales Model

BY ANDY KRAUSE AND MAX KUMMEROW, PH.D.

An earlier version of this article was presented at the American Real Estate Society (ARES) 25th Annual Meeting in Monterey, California, on April 3, 2009.

This paper describes a method for automating sales comparison valuations by choosing a small sample of comparable sales from a submarket of similar properties and *adjusting* their prices based on differences between sale and subject property characteristics. This logic is similar to that used in a traditional sales comparison adjustment grid approach using, for example, FNMA Form 1004. Traditional appraisal methods select, adjust, and reconcile a few comparable sales. This valuation algorithm follows the same steps and, in addition, computes summary price prediction error statistics useful for evaluating and improving the valuation protocol.

By calculating summary statistics on prediction errors, we can, in addition to automating the valuation process for samples of properties, automate aspects of the modeling improvement process. For example, we can implement what amounts to hill-climbing routines to improve submarket and comparable selection criteria as well as other aspects of the price-modeling process. The aim is to find valuation processes that minimize errors within a certain data set, submarket, or even specific properties.

We begin with a brief outline of how the literature and practice took us to this approach. Next we discuss the methodology of our iterative comparable sales-based model. Finally, we present examples of the application of the method by comparing it to a conventional hedonic regression model (a two-stage spatial error model segmented into submarkets) using data from Baltimore County, Maryland.

Literature Review

Although the widespread use of commercial automated valuation models (AVMs) is only 10 to 15 years old, the use

Andy Krause *consults for Greenfield Advisors, Seattle, Washington, on complex mass valuation assignments. Mr. Krause's primary line of work involves the development of mass appraisal models for environmental contamination litigation, with a focus on implementation of spatially cognizant models. He is pursuing a Ph.D. in urban design and planning at the University of Washington in Seattle. His research centers on land valuation modeling.*

Max Kummerow, Ph.D., *joined Greenfield Advisors in 2006 after two decades as a university faculty member. Dr. Kummerow studied real estate under James Graaskamp at the University of Wisconsin. He received his Ph.D. in property studies from Curtin University in Perth, Australia. His published research includes work on office market cycles, valuation theory, and mass appraisal methods.*

and publication of techniques for mass appraisal of real estate can be traced to the 1960s (Moore 2009). The Cole Layer Trumble (CLT) company began developing its version of computer-assisted mass appraisal (CAMA) methods about this time. Since then the volume of literature on regression modeling for appraisal and tax assessment purposes, often referred to as hedonic price modeling, has continued to expand unabated. And, the expansion in the academic literature has been paralleled by development of proprietary AVMs and further development of CAMA methods.

When CLT was acquired by Tyler Technologies, Inc. in 1999, the firm had 450 employees and had appraised some 50 million parcels in 46 states, greatly assisting tax assessors with property tax roll valuations. More recently, AVM methods have become more widespread for private sector valuations, even for individual properties. A good example is Zillow.com, a company that offers free online single-family and condominium residential value estimates for more than 93 million homes. Launched in 2006, Zillow has become one of the most visited real estate Web sites.

Early versions of mass appraisal software were sometimes criticized as inaccurate, and certainly they are all no better than the data used to drive their algorithms. Improvements in data and valuation methods have led to improvements in CAMA and AVM performance. Over time a significant portion of this progress in price modeling has been made through improvements in the actual model-fitting techniques, the recognition and modeling of spatial dependency, and the testing of models on more homogeneous samples (submarkets and/or data disaggregation) or, often, some combination of the three.

The use of sophisticated fitting techniques can account for more diverse functional forms or, in some cases, complete lack of functional relationships. Spline regressions, nonparametric regressions, and autoregressive techniques are some examples. Incorporating spatial information in pricing models through the use of direct spatial modeling with Cartesian coordinates (Fik, Ling, and Mulligan 2003), geostatistical models (Dubin 1998), or response surfaces (O'Connor 2008) has improved the precision of price estimates. Other studies (Goodman and Thibodeau 1998, 2003, 2007; Bourassa, Cantoni, and Hoesli 2007; Borst and McCluskey 2008) have focused on improving sample selection by delineating submarkets of homes in which the marginal price contributions of independent variables are more likely to be similar.

Predicted residuals from nearby sales (spatial errors) have been used in two separate but related ways in the literature. Case et al. (2004; in particular, see Case's model on page 178) developed a two-stage method in which errors from a single-stage ordinary least squares (OLS) model are used as predictors in the two-stage model; conversely, Pace and Gilley (1997), among others, used a simultaneous autoregressive (SAR) model to account for nearby residuals in a single-stage model. (See Bourassa, Hoesli, and Cantoni [2010] for a thorough literature review of recent uses of autoregressive models.) In sum, the spatial and the submarket techniques discussed both attempt to account for a phenomenon long recognized by appraisers and assessors—that home values are related to similar homes and/or nearby homes.

Connecting Regression to Traditional Appraisal

The heterogeneity of properties motivated traditional sales comparison approach appraisal methods to use very small samples of comparable sales to estimate values. Since the samples are as small as three in FNMA Form 1004, there is no way to estimate stable hedonic coefficients from the comparable sales hedonic characteristics. Appraisers, therefore, choose ad hoc adjustment factors based on various kinds of often

rather arbitrary evidence, sometimes summarized as *judgment and experience*. However, since the adjustments are small, assuming very similar comparable sales are available, even rather large percentage errors in the adjustments result in small errors in the value conclusion.

In an early insightful paper, Colwell, Cannaday, and Wu (1983) pointed out the mathematical similarity of hedonic regression and sales comparison adjustment grid pricing models. A regression model can be written as

(1)

$$P_s = BX + \varepsilon$$

where

P_s = the estimated price of a subject property

B = a $k \times 1$ vector of coefficients

X = an $n \times k$ matrix of property data (n properties and k property characteristics)

ε = a random error term, ideally normally distributed with mean zero.

Colwell, Cannaday, and Wu (1983) pointed out that a sales comparison adjustment grid is mathematically equivalent to an additive hedonic price model in the form:

(2)

$$P_s = P_o + B(X_s - X_o) + \varepsilon$$

where P_o = the observed price of a *comparable* property. Note that this observed price incorporates the effects of many hedonic variables, buyer and seller characteristics, market conditions, and so on that may not be explicitly included in the available data set. The fact that these are assumed to be similar for both P_s, the subject property price, and P_o, the comparable sale price, can make this model more accurate than a regression model (as pointed out by Colwell, Cannaday, and Wu [1983]). This equation can also be written as a price differences model:

(3)

$$P_s - P_o = B(X_s - X_o) + \varepsilon$$

As an example of this approach, consider a single comparable sale (sold two months prior) located in the same subdivision as the subject property that has identical attributes with the exception of 100 extra square feet of living space. Because of the proximity of these sales, the microspatial variations in price between them are nonexistent. If both are at essentially the same location, prices do not have to be adjusted for location differences. Also, because of the near identical structural features of close comparables, the only differences to be accounted for are market changes over the past two months and the premium paid for the extra living area. The right-hand side of this equation reduces to the sum of the adjustments for the market and for the additional square footage. Solving for the subject property value estimate (P_s) simply becomes a matter of adding the adjustments to the comparable sale prices—as in a traditional sales adjustment grid.

This sales comparison/price differences approach, therefore, can finesse omitted variables bias via the assumption that comparable properties do not differ much with respect to the values of omitted variables, and therefore the omitted variables contribute equally to the prices of subject and comparable properties, saving the appraiser from the problem of assessing their effects on prices. The list of characteristics that must be accounted for in equation 1 can be much larger than the list in equations 2 and 3, since by assumption, most factors are identical between the subject and comparable property and hence can be omitted from the price differences model. Both included and omitted variables are proxied by the price of the comparable. Only characteristics that vary between properties need be included in equations 2 and 3. By using table 1 as an example, rather than creating an equation that accounts for all nine variables, only the characteristics that differ—gross living area, age, and lot size—need to be considered.

Table 1. Grid example

Property	Gross Living Area	Age	Condition	Lot Size	Bathrooms	Location	Access	View	School District
Subject	1,800	18	Good	10,000	2.5	Average	Average	Regional	Jefferson
Comparable 1	1,790	20	Good	9,500	2.5	Average	Average	Regional	Jefferson
Comparable 2	1,820	20	Good	10,000	2.5	Average	Average	Regional	Jefferson
Comparable 3	1,890	18	Good	10,500	2.5	Average	Average	Regional	Jefferson
Comparable 4	1,680	24	Good	9,800	2.5	Average	Average	Regional	Jefferson

By reducing the dimensionality of the adjustments, the predictive variance can be minimized.

As pointed out by Kummerow and Galfalvy (2002), there is an error trade-off whereby the effect of the law of large numbers in reducing random variation, ε, is offset by increasing variance of prices in the submarket as well as increasing misspecification and measurement error biases as sample size increases. They conclude, therefore, that small samples can give the smallest price prediction errors. Results are data dependent because heterogeneity of properties and number of comparable sales vary, but in housing data the smallest errors regularly come with $n < 20$ and sometimes with n as small as one comparable sale. The n = 3–5 sample sizes common in traditional practitioner sales comparisons often perform best. This error trade-off provides a theoretical rationale for appraisers' use of small samples in sales comparison practice.

Early regression methods assumed, on the other hand, that larger samples are best. However, the papers cited demonstrate that the law of large numbers does not necessarily hold. Real estate price estimation is a law of medium numbers problem. Too small a sample results in unstable estimates, and too large a sample tends to result in less precise estimates, aggregation bias, and other biases and therefore poor out-of-sample prediction across submarkets in which responses and relevant variables may differ.

If the largest or smallest samples are not necessarily best, how are sample points chosen? Isakson (1986) recommended minimizing the hedonic distance between sample points when a subsample is selected for sales comparison. He used Mahalanobis distance as the metric of nearness, a measure that considers both differences and covariances in the hedonic variable values. However, Isakson did not weight these characteristics by the dollar value placed on them by consumers. If a characteristic does not matter to consumers, it does not matter whether it is similar or dissimilar between properties. A circular reference problem arises in that to calculate the best coefficients (dollar values per unit of hedonic characteristic), the best sample must be chosen, but simultaneously the sample is needed to calculate the coefficients.

Another obvious measure of distance is geographical distance between properties. Some AVM models include geocoordinates in databases and calculate distances between properties to choose closest comparables. But what if the house next door differs in its hedonic characteristics such as size, views, and the like? In light of this, a hybrid approach giving weight to both geographical and hedonic proximity would identify the best comparable sales. Pace and Gilley (1998) and Pace, Sirmans, and Slawson (2002) referred to statistical methods that mimic adjustment grid methods as grid estimators. Pace, Sirmans, and Slawson (2002) minimized hedonic distances with comparable choice algorithms that incorporate the coefficients to weight the hedonic characteristics.

Nonregression-based Mass Appraisal

The idea of automating the sales comparison approach has appeared sporadically in the real estate literature over time and has been used in some CAMA methods. Shenkel and Eidson (1971) presented a well-developed methodology for a comparable sales retrieval system. Dilmore (1974) created a simple mass appraisal model by "matching property attributes." His comparison shows this model to outperform a regression analysis and a proximity-based model for selected neighborhoods in Birmingham, Alabama. Lecture transcripts from the late 1980s reveal that Graaskamp and Robbins (1987) at the University of Wisconsin were developing an automated sales comparison system that they referred to as MarketComp based on choosing small samples of the most similar properties—an approach similar to sales comparison adjustment grids used in conventional individual property appraisals. In the late 1990s, Detweiler and Radigan (1996, 1999) published an original article and a follow-up describing their computer-assisted real estate appraisal system (CAREAS). Their work describes a statistically derived dissimilarity index used to select comparables and a regression model to create adjustment factors. Additionally, the authors give the CAREAS a limited expert system user control module.

More recently, Todora and Whiterell (2002) produced an automated sales comparison method for validating results from regression-based mass assessment models. This model selects comparable properties based on the Minkowski metric and then uses regression model coefficients to make standard, adjustment-grid-type adjustments to the comparables. The authors proposed an "override" system that allows modelers to use their knowledge to calibrate adjustments and comparable selections.

Borst and McCluskey (2007) presented the most complete explanation and testing of a comparable sales model to date. Their introductory discussion reviews and strengthens the relationship between multiple regression models—specifically the geographically weighted regression (GWR) models (Fotheringham, Brunsdon, and Charlton 2002)—and the sales comparison approach. The authors then developed a comparable sales method (CSM) and compared it to an OLS model, a segmented (submarketed) model, and a GWR model. Their results show that the CSM outperforms all other models in both prediction errors and spatial autocorrelation measures. A 2008 updated version of the Borst and McCluskey (2007) paper describes a modified comparable sales method (MCSM), which blends traditional regression estimates with the CSM. The model we ultimately developed is based on the academic work by Todora and Whiterell (2002) and by Borst and McCluskey (2007, 2008).

A Comparable Sales Model (CSM)

Our goal is to develop a mass appraisal model, based on the sales comparison approach, which is able to handle large sets of heterogeneous properties, respond to the modeler's knowledge of the data and the market, report error statistics, and, finally, offer an objective methodology to mine the data in order to facilitate continual model improvement over time. Toward this end, we developed an automated CSM, which functions as an expert system that allows the modeler to vary three parameters:

1. The size of the submarket, s
2. The number of comparables, c
3. The model's sensitivity to geographic distance, k.

Implicit in this method is a prior step whereby a price differences hedonic specification is chosen. This model specification too could be varied to seek better price prediction.

We developed a data-mining approach (similar to Borst and McCluskey's [2008] search for the optimal ratio weighting) that systematically searches for the

optimum values for each of the three parameters through a process of minimizing prediction errors. By focusing on improving error statistics, we can objectively guide the improvement process of the model while providing flexibility to apply this model to data sets of varying location, size, and time.

One insight of this approach is that it is the complete valuation protocol: data collection and quality, definition of submarkets, comparable selection, model specification, sample size, and choice of model coefficients that together produce a set of valuations. Error statistics are the ultimate test of how the interconnected aspects of the valuation protocol perform in a particular set of data. Results are data dependent—the same model specification does not work everywhere and optimum sample size and comparable selection criteria (and other aspects of the protocol) depend upon available data and data quality, the homogeneity of the data, and the actual hedonic characteristics as they are used by buyers in determining bid and offer prices.

Valuation Sequence

The general flow of the model can be described as follows:

1. Select submarket.
2. Determine implicit values of independent variables in the submarket.
3. Select best comparables.
4. Adjust comparables via a sales comparison grid.
5. Weight comparables and produce an estimated value.
6. Report error and repeat for next subject sale.

To account for market movements, all sales are time-adjusted to a single date based on a predetermined appreciation index specific to the geographic area in question.

The process begins by selecting the first transaction as the subject property to be valued. The remaining $n - 1$ sales are considered possible comparable sales. First, we cut this set of $n - 1$ potential comparables down to a submarket of more similar potential comparables. Similarity can be expressed many ways. We follow the logic of Isakson (1986) and consider it to be a measure of nearness in characteristic space; that is, the two sales that are nearest to each other in terms of size, age, quality, and so on are considered the most similar. Isakson used Mahalanobis distance, a scale-invariant measure that takes into account the correlations of the data set when calculating distance between two points. The formula for Mahalanobis distance is

(4)

$$Dm(x_{i,j}) = [(x_i - x_j)^t \times S^{-1} \times (x_i - x_j)]^{1/2}$$

where

$Dm(x_{i,j})$ = the Mahalanobis distance between subject i and comparable j

x_i = a vector of the subject property characteristics

x_j = a vector of the comparable j property characteristics

S^{-1} = the inverse covariance matrix of all the properties

t = a transposed matrix.

We, however, think that a traditional Mahalanobis calculation is not the best measure because physical distance does not offer a monotonic difference in price; that is, some areas far away are often more similar than areas closer, whereas it is hard to envision a scenario in which greater distance (difference) in square footage or quality would lead to more similar homes. Additionally, our initial testing found that including the *X, Y* coordinates in the Mahalanobis calculation produced undesirable results. To correct for this perceived fault, we create a dissimilarity value (DV) by adding a geographic distance penalty to the Mahalanobis distance; thus the larger the

sum of the Mahalanobis distance (without the X,Y) plus the distance metric, the more dissimilar the two properties. The DV can be expressed by expanding equation 4 to

(5)

$$DV(x_{i,j}) = [(x_i - x_j)^t \times S^{-1} \times (x_i - x_j)]^{1/2} + (\text{Dist}_{i,j} \div k)$$

where

$\text{Dist}_{i,j}$ = the distance between subject property i and comparable property j

k = distance metric (e.g., a k value of .25 would add 1 point to the DV for each quarter-mile distance between properties).

First, we calculate the DV between the subject property and all other properties in the data set. The list of potential comparable sales is then sorted by DV, and the s least dissimilar (or most similar) comparables are chosen as the *submarket.* Second, the sale prices of these s submarket sales are adjusted to the date of the subject sale based on a predetermined house price index for the area. Third, a simple semilog OLS regression is run to determine the implicit contribution (sales grid adjustments) of each of the variables in the data set. These adjustments are applied to the submarket sales identically to a traditional adjustment grid. Because the OLS model is in semilog format, the adjustments are in percentage form.

To determine which of the submarket sales to use as comparables, we total the gross adjustment percentage; add a distance penalty, which is calculated by dividing the distance between the submarket sale and the subject sale by the distance parameter, k; and further divide this by 100 to turn it into a percentage. This value is the comparability index. Note that the DV is based on a statistical determination of distance, whereas the comparability index is based on estimated implicit price differentials (the gross adjustment percentage) generated by the sale prices of the submarket sales plus an additional penalty for physical distance. The submarket sales are then sorted by this comparability index. The c number of most comparable sales are chosen as the comparables for the final adjustment grid. Finally, the c adjusted values indicated by the adjustment grid for the final comparables are weighted according to the formula:

(6)

$$(1 \div CI_j) \div [\Sigma_{j=1} (1 \div CI_j)]$$

where CI_j = the comparability index for each of the c comparables. The weights obtained by this formula are then applied to the time-adjusted price to create a contributory value from each comparable. The contributory values are then summed to create a value estimate for the subject property. This value estimate is compared to the actual sale price to produce the error of the estimate.

After reporting the error, the CSM then resets itself, reads in the values of the second subject sale, and repeats the processes until it has valued and calculated errors for all N of the sales. Obviously, this error calculation would not be possible when a property that has not yet been sold is being valued—that is, in an actual assessment or appraisal assignment—but it is reasonable to assume similar error distributions for valuations by the same method for similar properties using similar data. That is how all *revealed preference* or *inference from past sales* valuation methods work.

After all sales have been valued by using the CSM, the average error, the average absolute error, the number of predictions within 10 percent of the sale price, and the number of predictions more than 30 percent from the sale price for all the sales in the data set are calculated. No holdout sample is necessary since the model functions as a cross-validation model—it values each property using all other properties except the subject.

Parameter-changing Iterations

If only one set of valuation parameters—the submarket size (s), number of comparables (c), and distance sensitivity (k)—is being tested, the CSM is finished at this point. However, to test a set of differing parameters, the model loops back to the beginning and starts the process again, completing these procedures with different parameters selected by the user.

Because there are error statistics for a sample of valuations generated by looping through a data set, valuation performance can be compared as protocol parameters are changed. To avoid arbitrary modeling protocol decisions, all decisions are subjected to the test of comparison to market sale prices. Error summary statistics indicate whether the valuation protocol is acceptable. More precise and robust estimates provide evidence for the superiority of better protocols.

The remainder of this paper outlines an empirical example of valuation in Baltimore County, Maryland, using our CSM. The results of our CSM testing are compared to those of a standard hedonic model, using statistically defined submarkets and a spatial error correction method, to determine the relative performance of the CSM.

Data

The data used in this study cover the north Baltimore suburban/exurban area, specifically, the region north of Interstate 695, east of Reisterstown Road (Highway 30), and west of Bel Air Road (Highway 1). Figure 1 shows the extent of the study area with points representing the locations of the sales in the data set. The final, cleaned transactions database contains 5,787 sales of single-family detached homes. The data-cleaning process included elimination of suspected non-arm's-length sales and atypical outlier properties. The mean house in the data set is slightly more than 32 years old, 2,300 square feet in size, with 2.6 bathrooms on 0.8 acres of land. This *mean* house, which sold for $476,764, has a quality grade of more than 4.32 (table 2). As shown by the 10th and 90th percentile figures, this area contains a fairly heterogeneous housing stock, with prices and home sizes ranging from $113,000 and 720 square feet on the low end to $2,400,000 and 9,450 square feet at the upper bound.

Figure 1 illustrates a surface of the average housing price generated through a Kriging method. A point for each of the 5,787 sales overlays this price surface to show the spatial pattern of the sales. This map shows that high-value homes are concentrated in the exurban areas and low-value properties are primarily located in the established cities of Reisterstown (west), Lutherville–Timonium (south-central), and Perry Hall (east).

Results

The goal of the iterative parameter search process is to find the values that produce model results with the lowest prediction errors. We attempt to create this set of parameters in two ways: (1) *a priori*, based on knowledge of the area and some initial real estate theory, and (2) a *stepwise* testing mechanism, using a range of each parameter to find the lowest error for each (see table 3).

Table 2. Summary statistics

	Price	Grade	Age	Bathrooms	Square Footage	Acreage
Mean	$476,794	4.32	32.31	2.60	2,325	0.80
Median	$395,000	4.00	35.00	2.50	2,030	0.34
Minimum	$113,000	2	0	1	720	0.05
Maximum	$2,400,000	9	157	9	9,450	9.70
Standard deviation	$278,110	1	22	1	1,212	1.03
10th percentile	$230,000	3	3	2	1,120	0.17
90th percentile	$825,000	6	53	4	4,018	1.89

Figure 1. Study area price surface and sale locations

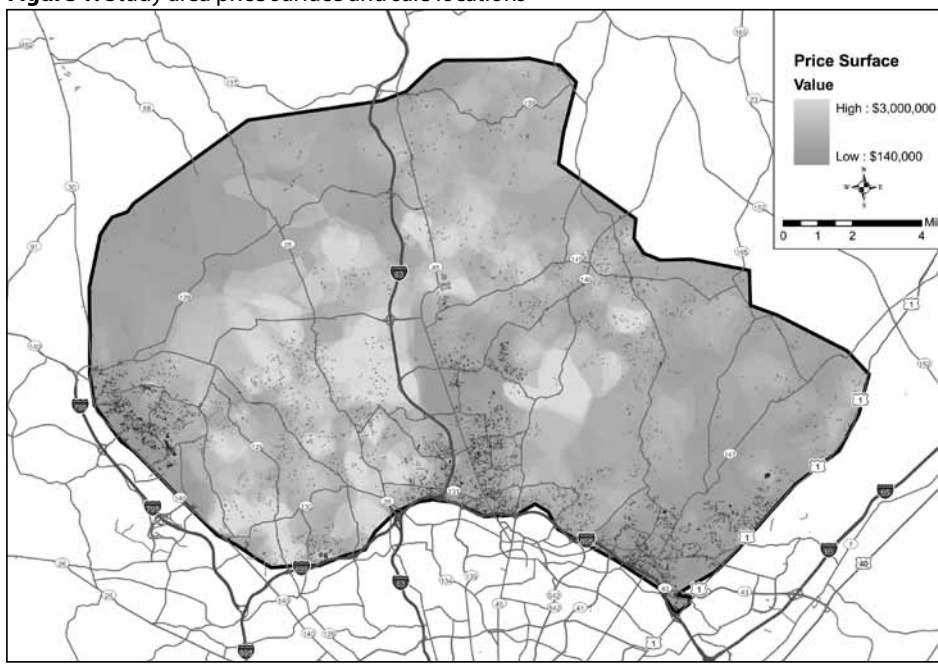

Table 3. Parameters for a priori and stepwise best models

Parameter	Model	
	A Priori	Stepwise Best
Submarket size	500	140
Number of comparables	3	9
Distance parameter	400 meters	150 meters

This research concentrates on lowest average absolute error. An average error closest to 0, lowest maximum error, the highest percentage of sales within 10 percent, the lowest percentage of errors greater than 30 percent, or other error statistics could also be considered candidates for ranking valuation accuracy.

CSM Results

First, a CSM protocol is tested by using a set of parameters based on knowledge of the area and results from preliminary testing during development of the model. The a priori model results in an average prediction error of 10.30 percent.

Then each parameter is tested sequentially (while the others are held constant) through a range of values to determine the *best* value based on lowest average errors. Submarket size variation, then number of comparables, and finally the distance sensitivity value are tested. After a parameter has been tested, its best value is selected and the next parameter is tested. This data-mining, stepwise, procedure produces a set of parameters that gives a mean absolute error of 9.72 percent, which is 5.6 percent (0.58 percentage points) lower than the a priori set of parameters (see table 4).

Table 4. Results for a priori and stepwise best models

Parameter	Model	
	A Priori	Stepwise Best
Mean error	−1.18%	−1.48%
Mean absolute error	10.30%	9.72%
Standard error	13.61%	12.79%
Percentage with absolute error less than 10%	59.62%	61.24%
Percentage with absolute error greater than 30%	3.73%	2.73%

Econometric Modeling Test

House price prediction is highly data dependent and comparing results from different time periods and/or geographic areas does not represent a just comparison. Thus the CSM is compared to other methods by using the identical sales data to estimate three hedonic pricing models:

1. A global model
2. A submarket segmented model
3. A submarket segmented model with a two-stage spatial error correction.

First, by using the 5,787 sales from the data set, quadratic variables for square footage, age, and acreage, as well as dummy variables for each grade quality, are created (see table 5). The primary form of development in the area is subdivisions containing very similar housing product. Subdivisions of very-high-quality homes are often located directly adjacent to older, less expensive subdivisions or even older stand-alone homes. To account for this development pattern, a predictor, SubMean, a variable that represents the average assessed value of all the properties in a subdivision, is added.

The sales data range over four years, from the beginning of 2004 to the end of 2007. Prices in the Baltimore area rose considerably over the first 2 years of this period, reached a plateau for about 18 months, and then began to decline in the latter part of 2007. To account for the changing market, the sale prices of all sales are adjusted by the price index developed in the CSM discussion. Following the published literature, a semilog OLS regression model is estimated, with the dependent variable as the natural logarithm of sale price.

Table 5. List of econometric variables

SF	Square footage of the dwelling
Age	Age of dwelling
Grade	Quality grade of dwelling
Bathrooms	Number of full bathrooms
GSF	Square footage of garage
Pool	Dummy for presence of pool
Acreage	Acreage
X	*X* coordinate of centroid of property
Y	*Y* coordinate of centroid of property
SubMean	Mean of assessed value in subdivision
Gr3–Gr 9	Dummy for quality grade 3 to grade 9
SE1–SE10	Nearest *n*-th spatial error (1 to 10)

The CSM values each sale with all other sales in the database, similar to a cross-validation method used to test *predictive* ability in regression modeling. Thus a cross-validation test is used to gauge the predictive ability of the regression models here and compare them to the CSM results presented previously. One benefit of cross-validation is that no holdout sample is needed; rather, each sale is valued by a model composed of all the other sales minus itself.

The global model specification produced an adjusted R^2 of .9101. All variable signs are as expected, including the steadily increasing values associated with the assessor's grade rankings. The positive Y coefficient and the negative Y^2 coefficient indicate that, all other things held equal, it is good to be away from the city to a point, but after that it becomes a disamenity. The same effect on the X coefficient shows convenient access to Interstate 83 adds a premium to house prices. This is affirmed by the location of the highest values homes in figure 1. The global model produces an average absolute error of 11.33 percent, with 55 percent of the predictions within 10 percent of the actual price and 4.9 percent with errors of greater than 30 percent.

Submarket Delineation

It has been well-established in the literature that division of property into submarkets can have a significantly positive effect on prediction accuracy. We grouped the sales into submarkets with a clustering technique known as K-means. The sales are clustered based on the standardized values of three variables: square footage, age, and quality grade. The K-means algorithm requires a preset

number of clusters to be determined by the user. Visits to the area revealed four major product types that often correlate with location, and as a result, we use a predetermined grouping of four clusters for the K-means algorithm.

The K-means algorithm produces different clusters depending on the initial random starting points. Therefore, we ran 100 iterations of the K-means procedure and selected the clustering breakdown that occurred most often. We then estimated four separate equations, one for each of the submarkets. Realizing that spatial differences still likely exist within a geographic area this large, we ran a third set of models utilizing spatial lags of the errors from the segmented second model as predictors.

In each of the submarket models—both nonspatial and spatial error—the set of significant predictors for each specific submarket differed slightly between the models. For instance, in submarket 1, the variable for very high assessor's grade homes, the *Y* variable, and its quadratic showed no significance and were removed from the model. Also, a number of the submarkets had no homes of certain quality grades, rendering those variables nonapplicable.

Comparing the nonspatial to the spatial error model for each submarket shows an improvement in explanatory ability, an increase in R^2, due to the use of spatial error predictors. Cross-validating the predictions in each submarket creates an average absolute error of 10.76 percent, an improvement over the 11.79 percent figure from the global model. Using the residuals as predictors from the (two-stage spatial error) submarket model (the total of the four equations) produces an average absolute error of 9.96 percent, 7.4 percent (0.8 percentage points) better than the segmented model without the spatial error correction.

Comparison of Results

Table 6 shows several error statistics for each method: the a priori and stepwise CSMs and the three set econometric models (the global, the submarket segmented, and the submarket segmented with spatial error correction). Also, in addition to the error statistics, we test for spatial autocorrelation between errors in each of the models using the Moran's I statistic. A Moran's I Z-score of greater than 1.96 indicates spatial autocorrelation at the 95 percent significance level. All three econometric models exhibit spatial autocorrelation in their errors term, though a decreasing amount as model complexity increases. The CSMs, on the other hand, show no statistically significant spatial autocorrelation in the error terms. These results are shown in the last two rows of table 6.

The stepwise CSM outperforms the best of those particular econometric models—the two spatial error correction models—in terms of lowest overall prediction error and spatial autocorrelation of the error terms. The differences in the spatial configuration of the errors favor

Table 6. Comparison of results

	Econometric Models			Comparable-based Models	
	Global	Segmented			
Average Error	Nonspatial	Nonspatial	Spatial Error	A Priori	Stepwise
Average absolute error	11.33%	10.76%	9.96%	10.30%	9.72%
Average error	1.07%	−0.96%	−0.84%	−1.18%	−1.48%
Percentage with absolute error less than 10%	4.91%	4.13%	3.53%	3.73%	2.73%
Percentage with absolute error greater than 30%	54.64%	56.45%	60.26%	59.62%	61.24%
Standard deviation of error	14.91%	14.16%	13.26%	13.61%	12.79%
Moran's I (Z-score)	49.1978	36.4254	9.4538	0.5735	0.9119

more heavily both CSMs over the set of econometric models. In sum, the stepwise CSM outperforms the econometric models in prediction error statistics and significantly outperforms them in terms of the spatial autocorrelation of the error terms. However, note that these comparisons are based on particular realizations of the two approaches. It is likely that both could be improved with further data exploration, model specifications, or other refinements of the techniques used. Moreover, results are data dependent. Our theoretical understanding suggests that the performance ranking of the models could vary depending on characteristics of the data—especially the degree of homogeneity among the properties.

Discussion

Statistical AVM and hedonic regression approaches to real estate valuation have developed separately from the traditional three approaches to value appraisal methods, mostly in academic work, public sector tax assessor evaluation methods, and dot.com information technology, and as a means of valuation cost control by lenders, courts, or other users of mass appraisals. This paper integrates statistical thinking about appraisal methods with the traditional sales comparison approach. Sales comparison can be seen as a small sample method in which the key is to improve the efficiency (and accuracy and precision) of estimates by choosing and inferring subject property valuations from a very homogeneous sample, while evaluating and modifying other aspects of the valuation protocol to optimize performance (i.e., to improve error distribution statistics).

We have used an AVM protocol that explores data via prediction error distributions to improve sample selection and other modeling parameters. Choice of submarket sample size, number of comparable sales, and distance sensitivity are data dependent and vary across samples (and likely within as well). Thus, further development of valuation process protocols (Kummerow 2006) offers a means of producing better valuations.

This paper offers two innovations, both tools for improving protocols:

- A simple algorithm to automate the sales comparison approach, increasing both objectivity and speed of sales comparison valuations
- Use of prediction error statistics and iterative modification of the process, providing a quick means for testing and improving the pricing protocol.

With the Maryland data, an automated CSM competes with and possibly outperforms a regression model using submarkets and a spatial error correction technique. Future research in this area could head in a number of directions to further explore various aspects of valuation protocols such as the following:

- Improvement of algorithms to select the best parameters
- Better comparable selection and weighting mechanisms
- Individual sale-based *best* modeling parameters for each property (similar to GWR's geographically changing coefficients).

Many AVM and CAMA valuation methods have found that while mass appraisal methods predict prices well for a majority of properties, some unusual or outlier properties prices are not well predicted by these methods. We think our approach is well adapted to exploring various ways to identify and deal with these outlier properties and doing so is an avenue to further improve mass appraisal methods. In other words, the trade-offs between hit rate (percentage of properties valued by the model) versus accuracy can be explored by varying treatment of outliers.

There are surprisingly few papers on appraisal errors, considering the power

of error analysis to assist in improving valuation protocols. The tools of mass appraisal methods make it convenient to generate large samples of valuations and hence prediction errors, which in turn makes it possible to explore ideas for improving valuation protocols.

References

Borst, R., and W. McCluskey. 2007. Comparative evaluation of the comparable sales method with geostatistical valuation models. *Pacific Rim Property Research Journal* 13 (1): 106–129.

Borst, R., and W. McCluskey. 2008. The modified comparable sales method as the basis for a property tax valuations system and its relationship and comparison to spatially autoregressive valuation models. In *Mass appraisal methods: An international perspective for property valuers,* ed. T. Kauko and M. d'Amato, 49–69. Chicester, UK: Wiley Blackwell.

Bourassa, S., E. Cantoni, and M. Hoesli. 2007. Spatial dependence, housing submarkets, and house price predictions. *Journal of Real Estate Finance and Economics* 35:143–160.

Bourassa, S., E. Cantoni, and M. Hoesli. 2010. Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research* 32:139–159.

Case, B., J. Clapp, R. Dubin, and M. Rodriguez. 2004. Modeling spatial and temporal house price patterns: A comparison of four models. *Journal of Real Estate Finance and Economics* 29:167–191.

Colwell, P., R. Cannaday, and C. Wu. 1983. The analytical foundations of adjustment grid methods. *Journal of the American Real Estate and Urban Economics Association* 11 (1): 11–29.

Detweiler, J., and R. Radigan. 1996. Computer-assisted real estate appraisal. *Appraisal Journal* January:91–101.

Detweiler, J., and R. Radigan. 1999. Computer-assisted real estate appraisal: A tool for the practicing appraiser. *Appraisal Journal* July:280–286.

Dilmore, G. 1974. Appraising houses. *Real Estate Appraiser* July–August:21–32.

Dubin, R. 1998. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics* 70 (3): 466–474.

Fik, T., D. Ling, and G. Mulligan. 2003. Modeling spatial variation in housing prices: A variable interaction approach. *Real Estate Economics* 31 (4): 623–646.

Fotheringham, A, C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships.* Chichester, UK: Wiley.

Goodman, A., and T. Thibodeau. 1998. Housing market segmentation. *Journal of Housing Economics* 7:121–143.

Goodman, A., and T. Thibodeau. 2003. Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics* 12 (3): 181–201.

Goodman, A., and T. Thibodeau. 2007. The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics* 35:209–232.

Graaskamp, J., and M. Robbins. 1987. Business 868 lecture notes. University of Wisconsin–Madison.

Isakson, H. 1986. The nearest neighbors appraisal technique: An alternative to the adjustment grid methods. *Journal of the American Real Estate and Urban Economics Association* 14:274–286.

Kummerow, M. 2006. Protocols for valuations. *Appraisal Journal* Fall:358–366.

Kummerow, M., and H. Galfalvy. 2002. Error trade-offs in regression appraisal methods. In *Real estate valuation theory,* ed. K. Wang and M. Wolverton. Boston: Kluwer.

Moore, J. 2009. A history of appraisal theory and practice: Looking back from IAAO's 75th year. *Journal of Property Tax Assessment and Administration* 6 (3): 23–49.

O'Connor, P. 2008. Automated valuation models by model-building practitioners: Testing hybrid model structure and GIS location adjustments. *Journal of Property Tax Assessment and Administration* 5 (2): 5–24.

Pace, R., and O. Gilley. 1997. Using the spatial configuration of data to improve estimation. *Journal of Real Estate Finance and Economics* 14:333–340.

Pace, R., and O. Gilley. 1998. Generalizing the OLS and grid estimators. *Real Estate Economics* 1:331–346.

Pace, R., C. Sirmans, and V. Slawson. 2002. Are appraisers statisticians? In *Real estate valuation,* ed. K. Wang and M. Wolverton. Boston: Kluwer.

Shenkel, W., and A. Eidson. 1971. Comparable sales retrieval systems. *Appraisal Journal* October:540–544.

Todora, J., and D. Whiterell. 2002. Automating the sales comparison approach. *Assessment Journal* January–February:25–33.