An Analysis of Implicit Value of Property Characteristics in Residential Property Prices using a Hedonic Value Approach

Zulkarnain[1*], Tania Arvianti[1]

[1]*Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*

Abstract. Residential property prices increase every year without an equivalent increase in people's purchasing power. To suppress these price fluctuations, the government has implemented the land banking concept, for which quantitative knowledge about factors affecting property prices could optimize its implementation. Therefore, this study aims to identify the implicit value and influence of public facilities and landscapes located around residential properties. The implicit value can be identified by taking a hedonic value approach, which is used to estimate the value of residential properties by considering the distance to public facilities and landscapes in Jakarta. The regression methods used are multiple linear regression (MLR), quantile regression (QR), and support vector regression (SVR). This study found that MLR offered the best accuracy (R^2 = 0.798), followed by QR (R^2 = 0.687), and SVR (R^2 = 0.563). The study also found that the proximity of residential properties to Mass Rapid Transit (MRT) stations, green open spaces, reservoir lakes, health service facilities, vocational schools, marketplaces, high schools, and mosques has a significant influence on residential property prices, with MRT stations, health service facilities, marketplaces, and high schools conveying the largest implicit value.

Keywords: Hedonic price method; Implicit value; Ordinary least square; Quantile regression; Support vector regression

1. Introduction

In addition to food (including water) and clothing, shelter is one of the three basic needs of human beings. However, every year in Indonesia, there is an increase in residential property prices that is not followed by an increase in people's purchasing power. In 2020, approximately ten million households did not own private housing, despite members of generation X (born in 1965–1980) having been of a productive age for a relatively long time. Therefore, the issue of residential property ownership in Indonesia is deemed critical.

To suppress fluctuations in residential property prices, the Indonesian government has implemented the concept of land banking, by which the government buys and invests in land for future infrastructure development. To optimize the land banking concept, it is necessary to identify factors that significantly affect changes in land and/or residential property prices. However, in Indonesia, few studies have been conducted to identify these factors using a quantitative approach. One way to quantitatively identify the implicit values of these factors is to take the hedonic price approach to consider residential property

prices.

The hedonic price model has been used in various studies to analyze the implicit marginal value of a particular characteristic. The basis of the hedonic price approach is the consumer demand theory formulated by Lancaster in 1966. Based on the theory, Lancaster showed how consumers see products as a group of characteristics. The characteristics of the product determine whether the consumer prefers to buy a particular product or not. In other words, consumer preferences can be represented as the fulfillment of customer satisfaction based on the characteristics of the purchased products (utility). In Lancaster's theory, consumers receive the characteristics they are aiming for by purchasing products or services with the desired characteristics. When a market price is observed, the price represents several characteristics grouped together, making it impossible to know the specific value of each characteristic (Thomsen, 2021). An approach was later developed by Rosen (1974) to determine the implicit value contained in each characteristic of a product, which is known as the hedonic price model. Rosen's theory involves two steps. First, the hedonic equation is estimated. Then, the price or implicit value of a characteristic is yielded by the derivative of the hedonic equation generated. The hedonic value of an object is found by representing different levels of characteristics, each with their own numeric values, to further constitute the price of the product. The hedonic value approach can be implemented to obtain the implicit marginal value of a characteristic (Herath and Maier, 2010).

Against this backdrop, the current research is carried out to quantitatively identify previously unidentified public amenities and landscapes that are closely related to changes in residential property prices. In Indonesia, especially in Jakarta, such quantitative research has not yet been conducted. This research also compares three regression methods, namely multiple linear regression (MLR), quantile regression (QR), and support vector regression (SVR), to obtain the best model to represent the condition in Jakarta. The variables included in this research are derived from several previous studies adjusted to the existing situation in Jakarta. Therefore, this research aims to generate an exploratory model to analyze the relationship between public amenities or landscapes and residential properties, as well as the influence of these features on changes in residential property prices in the Jakarta area.

2. Methods

In this research, the research topic, problem statement, objective, and research output were first determined, and a literature review regarding the topic and methods used in this study was conducted. Then, the variable selection is made by referring to previous studies adjusted to Jakarta's geographical condition (Hartoyo et al., 2014; McCord et al., 2018). This resulted in the use of the following 19 independent variables: lot area, building area, bedroom, bathroom, closest distance to MRT station, healthcare facilities, marketplaces, green open spaces, middle school, high school, vocational school, reservoir lakes, security facilities, urban forests, Transjakarta bus stops, commuter line or *kereta rel listrik* (KRL) stations, mosques, churches, toll gates, and a dependent variable, residential property prices. Overall, the MLR, QR, and SVR regression methods were used to generate models in this study.

2.1. Data Collecting

The data collected on the residential properties were obtained from a property agent and a bank that provides loans to purchase residential properties, as stated in the profile of data in Table 1. Overall, 495 data were obtained, with 100 data acquired from the bank and 395 from the property agent. However, these data do not contain detailed information about the respondents, which could potentially cause limitations in the findings. In addition

to the data obtained from third parties, open data from official online websites were utilized
to provide information on the location of public amenities and landscapes in DKI Jakarta.

Table 1 Profile of data

Data	Year	Sample	Source
Residential property sales	2021	100	Bank XYZ
Residential property sales	2021	395	Property Agent XYZ

2.2. Data Preprocessing

Next, data preprocessing was conducted to assure that the data were appropriate for
processing. It was conducted by screening for missing values and deleting duplicate data.
Then, the mapping of coordinates was performed using Google API as a basis to calculate
the distance variable between residential properties and other independent variables. The
distance was calculated using Python's Vincenty formula.

2.3. Data Processing

2.3.1. MLR

MLR is the primary method used in a hedonic value approach. The use of this method
is preceded by assumption testing to fulfill several assumptions, such as linearity,
normality, homoscedasticity, and multicollinearity. The assumption testing in this study
resulted in violations of the linearity, normality, and homoscedasticity assumptions. This
led to the need to transform both the independent and dependent variables, which was
carried out by determining the natural log of each variable. These newly transformed data
were then retested for the assumptions, which resulted in the fulfillment of every
assumption. Therefore, the data were appropriate for processing by the MLR method, for
which the mathematical equation is shown in Equation 1 with the objective of minimizing
the mean square error shown in Equation 2. The MLR model was generated by applying the
stepwise method, with the independent variables inserted into the model one by one based
on the significance of each toward the predicted value in the model. The stepwise method
was used because it allows the model to be generated only with independent variables,
which significantly affects the target value (Hair, 2014).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \epsilon \tag{1}$$

where β is the coefficient of the variable and X is the independent variable.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}\right)\right) \tag{2}$$

where Y is the observed value, β is the coefficient of the variable, and X is the independent
variable.

2.3.2. QR

To overcome the limitations of MLR, a different type of regression method, QR, was
used. This method divides the target value into several different quantiles, which makes the
method robust to outliers. It also allows the QR to be conducted without having to fulfill any
specific assumptions, unlike the MLR method. For this method, the data processing was
preceded by the data being partitioned into training and testing data, with a ratio of 80:20.
The data partition was performed to allow the model's accuracy to be tested. Next, the
regression model was generated, which yielded several regression equations based on the
quantile levels of the data, as shown in Equation 3. The regression model in QR was created
with the objective of minimizing Equation 4.

$$Q_\tau(y_i)=\beta_0(\tau)+\beta_1(\tau)x_{i1}+\ldots+\beta_p(\tau)x_{ip} \tag{3}$$

where β is the coefficient of the variable, X is the independent variable, and τ is the quantile.

$$\text{MIN } \sum_{i=1}^{n}\rho_\tau\left(y_i-\left(\beta_0(\tau)+\beta_1x_{i1}(\tau)+\ldots+\beta_p(\tau)x_{ip}\right)\right) \tag{4}$$

where β is the coefficient of the variable, X is the independent variable, and τ is the quantile.

To elucidate the characteristics possessed by residential properties, coefficient visualization was conducted for each quantile level. In the end, the model accuracy was tested utilizing pseudo-R^2, which can be assumed to be adjusted R^2 in other regression models. Accuracy testing in QR was conducted to determine which quantile level should be used for further analysis and comparison with the other two regression methods.

2.3.3. SVR

The last regression method used was the SVR algorithm. For this method, the data were partitioned in the ratio of 80:20, with the purpose of testing the trained data to obtain the accuracy of the model. The training data were used to generate the SVR model using the base parameters (kernel = '*linear*'; gamma = '*scale*'; epsilon = 0.1; C = 1). The SVR algorithm was applied based on the objective function shown in Equation 5 with the constraints described in Equation 6.

$$\text{MIN}\frac{1}{2}\left|\left|\mathbf{w}\right|\right|^2+C\sum_{i=1}^{n}\left|\xi_i\right| \tag{5}$$

where w is weights, C is the constant, and ξ is the loss function.

$$\sum_{i=1}^{n}\left|\xi_i\right|\left|y_i-w_ix_i\right|\leq\varepsilon+\left|\xi_i\right| \tag{6}$$

where y is the target value, x is the observed value, w is weights, C is the constant, and ξ is the loss function.

3. Results and Discussion

3.1. Results

The data processing yielded three different regression models, which portray the residential property prices represented by the property characteristics possessed. The results of the three regression models are similar. Due to the result being a regression equation, the positive value of the independent variable means that the longer the distance between residential properties and public amenities, the larger the increase in price. This indicates that residents of Jakarta do not prefer to own residential properties near those independent variables and vice versa. The MLR model in this study indicated that MRT stations, healthcare facilities, marketplaces, and high schools cause an increase in residential property prices as the proximity between the property and amenities gets closer. However, green open spaces, mosques, reservoir lakes, and vocational schools cause a decrease in residential property prices.

In the QR model, several other independent variables were included in the model. Among these variables, the amenity resulting in an increase in residential property prices is Transjakarta bus stops, while urban forest, KRL stations, and churches cause a decrease in residential property prices.

The SVR model included even more independent variables. This model indicated that in addition to the variables included in the MLR model, a closer distance to Transjakarta bus stops and toll gates cause an increase in residential property prices. By contrast, middle school, urban forests, KRL stations, and churches cause a decrease in residential property prices as the proximity between the property and the amenities closes.

Table 2 MLR model results

Model	B	Std. Error	t	Sig.	Collinearity Statistics	
					Tolerance	VIF
(Constant)	17.288	0.200	86.367	0.000		
MRT stations	-0.346	0.033	-10.427	0.000	0.642	1.557
Green open spaces	0.129	0.040	3.202	0.001	0.493	2.030
High schools	-0.059	0.013	-4.449	0.000	0.880	1.136
Healthcare facilities	-0.096	0.022	-4.289	0.000	0.539	1.856
Reservoir lakes	0.092	0.024	3.759	0.000	0.514	1.944
Marketplaces	-0.076	0.024	-3.130	0.002	0.459	2.179
Mosques	0.026	0.010	2.478	0.014	0.701	1.426
Vocational schools	0.084	0.036	2.348	0.019	0.692	1.445

Table 3 0.75 QR model results

Variable	Coefficient	Upper-bound	Lower-bound
MRT stations	-687191229	-872657467	-501724992.1
Healthcare facilities	-586731482.3	168644720.8	1004818244
Marketplaces	-7231872.588	-440361385	454825130.3
Green open spaces	526566559.5	238088474.6	815044644.5
Middle schools	1753704896	790751113.4	2716658678
High schools	-1605481044	-2281164606	-929797481.8
Vocational schools	779726356.7	-967464.988	1560420178
Reservoir lakes	217914679.7	-24152002.2	459981361.5
Security facilities	44869376.4	-249187264	338926016.4
Urban forest	206345207.9	65244038.2	347446377.6
Transjakarta bus stations	-879726264	-1325471457	-433981071.3
KRL stations	229004304.9	-18961512.9	476970122.7
Mosques	109040443.9	-326563939	544644826.7
Churches	314199194.4	144839071.6	483559317.2

Table 4 SVR model results

Variable	Coefficient	Variable	Coefficient
MRT stations	-0.150851162	Security facilities	0.015412216
Healthcare facilities	-0.013639008	Urban forest	0.075937518
Marketplaces	-0.00477147	Transjakarta bus stations	-0.083883245
Green open spaces	0.033674868	KRL stations	0.090278379
Middle schools	0.083736769	Mosques	0.043460916
High schools	-0.089346005	Churches	0.085623188
Vocational schools	0.029273212	Toll gates	-0.083883245
Reservoir lakes	0.052958148		

This study obtained different results from previous studies, in which it was more common for the QR and SVR models to yield higher accuracy than the MLR model. However, in this study, the MLR model provided the highest accuracy, indicated by the highest score of the adjusted R^2, as shown in Table 5 (Wang and Hong, 2014; Yoo et al., 2012).

The moderate accuracy of the SVR model may be caused by the application of the 'linear' kernel type instead of the 'rbf' type, which is mainly used. In general, 'rbf' is able to provide better results because of the nonlinearity it possesses. However, the objective of this research required the coefficients of each independent variable, which can only be acquired using the 'linear' kernel type. The higher accuracy of the MLR than the QR may also have been caused by the minimum amount of data used. The sample size requirements

for linear regression met the rule of thumb set for the model. However, the QR divides the target variables into several quantiles, which require a larger variety of data.

Table 5 Regression models accuracy comparison

Methods	R^2
MLR	0.798
QR (0.75)	0.687
SVR	0.563

3.2. Discussion

Regarding accuracy testing, as the MLR model gave the highest adjusted R^2 value, it was used for a more thorough analysis in this study. As shown in Table 6, the MRT station had the highest implicit value of Rp 2.444.553,64, followed by healthcare facilities, marketplaces, and high schools, consecutively. These four independent variables, therefore, result in an increase in property prices as the distance between these amenities and the property is close. The increase of residential property prices in parallel with the nearing distance between the property and MRT station indicates that residents of DKI Jakarta highly prioritize and prefer transportation accessibility, which causes the price of lots around this particular amenity to fluctuate. In addition, MRT stations are mostly located near the city center, which provides accessibility to other public facilities. Nearly located residential properties to MRT stations reduce the cost of travel which induces higher property value (Berawi et al., 2018; Berawi et al., 2019). Based on the survey conducted by the Communication, Information, and Statistics Division of the DKI Jakarta Government in 2019, 76.83% of Jakarta citizens value MRT and Light Rail Transit (LRT) highly, which supports our hypothesis (Dinas Komunikasi Informasi dan Statistik, 2019).

As well as transportation, a shorter distance to education facilities (nonvocational schools) was found to cause an increase in residential property prices. In the selection process of the high school variable, the data were filtered in a way that resulted in the high schools included in this research having an A for accreditation. The study results suggest that residents of DKI Jakarta highly value high schools with a good reputation. This hypothesis is supported by a survey by the Jakarta Government, which indicated that the most serious and urgent problem in Jakarta's public service sector is the unequal quality of education (Dinas Komunikasi Informasi dan Statistik, 2019).

Table 6 Property characteristics implicit value

Property Characteristics	Constant	Coefficient	LN Inverse	Implicit Value (Rp)*
Mosques	17.288	0.02563357	1.02596494	1.773.688,18
High schools	17.288	-0.0585284	1.06027509	-1.833.003,57
Marketplaces	17.288	-0.0764928	1.07949445	-1.866.230,01
Vocational schools	17.288	0.08394327	1.08756719	1.880.186,16
Reservoir lakes	17.288	0.09164979	1.09598093	1.894.731,83
Healthcare facilities	17.288	-0.0957249	1.10045628	-1.902.468,81
Green open spaces	17.288	0.12948174	1.13823833	1.967.786,42
MRT stations	17.288	-0.346435	1.41401761	-2.444.553,64

*Change in property price as the distance grows by 100 m

Accessibility to healthcare facilities and marketplaces is also deemed important to Jakarta residents. The high value placed on healthcare facilities was expected and is supported by the sustainable development goals set forth by the Ministry of Health, which

aims to promote the health of Indonesian citizens as much as possible. It is also supported by the survey, which stated that 86.44% of Jakarta citizens rank healthcare services as highly important (Dinas Komunikasi Informasi dan Statistik, 2019). The proximity between marketplaces and residential properties is also valued highly by DKI Jakarta citizens, which may be caused by the need to obtain daily goods and services. Based on its analysis of the increase in residential property prices, this study found that the residents of DKI Jakarta highly value mobility and accessibility in their daily lives. These findings align with those from research conducted in Bogor, Indonesia, in which nearing distance to healthcare facilities resulted in an increase in residential property prices. However, the reasons why healthcare facilities produced positive effects on property prices were not elaborated on in the research (Hartoyo et al., 2014).

However, the opposite phenomenon applies to several other variables, namely green open spaces, reservoir lakes, vocational high schools, and mosques. The decrease in residential property prices when the property is located near these variables may be caused by mismatched preferences between the people and amenities, especially the people who own residential properties in a price range similar to the observed data of this study. However, this matter cannot be analyzed thoroughly due to the lack of respondents' profiles in the collected research data. Regarding the reservoir lake variable, the decrease in property prices may be a result of the existence of a high risk in the residential area; the risk of flooding in the neighborhood during the rainy season increases because the residential property is located near reservoirs. Hino and Burke (2021) found that floods would temporarily lower property values as risk-aware buyers responded to floodplain information. Regarding vocational schools, the decrease in property prices may be a result of the school simply not being a main interest of respondents of this research. However, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (The Ministry of Education, Culture, Research, and Technology) and MarkPlus, Inc. (2021) found that 82.05% of the respondents across West, East, and Central Java, North and South Sumatera, North and South Sulawesi, DKI Jakarta, West Kalimantan, and Bali were interested in continuing their secondary education in vocational schools. Therefore, further research regarding this hypothesis should be conducted.

An interesting finding is obtained in this study. Green open spaces and mosques cause a decrease in residential property prices as the distance between these variables and residential properties closes. One reason for the decrease caused by green open spaces may be the uneven quality of amenities in DKI Jakarta. In the data collection stage of this study, the specific condition of green open spaces was unknown, which creates the possibility that the residents were not satisfied with most of the green open spaces in Jakarta, causing a drop in the implicit value of this amenity. In a survey, the Jakarta Government found that a gap still exists between the expectation and actual performance of the government regarding Jakarta public amenities (Dinas Komunikasi Informasi dan Statistik, 2019).

The next unexpected finding was that mosques cause a decrease in residential property prices, despite 84% of DKI Jakarta residents being registered as Muslims. Respondents' characteristics and traits may have an impact on the results of this study. Age and beliefs factors may play a significant role in this study and are assumed to cause variations in the value placed on places of worship. As the data collected in this research do include respondents' characteristics, a thorough analysis regarding this matter cannot be conducted. However, it may be assumed that most of the research respondents in this study are not Muslim or are young adults. This assumption is based on the theory suggested by Hayward and Krause (2015), who found that the religiosity of a person increases as one ages.

4. Conclusions

Analyzing the implicit value of property characteristics using the hedonic price method can be implemented by employing several regression methods, such as MLR, QR, and SVR. This study found that the MLR model gave the best accuracy (adjusted $R^2 = 0.798$) when compared to the other two, followed by QR (adjusted $R^2 = 0.687$) and SVR (adjusted $R^2 = 0.563$), consecutively. Despite the accuracy differences, the three methods yielded similar results regarding the property characteristics that significantly affected residential property prices. An increase in residential property prices was found to be influenced by close proximity between the property and MRT stations, healthcare facilities, marketplaces, and high schools. By contrast, a decrease in price was caused by close proximity between the property and green open spaces, reservoir lakes, vocational schools, and mosques.

The results yielded by the MLR model indicate that the public facilities with the highest implicit value are MRT stations, healthcare facilities, marketplaces, and high schools, respectively. These implicit values imply that residents of DKI Jakarta value these facilities the most. Hence, land banking can be implemented to suppress the residential property price fluctuations in lots located near these four public facilities in particular. However, further research must be conducted to develop and provide validation for the hypotheses of this research.

Acknowledgements

References

Berawi, M.A., Suwartha, N., Kurnia, K., Gunawan., Miraj, P., Berawi, A.R.B., 2018. Forecasting the Land Value around Commuter Rail Stations using Hedonic Price Modeling. *International Journal of Technology*, Volume 9(7), pp. 1329–1337

Berawi, M.A., Suwartha, N., Fathiya Salsabila., Gunawan., Perdana Miraj., Woodhead, R., 2019. Land Value Capture Modeling in Commercial and Office Areas using a Big Data Approach. *International Journal of Technology*, Volume 10(6), pp. 1150–1156

Dinas Komunikasi Informasi dan Statistik, 2019. *Survei Kepuasan Kinerja Pemerintah Provinsi DKI Jakarta* (Satisfaction Survey on the Performance of DKI Jakarta Provincial Government). Available Online at https://statistik.jakarta.go.id/media/2020/05/Buku-Survei-Kepuasan-Kinerja-Pemerintah-Provinsi-DKI-Jakarta-tahun-2019.pdf, Accessed on June 25, 2021

Hair, J.F., 2014*. Multivariate Data Analysis.* 7th Edition. Edinburg: Pearson

Hartoyo, D., Saefuddin D.S., Gunadi, A., 2014. The Model of Property Tax Assessment based on Straight Line Distance and Distance of Travel: Case Study in the West of Bogor City. *Kajian Ekonomi Dan Keuangan,* Volume 18(1), pp. 1–14

Hayward, R.D., Krause, N., 2015. Aging, Social Developmental, and Cultural Factors in Changing Patterns of Religious Involvement Over a 32-Year Period: An Age–Period–Cohort Analysis of 80 Countries. *Journal of Cross-Cultural Psychology*, Volume 46(8), pp. 979–995

Herath, S., Maier, G., 2010. *The Hedonic Price Method in Real Estate and Housing Market Research: A Review of the Literature, SRE-Disc sre-disc-2010_03*. Institute for Multilevel

Governance and Development, Department of Socioeconomics, Vienna University of Economics and Business

Hino, M., Burke, M., 2021. The Effect of Information about Climate Risk on Property Values. *Proceedings of the National Academy of Sciences,* Volume 118(17), pp. 1–9

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi., 2021. Indonesian's Interest in Vocational Schools. Available Online at https://www.kemdikbud.go.id/main/blog/2021/04/minat-masyarakat-terhadap-pendidikan-vokasi-cukup-tinggi, Accessed on June 25, 2021

McCord, M.J., Davis, P.T., Bidanset, P., McCluskey, W., McCord, J., Haran, M., MacIntyre, S., 2018. House Prices and Neighbourhood Amenities: Beyond the Norm? *International Journal of Housing Markets and Analysis*, Volume 11(2), pp. 263–289

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, Volume 82(1), 34–55

Thomsen, M.R., 2021. An Interactive Text for Food and Agricultural Marketing. *Open Educational Resources.* Available Online at https://scholarworks.uark.edu/oer/1

Wang, H., Hong, M., 2014. Study on Residential Hedonic Price Classification Model Based on MDLP Binning and Support Vector Machine. *Information Technology Journal*, Volume 13(18), pp. 2710–2719

Yoo, S., Im., J., Wagner, J.E., 2012. Variable Selection for Hedonic Model using Machine Learning Approaches: A Case Study in Onondaga County, NY. *Landscape and Urban Planning*, Volume 107(3), pp. 293–306