**13th Pacific-Rim Real Estate Society Conference**
**Fremantle, Western Australia, 21 to 24 January 2007**

# PRICE DIFFERENCES MODELS

Max Kummerow
*Greenfield Advisors*
*2601 4th Ave Suite 650*
*Seattle, Washington, USA*
*Greenfieldadvisors.com*
max@greenfieldadvisors.com

**Abstract:**
Appraisal practice often relies on small samples—three comparable sales, for example—from which to infer likely selling price of a subject property. There is good reason for this small sample approach, namely the heterogeneity of properties, market participants, locations and sub-markets that tends to increase misspecification, measurement errors and population variance as sample size increases. Therefore, the law of large numbers may not hold and estimates may become less efficient as sample size increases. The logic of all three "approaches" to value involves inferring price differences from a small sample of similar transactions using a hedonic model in the form Price = Price of comparable sale +/- price differences due to differing property characteristics. This paper offers an example of a small sample model, estimated from Seattle housing sales data analogous to a time series model in differences. Results are data dependent and will be the same as results of a regression with data in price and property characteristic levels only if the functional relationships are linear with constant slope throughout the range of the predictor variables. The data indicate that this is not the case, arguing for small samples and spatial models.

Keywords: Valuation, appraisal, hedonic models, sales comparison, appraisal errors, AVM models.

# Introduction

This paper compares empirical estimates of "price differences model" coefficients versus regression "model in levels" coefficients. Colwell, Cannaday & Wu (CC&W 1983) pointed out the mathematical equivalence of sales comparison "adjustment grids" to an additive regression model. Pace (1988b) refers to "grid estimators" where prices are estimated from a sample of similar properties.

The widely used Fannie Mae Uniform Residential Appraisal Report (URAR Form 1004, Table 1) compares the subject property to three comparable sales. Transaction prices of the sales are "adjusted" to reflect differences of hedonic characteristics between the sold property and the subject property. These adjustments, because they account for differences between the sold and subject properties, could be called a "price differences model."

**Table 1 Uniform Residential Appraisal Report**

| Freddie Mac Form 70 | Page 2 | | Fannie Mae Form 1004 March 2005 |
|---|---|---|---|
| FEATURE | SUBJECT COMPARABLE SALE # 1 | COMPARABLE SALE # 2 | COMPARABLE SALE # 3 |
| Address | | | |
| Proximity to Subject | | | |
| Sale Price $ $ $ $ | | | |
| Sale Price/Gross Liv. Area | $ sq. ft. | $ sq. ft. | $ sq. ft. |
| VALUE ADJUSTMENTS | | | |
| | DESCRIPTION +(-) $ Adjust | DESCRIPTION +(-) $ Adjustt | DESCRIPTION +(-) $ Adjust |
| Sale or Financing Concessions | | | |
| Date of Sale/Time | | | |
| Location | | | |
| Leasehold/Fee Simple | | | |
| Site | | | |
| View | | | |
| Design (Style) | | | |
| Quality of Construction | | | |
| Actual Age | | | |
| Condition | | | |
| Above Grade Total Bdrms. Baths Total Bdrms. Baths Total Bdrms. Baths Total Bdrms. Baths | | | |
| Room Count | | | |
| Gross Living Area | | | |
| Basement & Finished | | | |
| Rooms Below Grade | | | |
| Functional Utility | | | |
| Heating/Cooling | | | |
| Energy Efficient Items | | | |
| Garage/Carport | | | |
| Porch/Patio/Deck | | | |
| Net Adjustment (Total) | + - $ | + - $ | + - $ |
| Adjusted Sale Price | | | |
| of Comparables Net Adj. % | | | |
| Gross Adj. % | | | |
| Net Adj. % | | | |
| Gross Adj. % | | | |
| Indicated value of subject property | $ | $ | $ |

Even where valuers do not explicitly adjust prices of a sold property for specific hedonic characteristics, but instead take an overall of "gestalt" approach to inferring overall value differences, they must have in mind (perhaps subconsciously) some basis for their estimate of price differences. Property characteristics, neighborhood characteristics, buyer/seller characteristics, location variables and circumstances of sale may all affect prices. This paper extends the price differences modeling concept by exploring sample selection and "price differences model" specification and estimation issues. The paper presents an empirical example comparing price differences model coefficients estimated from a sample of 505 Seattle suburban house sales to coefficients estimated from a regression model. If the relationships between hedonic coefficients and price are linear throughout the price range, then the coefficients should be roughly equal. If relationships between hedonic characteristics and price are non-linear, then coefficients should differ.

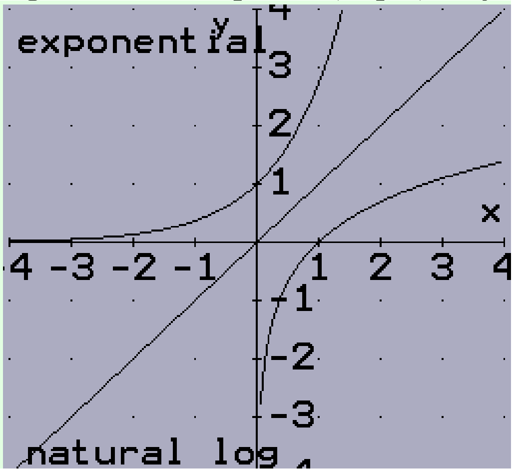
Appraisal textbooks often recommend "paired sales" as a way of estimating the price impact of particular property features, ignoring the possible random variation of prices that could lead to large errors in estimating the value of hedonic characteristics from small samples. Kummerow (2002) points out that possible price of a property is a random variable, with observed price a draw or event from a possible price distribution.

By the way, all of these efforts to infer property prices from property characteristics stem from Kelvin Lancaster's (1966) consumer demand theory stating that prices paid for complex goods (such as real estate) can be interpreted as a sum of prices paid for various property characteristics. So to understand why houses sell for more or less, we talk about value enhancing (or detracting) property hedonic characteristics.

The CC&W paper recommends estimating coefficients for price adjustments from a larger sample of sales, since statistically stable coefficient estimates cannot be made from small samples. However, there are at least two problems with using regression coefficients in a small sample price differences model. The first is possible non-linear responses, the second that responses may vary across time and space—there may be spatial or serial autocorrelation of model residuals.

Thinking of this in time series jargon, if relationships between price and hedonic variables are linear, then a model in levels, $Ps=\Sigma b1X$, will have the same coefficients as a model in "differences" $Ps-Po=\Sigma b2(Xs-Xo)$, where Ps is the price of a subject property, Po is the price of a comparable property, X are hedonic characteristics and b1 and b2 are hedonic coefficients. However, if the hedonic functions are not linear, then the coefficients in a price differences model would differ from regression coefficients, because they would reflect responses along a small section of the hedonic function that might have a different slope than a linear approximation of the price/hedonic characteristic relationship across the whole range of the hedonic characteristic. Figure 2 shows changing slopes for log and exponential functions versus a linear relationship.

**Figure 2 Price responses (slopes) varying at different values of non-linear functions**




Linear response the same across the range of X and Y. Nonlinear response varies with X.

http://www.mathnstuff.com/

All three traditional "approaches to value" used by appraisers infer price differences between sold and subject properties from a small sample of transactions. The "direct sales comparison approach" estimates from transactions data the market value of differences in characteristics between properties using a hedonic model in the form Price of subject = Price of comparable sale + Price differences due to differing property characteristics. The "income approach" uses transactions data to identify cap rates or discount rates and property cash flows from sale prices and rent and expenses transactions. Cap rate = Cap rate observed in a sale +/- differences due to differing property characteristics. The "cost approach" uses transactions data to identify building costs, land costs and depreciation allowances. Cost=Land value (from transactions) + Cost to construct (from transactions) – depreciation (from transactions). So all three approaches infer prices from a sample of transactions data, it is only the types of transactions and data items considered that differ.

This is basic Marshallian neoclassical economics: Transactions result from the intersection of supply and demand and reveal prices. Absent transactions, prices can be inferred from other information—stated preferences (also called "contingent valuations" or "willingness to pay" or "willingness to accept")—but not as reliably. An issue not considered here is whether market prices (transactions data) represent efficient prices, that is, prices based on adequate information, self-interest and other conditions requisite to market efficiency. That is an important issue since if markets are efficient, that implies that future prices are a random walk (since all information is already incorporated in prices), so forecasters cannot "pick winners." Luckily, this quasi-religious and difficult to resolve dispute between "Chicago School" and Institutional Economists is not the topic of this paper.

Automated Valuation Methods (AVMs or "mass valuations") comprise a "fourth approach" family of methods that are explicitly statistical in inferring prices, often from large samples of transactions data. Calling this another approach to value is useful because the various approaches are regarded as check methods for each other. Particular approaches may be more or less significant depending upon the purpose of the appraisal and available data. In a world of larger databases and easy to use statistics packages, valuations may in the future end up including a statistical inference exercise very much like an AVM. Banks using AVMs as a check method or means of detecting fraud or errors have in effect already added this fourth approach check on valuations.

Many lenders already use AVM valuations as their primary valuation in cases with low perceived risk such as low loan to value ratio loans, equity loans or refinancing of seasoned loans. The traditional “three approaches” might be called “case study” methods for use where the appraiser has too little data for statistical reliability and instead attempts to infer price from anecdotal data.

Appraisal practice prefers to rely on small samples—often, as mentioned, three comparable sales in U.S. practice—from which to infer likely selling price of a subject property. There is good reason for this approach, namely the heterogeneity of properties, locations and sub-markets that tends to increase misspecification, measurement errors and population variance as sample size increases. Therefore, the law of large numbers may not hold and estimates may become less efficient as sample size increases.

Nevertheless, most academic hedonic price models, and probably most AVM methods, have used regression to estimate coefficients based on large data sets. Fotheringham, et al. (2002) provide an empirical demonstration that this produces an “average” coefficient value across a sample space that does not necessarily represent responses at particular sample points (properties) within the sample. Pace, et al. 2002, provide another empirical study arguing that small sample or “grid estimator” methods as they call them, provide better price estimates than regression. They remark that “statistically challenged” valuers regularly produce estimates of value with smaller standard errors than the regression models published by PhDs at universities. Sirmans, et al.’s meta-analysis examined “square footage, lot size, age, bedrooms, bathrooms, garage, swimming pool, fireplace and air conditioning.” in published hedonic models, demonstrating considerable variation in the estimates of these variables effects on prices in different contexts. (Sirmans, et al. 2005, p. 2) (1999) and others have demonstrated the existence of distinct sub-markets with varying coefficient values between submarkets.

Kummerow and Galfalvy (2002) present another empirical demonstration of the smaller errors obtainable through inferences from small samples, noting that since the market has concluded that small samples “work best” it is good to look for a theoretical explanation. They argue that random errors decrease with sample size, but misspecification and measurement errors increase, resulting in optimum sample sizes that are quite small—ranging between 1 and 20 in the Perth data they analyzed. They explain this as an “error trade-off” with random errors of sampling distributions decreasing with sample size, but other errors increasing. Instead of the law of large numbers, a “law of medium numbers” applies, with a data dependent variable optimum (small) sample size providing the most accurate price inferences.

Another conclusion of Kummerow and Galfalvy, 2002, is that all results are data dependent. They point out that the appropriate test of valuation model results is prediction errors since mispecified models can lead to misleading and biased standard error statistics and coefficients. Although some structural relationships usually hold (people often pay more for larger properties), here are no general theoretical models or functional forms to "explain" what people pay for houses. Hundreds of variables have been found significant and empirical studies and meta-analyses indicate clearly that responses to the same variables differ in different contexts and circumstances. Pricing processes are complex, variable and evolving over time. As Fotheringham, et al. point out, if the objective is to produce price estimates for particular houses, or estimates of the response of prices to a particular variable in a given location, there is little value in calculating average responses from a large sample. Valuers are usually not interested in what another square meter of floor area is worth on average; rather we want to estimate the value of a particular property.

## Cross Sectional Models in Differences

In time series econometrics the possibility of spurious models when processes are integrated, that is, when variables are non-stationary, has often been addressed by taking differences until variables are stationary.[1] For example, the model in levels, say level of consumption, for example, explained by levels of wealth and incomes might be replaced by a model in differences, that is change in consumption explained by change in wealth and change in incomes.

Model in levels: $C = f^*(Y,W)$

Model in differences: $\Delta C = f(\Delta Y, \Delta W)$

Where $\Delta C = Ct - C(t-1)$

One of the symptoms of problems due to non-stationarity is likely to be serial auto-correlation of errors.

In a cross sectional model, there may also be problems when modeling in levels, but often this might best be explained in a different way, although I think the issues are identical. In a cross sectional model, there may be omitted variables. So if we model a process as (stylized representation):

$Y = b1X1+b2X2\ldots biXi + C + e$

The data generating process is likely to actually be more complex, say,

Equation 1) $Y = b1X1+b2X2\ldots biXi + C+ d1Z1+d2Z2+\ldots dkZk+ C + e$

In modeling house prices, non-stationarity of the data generating process may be reflected by different variables and different responses to variables in different parts of

[1] Integrated processes contain a "unit root"

the sample. The commonly used "sales comparison approach" finesses that problem by inferring prices from a few sales of similar properties. The model is really a model of price differences:

Equation 2) $Ps = Po + \Sigma bi(Xs\text{-}Xi)+e$

Here the price of a subject property, Ps, is inferred from a transaction price of a similar property, Po, "adjusted" for hedonic characteristics differences where Xs-Xi is the difference between two similar properties in amount of the characteristic and bi are pricing weights or reveal preference market responses (i.e. price differences) due to these differences in characteristics. Example, if b for square feet of floor area is $100, then an example, for two houses with 2500 and 3000 square feet of floor area would be:

Ps= 300,000 +100*(2500-3000) = $250,000

The estimate of probable selling price for a house with 2500 sq. ft. assuming an otherwise similar but larger (3000 sq. ft.) house has sold for $300,000 is found by evaluating the difference between the two properties. In this example, the 500 sq. ft. difference multiplied by $100/sq.ft. equals a $50,000 price difference.

I've not seen a model actually estimated this way, possibly a reflection of my ignorance of spatial modeling literature. One problem is that while in a time series model the proper differencing technique is obvious (for example, "subtract last quarter's result from this quarter's result) due to the regular unidirectional time increments at which data is collected. The appropriate lags to difference are straightforward.

In cross sectional data, which data points to difference may not be so obvious. Translating that sentence into "3 approaches to value" jargon, the choice of comparable sales may not be unambiguous. The "most similar" sales might be geographically closest, but heterogeneous properties can exist side by side.

And, similarity in terms of price depends on how consumers value hedonic characteristics. So properties differing on characteristics not important to consumers might still sell for similar prices. To choose the best comparables we need to consult the price differences model to find out which are most similar from consumer's pricing models point of view, but that gives a circularity problem in that we need the sample of comparables to estimate the pricing model.

Spatial autocorrelation measures may be helpful in choosing comparables—in the sense of "how far apart" they should be. Note that when sales comparison approach language says "comparable sale" a time series modeler would simply say "lag of the dependent variable." What is the appropriate spatial lag for differencing house sale data points in a cross sectional analysis? Is it differences in geographical space or hedonic space we need to minimize? The reason models in differences are not common in cross sectional data as in time series data analysis may be the messy nature of these decisions. There is no absolutely right or wrong way to difference cross sectional data, although there are better and worse price differences models and sample selection protocols.

One benefit of the sales comparison/price differences approach is that many omitted variables may be less different in the near neighbor lags than throughout the sample. So that will mean the price differences model will have fewer non-included confounding factors to increase standard errors. Say, for example, that in the pricing process represented by equation 1, all of the variables had identical values except b1, b3 and Z4. This is entirely possible: They are in the same neighborhood, same size, same builder, same location, etc.

So the differencing much simplifies the pricing model we need to estimate:

Ps-Po = price difference = b1(X1s-X1o)+b3(X3s-X3o)+c4(Z4s-Z4o)

Because we do not include data on Z4, the actual model becomes simply:

Ps-Po = b1(X1s-X1o)+b3(X3s-X3o)

Fewer omitted variables are ignored in the latter case than in equation 1. The result should be better out of sample price predictions.

A final point before proceeding to some data analysis—the coefficients in a cross sectional price differences model would usually differ from the coefficients in a model in levels. Only if all responses are linear would the coefficients be equal. But we do not expect linear effects, but rather diminishing returns. Most investigators prefer semi-log specifications that tend to make relationships more nearly linear. Nevertheless, the dependent variable, price differences, is not the same as the price levels dependent variable, so with non-linear relationships, the coefficients will differ between regression and price differences models, even with the same included variables.

## Empirical estimates

A sample of October/November 2006 sales was obtained through Metroscan, a data vendor. The sales were from six eastern Seattle metropolitan area suburbs. After some data cleaning and trimming to eliminate sales with missing data and outliers, 505 sales remained. The data included geographical coordinates, square feet of finished floor area, year built and other property characteristics.

### Regression statistics

Several regression models were estimated from the sample, for example, one with adjusted R^2 of .65 and standard error of the estimate $138,700, produced the following coefficient estimates. The omitted suburb is Bothell. Note that the model included dummies for high property grade and a property condition dummy, as well as distance from the central business district of Seattle.

**Table 2 Hedonic price regression of the full six suburb sample**

| Model | | Unstandardized Coefficients | | t | Sig. |
|---|---|---|---|---|---|
| | | B | Std. Error | | |
| 1 | (Constant) | 162657 | 57966 | 2.8 | .005 |
| | finishSF | 200 | 10 | 18.7 | .000 |
| | distcbd | -6407 | 2344 | -2.7 | .007 |
| | conddum | 58009 | 18033 | 3.2 | .001 |
| | hseage | -518 | 508.261 | -1.0 | .308 |
| | higradum | 51183 | 17005 | 3.0 | .003 |
| | viewdum | 167134 | 24182 | 6.9 | .000 |
| | Issaqdum | 47209 | 31929 | 1.4 | .140 |
| | Kirkdum | 92171 | 28726 | 3.2 | .001 |
| | Redmdum | 79626 | 28156 | 2.8 | .005 |
| | Sammdum | 64411 | 31211 | 2.0 | .040 |
| | Wooddum | 105814 | 32994 | 3.2 | .001 |

a Dependent Variable: PRICE

This is not a satisfactory model according to normal probability and residual plots (it tends to overestimate low value homes and underestimate high value homes), but it will serve to illustrate the issues addressed in this paper.[2]

Fotheringham, et al.'s insistence that overall statistics mask variation in responses within samples, was confirmed by estimating a somewhat abbreviated (to preserve degrees of freedom) model separately for each of the six suburbs. R^2 varied from .57 to .78 and standard errors of estimate from 73,000 to 172,000 (and that in two adjacent suburbs).

Measurement errors, misspecification or under-specification in this model (or, if you interpret another way—missing qualitative variables and variation in responses or pricing variables across a heterogeneous sample space) is demonstrated by variation in coefficients (including "wrong signs" and sign changes in a few cases) when each suburb was estimated separately:

**Table 3 Variation of coefficients in suburb sub-samples**

| suburb | Model | | Unstandardized Coefficients | | t | Sig. |
|---|---|---|---|---|---|---|
| | | | B | Std. Error | | |
| Bothell | 1 | (Constant) | -89077 | 139822 | -.637 | .530 |
| | | FINISHSF | 209 | 24 | 8.531 | .000 |
| | | distcbd | 3989 | 7515 | .531 | .600 |
| | | Conddum | -23443 | 40007 | -.586 | .563 |

[2] Ok, the real story is I ran out of time to figure out a better model. Consider this a draft version.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Hseage | 1889 | 982 | 1.924 | .066 |
| | | higradum | 39882 | 36229 | 1.101 | .281 |
| | | VIEWDUM | 31220 | 78071 | .400 | .693 |
| Issaqua | 1 | (Constant) | 453824 | 95204 | 4.767 | .000 |
| | | FINISHSF | 143 | 19 | 7.348 | .000 |
| | | distcbd | -8077 | 3456 | -2.337 | .022 |
| | | conddum | 26727 | 38936 | .686 | .494 |
| | | hseage | -1566 | 800 | -1.956 | .054 |
| | | higradum | 2449 | 31101 | .079 | .937 |
| | | VIEWDUM | 109977 | 34664 | 3.173 | .002 |
| Kirklan | 1 | (Constant) | 784316 | 166330 | 4.715 | .000 |
| | | FINISHSF | 221 | 28 | 7.734 | .000 |
| | | distcbd | -42816 | 9133 | -4.688 | .000 |
| | | conddum | 119885 | 33179 | 3.613 | .000 |
| | | hseage | -1662 | 941 | -1.766 | .080 |
| | | higradum | 14807 | 38782 | .382 | .703 |
| | | VIEWDUM | 218347 | 45674 | 4.780 | .000 |
| Redmond | 1 | (Constant) | -23082 | 86501 | -.267 | .790 |
| | | FINISHSF | 220 | 15 | 14.096 | .000 |
| | | distcbd | 3178 | 3222 | .986 | .326 |
| | | conddum | 45425 | 29107 | 1.561 | .121 |
| | | hseage | 860 | 1046 | .822 | .413 |
| | | higradum | 57596 | 26427 | 2.179 | .031 |
| | | VIEWDUM | -132317 | 85087 | -1.555 | .123 |
| Sammami | 1 | (Constant) | 317572 | 205158 | 1.548 | .126 |
| | | FINISHSF | 177 | 27 | 6.467 | .000 |
| | | distcbd | -8151 | 7536 | -1.082 | .283 |
| | | conddum | 61924 | 73275 | .845 | .401 |
| | | hseage | -968 | 1621 | -.597 | .552 |
| | | higradum | 71134 | 50354 | 1.413 | .162 |
| | | VIEWDUM | 191581 | 59460 | 3.222 | .002 |
| Woodinv | 1 | (Constant) | 112603 | 264605 | .426 | .672 |
| | | FINISHSF | 219 | 41 | 5.233 | .000 |
| | | distcbd | -746 | 10201 | -.073 | .942 |
| | | conddum | -30494 | 66923 | -.456 | .651 |
| | | hseage | -75 | 2413 | -.031 | .975 |
| | | higradum | 33033 | 72831 | .454 | .652 |
| | | VIEWDUM | 17000 | 97915 | .174 | .863 |

a Dependent Variable: PRICE

This is not quite Fotheringham, et al.'s "geographically weighted regression" but does have the same flavor of complexity in changing market responses across space. The finished square feet coefficient remains relatively stable (although varying from $143 to $221), but other coefficients are all over the place.

I interpret this as due to omitted variables and measurement errors, even in this relatively rich data set. The residual plots looked better in these smaller more uniform data sets, but there were more likely to be influential data points in these smaller samples, another partial explanation for unstable coefficients across samples.

The objective of the paper was to compare these regression coefficients with estimates of coefficients from a price differences model. An Excel macro was written to select the three nearest (in distance) sales to each property and difference prices and hedonic characteristic variable values between each of the three pairs of subject property and nearest sales. From the 505 members of the sample, this generated 1515 data points consisting of differences between prices (Ps-Po) and hedonic characteristics (Xs-Xo) for each property and its three nearest neighboring sales.

This new artificially constructed data set was used to estimate price differences model coefficients. This model was as poorly specified as the model in levels. R^2 was .64 and standard error of estimate $155,000. Coefficients were reasonably similar to coefficients of the "model in levels," reflecting averaging across the sample space pointed out by Fotheringham, et al.

**Table 4 Coefficients from price differences model, n=1515**

| Model | | Unstandardized Coefficients | | T | Sig. |
|---|---|---|---|---|---|
| | | B | Std. Error | | |
| 1 | (Constant) | 6835 | 6362 | 1.074 | .283 |
| | distcbd | -11804 | 8764 | -1.347 | .178 |
| | hseage | -797 | 252 | -3.163 | .002 |
| | distcomp | -10091 | 9439 | -1.069 | .285 |
| | FINISHSF | 207 | 4 | 45 | .000 |

a Dependent Variable: PRICEdifference

These estimates based on the entire sample mask variation in responses—and that the properties may not be comparable enough to give good sales comparison methods results.

Restricting the sales to those where the difference in house size was less than plus or minus 300 sq. feet, resulted in very low adjusted R^2, but smaller standard errors of estimate (.03 and $120,000 respectively). And the coefficients changed dramatically when estimated from this sample of similar sized houses. Now that the houses are of similar size, finished square feet is less important in explaining house price variation.

**Table 5 Coefficients with sample restricted to houses less than 300 sq. ft. different in area**

| Model | | Unstandardized Coefficients | | t | Sig. |
|---|---|---|---|---|---|
| | | B | Std. Error | | |
| 1 | (Constant) | 2143 | 5721 | .37 | .708 |
| | distcbd | -17304 | 14129 | -1.22 | .221 |
| | hseage | 1461 | 369 | 3.95 | .000 |

| | | B | Std. Error | t | Sig. |
|---|---|---|---|---|---|
| | FINISHSF | 89 | 36 | 2.45 | .014 |

a Dependent Variable: PRICEdifference

Further restricting the sample to houses within 200 sq. feet of identical size reduced the sample to 336 sales. Standard errors of estimate fell to $116,000.

**Table 6 Coefficients with sample restricted to houses less than 200 sq. ft. different in area**

| Model | | Unstandardized Coefficients | | t | Sig. |
|---|---|---|---|---|---|
| | | B | Std. Error | | |
| 1 | (Constant) | -1532 | 6441 | -.238 | .812 |
| | FINISHSF | -4 | 56 | -.074 | .941 |
| | distcbd | -15262 | 15639 | -.976 | .330 |
| | hseage | 1598 | 443 | 3.605 | .000 |

a Dependent Variable: PRICEdifference

The model has nearly "disappeared, in that the only significant variable left to explain price differences in this sample of similarly sized houses is house age differences and with older houses worth more in this sample.

When the sample was divided into subsets, coefficients changed dramatically in predictable ways, with less range in X resulting in smaller price explanatory power.

**Table 7 Dividing sample into homes similar in size versus homes different in size**

| finishsfcat | | Unstandardized Coefficients | | t | Sig. |
|---|---|---|---|---|---|
| | | B | Std. Error | | |
| differsize | (Constant) | -50713 | 12078 | | .000 |
| | FINISHSF | 246 | 11 | .676 | .000 |
| | distcbd | 2203 | 13564 | .005 | .871 |
| | hseage | -913 | 408 | -.068 | .026 |
| samesize | (Constant) | 13897 | 6605 | | .036 |
| | FINISHSF | 214 | 7 | .685 | .000 |
| | distcbd | -17313 | 11608 | -.036 | .136 |
| | hseage | -172 | 306 | -.014 | .573 |

a Dependent Variable: PRICEdifference

## Discussion

The overall model was suspect in this example, so the conclusions may be as well. Perhaps the strongest lesson is the need to incorporate spatial methods into hedonic models and to test various levels of aggregation in testing submarkets. By estimating the price differences coefficients across the whole data set, the author inadvertently recreated the large sample "average values" outcome, even though the data was price differences data. But it is clear that better price estimates can be obtained by selecting more uniform, smaller samples, where data are matched on important included variables. This matching helps proxy for unobserved, non-included values that will also tend to be similar for houses of similar size, age and distance from the CBD for example. The omitted variables thereby accounted for include location characteristics and qualitative variables that will tend to be similar in houses of similar size, age and location.

The method of using Excel macros to generate data sets and prediction error statistics makes generation of these price differences data sets quite convenient. It may be that valuers will soon be able to add a "fourth approach" to value by simply gluing a sample of sales data into a spreadsheet template and pressing a "run" button.

## References


Cho, Man, House Price Dynamics: A survey of Theoretical and Empirical Issues, *Journal of Housing Research*, 1996, 7:2, 145-172.

Colwell, P. F., R. E. Cannaday, and C. Wu, The Analytical Foundations of Adjustment Grid Methods, *Journal of the American Real Estate and Urban Economics Association,* 1983, 11:1, 11-29.

Des Rosiers, F. and M. Theriault, Rental Amenities and the Stability of Hedonic Prices: A Comparative Analysis of Five Market Segments, *The Journal of Real Estate Research*. 1996, 12:1, 17-36.

Dilmore, G., Appraising with Regression Analysis, A Pop Quiz, Appraisal Journal, October 1997, 403-404.

Dubin, R., R. K. Pace and T. G. Thibodeau, Spatial Auto regression Techniques for Real Estate Data, *Journal of Real Estate Literature*, 1999, 7, 79-95.

Ekeland, Ivar James J. Heckman, Lars Nesheim "Identifying Hedonic Models," *American Economic Review*, Vol. 92, No. 2, Papers and Proceedings of the One Hundred Fourteenth Annual Meeting of the American Economic Association (May, 2002), pp. 304-309

Epley, D. R., A Note on the Optimal Selection and Weighting of Comparable Properties, *Journal of Real Estate Research*. 1997, 14:2, 175-181.

Fotheringham, Stewart, Martin Brunsdon and Chris Charlton. *Geographically Weighted Regression: The Study of Spatially Varying Relationships* New York, John Wiley & Son, 2002.

Gloudemans, Robert. *Mass Appraisal of Real Property*. Chicago, International Association of Appraising Officers (IAAO), 1999.

Halvorsen, R. and H. Pollakowsky. "Choice of Functional Form for Hedonic Price Equations." *Journal of Urban Economics* 10(1), 37-49.

Isakson, H. R., The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Methods, *Journal of the American Real Estate and Urban Economics Association*, 1986, 14:2, 274-286.

Kang, H.B. and A. K. Reichert, An Empirical Analysis of Hedonic Regression and Grid-Adjustment Techniques in Real Estate Appraisal, *Journal of the American Real Estate and Urban Economics Association*, 1991, 19:1, 70-91

Keuzenkamp, H. , M. McAleer, and A. Zellner (eds.), Simplicity, Inference, and Econometric Modelling, Cambridge, Cambridge University Press, 2002, pp. 13-32

Kummerow, Max. "A Statistical Definition of Value" *Appraisal Journal*, 2002, V70#3, pp. 407-416

Kummerow, Max. "Protocols for Valuations" *Appraisal Journal*, 2006, V74 # 4, 358-366.

Kummerow, Max and Hanga Galfalvy. "Error Trade-Offs in Regression Appraisal Methods" in *Real Estate Valuation Theory*, Ko Wang and Marvin Wolverton, eds. Boston, Kluwer Academic Press, 2002, pp.105-132

Lancaster, K., 1966, A New Approach to Consumer Theory, Journal of Political Economy 74, 132-157.

Laurice, Jennifer and Radha Bhattacharya. Prediction Performance of a Hedonic Pricing Model for Housing. *The Appraisal Journal***.** Chicago: Spring 2005. Vol. 73, Iss. 2; p. 198 (12 pages)

Lentz, G. H. and K. Wang, Residential Appraisal and the Lending Process: A Survey of Issues, *Journal of Real Estate Research*, 1998, 15:1/2, 11-39.

Lessinger, J., A 'Final' Word on Multiple Regression and Appraisal, *Appraisal Journal*, July 1972, 449-458.

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

Malpezzi, Stephen. "Hedonic Pricing Models: A Selective and Applied Review, Center for Urban Land Economics Research, 2002. http://www.bus.wisc.edu/realestate/pdf/pdf/Hedonic %20 Pricing %20Models%20Survey%20for%20Maclennan.pdf (Retrieved 12/8/2006)

Pace, R. K., "Total Grid Estimation," *Journal of Real Estate Research*, 1998b, 15:1/2, 101-114.

Pace, R. K. and Gilley, O. W., "Using the Spatial Configuration of Data to Improve Estimation", *Journal of Real Estate Finance and Economics*, 1997,14, 333-340.

Pace, Sirmans and Slawson, 2002

Rogers, Warren. "Errors in hedonic modeling regressions: Compound indicator variables and omitted variables." *The Appraisal Journal***.** Chicago: Apr 2000. Vol. 68, Iss. 2; p. 208

Schiller, Robert J. and Allan Weiss, "Evaluating Real Estate Valuation Systems" proceedings of the Business and Economic Statistics Section of the American Statistical Association, 1996.

Sirmans, G. Stacy, L. McDonald, D. MacPherson and E. Zietz. "The Value of Housing Characteristics: A Meta Analysis."Journal of Real Estate Finance and Economics, Vol. 33, no. 3, (November 2006): 215-40.

Tashman, L. J., et al., Effect of Regressor Forecast Error on the Variance of Regression Forecasts, *Journal of Forecasting*, 2000,19, 587-600.

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

Watkins, C., Property Valuation and the Structure of Urban Housing Markets, *Journal of Property Investment and Finance*, 1999, 17:2, 157-175.

Weirick, W. N. and F. J. Ingram, Functional Form Choice in Applied Real Estate Analysis, *Appraisal Journal*, January 1990, 57-73.