SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Robust Regression on Simple Housing Environmental Performance Measurement

Hafizah Miranda Putri Hrp[1)], Esther Sorta Mauli Nababan[2)*], Tulus[3)]
[1)] Magister of Mathematics, Universitas Sumatera Utara, Indonesia
[2)3)] Department of Mathematics, Universitas Sumatera Utara, Indonesia
[1)] hafizahmiranda@gmail.com, [2)] esther@usu.ac.id, [3)] tulus@usu.ac.id

Submitted :May 1, 2024| **Accepted** : June 3, 2024| **Published** : Jun 6, 2024

Abstract: Robust regression in residential environmental performance measurement can provide significant benefits. This method can help identify and overcome uncertainties in measurement results so that decision making related to environmental protection and management can be done more accurately. This type of research is quantitative with a descriptive approach. The results of this study indicate that the results of the M-estimate robust regression estimation on simple housing environmental performance obtained the M-estimate robust regression equation, namely $\hat{y}$ = 15.562+0.476X1-0.453X2+0.222X3-0.427X4. Based on the p-value test results, it shows that population attributes, house attributes, energy consumption volume and utilities and services have a significant effect on environmental performance. The results of the hypothesis testing of the M-estimation robust regression method on the environmental performance of simple housing obtained that the variables of population attributes and energy consumption volume have a positive and significant effect on environmental performance while the variables of house attributes and utilities and services have a negative and significant effect on environmental performance.

Keywords: Environmental Performance, Population Attributes, House Attributes, Energy Consumption Volume, Utilities and Services

INTRODUCTION

Robust regression is an estimator to analyze data contaminated with outliers. There are several estimates in robust regression that are used to handle outliers and provide stable results in the presence of outliers (Bertsimas et al., 2011). Robust in performance systems generally refers to the ability of the system to continue to maintain environmental performance in the face of disturbances. Suppose someone has to make constant decisions and faces the difficulty of a multidimensional problem with some degree of ambiguity or error in the parameters to be analyzed and some sort of stochastic uncertainty of the process and its environment. Then this engineer must also determine if the proposed solution is robust. This means that the solution is feasible to implement for any scenario of stochastic parameters and uncertainties and that this feasibility remains close to the optimality condition. Then two basic concepts emerge: the fuzziness of the feasibility of the solution and the uncertainty of the objective value of the function (Dubois & Prade, 2015; García & Peña, 2018).

The environmental performance of simple housing development can be defined as the success of housing management in managing the interaction between resident activities and the environment. Measurement of environmental performance can be done through the use of robust regression methods. This method allows for increased robustness to disturbances or uncertainties in environmental data, making the measurement results more reliable and relevant. Using a robust regression approach can better identify impacts on the environment. In addition, robust regression also helps in evaluating the effectiveness of the environmental programs that have been implemented, making it possible to make necessary repairs and improvements (Asiaei & Jusoh, 2017).

The robust regression method has been proven to be effective in dealing with data variations and disturbances in various fields, but its application in the measurement of housing environmental performance is still limited. In this research, a case study will be conducted in Martubung Simple Housing to analyze the parameters of robust regression method in environmental performance measurement. This case study will involve the measurement of resident attributes, house attributes, energy consumption volume, utilities and services to environmental performance (Nababan et al., 2018). By using robust regression methods, it is expected that accurate, reliable measurements can be obtained, and can provide useful information for housing managers in improving the environmental quality of simple housing (Malmqvist & Glaumann, 2006).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The application of robust regression methods in residential environmental performance measurement can provide significant benefits. This method can help identify and overcome uncertainty in measurement results, so that decision making related to environmental protection and management can be done more accurately (Smith et al., 2022). This research is also expected to contribute to the development of science and technology in the field of environmental measurement of housing and provide guidance for further research and development in an effort to improve the performance of the housing environment in general. In addition, the use of robust regression methods can also increase measurement robustness to environmental variations that may occur within a certain period of time. Thus, the measurement of housing environmental performance can remain consistent and reliable, despite changes in environmental conditions (Shameem et al., 2017).

In this study, robust regression, specifically M-estimation, will be used to estimate the performance parameters of residential neighborhoods. By using robust M-estimation, the results of system performance measurement become more stable and reliable, even when there are unusual or inconsistent data. This helps in producing more accurate and trustworthy analysis, thus enabling better decision-making in improving overall system performance. In addition, robust M-estimation can also reduce the risk of misinterpretation due to extreme data, thus strengthening the confidence in the system performance measurement results.

LITERATURE REVIEW

M-estimation (Maximum likelihood type) was introduced by Huber in 1973 and is a simple method both in calculation and theoretically. In principle, M-estimation is an estimate that minimizes an objective function (Huber, 1973, 1981).

$$min_\beta \sum_{i=1}^{n} \rho(e_i) = min_\beta \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{k} X_{ij}\beta_j\right) \tag{1}$$

$$\sum_{i=1}^{n} \rho(u_i) = \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) \tag{2}$$

Where $\rho(u_i\ )$ is a symmetric function of the residuals or a function that contributes each residual to the objective function. In general, a robust scale estimate needs to be estimated and $\hat{\sigma}$ is the scale of the robust estimate. To find $\hat{\sigma}$ in robust regression with M-estimation often used Eq:

$$\hat{\sigma_l} = \frac{median|e_i - median(e_i)|}{0{,}6745} \quad l = 1,2,\dots,n \tag{3}$$

To minimize the objective function ρ the first partial derivative of ρ with respect to β_j where $j = 0,1,\dots,k$, must be equalized to 0. This results in a necessary condition for the minimum. This results in the system of Equations (Huber, 1981):

$$\frac{\vartheta\rho}{\vartheta\beta} = 0 \tag{4}$$

$$\sum_{i=1}^{n} X_{ij}\varphi\left(\frac{e_i}{\hat{\sigma}}\right) = \sum_{i=1}^{n} X_{ij}\,\varphi\left(\frac{y_i - \sum_{i=1}^{n} X_{ij}\,\beta_j}{\hat{\sigma}}\right) = 0 \tag{5}$$

Where $\varphi = \rho'$ and X_{ij} is the i observation on the j parameter and $X_{i0} = 1$ is defined as the weighting function:

$$W(u_i) = \frac{\varphi(\frac{e_i}{\hat{\sigma}})}{\left(\frac{e_i}{\hat{\sigma}}\right)} \tag{6}$$

According to (Healy, 2005) the objective function is used to obtain the value of the weighting function in robust regression. The weighting function in M-estimation robust regression that is often used is the Huber weighting function. Huber's weighting function criteria are as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$W(u_i) = \begin{Bmatrix} 1, \\ \frac{c}{u_1}, \end{Bmatrix}, |u_i| \le c \qquad (7)$$

$$|u_i| > c$$

u_i is the i residual, while the value of c is expressed as the tuning constant. If $\alpha = 5\%$ is taken, then Huber's M-estimation will effectively use a value of $c = 1.345$ (Healy, 2005). According to (Montgomery et al., 2021) parameter estimation in M-estimation robust regression is done by Iteratively Reweighted Least Square (IRLS) estimation. This iteration requires an iteration process where the value of W_i will change its value in each iteration. To use IRLS, it is assumed that $\widehat{\beta_0}$ exists and $\widehat{\sigma_l}$ is the scale of the robust estimate. Then $p = 1 + k$ is written in the system Eq:

$$\sum_{i=1}^{n} X_{ij} W_i \left(y_i - \sum_{j=0}^{k} X_{ij}\beta_j \right) = 0 \qquad (8)$$

W is an $n \times n$ diagonal matrix with diagonal elements $W_{1,1}, W_{1,2}, \dots, W_{1,n}$. So, the robust regression parameter estimation with IRLS, for $l + 1$ iteration is:

$$\hat{\beta} = \left(X^T W^1 X\right)^{-1} X^T W_y^1, \quad l = 0,1,2,3, \dots. \qquad (9)$$

METHOD

This method aims to gain a deeper understanding of performance measurement using robust regression analysis on Martubung housing. Robust regression analysis was chosen to overcome problems that often arise, which can affect the accuracy of ordinary regression models. With this method, it is expected to produce a regression model that is more stable and reliable in describing the relationship between the variables that affect the performance of the housing (Cabana et al., 2020). In this study, environmental measurement methods were used to measure parameters relevant to the environmental performance of housing. The measurement methods include measuring resident attributes, housing attributes, energy consumption volume, utilities and services. The data analysis methods used in this study are Validity Test, Respondent Characteristics, Descriptive Analysis, Correlation Analysis, Parameter Estimation Using M-Estimation Robust Regression, Coefficient of Determination, and Partial Test.

this section, each researcher is expected to be able to make the most recent contribution related to the solution to the existing problems. Researchers can also use images, diagrams, and flowcharts to explain the solutions to these problems.

RESULT

The results are based on surveys and questionnaires of residents in each block. With a sample size of 144 randomly selected residents involved in the research process.

Validity Test

Through the validity test of 144 respondents, the results show that all are valid where the value of r-count > 0.1637.

Respondent Characteristics

The characteristics of respondents in this study include occupation, education, monthly income, and home ownership status, as shown in Figures 1-4.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

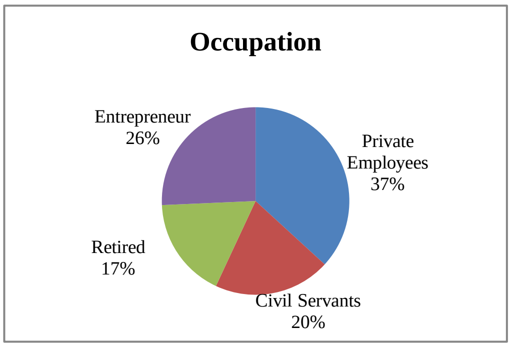

Figure 1. Characteristics of Respondents Based on Occupation

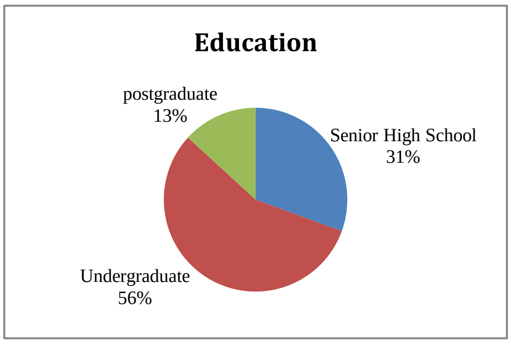

Figure 2. Characteristics of Respondents Based on Education

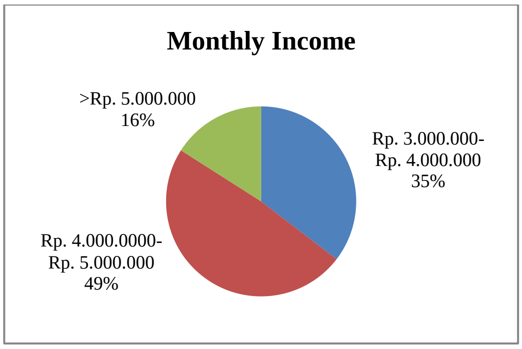

Figure 3. Characteristics of Respondents Based on Monthly Income

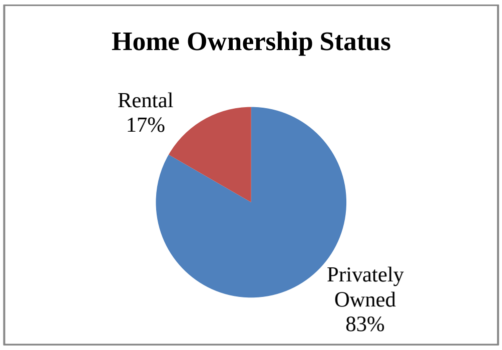

Figure 4. Characteristics of Respondents Based on Home Ownership Status

Descriptive Analysis

The standard deviation value illustrates the extent of distribution of a group of data relative to the average, or in other words, it provides a description of the heterogeneity of the data group, as described in the table below.

Table 1. Descriptive Statistics Results

Variable	Obs	Mean	Std. Dev.	Min	Max

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Y	144	16.25694	4.625479	6	25
$X1$	144	16.75	4.555025	7	25
$X2$	144	12.47222	1.490973	6	15
$X3$	144	13.47917	4.145886	5	20
$X4$	144	10.84028	1.589578	5	14

Correlation Analysis

If the significance value is $p < 0.05$, then there is a relationship between population attributes, house attributes, energy consumption volume, utilities and services, and environmental performance, as described in the table below.

Table 2. Pearson Correlation Test Results

	Y	$X1$	$X2$	$X3$	$X4$
Y	1.000				
$X1$	0.5640 0.0000	1.0000			
$X2$	-0.1789 0.0319	-0.0010 0.9902	1.0000		
$X3$	0.4100 0.0000	0.4278 0.0000	0.0197 0.8147	1.0000	
$X4$	-0.2312 0.0053	-0.0654 0.4359	0.2474 0.0028	-0.0891 0.2882	1.0000

Parameter Estimation Using M-Estimation Robust Regression

Parameter estimation using M-estimation robust regression will produce parameters that are resistant to outliers. The resulting M-estimation robust regression model is $\hat{Y} = 15.562 + 0.476X_1 - 0.453X_2 + 0.222X_3 - 0.427X_4$, as described in the table below.

Table 3. M-estimation Robust Regression Results

Y	Coef.	Robust Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]	
$X1$	0.4763178	0.0715992	6.65	0.000	0.3347535	0.6178821
$X2$	−0.4532653	0.1555867	−2.91	0.004	−0.7608879	−0.1456428
$X3$	0.2221665	0.0847478	2.62	0.010	0.054605	0.389728
$X4$	−0.4266255	0.1846968	−2.31	0.022	−0.7918039	−0.0614471
_cons	15.56197	3.09458	5.03	0.000	9.443433	21.6805

Coefficient of Determination

Based on the results of the coefficient of determination (R^2) test, the Adjusted R Square value is 0.406. This indicates that population attributes, house attributes, energy consumption volume, and utilities and services explain 40.6% of the variance in environmental performance, while the remaining 59.4% is influenced by other variables not included in the model, as described in the table below.

Table 4. Determination Coefficient Results (Robust Regression)

Linear Regression	Number of obs	=	144
	F(4,139)	=	38.79
	Prob > F	=	0.0000
	R-squared	=	0.4062
	Root MSE	=	3.6152

DISCUSSIONS

Partial Test

The table below describes the following:

1. The population attribute variable obtained a calculated t-value of 6.65, which is greater than the t-table value of 1.97718 (6.65 > 1.97718), with a significance value of 0.000, which is less than 0.05 (0.000 < 0.05). Therefore, it can be concluded that population attributes have a positive and significant effect on environmental performance.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

2. The house attribute variable obtained a calculated t-value of -2.91, which is less than the t-table value of 1.97718 (-2.91 < 1.97718), with a significance value of 0.004, which is less than 0.05 (0.004 < 0.05). Therefore, it can be concluded that house attributes have a negative and significant effect on environmental performance.
3. The energy consumption volume variable obtained a calculated t-value of 2.62, which is greater than the t-table value of 1.97718 (2.62 > 1.97718), with a significance value of 0.010, which is less than 0.05 (0.010 < 0.05). Therefore, it can be concluded that the volume of energy consumption has a positive and significant effect on environmental performance.
4. The utilities and services variable obtained a calculated t-value of -2.31, which is less than the t-table value of 1.97718 (-2.31 < 1.97718), with a significance value of 0.022, which is less than 0.05 (0.022 < 0.05). Therefore, it can be concluded that utilities and services have a negative and significant effect on environmental performance.

Table 5. Partial Test Results

Linear Regression		Number of obs	=	144		
		F(4,139)	=	38.79		
		Prob > F	=	0.0000		
		R-squared	=	0.4062		
		Root MSE	=	3.6152		
Y	Coef.	Robust Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]	
$X1$	0.4763178	0.0715992	6.65	0.000	0.3347535	0.6178821
$X2$	−0.4532653	0.1555867	−2.91	0.004	−0.7608879	−0.1456428
$X3$	0.2221665	0.0847478	2.62	0.010	0.054605	0.389728
$X4$	−0.4266255	0.1846968	−2.31	0.022	−0.7918039	−0.0614471
_cons	15.56197	3.09458	5.03	0.000	9.443433	21.6805

Differences of M-estimation Robust Regression from Other Regressions

M-estimate robust regression has several advantages and distinct differences compared to other regression methods. One of the main advantages is its ability to handle outliers in the data more effectively. Additionally, M-estimate robust regression is more stable in the face of assumption violations, such as heteroscedasticity and non-normality of residuals. This stability results in more consistent and reliable estimates, particularly in cases where the data does not meet the classical assumptions of ordinary linear regression. However, it is important to note that the use of M-estimation robust regression also has some drawbacks, such as the tendency to produce less efficient estimates in datasets that are relatively free of outliers.

CONCLUSION

Based on the results of the data analysis that has been stated previously, it can be concluded that the results of the M-estimate robust regression estimation on simple housing environmental performance obtained the M-estimate robust regression equation, namely $\hat{Y} = 15.562 + 0.476X_1 - 0.453X_2 + 0.222X_3 - 0.427X_4$. Based on the p-value test results, it shows that population attributes, house attributes, energy consumption volume and utilities and services have a significant effect on environmental performance. Anda the results of the hypothesis testing of the M-estimation robust regression method on the environmental performance of simple housing obtained that the variables of population attributes and energy consumption volume have a positive and significant effect on environmental performance while the variables of house attributes and utilities and services have a negative and significant effect on environmental performance.

REFERENCES

Asiaei, K., & Jusoh, R. (2017). Using a robust performance measurement system to illuminate intellectual capital. *International Journal of Accounting Information Systems*, *26*, 1–19. https://doi.org/https://doi.org/10.1016/j.accinf.2017.06.003

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, *53*(3), 464–501.

Cabana, E., Lillo, R. E., & Laniado, H. (2020). Robust regression based on shrinkage with application to Living Environment Deprivation. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 293–310. https://doi.org/10.1007/s00477-020-01774-4

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13725
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Dubois, D., & Prade, H. (2015). Possibility Theory and Its Applications: Where Do We Stand? In *Springer Handbook of Computational Intelligence* (pp. 31–60). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43505-2_3

García, J., & Peña, A. (2018). Robust Optimization: Concepts and Applications. In *Nature-inspired Methods for Stochastic, Robust and Dynamic Optimization*. InTech. https://doi.org/10.5772/intechopen.75381

Healy, K. (2005). Book Review: An R and S-PLUS Companion to Applied Regression. *Sociological Methods & Research*, *34*(1), 137–140. https://doi.org/10.1177/0049124105277200

Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, *1*(5), 799–821. http://www.jstor.org/stable/2958283

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons, Inc.

Malmqvist, T., & Glaumann, M. (2006). Selecting problem-related environmental indicators for housing management. *Building Research & Information*, *34*(4), 321–333. https://doi.org/10.1080/09613210600733658

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (Sixth). Wiley.

Nababan, E., Siahaan, N., Bangun, P., & Rosmaini, E. (2018). Environmental Performance Measurement of the Simple Urban Housing in Martubung Medan. *IOP Conference Series: Materials Science and Engineering*, *288*, 012059. https://doi.org/10.1088/1757-899X/288/1/012059

Shameem, M., Kumar, C., Chandra, B., & Khan, A. A. (2017). Systematic Review of Success Factors for Scaling Agile Methods in Global Software Development Environment: A Client-Vendor Perspective. *2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*, 17–24. https://doi.org/10.1109/APSECW.2017.22

Smith, C., Guennewig, B., & Muller, S. (2022). Robust subtractive stability measures for fast and exhaustive feature importance ranking and selection in generalised linear models. *Australian & New Zealand Journal of Statistics*, *64*(3), 339–355. https://doi.org/10.1111/anzs.12375

*name of corresponding author