On the Weights for Characteristics and Comparables for Property Valuation using Quality Rating Valuation Estimation

by Karina Agustin

Submission date: 16-May-2023 03:33PM (UTC+0700)
Submission ID: 2094485986
File name: n_the_Weight_of_Variables_Comparable_in_Valuation_anonymous.docx (94.88K)
Word count: 7334
Character count: 39802

On the Weights for Characteristics and Comparables for Property Valuation using Quality Rating Valuation Estimation (QRVE)

Abstract

This study considers the problem of finding weights for building characteristics and compares buildings in property valuation to provide a more rigorous analytical foundation for a simple yet practical valuation technique knowns as Quality Rating Valuation Estimation (QRVE). Mathematically, we prove that the "best" characteristic weights can be obtained from Multiple Linear Regression Analysis (MRA) coefficients. Furthermore, by applying the Gower Similarity index and the Partition Around Medoid (PAM) clustering technique, the proposed algorithm provides an appropriate similarity of the weighing of compared buildings. The case studies illustrate a way to select a subset of characteristics when there are many of them with two numerical examples, as well as a complete modification of QRVE in conjunction with the grid adjustment technique. The modified QRVE proposal results in a very reasonable and high valuation performance of the building value estimate.

Introduction

Property (real estate) valuation methods have been studied extensively worldwide, and researchers from various countries have published their results recently [1][2][3]. From a practitioner's standpoint, there are five classifications of valuation methods: comparable, investment, profit, depreciated replacement, and residual methods [4]. In Indonesia, Komite Penyusun Standar Penilaian Indonesia (KPSPI) & Masyarakat Profesi Penilai Indonesia (MAPPI) via "Kode Etik Penilaian Indonesia dan Standar Penilaian Indonesia Edisi VII" (2018) explained that there are three approaches, namely: market data approach, income approach, and cost approach. This paper focuses primarily on the (sales) comparison approach defined by Lemen or the market data approach specified by KPSPI/MAPPI [5]. The primary measurement technique used in the market data approach that a valuer tries to decide the fair market price of a subject property (either for sale or rent) is by comparing the subject property with the prices of several other subjects for a particular set of variables [6].

In Indonesia and several other Southeast Asia countries as well as Australia, one of the most widely used (sales) comparison methodology in real estate valuation is commonly referred to as Quality Rating Valuation Estimation (QRVE) technique. QRVE is a property valuation technique that assesses the quality of a property based on a set of pre-defined criteria. The technique involves assigning a rating or score to different criteria of the property [7]. In Indonesia, this QRVE technique is officially approved by the Indonesian Tax authority (Direktur Jendral Pajak) via Appendix 1 SE-54/PJ/2016 (other adjustment techniques include: percentage adjustment, cost adjustment, pair-wise data comparison, & other Statistical techniques) [8]. There are several general steps in QRVE procedures [9], such as defining the weight and score of the characteristic building based on the subjectivity of the appraiser that will influence the market value estimate (of the subject). This subjectivity is because of the expert appraiser's assumption to produce an estimation that probably results inevitably vary estimate, both from case to case and from different appraisers [10]. Therefore, this study will determine the basis of weight assignments on those building characteristics using an analytical (statistical or mathematical) foundation.

The next part of our research is about the applicability of comparables data (sales & their variables) other than forming the simple linear regression equation. In other words, are these comparables' values (sales or rent) data and their variables just needed to create a linear line, or could they be used any other way in doing a better valuation? For example, is it better to use the comparables with a wide range of values, e.g., the larger or a smaller ratio of (maximum sales/minimum sales), or are they irrelevant? More importantly, if we need to compare several comparables building, should they be assumed equally important, or are a set of comparables more relevant to be considered as the comparables?

In this paper, we present our findings about the above two research questions, i.e., the weights of variables and the weights of comparables, as well as our proposal. In the next section, we started with our effort to trace the history (and presented it in the literature review). Then in the next section, we present a more robust analytical foundation for QRVE based on Machine Learning and Clustering techniques yet simple enough so that they can be easily adopted in practice. Next, we illustrate our proposal with several examples from previously published research and demonstrate that it produces a similar valuation—finally, our closing remark and further research direction. An R function is also being developed to be used by the practitioner.

Literature Review

Real property is the foremost real estate appraisal theory discussion as the community believes it to be an excellent investment. Four interactive and significant factors that affect actual property values are social, economic, governmental, and environmental factors. Therefore, the appraiser must continuously consider those factors and update the recent information to interpret the market trends. Based on several approaches that are generally implemented, the market approach, income approach, and cost approach that constitute the three traditional assessment procedures, the market approach is the most organized approach for value estimation due to the possibility of directly comparing the assessed value of the subject property and the selling prices of the comparable within the same market area, when data are available. Statistical analysis plays an important role in the market approach, and there are several steps involved in the data analysis process, such as data collection, data exploration, model specification, model estimation, and model evaluation. In the market approach, the QRVE concept is used to adjust the values of comparable assets based on their quality ratings. Quality ratings are typically based on several factors that can affect the value of an asset, such as its age, condition, location, and amenities. By adjusting the values of comparable assets based on their quality ratings, the appraiser can obtain a more accurate estimate of the value of the subject asset [11]

In terms of the QRVE itself, it seems that the method was popularized in Indonesia by Hartoyo [7] after Cooper [12], during "Indonesia – Australia Specialized Training Project 2002," declared that QRVE is a technique that appears systematic and a bit scientific – that was first written by Ratcliff [13]. We later found an online note (article) written by Cooper about Multiple Regression Analysis:

> *"No matter how attractive MRA modeling might look in principle, we must face the fact that the number of property characteristics that are expected to have a significant influence on market value will, in the majority of cases, far exceed the number of transactions available for analysis."*

In other words, the QRVE technique (with simple linear regression) is used as opposed to multiple regression since the degree of freedom in multiple linear regression shrinks quickly. QRVE technique can be described by the following steps [9]:

- Decide on a set of m characteristics (x_i) that the valuer believes can, influence the market value (of the subject), and assign weights to each one of them $(w_i)\ \forall i = 1, \dots, m$.
- Assign scores (Likert scale, usually between 1 – 5 or 1 – 10) to the subject of evaluation as well as its comparables (assume: there are n comparables, and they are usually subscripted with $j = 1, \dots, n$ – furthermore, the subject is usually subscripted as index 0).
- Calculate the quality rating/score for the subject as well as its comparables $qs_j = \sum_{i=1}^{m} x_i w_i$
- Run a simple linear regression in the form $v_j = b_0 + b_1(qs_j) + \varepsilon_j$ to obtain b_0 and b_1 using the n comparables.
- Use the value of b_0 and b_1 together with the quality score of the subject, i.e., qs_0 to determine the estimated value of the subject $v_0 = b_0 + b_1(qs_0)$ – recall that the subject of evaluation is subscripted with index 0.

In other country, such as Thailand, the Weighted Quality Score (WQS) technique is prevalent. Sriboonjit & Rattanaprichavej proposed a stepwise regression to select the characteristics used in valuation. However, they did not suggest anything about the weights [14].

Given the previous study, how the weights are assigned to those characteristics still needs to be clarified. Our further research about the weight of building characteristics related to physical and environmental factors led us to several studies provided by the Real Estate division of Sauder Business School at the University of British Columbia, namely the Foundations of Real Estate Appraisal course & Statistical and Computer Applications in Valuation course about the technique called Quality Point (QP) method, prepared by Zaric [15][16]. He suggested the "best" set of weights $(w_i's)$ is interpreted as the values for the weights $(w_1, w_2, \dots, w_m)$ so that *the price per quality point of all of the index properties (comparables)* are as close together as possible, *i.e.*, $\frac{v_1}{\sum_{i=1}^{m} x_{i,1} w_i} = \frac{v_2}{\sum_{i=1}^{m} x_{i,2} w_i} = \cdots = \frac{v_j}{\sum_{i=1}^{m} x_{i,j} w_i}$. He further recommended the use of (non)linear programming to find the weights for characteristics.

The second part of the literature review is the selection of comparables and the weight used on the comparables. The analytical foundation for Adjustment Grid Comparison (AGM) was first proposed by Coldwell *et al* [17]. They gave a complete statistical (mathematical) foundation in explaining the sales comparison & grid adjustment technique in appraisal methodology. They pointed out that the estimated fair market price of a subject property ($\boldsymbol{v}_s$ = a vector of $n \times 1$ constant value, *i.e.*, the same estimated fair market price) can be represented in the following mathematical form:

$$\boldsymbol{v}_s = \boldsymbol{v}_o + (\boldsymbol{X}_s - \boldsymbol{X}_o)\boldsymbol{b} + \boldsymbol{\varepsilon} \qquad [1]$$

Where: $\boldsymbol{v}_o$ = a vector of $n \times 1$ observed price of a comparable properties, $\boldsymbol{b}$ = a $k \times 1$ vector of coefficients, $\boldsymbol{X}_s$ and $\boldsymbol{X}_o$ are an $n \times k$ matrices of comparable properties data & their characteristics/attributes (*n* properties and *k* attributes/characteristics). Of course, ideally the error term $n \times 1$ vector $\boldsymbol{\varepsilon}$ should be normally distributed.

Furthermore, the first significant paper about the weights of the comparables building was by Vandell that proposed a method that is theoretically equivalent to a minimum-variance estimate [18]. However, he admitted that his approach would depend heavily upon how well the variance/covariance matrix is estimated from the underlying distribution. Guijarro recently formulated a Quadratic Programming problem similar to the mean-variance portfolio optimization problem [19]. Interestingly, he also argued that the sales comparison model should minimize the adjusted prices' variance, not their coefficient of variation. We approach the problem from a slightly different perspective, i.e., as a clustering problem. The use of clustering in Sales Comparison is not new. Isakson was the first to propose using Mahalanobis distance as a weight in a technique called Nearest Neighbor Appraisal Technique (NNAT) [20][21]. Recently much research along this line has been quite popular. For example, Cajias et al. analyze residential real estate in Germany using the clustering technique [22]. Calka proposed a two-stage approach where a clustering method is used to group properties in the first stage [23]. Lastly, Rahim & Razali compared Eucledian, Minskowsky, & Cosine similarities as the basis for the Sales Comparison method [24]. We propose to utilize Gower distance with Podani extension because the QRVE technique uses mainly ordinal data, and Gower distance can handle more general data scales.

More Scientific Analytical Valuation with QRVE

Given the popularity of the QRVE/WQS/QP technique, it is imperative to build a more robust analytical foundation. Let's start with some assumptions, and then we can derive a more solid analytical foundation:

- Assume that we must choose m out of p available characteristics to be used in valuation (of course, $p > m$). We further assume that these characteristics are additive.
- Assume that there are j (index) properties with values $[v_1 \quad v_1 \quad \cdots \quad v_n]^T$ that are being used as comparables to decide the value of the subject, i.e., v_0.
- Of course, we know the vector characteristic (ordinal) values for the subject, *i.e.*, $\mathbf{x}_0^T = [x_{1,0} \quad x_{2,0} \quad \cdots \quad x_{m,0} \quad \cdots \quad x_{p,0}]$ as well as the characteristics (ordinal) values of these comparables, *i.e.*, the matrix of comparables × characteristics are given below:

$$\mathbf{X} = \begin{array}{c c} & [w_1 \quad \cdots \quad w_p] \\ \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} & \begin{bmatrix} x_{1,1} & \cdots & x_{p,1} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{p,n} \end{bmatrix} \end{array}$$

Subset of Regression problem

The first problem we need to tackle is that there could be a situation where the number of available characteristics is greater than the available comparables; that is the classic problem in Multiple Regression Analysis that Cooper mentioned [12]. Luckily, the selection of subsets of regression variables has been considered by many researchers [24][25][26][27][28]. Together with various techniques such as Leaps & Bounds that had been implemented in R (*via* R-package: leaps) [29]. Together with Bayesian Information Criteria (BIC), Adjusted R^2, and/or Mallow's C_p can be used to figure out the most suitable numbers of characteristics m. Mallow's C_p is almost identical to Akaike Information Criterion (AIC) – all are mathematical methods for evaluating how well a model fits the data it was generated from.

On the Weights of Characteristics

Given that we have obtained the most suitable selection of m subsets of regression variables with $m < n$, we can proceed in similar manner as suggested by Zaric. However, we do it differently. First, consider the simple linear regression in QRVE technique (we let z_j be the quality score for comparable j):

$$v_j = b_1 z_j + e_j \qquad [2]$$

Let $\boldsymbol{v}^T = [v_1 \quad \cdots \quad v_n]$ be the $n \times 1$ vector of Value, $\mathbf{z}^T = [z_1 \quad \cdots \quad z_n]$ be the $n \times 1$ vector of Quality Score, and $\boldsymbol{e}^T = [e_1 \quad \cdots \quad e_n]$ be the $n \times 1$ vector of error term. Note that we use small letter bold to indicate it is vector, and capital bold letter to indicate it is matrix. Please note that by definition, b_1 is the quality point that we are looking for since it gives the same ratio $\frac{v_j}{qs_j}$ $\forall j = 1, \cdots, n$.

Lemma 1:

The most linear ("best" as defined by Zaric) quality score for comparable j (*i.e.*, qs_j) in [2] can be obtained by defining the weight w_i of characteristic i as:

$$w_i = \frac{c_i}{\sum_{i=1}^{m} c_i} \quad \forall i = 1, \dots, m \qquad \text{if all } c_i > 0 \tag{3a}$$

Or

$$w_i = \frac{c_i}{\max_i\{c_i\} - \min_i\{c_i\}} \quad \forall i = 1, \dots, m \qquad \text{if } \exists\, c_i \le 0 \tag{3b}$$

Where: $\boldsymbol{c}^T = [c_1 \quad \cdots \quad c_m]$ is the Ordinary Least Square (OLS) estimate for $\boldsymbol{v} = \boldsymbol{Xc} + \boldsymbol{u}$ with m variables.

Proof:

First, note that R^2 of equation [2], denoted by R_{qs}^2 is given by:

$$R_{qs}^2 = 1 - \frac{\sum_i e^2}{\sum_i (v_i - \bar{v})^2} = 1 - \frac{\boldsymbol{e}^T \boldsymbol{e}}{\overline{\boldsymbol{v}}^T \overline{\boldsymbol{v}}} \tag{4}$$

In [3] the denominator $\overline{\boldsymbol{v}}^T \overline{\boldsymbol{v}}$ is constant.

Now, for multiple linear regression with the most suitable m characteristics, we have the following in vector (matrix) notation:

$$\boldsymbol{v} = \boldsymbol{Xc} + \boldsymbol{u} \tag{5}$$

where: $\boldsymbol{X}$ is an $n \times m$ matrix of n building where each property has m characteristics, $\boldsymbol{c}^T = [c_1 \quad \cdots \quad c_m]$ be the $m \times 1$ vector of coefficient for each characteristic, and $\boldsymbol{u}^T = [u_1 \quad \cdots \quad u_n]$ be the $n \times 1$ vector of error term. The R^2 of equation [5], denoted by R_{MLR}^2 is given by:

$$R_{MLR}^2 = 1 - \frac{\sum_i u^2}{\sum_i (v_i - \bar{v})^2} = 1 - \frac{\boldsymbol{u}^T \boldsymbol{u}}{\overline{\boldsymbol{v}}^T \overline{\boldsymbol{v}}} \tag{6}$$

Furthermore, define $\boldsymbol{z} = \boldsymbol{Xw}$ where $\boldsymbol{w}^T = [w_1 \quad \cdots \quad w_m]$ is an $m \times 1$ vector weight that we need to decide. Of course, $\sum_i w_i = 1$. Therefore, from [2] we have:

$$\boldsymbol{v} = b_1 \boldsymbol{Xw} + \boldsymbol{u} \tag{7}$$

Given that b_1 and $\boldsymbol{c}$ are both OLS estimate, we have:

$$\boldsymbol{e}^T \boldsymbol{z} = \boldsymbol{u}^T \boldsymbol{Xw} = 0$$

Since $\boldsymbol{z} = \boldsymbol{Xw}$, it must follow that $\boldsymbol{e} = \boldsymbol{u}$ and $b_1 \boldsymbol{w} = \boldsymbol{c}$ in order for $R_{qs}^2 = R_{MLR}^2$. This completes our proof.■

It should be noted here that the denominators in [3a] & [3b] are essentially a constant. It can be replaced by any constant value.

On the Weights of Comparables

Given that QRVE/WQS/QP technique is essentially dealing with ordinal data, we propose a practical Gower distance (i.e., Gower similarity/dissimilarity) together with a very simple Partition Around Medoid (PAM) clustering technique. One of the most commonly used is the one known as Gower distance (i.e., Gower dissimilarity), first proposed in 1971 [30], and it was proven to satisfy the following (Mathematical) distance properties:

- Non-negative: $d(P_1, P_2) \geq 0$ (with $d(P_j, P_j) = 0\ \forall j$)
- Symmetry: $d(P_1, P_2) = d(P_2, P_1)$
- Triangle Inequality: $d(P_1, P_2) \leq d(P_1, P_3) + d(P_3, P_2)$

Dissimilarity is a more flexible way to measure unlikeliness between objects. It is essential to understand that the original proposal by Gower does not consider ordinal data. However, Podani, Kaufman & Rousseeuw are the ones who propose some modifications to cover the ordinal scale as well [31][32]. We use the amendment proposed by Podani in this article.

Within the context of appraisal of a real estate (building), for a particular characteristic (attribute/trait), the Gower Distance (i.e., Gower dissimilarity index) with Podani's extension for building s and building t is defined as follows:

- For data with Interval/Ratio scale: $d_i(s,t) = \frac{|x_{s,i} - x_{t,i}|}{\max_k\{x_{k,i}\} - \min_k\{x_{k,i}\}}$ [8]
- For data with Ordinal scale: $d_i(s,t) = \frac{|\mathrm{Rank}(x_{s,i}) - \mathrm{Rank}(x_{t,i})| - \frac{1}{2}(T_{i,s} - 1) - \frac{1}{2}(T_{i,t} - 1)}{\max_k\{\mathrm{Rank}(x_{k,i})\} - \min_k\{\mathrm{Rank}(x_{k,i})\} - \frac{1}{2}(T_{i,max} - 1) - \frac{1}{2}(T_{i,min} - 1)}$ [9]
- For data with Binary/Nominal scale: $d_i(s,t) = \begin{cases} 1 & \text{if } x_{s,i} = x_{t,i} \\ 0 & \text{if } x_{s,i} \neq x_{t,i} \end{cases}$ [10]

In equation [9], $T_{i,s}$ is the number of objects that have the same ranking score for characteristic i as object s (including s itself), while $T_{i,min}$ & $T_{i,max}$ are the numbers of objects that have minimum and maximum rank respectively.

Equations [8] – [10] can be used to construct dissimilarity index between the subject (*e.g.*, Building 00) with its comparables (*e.g.*, Building 1 – Building n) depending on the characteristics that are chosen by appraisal. Notice that all values of dissimilarity index $d_i(s,t)$ are between 0 and 1, *i.e.*, $0 \leq d_i(s,t) \leq 1$.

Obviously, the similarity index between 2 objects (s,t) for an attribute i is then defined as:

$$s_i(s,t) = 1 - d_i(s,t) \qquad [11]$$

Also, if we have weights for every characteristic i (w_i) – as obtained by Lemma 1, we can easily multiply the (dis)similarity index(es) for each characteristic i by the corresponding weight to obtain a singular value (of Gower Distance $= GD$ or Gower Similarity $= GS$) that represents (dis)similarity between 2 objects (in particular, we just need to focus dissimilarity between our subject (Building 0) with its comparables, Building j, *i.e.*,

$$GS(0,j) = \frac{\sum_{i=1}^{m} w_i s_i(0,j)}{\sum_{i=1}^{m} w_i} \quad [12a]$$

$$GD(0,j) = \frac{\sum_{i=1}^{m} w_i d_i(0,j)}{\sum_{i=1}^{m} w_i} \quad [12b]$$

Notice that in [12a] and [12b], equation [11] still holds, namely: $GS(0,j) + GD(0,j) = 1\ \forall j = 1, \dots, n$ because $\sum_{i=1}^{m} w_i = 1$. Given that we are interested to find out the fair market price of real estate (building) 0, we can simply calculate $GS(0,j)\ \forall j = 1, \dots, n$. This is our first step in calculating weights of comparables, and we can then choose a reasonable cut-off point for $GS(0,j)$, for example: $GS(0,j) \geq 0.75$ to select those comparable buildings (or to choose the top q out of n comparables according ot their similarity).

Alternatively, we can run PAM clustering technique (and use shilouette index) to see which are q comparables that belong to the same cluster as the subject 0. Furthermore, once we obtain q out of n comparables according to their similarities (either via PAM clustering or by selection for $GS(0,j)$, we can use the Gower Similarity measure as weights for the comparables because Gower Similarity satisfies distance properties.

Proposed Algorithm for QRVE/WQS/QP & its Rationale

Instead of using trial & error and (non)linear programming, we can solve the weighing for characteristics as well as weighing for comparables as follows:

Step 1: Use BIC, Adjusted R^2, Mallow's C_p, or Stepwise Regression to select a subset of characteristics used, *i.e.*, choose $m(< n)$ characteristics from p possible charactersitics that may impact value (price).

Step 2: Solve the multiple linear regression: $\boldsymbol{v} = \boldsymbol{X}\boldsymbol{c} + \boldsymbol{u}$ for the coefficient vector $\boldsymbol{c}^T = [c_1 \quad \cdots \quad c_m]$.

Step 3: To get the weight $\boldsymbol{w}^T = [w_1 \quad \cdots \quad w_m]$, we can easily normalize $\boldsymbol{c}^T = [c_1 \quad \cdots \quad c_m]$ by dividing each value with the summation or the range, namely: $w_i = \frac{c_i}{\sum_{i=1}^{m} c_i}$ or $w_i = \frac{c_i}{\max_i\{c_i\} - \min_i\{c_i\}}\ \ \forall i = 1, \dots, m$ – Lemma 1, equation [3a] or [3b].

Step 4: Calculate the quality score for the subject and its comparables, $qs_j = \sum_{i=1}^{m} x_{i,j} w_i \quad \forall j = 0, 1, \dots, n$.

Step 5: Run a simple linear regression using n available comparables' prices v_j as dependent variable and quality score qs_j as independent variable to obtain simple linear regression: $v_j = b_1(qs_j) + \varepsilon_j$

Step 6: Calculate Gower Similarity indexes between the subject (*i.e.*, Building 0) to all comparables to produce the Gower Similarity matrix, and use PAM to form clustering. Then, use the silhouette index to decide how many clusters (or a simple cut-off points) to select the top q out of n comparables according to their similarities.

Step 7: Use the Grid Adjustment technique to produce the range-estimates for fair market price of the subject using the QRVE technique by comparing it with comparable j, namely:

$$\hat{v}_{0,j} = v_j + b_1(qs_0 - qs_j) \quad \forall j = 1, \dots, q \quad [13]$$

Step 8: Use the Gower Similarity $GS(0,j)$ as weights of comparables & equation [13] to produce a single point estimate of fair market price as:

$$\hat{v}_0 = \frac{\sum_{j=1}^{q} GS(0,j)\hat{v}_{0,j}}{\sum_{j=1}^{q} GS(0,j)} \qquad [14]$$

There are several benefits of proposing the above approach:

- The first step provides flexibility to valuer to select all characteristics that the valuer thinks impacts the value (fair market price). We can then use a more robust Statistical/Analytical technique to reduce the characteristics to a desire numbers – preventing the bias in the selection.
- The weights of the characteristics are no longer subjective, but it follows the principle of Quality Point as suggested by Zaric.
- In addition to the fact that it can be used as the weight for comparables, Gower Similarity between subject 0 and all comparables j, *i.e.*, $GS(0,j)$, provides a scientific foundation that gives more detailed picture of how similar each comparable to the subject. Gower Similarity is chosen since it is flexible to handle various scales (ratio/interval, ordinal, & nominal).

Next, we provide two numerical examples of our proposed approaches with detailed calculation steps. In addition, we also offered two R-functions as in the Appendix.

Some Data & Numerical Examples

To illustrate our proposed methodology, we consider the following two examples. To explain the process, we use a combination of R package and Spreadsheet.

Example 1 – Winanda [33]

In this example, the dependent variable is the price of the property/house, and all independent variables are ordinal scale. There are 8 independent variables, namely:

- Area = house & land sizes of the property,
- Cbd = distance of property to central business district of Sidoarjo (city in East Java)
- Lapindo = distance to Sidoarjo mud flow
- Location = location of the property/house with respect to public transportation
- Safety = safety feature of the property/house
- Water = clean water system provided for the property/house
- Design = additional design features offered for the property/house
- Drainage = drainage system of the property/house

Furthermore, there are 19 comparable properties/houses. The subject property is property #15, and the developer offered it for IDR 220,220,000. For completeness, we put the data as in Table 1 below.

Table 1. Property Valuation for Housing Type 36 – 45 in Sidoarjo

House	Area	Cbd	Lapindo	Location	Safety	Water	Design	Drainage	Price
1	1	4	2	2	1	4	1	1	IDR 79,000,000
2	2	4	2	2	1	4	3	1	IDR 90,000,000
3	1	4	1	1	2	1	2	1	IDR 93,500,000

House	Area	Cbd	Lapindo	Location	Safety	Water	Design	Drainage	Price
4	1	4	1	1	2	1	2	1	IDR 104,500,000
5	3	4	1	1	2	1	4	1	IDR 157,500,000
6	1	4	2	2	3	4	1	2	IDR 196,000,000
7	1	4	2	2	3	4	3	2	IDR 216,000,000
8	1	4	2	2	4	4	1	1	IDR 191,859,000
9	4	4	2	2	4	4	3	1	IDR 240,075,000
10	1	4	2	1	1	4	1	2	IDR 122,567,450
11	2	4	2	1	1	4	1	2	IDR 144,650,179
12	2	4	2	3	3	3	4	1	IDR 209,000,000
13	1	4	2	3	3	3	4	1	IDR 200,000,000
14	1	4	2	3	4	2	3	1	IDR 191,000,000
16	1	3	3	2	1	2	2	2	IDR 136,800,000
17	2	3	3	2	1	2	4	2	IDR 168,750,000
18	2	2	4	4	4	4	4	2	IDR 355,000,000
19	1	1	4	2	1	3	1	1	IDR 150,000,000
20	2	1	4	2	1	3	3	1	IDR 185,000,000
15	1	4	2	3	4	2	1	1	IDR 220,220,000

Step 1: Running the R-package "leaps" with "*harga*" as the dependent variable and all 8 independent variables produces the results in Table 2. Given that all Adjusted R^2, Mallow's C_p, and BIC are identical using 7 independent variables and the independent variables are: *luas*, *cbd*, *lapindo*, *keamanan*, *air*, *desain*, and *drainase*. These are different from the result obtained by Winanda (2013) *via* Multiple (Linear) Regression Analysis (read the signficiant of these variables in Table 3).

Table 2. The result of Step 1 running R-package "leaps" with "harga" as the dependent variable

# of Variables	Area	Cbd	Lapindo	Location	Safety	Water	Design	Drainage
1				*				
2			*		*			
3			*		*		*	
4		*			*		*	*
5	*	*		*	*			*
6	*	*			*	*	*	*
7	*	*	*		*	*	*	*
8	*	*	*	*	*	*	*	*

Criteria	# of Var.
Adjusted R^2	7
Mallow's C_p	7
BIC	7

Step 2: We solve the MLR for 7 independent variables above using OLS to produce the result in Table 3.

Step 3: Given that there is some negative coefficient (for C*bd*) from the result of OLS, we use the equation [3b] to get the weights.

Step 4: Using the weights that are produced from Step 3, we can calculate the quality score for all 19 comparables j as well as the subject property #15 to produce the following result in Table 4.

Table 3. The result of Step 2 running MLR on 5 independent variables & Step 3

Regression Statistics	
Multiple R	0.9983
R Square	0.9965
Adjusted R Square	0.9115
Standard Error	13,411,029
Observations	19

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	6.22219E+17	8.88884E+16	494.2205273	7.53846E-13
Residual	12	2.15827E+15	1.79856E+14		
Total	19	6.24377E+17			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Weight*
Area	10,689,567	4,472,651	2.3900	0.0341	16.62%
Cbd	-20,217,166	3,267,721	-6.1869	0.0000	-31.43%
Lapindo	8,800,404	4,697,296	1.8735	0.0856	13.68%
Safety	39,750,383	3,013,513	13.1907	0.0000	61.80%
Water	7,450,858	3,824,303	1.9483	0.0752	11.58%
Design	13,212,008	3,664,378	3.6055	0.0036	20.54%
Drainage	44,105,237	7,447,637	5.9220	0.0001	68.57%

Table 4. Quality Score for all comparables and subject

House	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
qs	1.155	1.732	1.494	1.494	2.237	3.077	3.487	3.009	3.918	1.841

House	**11**	**12**	**13**	**14**	**16**	**17**	**18**	**19**	**20**	**15**
qs	2.007	3.058	2.891	3.188	2.266	2.843	5.379	2.256	2.833	2.777

Step 5: Running a simple linear regression using qs as independent variable produce $b_1 = 64{,}322{,}403$. Therefore, a direct point estimate using QRVE with $qs = 2.7773$ produces the value IDR 178,642,202.

Step 6: Calculating the Gower Similarity (with Podani's extension) of Rumah #15 with comparables using R-package "FD" produces the following index showed in Table 5.

Table 5. Gower Similarity between subject (Rumah #15) and all 19 comparables

House	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
GS Index	0.625	0.375	0.375	0.375	0.250	0.500	0.375	0.750	0.500	0.500

House	**11**	**12**	**13**	**14**	**16**	**17**	**18**	**19**	**20**
GS Index	0.375	0.500	0.625	0.875	0.250	0.125	0.125	0.375	0.125

Notice that if we choose the top 2 properties to be compared, Gower Similarity suggested the use of comparable #8 (which is the same as Winanda) and comparable #14 (different from Winanda, which suggested comparable #12). Therefore, readers can judge our proposal using Gower Similarity compared with the original research article [33] by examining the table below:

Table 6. Comparisons between GS index *vs.* Winanda (2013)

House	Area	Cbd	Lapindo	Safety	Water	Design	Drainage	Price
8	1	4	2	4	4	1	1	IDR 191,859,000
12	2	4	2	3	3	4	1	IDR 209,000,000
14	1	4	2	4	2	3	1	IDR 191,000,000
15	1	4	2	4	2	1	1	

Step 7 & 8: Using grid adjustment technique using comparables #8 & #14 and Gower Similarity index as the weight, we can calculate the fair market value for the subject (rumah #15) in Table 7.

Table 7. Using Gower Similarity index as the weight in Grid Adjustment

House	Price	qs	Gower Similarity	Estimated Fair Market Value
8	IDR 191,859,000	3.0090	0.7500	IDR 176,957,284
14	IDR 191,000,000	3.1881	0.8750	IDR 164,575,985
15	IDR 178,642,202	2.7773		IDR 170,290,431

Notice that using just the top 2 comparables, our estimate produces much lower fair market value of IDR 170,290,431 $= \left(\frac{0.750\times 191{,}859{,}000+0.875\times 191{,}000{,}000}{0.750+0.875}\right)$. This demonstrates that the price for the subject (IDR 220,220,00) is relatively high compared to the comparables. We suspect that a missing factor in this exercise is the time value of money adjustment in the comparables. Nonetheless, this exercise illustrates how to use our proposed modification to QRVE/WQS/QP technique.

Example 2 – Shetty *et.al.* [34]

Shetty et al. (2020) recently wrote a comparison study between multiple linear regression and traditional valuation methods (such as land & building method, rental income approach, composite rate method, & detail estimation) for building in India. They found out the variations of using multiple linear regression range from 4.90% – 22.17%. We also use their data to illustrate our proposed approach – how to deal with type ratio/interval scale data using QRVE/WQS/QP. Again, for completeness, Table 8 contains the original data from Shetty et al. (2020). The dependent variable Value is INR (Indian Rupee). The first 5 independent variables (Area, Floors, Age, Rooms, & Parking) are ratio/interval scales, and the remaining 4 independent variables (Shape, Location, Access, & Near) are ordinal scales.

Table 8. The original data with 5 Ratio & *4 Ordinal* scales (Value is in INR)

Building	Area	Floors	Age	Rooms	Parking	*Shape*	*Location*	*Access*	*Near*	Value
P01	931	1	4	2	1	3	3	3	1	1,303,000
P02	975	1	6	1	0	2	3	3	1	1,779,375

Building	Area	Floors	Age	Rooms	Parking	*Shape*	*Location*	*Access*	*Near*	Value
P03	1135	1	3	2	0	2	3	2	1	2,071,375
P04	1140	1	10	2	1	1	3	2	1	2,125,750
P05	1365	1	4	2	1	1	2	2	1	2,661,750
P06	1443	2	14	3	1	2	2	1	1	2,743,000
P07	1453	2	1	4	1	2	2	1	2	2,853,265
P08	1545	1	11	3	1	1	2	2	2	3,167,250
P09	1776	2	8	3	1	1	2	1	1	3,263,000
P10	1780	2	8	3	2	2	2	1	1	3,711,300
P11	1960	2	7	3	2	2	1	1	1	4,116,000
P12	2185	2	13	4	2	2	2	2	3	4,643,125
P13	2350	2	4	4	2	1	1	1	1	5,052,500
P14	2485	2	2	5	2	1	1	1	2	5,380,025
P15	2680	2	1.5	5	3	1	1	1	2	5,963,000
P16	2725	2	3	5	3	1	1	1	1	6,144,875
P17	2880	2	5	6	3	1	1	2	1	6,580,800
P18	2950	3	6	6	3	1	1	1	3	6,873,500
P19	3125	3	4	7	3	2	1	1	2	7,296,875
P20	3350	3	2	7	3	1	1	1	3	7,872,500
P00	2545.42	2	1	5	2	2	2	1	2	

By using R-package "leap" to analyze the data [35], we use 2 independent variables to illustrate (since using 1 variable is not very interesting, even though Statistically it might be equivalent or even better).

Table 9. The result of Step 1 running R-package "leaps" on "Value" as independent variable

Criteria	# of Var.
Adjusted R^2	5
Mallow's C_p	2
BIC	2

# of Variables	Area	Floors	Age	Rooms	Parking	Shape	Location	Access	Near
1	*								
2	*						*		
3	*			*			*		
4	*	*		*			*		
5	*			*		*	*	*	
6	*			*		*	*	*	*
7	*		*	*		*	*	*	*
8	*		*	*	*	*	*	*	*
9	*	*	*	*	*	*	*	*	*

Notice that the two independent variables are: *Area* (which is of type Ratio/Interval scale) and *Location* (which is of type Ordinal scale). Running an OLS for: $\text{Value} = c_1\text{Area} + c_2\text{Access} + \varepsilon$ and produces the values of $c_1 = 2347.49$ with $t_{\text{Stat}} = 81.38$ and $c_2 = -246{,}283.30$ with $t_{\text{Stat}} = -7.60$ (both significant).

Again, notice that one coefficient is a negative. So, we use the range to normalize and produce weights for both characteristics: $w_1 = \frac{2347.49}{(2347.49+246283.30)} = 0.94\%$ (for Area) and $w_2 = \frac{-246283.30}{(2347.49+246283.30)} = -99.06\%$ (for Location). With these weights, the quality score for all comparables & subject are given as:

Table 10. Quality Score for all comparables and subject (P00)

Building	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10
qs	5.819	6.234	7.745	7.792	10.907	11.643	11.738	12.606	14.787	14.825

Building	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P00
qs	17.515	18.649	21.197	22.472	24.313	24.738	26.201	26.862	28.515	30.639	22.052

Running simple linear regression using OLS produces: $\text{Value} = 248630.79 \times \text{qs}$ with $t_{\text{Stat}} = 114.86$. A point estimate using quality score for P00, produce a fair market estimate, *i.e.*, $\text{Value}_{00} = \text{INR } 5{,}482{,}777$. A grid adjustment result using Gower Similarity index with $GS(00, j) \geq 0.70$ [36][37] and PAM clustering [38] is in Table 11. Both produce fair market estimate of INR 5,453,940 and INR 5,353,737 respectively for the subject.

Table 11. Grid Adjustment using ($GS(00, j) \geq 0.70$) and PAM clustering with 3 Cluster

Cluster	2	2	2	2	2	2	3	3	3	2
Building	P06	P07	P09	P10	P11	P12	P14	P15	P19	P00
GS (00, *j*)	0.6531	0.8943	0.6085	0.7569	0.6626	0.6401	0.7665	0.7303	0.7070	
Rank (GS(00, *j*))	7	1	10	3	6	8	2	4	5	
qs	11.6432	11.7376	14.7873	14.8250	17.5151	18.6489	22.4720	24.3131	28.5146	22.0519
Value	2,743,000	2,853,265	3,263,000	3,711,300	4,116,000	4,643,125	5,380,025	5,963,000	7,296,875	
Using GS(00, j) >= 0.70										
Adjustment		2,564,443		1,796,814			-104,448	-562,208	-1,606,841	
Estimated Value of P00		5,417,708		5,508,114			5,275,577	5,400,792	5,690,034	5,453,940
Using PAM with 3 clusters										
Adjustment	2,587,918	2,564,443	1,806,204	1,796,814	1,127,983	846,082				
Estimated Value of P00	5,330,918	5,417,708	5,069,204	5,508,114	5,243,983	5,489,207				5,353,737

Interestingly, PAM clustering with 3 clusters (with reasonably good Silhouette Width = 0.40) produces P00 is in the same cluster as P06, P07, P09, P10, P11, & P12 (*i.e.*, in Cluster #2). However, using the similarity cut-off $GS(00, j) \geq 0.70$ produces a comparison with P07, P14, P10, P15, & P19. It is also worth pointing out that Shetty's estimate using MRA is INR 5,652,770. Our modified QRVE proposal produces a very reasonable estimate.

Conclusion & Further Research

Generally, appraisal implements the market approach to estimate the fair price for the subject building. The fundamental concept for this approach is the estimated building value compared with similar and comparable buildings. From the case study, we have provided a more rigorous analytical foundation for the QRVE/WQS/QP technique (a simple technique that is very popular in Indonesia and Southeast Asia). Following Zaric's note, we have provided an option that optimizes the weights for characteristics that linearize the quality score. Furthermore, we propose using Gower Similarity to select the compared building and put weights on comparables in the grid adjustment technique. When the value of Gower similarity of the compared building is closer to one, the compared building is more similar to the subjected building. In the first numerical example, the Gower similarity selected the two most appropriate buildings among the 19 compared buildings to estimate the fair value of the subject building. For the second numerical example, five compared buildings among 20 buildings are selected to estimate the fair value of the subjected building. Gower similarity distance is a reasonable method to choose the most appropriate buildings for estimating the building value accountably. The overall approach remains simple and can be done using an open-source R package and spreadsheet, as illustrated with two numerical examples. Since this study only considered the physical and environmental issues in estimating the building value, this practical modified QRVE technique can now be extended to cover new areas for valuation consideration, such as social value. Nowadays, green building concepts (one of the factors of social value in environmental aspects) are more appropriate to be judged using an ordinal data scale; see: Agustin & Soewandi (2022 – to appear) could be captured. A further technique involving non-parametric statistics (e.g., Rank Regression) may also be an exciting area of further research. Of course, when the weights of attributes (characteristics or independent variables) are expert judgment, the use of MCDA techniques such as AHP, Electre, Promethee, Topsis, etc. are also an exciting topic to explore (Fischer (2003) started this subject). We believe combining them with non-parametric Statistics could yield some interesting findings. Of course, the practicality should always be considered.

References

1. Kim, Y., Choi, S., and Yong Yi, M. (2020), "Applying Comparable Sales Method to the Automated Estimation of Real Estate Prices", *Sustainability* 12, no. 14: 5679. https://doi.org/10.3390/su12145679
2. Tomić, H., Ivić, S. M., Roić, M., Šiško, J. (2021), "Developing an efficient property valuation system using the LADM valuation information model: A Croatian case study", *Land Use Policy*, Volume 104, 105368, ISSN 0264-8377, https://doi.org/10.1016/j.landusepol.2021.105368.
3. Deaconu, Adela, Buiga, A., and Tothăzan, H. (2022), "Real Estate Valuation Models Performance in Price Prediction", *International Journal of Strategic Property Management* 26 (2), pp. 86 – 105, https://doi.org/10.3846/ijspm.2022.15962.
4. Lemen, J. (2020), https://ww3.rics.org/uk/en/journals/property-journal/apc-5-valuation-methods.html
5. KPSPI, & MAPPI. (2018), *Kode Etik Penilaian Indonesia dan Standar Penilaian Indonesia*, Edisi VII, Jakarta: MAPPI.
6. Everett, C. H., Dowell, J. M., Bresnahan, C. A., Felthouse, W. W., Henry, E. E., Hutchinson, R. M., McNair, J. M., Meyer, H. F., Nisley, F., Rickert, G. E., Davis, W. D., Stahr, R. D., Tubbs W. A., Wilkinson, J. M., Easley, C. H., Miles, J. A., Garrett, H. J., Green, J. C., Heid, L. V., Orbaker, C. L., Scott, H. A., Thorkelson, G. F., Davis, W. D. (1966), "Professional Techniques Recommended in the Use of the Market Data Approach", Journal of ASFMRA 30 (1), pp. 108-111. https://www.jstor.org/stable/43757610
7. Hartoyo (2003), "Quality Rating Value Estimation – Suatu Teknik Penilaian Properti", *Jurnal Survey dan Penilaian Properti*, pp. 71 – 81. https://www.scribd.com/document/597999298/Quality-rating-value-estimation-Hartoyo
8. Surat Edaran Direktur Jenderal Pajak Nomor: SE-54/PJ/2016

9. Sudibyanung & Dewi, A. R. (2019), *Modul Praktik Penilaian Tanah*, Program Studi Diploma IV Pertanahan, Kementrian Agraria dan Tata Ruang/BPN, Sekolah Tinggi Pertanahan Nasional, Yogyakarta, Indonesia.
10. Isaac, D., O'Leary, J. (2012), *Property Valuation Principles,* Macmillan International Higher Education and Red Globe Press
11. Munneke, H. J., & Almy, R. (2016). Market Analysis for Real Estate: Concepts and Applications in Valuation and Highest and Best Use. Chicago: The Appraisal Institute.
12. Cooper (2022), *The Future of Valuation Profession*, AREA Press Release No. 553/2022: August 04, 2022. https://board.postjung.com/1415167
13. Ratcliff, Richard U (1972), *Valuation for Real Estate Decisions*, Democrat Press.
14. Sriboonjit, J. & Rattanaprichavej, N. (2013) "Factor's Weight of Importance in the Weighted Quality Score (WQS) Technique: CBD's Serviced Apartment", *International Journal of Information Business and Management* 5 (4). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2403736
15. BUSI 330 – Foundations of Real Estate Appraisal, Sauder Business School, University of British Columbia.https://professional.sauder.ubc.ca/re_creditprogram/course_resources/courses/content/330/qp_info.cfm
16. BUSI 344 – Statistical and Computer Applications in Valuation, Sauder Business School, University of British Columbia. https://professional.sauder.ubc.ca/re_creditprogram/course_resources/courses/content/344/qp_sample.pdf
17. Coldwell, P.F., Cannaday, R.E. and Wu, C. (1983), "The Analytical Foundations of Adjustment Grid Methods," *Real Estate Economics*, Vol. 11(1): pp. 11 – 29. https://doi.org/10.1111/1540-6229.00277.
18. Vandell, Kerry D. (1991), "Optimal Comparable Selection and Weighting in Real Property Valuation," *Real Estate Economics*, Vol. 19, pp. 213 – 239. https://doi.org/10.1111/1540-6229.00550.
19. Guijarro, Francisco (2021), "A Mean-Variance Optimization Approach for Residential Real Estate Valuation," *Real Estate Management and Valuation*, 29(3), pp. 13 – 28. https://doi.org/10.2478/remav-2021-0018.
20. Isakson, Hans R. (1986), "The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Methods," *Real Estate Economics*, American Real Estate and Urban Economics Association, Vol. 14(2), pp. 274 – 286. https://doi.org/10.1111/1540-6229.00387.
21. Isakson, Hans R. (1988), "Valuation Analysis of Commercial Real Estate Using the Nearest Neighbors Appraisal Technique," *Growth and Change*, Vol. 19, pp. 11 – 24. https://doi.org/10.1111/j.1468-2257.1988.tb00466.x.
22. Cajias, M., Freudenreich, P., Freudenreich, A. *et al.* (2020), "Liquidity and Prices: a Cluster Analysis of the German Residential Real Estate Market," *Journal of Business Econonomics*, Vol. 90, pp. 1021 – 1056. https://doi.org/10.1007/s11573-020-00990-2.
23. Calka, Beata (2019), "Estimating Residential Property Values on the Basis of Clustering and Geostatistics," *Geosciences*, Vol. 9 No. 3 143, https://doi.org/10.3390/geosciences9030143.
24. Rahim, F. N. F. & Razali, M. F. (2018), "Comparing the Performance of Similarity Measures in Sales Comparison Method," *Proceedings of the ASEAN Post Graduate Conference (APGC) 2018*, University of Malaya, Kuala Lumpur, 15 November 2018.
25. Miller, Alan J. (1984), "Selection of Subsets of Regression Variables," *Journal of the Royal Statistical Society, Series A (General)*, *147*(3), pp. 389 – 425. https://doi.org/10.2307/2981576.
26. Das, A. & Kempe, D. (2008), "Algorithms for Subset Selection in Linear Regression," In *Proceedings of the Fortieth Annual ACM symposium on Theory of computing (STOC '08)*,

Association for Computing Machinery, New York, NY, USA, pp. 45 – 54. https://doi.org/10.1145/1374376.1374384.

27. Park, Y.W., Klabjan, D. (2020), "Subset Selection for Multiple Linear Regression via Optimization," *Journal of Global Optimization*, Vol. 77, pp. 543 – 574. https://doi.org/10.1007/s10898-020-00876-1.
28. Zhu, J., Wen, C., Zhu, J., Zhang, H., and Wang, X (2020), "A Polynomial Algorithm for Best-Subset Selection Problem," *Proceedings of the National Academuy of Sciences (PNAS) of the United States of America* Vol. 117 No. 52, pp. 33117–33123. https://doi.org/10.1073/pnas.2014241117.
29. Furnival, G. M. & Wilson, R. W. (1974), "Regressions by Leaps and Bounds," *Technometrics*, 16:4, pp. 499 – 511. https://doi.org/10.2307/1267601.
30. Gower, J. C. (1971), "A General Coefficient of Similarity and Some of Its Properties," Biometrics, Vol. 27(4), pp. 857 – 871. https://doi.org/10.2307/2528823.
31. Podani, J. (1999), "Extending Gower's General Coefficient of Similarity to Ordinal Characters," Taxon, Vol 48(2), pp. 331 – 340. https://doi.org/10.2307/1224438
32. Kaufman, L. & Rousseeuw, P. (1990, 2005), "Finding Groups in Data: An Introduction to Cluster Analysis," John Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316801
33. Winanda, L. A. R. (2013), "Model Penilaian Harga Jual Rumah Kelas Menengah (Tipe 36-45) dengan Analisis Regresi (Studi Kasus Di Wilayah Kabupaten Sidoarjo)," *Jurnal PROKONS Politeknik Negeri Malang*, Vol. 7 No. 2, pp. 155 – 164. http://dx.doi.org/10.33795/prokons.v7i2.48
34. Shetty, D. V., Rao, B. P., Prakash, C., Vaibhava, S. (2020), "Multiple Regression Analysis to Predict the Value of a Residential Building and to Compare with the Conventional Method Values," *Journal of Physics: Conference Series*, 1706. 012118. https://10.1088/1742-6596/1706/1/012118.
35. Lumley, Thomas based on Fortran code by Alan Miller (2020), "leaps: Regression Subset Selection", *R package*, version 3.1, https://CRAN.R-project.org/package=leaps
36. Laliberté, E., and P. Legendre (2010), "A Distance-Based Framework for Measuring Functional Diversity from Multiple Traits", *Ecology*, Vol. 91, pp. 299 – 305.
37. Laliberté, E., Legendre, P., and B. Shipley. (2014), "FD: measuring functional diversity from multiple traits and other tools for functional ecology", *R package*, version 1.0 – 12.
38. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2021), "cluster: Cluster Analysis Basics and Extensions", *R package*, version 2.1.2.

Appendix – R Functions to calculate modified QRVE

```
############################################################################
# Function to get mode, multiple values will be printed out
# https://www.statology.org/mode-in-r/
find_mode <- function(x) {
  u <- unique(x)
  tab <- tabulate(match(x, u))
  u[tab == max(tab)]
}
############################################################################


############################################################################
# Function to produce the best weight for qs with reduced criteria
# Written by Karina Agustin, Susan Widjojo, Hani Soewandi
# criteria = 1 (for Adj R2), = 2 (for Cp), = 3 (for BIC)
# Output_Type = 1 (for weight), = 2 (for quality score),
# Output_Type = 3 (for QP coefficient), = 4 (for estimate of the subject)
ksh_qrve1 <- function(Input, criteria = 0, Output_Type){
   n <- dim(Input)[1]
   m <- dim(Input)[2]
   X <- Input[1:n-1, 1:m]
   X_all <- Input
   V <- Input[1:n-1, m]
   p00 <- Input[n, 1:m-1]

   models <- leaps::regsubsets(X[1:n-1, m]~., data = X[1:n-1, 1:m-1], nvmax =
      m-1, intercept=FALSE)
   res.sum <- summary(models)
   subset <- data.frame(
               Adj.R2 = which.max(res.sum$adjr2),
               Cp = which.min(res.sum$cp),
               BIC = which.min(res.sum$bic)
             )

   if (criteria == 1) {
      m_best <- unlist(subset)[1]
      } else if (criteria == 2) {
      m_best <- unlist(subset)[2]
      } else if (criteria == 3) {
      m_best <- unlist(subset)[3]
      } else {
      m_best <- max(find_mode(unlist(subset)))
   }

   t_weight <- coef(models, m_best)

   if (min(t_weight) > 0) {
       weight <- t_weight/sum(t_weight)
   } else {
       weight <- t_weight/(max(t_weight)-min(t_weight))
   }

   X_m <- X_all[ , names(X_all) %in% names(weight)]
   qs <- as.matrix(X_m) %*% weight
```

```
   qp_model <- lm(V ~ qs[1:n-1]+0)
   qp_b1 <- coef(qp_model)
   point_V0 <- qp_b1 * qs[n]

   if (Output_Type == 1) {
      answer <- weight
      } else if (Output_Type == 2) {
      answer <- qs
      } else if (Output_Type == 3) {
      answer <- qp_b1
      } else {
      answer <- point_V0
   }

   return(answer)
}
##############################################################################


##############################################################################
# This function produce estimate using Gower Similarity
# Written by Karina Agustin, Susan Widjojo, Hani Soewandi
# GS = 0 (for PAM), GS = n (for Top n)
# Output_Type = 1 (for set of comparables)
# Output_Type = 2 (for Gower Similarity index)
# Output_Type = 3 (for Adjustment of Comparables)
# Output_Type = 4 (for Estimate of the Subject)
# qs & qp can be obtained from ksh_qrve1
ksh_qrve2 <- function(Input, ord, qs, qp, GS = 0, Output_Type){
   n <- dim(Input)[1]
   m <- dim(Input)[2]
   k <- ord
   X <- Input
   V <- Input[1:n-1, m]
   p00 <- Input[n, 1:m-1]
   y <- colnames(Input)[m]

   X_input <- X[, 1:k-1]
   for (j in k:(m-1)) {
      X_input <- cbind(X_input, factor(X[ , j]))
   }
   colnames(X_input) <- names(X[1:m-1])

   gower_dist <- FD::gowdis(X_input, ord = c("podani"))
   gd_matrix <- as.matrix(gower_dist)
   gs_matrix <- 1 - gd_matrix

   # plot sil_width untuk Gower distance
   sil_width <- c(NA)
   sil_width[1] <- c(0)
      for(i in 2:5){
         pam_fit <- cluster::pam(gower_dist, diss = TRUE, k = i)
         sil_width[i] <- pam_fit$silinfo$avg.width
      }
   best_cluster <- which.max(sil_width)
   pam_fit <- cluster::pam(gower_dist, diss = TRUE, k = best_cluster)
```

```
    comparables <- rep(1, n-1)
    gs_comparables <- gs_matrix[n, 1:n-1]
    gs_rank <- rank(gs_comparables, ties.method="first")
    for (i in 1:(n-1)) {
       if (GS == 0) {
          if (pam_fit$clustering[i] != pam_fit$clustering[n]) {
             comparables[i] = 0
          }
       } else {
          if (gs_rank[i] <= (n - GS - 1)) {
             comparables[i] = 0
          }
       }
    }
    adj_comparables <- qp * (qs[n] - qs[1:n-1])
    t_estimate <- (adj_comparables + V) * comparables
    grid_estimate <-
       sum(t_estimate*gs_comparables)/sum(gs_comparables*comparables)

    if (Output_Type == 1) {
       answer <- comparables
       } else if (Output_Type == 2) {
       answer <- gs_comparables
       } else if (Output_Type == 3) {
       answer <- adj_comparables
       } else {
       answer <- c(t_estimate, grid_estimate)
    }

    return(answer)
}
```

On the Weights for Characteristics and Comparables for Property Valuation using Quality Rating Valuation Estimation

ORIGINALITY REPORT

1%	1%	1%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to HELP UNIVERSITY Student Paper	1%
2	Submitted to Liberty University Student Paper	1%

Exclude quotes Off

Exclude bibliography On

Exclude matches < 1%