Price Prediction for Residential Properties consistent with Sales Comparison Approach

Lin, Tzu-Chin
Associate Professor
Department of Real Estate and Built Environment
National Taipei University
No.67.Sec.3,Ming-Sheng E. Rd
104, Taipei, Taiwan
Tel: Int.-886-2-25009159
Fax: Int.-886-2-25074266
E-mail: tclin@mail.ntpu.edu.tw

This draft paper is circulated for comments only, not for citation.

Abstract

Sales comparison approach is frequently used in property appraisal. It, however, involves judgments on several matters and is criticised as subjective and heavily dependent upon experience. Regression method provides a more objective way of estimating values of property attributes. Nevertheless, comparable selections and weights allocation remain absent. The nearest neighbor method explicitly incorporates grid-adjustment process with the regression method, and thus retains the essence of sales comparison approach. We apply the nearest neighbor method to predicting prices of residential properties in Taipei City. The nearest neighbor method does not outperform the regression method in our study. We, however, believe that the information revealed through the nearest neighbor will be of great benefit to appraisers in justifying their judgments.

Keywords: grid-adjustment; nearest neighbor method; sales comparison approach

Sales Comparison Approach in Property Appraisal

Sales comparison approach is defined as "A comparative approach to value that considers the sales of similar or substitute properties and related market data and establishes a value estimate by process involving comparison…." (International Valuation Standards 2005: 405). Or it is said to be "A set of procedures in which a value indication is derived by comparing the property being appraised to similar properties that have been sold recently, applying appropriate units of comparison, and making adjustments to the sale prices of the comparables based on the elements of comparison…." (Appraisal Institute 2001: 63). The amount that an appraiser adds to or subtracts from the price of a comparable property is an estimate of the market value of attributes. And this estimate is done on the basis of experience, judgment, and knowledge of how individual buyers and sellers tend to price these attributes. (Brueggeman and Fisher 2001: 168) The appraiser then gives weights to adjusted prices of comparable properties and uses the weighted average price as the final indicated value (Corgel et al. 2001: 305). Therefore, to some extent, sales comparable approach is seen as a subjective process and serious errors can result if without justifiable adjustments (Brueggeman and Fisher 2001: 228).

The above discussion therefore suggests three essential elements involved in sales comparison approach which places high demands upon appraisers' professional judgments that often attract criticisms, namely selection of similar or comparable properties, price adjustments for attribute differences and allocation of weights among comparable properties.

Regression Analysis, Grid-Adjustment Method and Nearest Neighbors Technique

There has been a long history of applying regression analysis to property appraisal, at least dating back to 1922 by Haas (Colwell and Dilmore 1999), on agricultural land. Through regression analysis, respective price of individual attribute or characteristic is estimated on the foundation of hedonic pricing models. Sirmans et al. (2005) in a recent review of hedonic pricing models in real properties conclude that the most frequently included characteristics are lot size, square feet, age, the number of stories, and a time trend. These variables generally have the expected signs although in some instances they are not significant. A total of slightly less than 200 pieces of empirical work are cited. Through regression analysis, implicit prices of individual property attributes are estimated in an objective way. Appraisers do not need to make subjective guesses.

Despite the vast number of regression-type analysis in property prices, a significant part of them are not in full conformation with sales comparison approach in property appraisal. In principle, price of the subject property is derived as the sum of values of respective attributes. This regression process itself does not explicitly consider selection of comparable properties and assignment of weights to them. Colwell et al. (1983) identify three popular adjustment-grid methods in appraisal literature, namely additive dollar adjustment method, additive percentage adjustment method and multiplicative percentage adjustment method. These three methods respectively correspond to linear, semi-log and double-log regression models so far as the price adjustment is concerned. The authors conclude that grid-adjustment method, in contrast to traditional regression analysis, is more objective in deriving price adjustment between comparables and subject property. However, they suggest using empirical experiences in deciding weights allocation among comparables but do not offer objective criterion on this. Kang and Reichert (1991) provide an empirical study which compares the accuracy of price prediction of regression and grid-adjustment analysis. They examine sales prices of 1751 houses during 1986 for three Chicago suburban areas: Lombard, Wheaton and Naperville. Absolute value of the net adjustment factor is proposed to measure the similarity among potential comparable properties. This factor is the sum of attribute price obtained by regression, multiplied by the difference between subject property and potential comparables on values of individual attributes. The importance of individual attributes is represented by their regression coefficients. Based on the formula proposed by Kang and Reichert, the range of weighting values is between 0 and (1/n-1). The highest individual weight is 50% in the case of three comparables and 25% in the case of five comparables. There is consequently an intrinsic limit on the weights that might produce biases. The empirical results show that the forecasting errors of regression analysis are in the range of 6.25% and 10.04%. In contrast, the forecasting error of grid-adjustment method ranges from 3.3% through 11.64%. It is noted that the multiplicative percentage adjustment method is likely to prove more accurate in markets that are in equilibrium and housing and neighborhood characteristics are homogeneous.

Isakson (1986) criticizes grid-adjustment method for the amount of its adjustment is typically based upon subjective judgment. He proposes the nearest neighbor (NN) method instead. In the context of nearest neighbor method, the attributes of an individual property are viewed as coordinates along n-dimensional axis and every property is thus given a particular point in the m-dimensional space. Through calculation of Mahalanobis distances between the subject property and others enables the k-nearest properties to be identified. In addition, the weights for comparable properties are in inverse proportion to their respective Mahalanobis distances. Any appraiser who applies the nearest neighbor method to the same property at the same point will get the exact same estimate of value. Empirical study is undertaken to housing sales in Spokane, Washington from July 1, 1978 through September 30, 1978. Sales of 563 houses are employed to establish the prediction model by which prices of another 112 houses are estimated. It is concluded that the nearest neighbor method is statistically more accurate than any of the grid-adjustment methods in terms of sum-of-squared errors in price prediction. Despite the differences between these two methods, nearest neighbor method is best seen as an improvement to grid-adjustment method for their analytical foundation does not fundamentally differ. Isakson finally calls for more application of the nearest neighbor method on other markets Isakson (1988) extends nearest neighbors appraisal technique to a variety of commercial real estate in Dallas, Texas. NN estimates of value are significantly more accurate than OLS estimates for retail and miscellaneous properties, and are more accurate, but not significantly so, than OLS estimates for apartment buildings, industrial properties and office buildings. Isakson calls for more application of NN in other markets and for different types of properties.

Study Area, Empirical Data and Models

Review of previous studies reveals a series of efforts to make appraisal more objective and the needs for more empirical work. This article thus applies the nearest neighbor method to a suburban area of Taipei City to examine its prediction reliability and extend its application to apartments and high-rise apartments.

The study area is Wen-Sun district of Taipei City, a suburban area dominated by residential activities. The data set of 1468 apartments and 926 high-rise apartments that were sold in the market from January 2001 through January 2005 is supplied by the Land Administration Department of Taipei City. Apartments refer to residential buildings of or under 5 stories and without an elevator, and high-rise apartments refer to residential buildings over 5 stories, usually with elevators. As the data obtained includes address for individual property, all properties are located in space through address-matching function. In addition, their distances to major public facilities are calculated through networking analysis, both through geographical information system. We also keep 50 apartments and 50 high-rise apartments as the hold-out samples for later examination of prediction accuracy. In order to have a thorough examination in space, we adopt the quadrat analysis (Lee and Wong 2001: 62-72) to delineate Wen-Sun district into 50 quadrats and select one apartment and one high-rise apartment from each as the hold-out samples.

Distances to the nearest park, junior high school, public car park, power station, mass rapid transit, and hospital are calculated. A number of dummy variables are also included, including whether a property fronts onto a main road, a series of variables presenting the time trend as year 1999 being the base year, and a series of floor variables that present the effects of different floors on price. In Taiwan, ground-floor property is normally more expensive than on other floors for it has its own entrance and therefore with higher privacy. Top-floor property is expected to command a premium price because of its better access to the common areas at the top of a building. In contrast, the fourth-floor property is usually cheaper as the number four is in Taiwan often regarded as bad luck (similar to number 13 in the western culture) due to its similar pronunciation with the word death. Table 1 shows the summary of selected variables.

Table 1 Statistical Summary of Selected Variables

	Apartments				*High-rise apartments*			
	Mean	Min.	Max.	S.D.	Mean	Min.	Max.	S.D.
Sales price (New Taiwan Dollars)	5,244,346	2,220,000	22,000,000	1,707,418	6,428,456	1,600,000	17,500,000	2,681,682
Total floors (m^2)	101	37	304	25.6	108	21	345	43.2
Structure age (in years)	24	3	38	6.5	11	3	27	4.8
Distance to nearest park (m)	244	36	590	103.4	260	67	599	95.5
Distance to nearest school (m)	261	59	613	106.9	299	67	625	123.6
Distance to nearest public car park (m)	536	28	1658	256.2	666	38	1752	421.2
Distance to nearest power station (m)	691	52	1877	324.9	672	80	2731	398.9
Distance to nearest MRT (m)	1146	66	2946	659.1	1134	123	3033	692.8
Distance to nearest hospital (m)	1850	55	2326	703.3	1887	363	4573	702

No multicollinearity is found in any of the regressions with VIF test. Heterodasticity is found in some regressions, and where it is found, original standard error is replaced by White's corrected standard error (Gujarati 2003: 417-8). In addition, Dubin-Watson test shows no autocorrelation problem in any of them. Table 2 summarizes the regression results for both apartments and high-rise apartments in three functional forms.

Table 2 Regression Results for Apartments and High-rise Apartments

Variables	Apartment (linear)	Apartment (semi-log)	Apartment (double-log)	High-rise apartment (linear)	High-rise apartment (semi-log)	High-rise apartment (double-log)
Intercept	1999006(10.03)	14.90744(411.09)	12.17265(83.05)	2584413(11.62)	14.975(406.62)	12.01025(81.8)
Total floor areas (m^2)	50320***(54.16)	0.00836***(49.46)	0.88838***(51.68)	53045***(61.08)	0.00826***(57.47)	0.87028***(71.48)
Structure age (in year)	-42161***(-11.17)	-0.00644***(-9.37)	-0.15043***(-12.83)	-66938***(-8.06)	-0.01098***(-7.99)	-0.15422***(-10.83)
Front onto main road (yes:1; no:0)	-33865(-0.55)	-0.01438(-1.29)	-0.00969(-0.62)	-30840(-0.35)	-0.01719(-1.18)	-0.03683**(-2.12)
Distance to park (m)	-1460.95536**(-2.54)	-0.00022292**(-2.13)	-0.03787***(-1.9)	454.0666(0.48)	0.00004171(0.26)	-0.01582(-0.48)
Distance to school (m)	300.12282(0.55)	-0.00004251(-0.43)	0.00915(0.44)	-307.6866(-0.37)	0.00003588(0.26)	0.013(0.42)
Distance to public car park (m)	-325.95376***(-3.34)	-0.00008089***(-4.55)	-0.04262***(-5.49)	-10.30846(-0.07)	0.00001449(0.62)	0.02734***(2.93)
Distance to power station (m)	-189.54542**(-2.29)	-0.00002378(-1.58)	0.03419***(4.24)	-399.10291***(-3.44)	-0.00003996**(-2.08)	-0.01076(-1.22)
Distance to MRT (m)	-571.63997***(-11.85)	-0.00010728***(-12.22)	-0.10838***(-15.07)	-398.78172***(-6.02)	-0.00005018***(-4.57)	-0.06531***(-7.13)
Distance to hospital (m)	200.74544***(2.83)	0.00003400***(2.64)	0.05379***(3.72)	-183.56893**(-1.84)	-0.00004432***(-2.68)	-0.0816***(-4.23)
Sold in	-399682***(-5.49)	-0.06248***(-4.72)	-0.11514***(-6.19)	-1005213***(-7.65)	-0.1512***(-6.95)	-0.25072***(-9.61)

2000 (yes:1; no:0)						
Sold in 2001 (yes:1;no:0)	-955716***(-13.38)	-0.15959***(-12.28)	-0.24452***(-13.37)	-1553443***(-12.83)	-0.21588***(-10.77)	-0.38205***(-15.86)
Sold in 2002 (yes:1; no:0)	-816455***(-10.16)	-0.13878***(-9.49)	-0.21893***(-10.64)	-1640428***(-12.65)	-0.21445***(-9.99)	-0.3852***(-14.93)
Sold in 2003 (yes:1; no:0)	-673895***(-9.09)	-0.13541***(-10.04)	-0.20076***(-10.59)	-1433558***(-11.78)	-0.21181***(-10.52)	-0.372***(-15.37)
Sold in 2004 (yes:1; no:0)	-333319***(-3.06)	-0.04954***(-2.5)	-0.08255***(-2.96)	-1696303***(-8.61)	-0.22193***(-6.81)	-0.3893***(-9.95)
On ground floor (yes: 1; no: 0)	1298541***(18.53)	0.22250***(17.46)	0.34737***(19.4)	1396174***(9.41)	0.18919***(7.7)	0.30094***(10.25)
On fourth floor (yes:1; no:0)	-235599***(-3.33)	-0.02108(-1.64)	-0.02412(-1.33)	-247498**(-2.06)	-0.04172**(-2.1)	-0.0545**(-2.29)
On top floor (yes:1; no:0)	88178***(1.54)	0.01074(1.03)	0.02736**(1.87)	221427(1.35)	0.05187***(1.91)	0.06292*(1.93)
adj R^2	0.7391	0.7064	0.7205	0.8249	0.7064	0.8631
D-W value	1.965	2.041	1.786	1.865	1.977	2.014

***: significant at 0.01 significance level **: significant at 0.05 significance level *:: significant at 0.1 significance level

numbers in parentheses are t-values

The empirical results as a whole correspond to our prior expectation. Residential properties that front onto a main road are penalized likely due to the noise and congestion. Properties with a better access to MRT command a premium in price.

Besides, properties significantly depreciate with structure age. Properties on ground and top floors tend to be more expensive than those on other floors. In contrast, properties on the fourth floor are liable to suffer from price reduction. In addition, the series of time variables present a consistent price trend. However, apartments seem to be more sensitive than high-rise apartments to access to parks and public car parks. It is probably because high-rise apartments in Taiwan often have their own exclusive, though small, green areas and own underground car parks. The variable coefficients for access to power station and hospital are contrary to prior expectation. This discrepancy might be because the variations among power stations, such as in size and surrounding neighborhoods, are not fully explained by this distance variable. Also, only one, the primary district, hospital is considered and those local clinics are ignored in the models. However, the primary purposes of this paper are not on the individual coefficients but in the comparison of price prediction between methods. The two variables are therefore retained for later analysis.

The variable coefficients derived in Table 2 are then employed to estimate prices for the hold-out samples of 50 apartments and 50 high-rise apartments. They are first used to directly arrive at the estimated property prices as the conventional regression analysis does. As we argued earlier, the conventional regression method, through performing reasonably well, are not in full conformation with the principle of the sales comparison method. Selection of comparable properties, price adjustments and weights allocation are absent in derivation of the final predicted price. Therefore, in addition to the conventional regression method, the nearest neighbor method is also undertaken for comparison.

In compliance with the real property valuation standard published by the Taiwan government, three comparables for each apartment and high-rise apartment in the hold-out sample are chosen. These comparables are those with the three shortest Mahalanobis distances to the subject property. The weights allocated to these three properties are determined and the weighted average price, namely the indicated price, is derived as sales comparison method. Up to this stage, the sales comparison method is completed in conjunction with regression analysis. The Mahalanobis distances and weights allocated to comparable properties are derived by formulas 1 and 2.

$$D^2_{ij} = (X_i - X_j)\, E^{-1}\, (X_i - X_j)' \qquad (1)$$

$$W_{ij} = (1/D^2_{ij}) \,/\, \sum_{i=1}^{3} (1/D^2_{ij}) \qquad (2)$$

where D_{ij}: Mahalanobis between property i and j

X: a vector of the factor-coordinates of the property

E: the factor-coordinate covariance matrix of all of the properties

Mahalanobis distances between all properties in the hold-out sample and the remaining properties are calculated and through which the three most similar properties for each subject property are identified. Weights allocated to the three comparable properties are consequently determined. All these procedure are in its nature objective.

Accuracy of Models in Predicting Property Prices

In order to examine the accuracy of price prediction between the traditional regression method and the nearest neighbor method, Table 3 illustrates the results for the properties in the hold-out sample in terms of average prediction errors.

Table 3 Average Prediction Errors for Hold-out Sample

Apartments					
NN linear	NN semi-log	NN double-log	Regression linear	Regression semi-log	Regression double-log
10.22%	10.41	11.12%	12.18%	11.78%	11.27%
(13.09%)	(12.64%)	(11.82%)	(9.96%)	(10.83%)	(9.56%)
High-rise Apartments					
NN linear	NN semi-log	NN double-log	Regression linear	Regression semi-log	Regression double-log
10.15%	10.52%	13.92%	11.91%	13.75%	12.11%
(10.3%)	(10.68%)	(15.4%)	(10.7%)	(11.65%)	(12.01%)

Note: numbers in the parentheses are standard deviation.

In regards to apartments, the average errors for the nearest neighbor method are lower in all functional forms than the regression method, although not by a significant margin. It seems that the nearest neighbor method performs better in predicting property prices. However, the standard deviations of prediction errors for the nearest neighbor method are clearly higher than the regression method. Thus the predicting reliability of nearest neighbor method appears to be more unstable. As for high-rise apartments, except for double-log function, the nearest neighbor method predicts price better than the regression method in both accuracy and stability. Overall, we cannot conclude definitively which method is better than another. We further calculate the coefficient of variation (Eckert 1990: 539) for prediction errors. This coefficient expresses the standard deviation as a percentage of average errors and thus makes comparison between groups easier. Table 4 details the results of coefficient of variations for both apartments and high-rise apartments in the hold-out samples.

Table 4 Coefficient of Variations for Hold-out Samples

Apartments					
NN-linear	NN semi-log	NN double-log	Regression-linear	Regression semi-log	Regression double-log
14.58%	14.11%	13.3%	11.34%	12.27%	10.77%
High-rise Apartments					
NN-linear	NN semi-log	NN double-log	Regression-linear	Regression semi-log	Regression double-log
11.47%	11.94%	17.89%	12.15%	13.5%	13.67%

The smaller the coefficient of variations, the better it performs. It is clear that in terms of coefficient of variations, the traditional regression method performs better than the nearest neighbor method on apartments. In contrast, the nearest neighbor method seems to overall outperform regression method on high-rise apartments.

Despite that the average and variation figures are useful measures, the accumulation of prediction errors provides other insights into the prediction accuracy of models. The prediction error, the difference of predicted value and true value divided by true value, in percentage terms is on the X-axis and the accumulation percentage is on the Y-axis on Tables 5 through 10. All figures other than those in Table 10 indicate that nearest neighbor method outperforms the traditional regression model. Taking 10% of prediction error as an example, apart from the double-log function for high-rise apartments, the accumulative proportion of prediction errors using the nearest neighbor method is substantially higher than the regression method. As far as the accumulative prediction errors are concerned, the nearest neighbor method seems to be superior to the traditional regression model.

Table 5 Accumulated Prediction Errors- Apartments (linear)

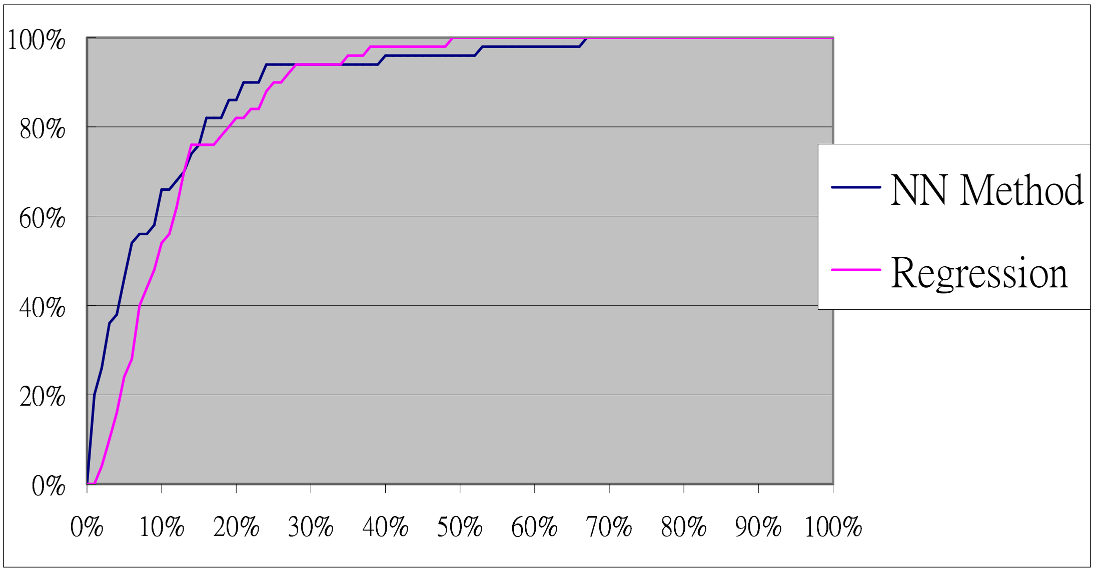


Table 6　Accumulated Prediction Errors- Apartments (semi-log)

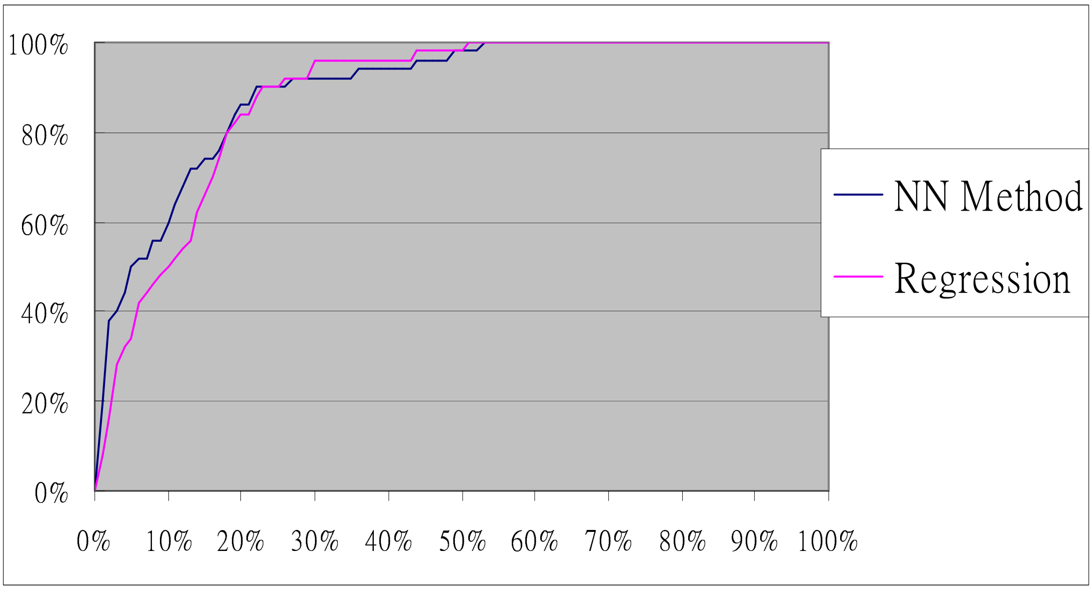


Table 7　Accumulated Prediction Errors- Apartments (double-log)

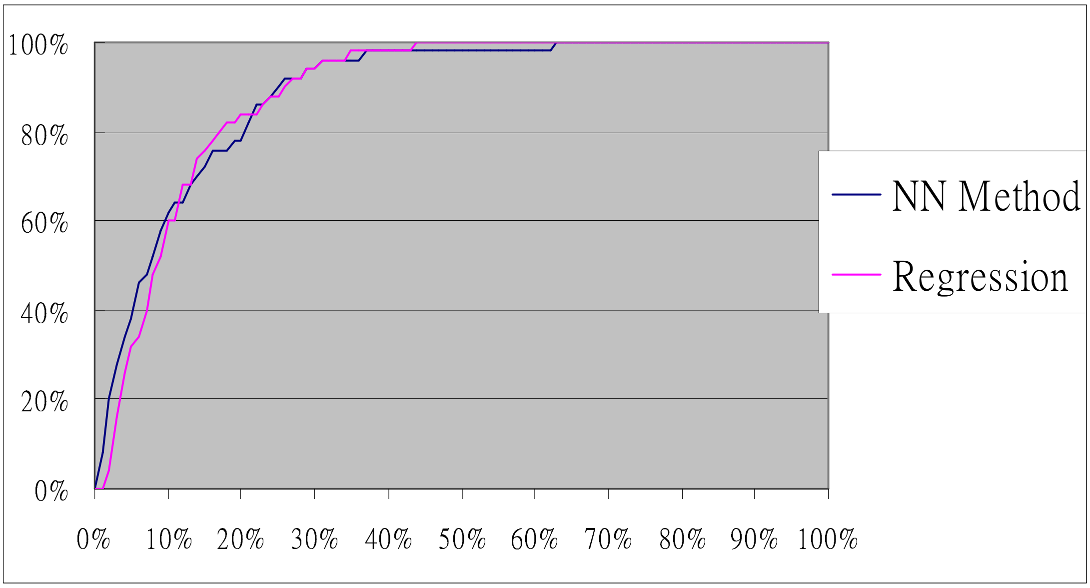


Table 8 Accumulated Prediction Errors- High-rise Apartments (linear)

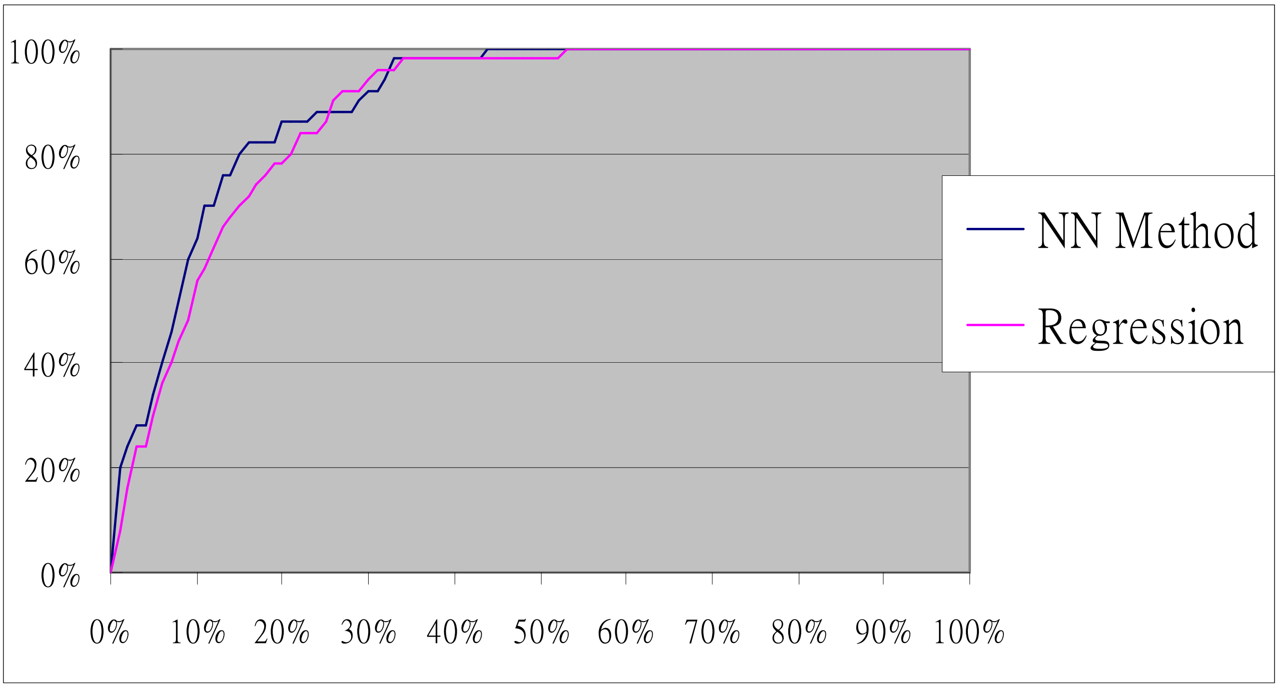


Table 9 Accumulated Prediction Errors- High-rise Apartments (semi-log)

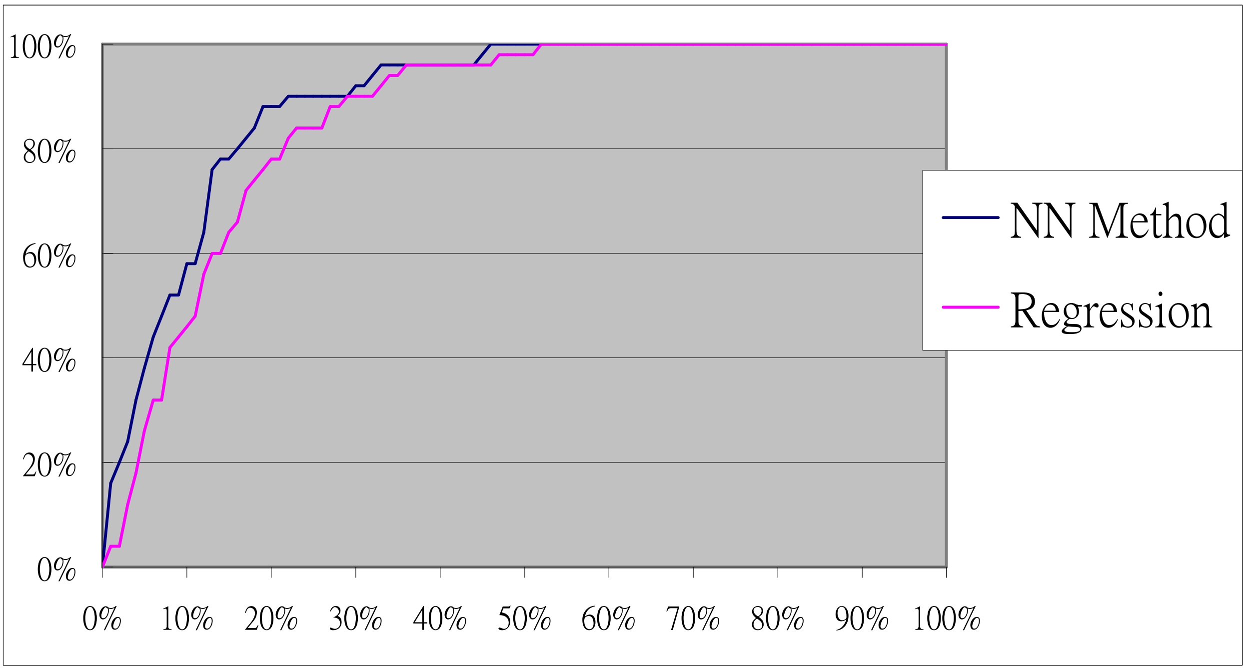
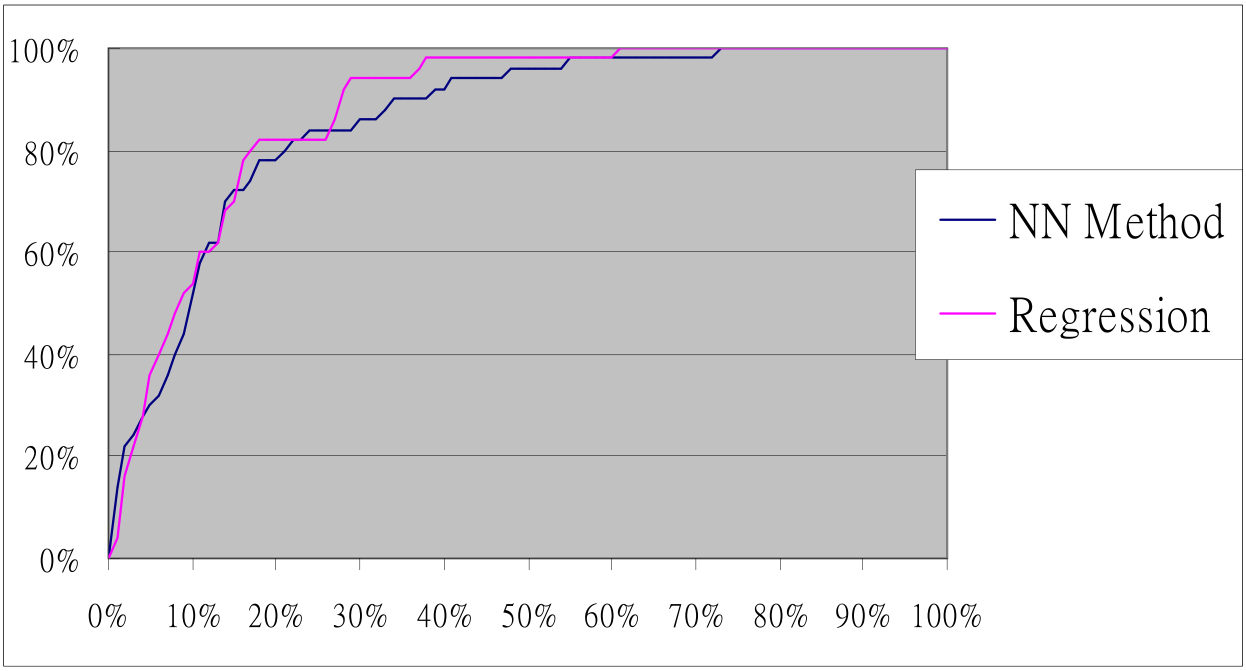


Table 10 Accumulated Prediction Errors- High-rise Apartments (double-log)



Another related issue in practice is the geographical areas within which the comparable properties might be located. In other words, what is the possible boundary from the subject property that an appraiser might pick the comparables? Table 11 illustrates the results in our study.

Table 11 Distances of the Comparable Properties from the Subject Property

	Apartments		*High-rise apartments*	
Distances	No. of comparables	Accumulated percentage	No. of comparables	Accumulated percentage
0-100 m	164	65.6 %	171	68.4 %
100-200 m	32	78.4 %	27	79.2 %
200-300 m	11	82.8 %	18	86.4 %
300-400 m	9	86.4 %	8	89.6 %
400-500 m	4	88 %	7	92.4 %
500-600 m	5	90 %	0	92.4 %
600-700 m	5	92 %	0	92.4 %
700-800 m	5	94 %	2	93.2 %
800-900 m	4	95.6 %	0	93.2 %
900-1000 m	1	96 %	3	94.4 %
1000-1500 m	1	96.4 %	0	94.4 %
1500-2000 m	2	97.2 %	2	95.2 %
2000-2500 m	5	99.2 %	3	96.4 %
2500-3000 m	2	100 %	8	99.6 %
>3000 m			1	100 %

Likely because Taipei is a very high-density city, respectively 65.6 and 68.4 percent of the three comparables for apartments and high-rise apartments are found within 100 meters of the subject property. Ninety percent of comparables are within 600 meters of the subject property. Moreover, except for one case, comparables are no more than 3000 meters away from the subject property. A practical implication might be that in this district a comparable with a distance of over 3000 meter from its subject property demands more examination and explanation. Finally, 31 comparable apartments are in the same building with the subject apartment. Also, 28 comparable high-rise apartments are in the same building with the subject high-rise apartments. The evidence suggests that the nearest neighbor method has a tendency to find comparables with very similar attributes, and this tendency corresponds to the rule-of-thumb in practice.

Concluding Remarks

Sales comparison is probably the most frequently used method in property appraisal. The application of regression method with a large number of sales data enables a more objective estimate of attribute values. However, even when regression method is applied, some essential elements of sales comparison method are missing, such as selection of comparable properties and weights allocation.

The current study, after reviewing the incorporation of regression method with grid-adjustment concept, extends the nearest neighbor method to high-density residential properties that have not been examined. The results of prediction accuracy are mixed but overall are in favour of the nearest neighbor method. These results might be caused by the functional forms adopted. What is more, the process of selecting comparables uncovers useful information for valuation. We can through the comparable selections understand which properties are similar and to what extent. This information can serve as fairly useful reference knowledge to appraisers. Paired-sales method can derive value of only one attribute at one time. But traditional regression method considers all sales equal in similarity which is contrary to the essence of the sales comparison method. The nearest neighbor, or improved grid-adjustment, method not only estimates the attribute values through regression analysis, but also remains the philosophy of the sales comparison method in comparable estimation and weights allocation.

Although the nearest neighbor method does not predict property prices more accurately than the regression method in this study, we believe the whole process itself provides useful information from which can substantially benefit appraisers. More empirical studies along this line therefore shall be called for.

References

Appraisal Institute (2001) The Appraisal of Real Estate. Chicago: Appraisal Institute.

Brueggeman, W. and Fisher, J. (2001) Real Estate Finance and Investments. New York: McGraw-Hill Irwin.

Colwell, P., Cannaday, R. and Wu, C. (1983) The Analytical Foundations of Adjustment Grid Methods. AREUEA Journal. 11(1): 11-29.

Colwell, P. and Dilmore, G. (1999) Who was First: An Examination of an Early Hedonic Study. Land Economics. 75(4): 620-26.

Corgel, J., Ling, D. and Smith, H. (2001) Real Estate Perspectives. New York: McGraw-Hill Irwin.

Eckert, J. (1990) Property Appraisal and Assessment Administration. Chicago: The International Association of Assessing Officers.

Gujarati, D. (2003) Basic Econometrics. New York: McGraw-Hill Irwin.

Isakson, H. (1986) The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Methods. AREUEA Journal. 14(2): 274-86.

Isakson, H. (1988) Valuation Analysis of Commercial Real Estate Using the Nearest Neighbors Appraisal Technique. Growth and Change. 19(2): 11-24.

Kang, H-B and Reichert, A. (1991) An Empirical Analysis of Hedonic Regression and Grid-Adjustment Techniques in Real Estate Appraisal. AREUEA Journal. 19(1): 70-91.

Lee, J. and Wong, D. (2001) Statistical Analysis with Arcview GIS. New York: John Wiley & Sons, Inc.

Sirmans, G., Macpherson, D. and Zietz, E. (2005) The Composition of Hedonic Pricing Models. Journal of Real Estate Literature. 13(1): 3-43.