

THE APPLICATION OF WEIGHTED LEAST SQUARE (WLS) IN PARAMETER ESTIMATION OF A PATH ANALYSIS

Ni Wayan Suryawardhani¹, Waego Hadi Nugroho², Adji Achmad Rinaldo Fernandes³

***ABSTRACTS---**Path analysis is one of the multivariate analysis techniques and development of regression analysis developed by Sewall Wright in 1934. The vertical garden technique is one of the greening techniques that can be applied in green building. Vertical garden is an upright garden concept, namely plants and other garden elements arranged in such a way in an upright field. In this study the author wants to find out what factors influence the interest of the people of Bendosari Village in building a vertical garden. The data used in this study are primary data with respondents from the Bendosari Village community from Cukal, Dadapan Wetan, Dadapan Kulon, Ngprih and Tretes. The method used is path analysis with the Weighted Least Square (WLS) method as an estimator of its parameters. The WLS method is used to overcome (heterogeneous) error constants in the data obtained. The results of this study indicate that the attitudes, intentions and behavior of the Bendosari Village community in constructing vertical gardens are significantly influenced by Perception of Benefits, Influence of Social Environment and Motivation. Whereas the Perception of Ease of Development does not affect the attitudes, intentions, and behavior of the Bendosari Village community in building a vertical garden significantly. This shows that the Bendosari Village community is a reliable and competent community because the Bendosari Village community tends to ignore the view that whether a facility is easy or difficult to make before building it.*

***Keywords---**Path Analysis, Multivariate Analysis, WLS, Vertical Garden.*

I. INTRODUCTION

A path analysis is one of the techniques in multivariate analysis and is advancement of regression analysis developed by Sewall Wright in 1934. [1] Wright developed the method as a means to study the direct and indirect effects of several variables, in which several variables are seen as causes and other variables are seen as a result. The Ordinary Least Square (OLS) method is used if the model is linear in parameters, i.e. by minimizing the number of error squares. The OLS method has not been able to accommodate the correlation between errors in the equation. To overcome the correlation between equations, the Weighted Least Squares (WLS) method is used to get a better beta coefficient [2].

The path analysis is often used in various cases, for example, in this study, which will examine the application of it to the factors that influence the interest of the community in Bendosari Village in building a vertical garden. In the era of globalization, the current increase in population causes an increase in housing needs and land narrowing for areas of vegetation or greenery. One effort to expand the vegetation area on narrow land is by having green buildings. A vertical

¹Department of Statistics. Brawijaya University, 65145, Indonesia E-mail: wwardhani@ub.ac.id

garden technique is one of the greening techniques that can be applied in green buildings. A vertical garden is an upright garden concept, consisting of plants and other garden elements arranged in such a way in an upright field.

The vertical garden is one of the solutions to greening efforts as well as farming on limited land to minimize global warming effects while increasing the benefits and aesthetic value of home and building environment [3]. The construction of a vertical garden is an innovative effort to maximize the potential of food production in a region. Bendosari Village as one of the largest vegetable crop producer in Malang also experiences an increase in population every year. This vertical garden is very well applied by the village community making it interesting to study.

II. LITERATURE REVIEW

Regression Analysis

A regression analysis underlies a path analysis in the application of the comparison of Ordinary Least Square (OLS) and Weighted Least Square (WLS) in this study [4]. The regression equation can be written with the following model:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \varepsilon_i \quad (2.1)$$

with

$$i = 1, 2, 3, \dots, n$$

in which:

Y_i : dependent variable i

β_0 : intercept

β_i : regression coefficient i

X_i : independent variable i

ε_i : error i which is assumed to be mutually independent with zero mean and constant variance

n : observation size [5].

Path Analysis

In 1934, Sewall Wright developed path analysis, a method used to study the direct and indirect effects of a variable, where several variables are seen as causes and other variables as a result [1]. According to [6], path analysis is used when researchers want to analyze the relationship between complex variables that cannot be done using a multiple regression analysis.

Integration of TPB (Theory of Planned Behavior) and TAM (Theory of Acceptance Model)

In TPB, behavioral intention is a determinant of a person's behavior that is influenced by attitudes, subjective norms, and perceived behavioral control. TAM is often applied because one of the main variables is behavioral intention, which is influenced by two other variables, namely perceived benefit, perceived ease, perceived environmental influence, and motivation—all of these are proven to influence the attitudes and behavior of individual users of information technology systems. In the integration of TAM and TPB, the TPB model includes the four constructs to overcome TAM weaknesses that cannot control individual behavior. This shows that TAM and TPB can be used together to analyze the factors that

influence individual behavior to build a vertical garden. The relationship between variables in the TPB and TAM integration models is presented in Figure 1 below:

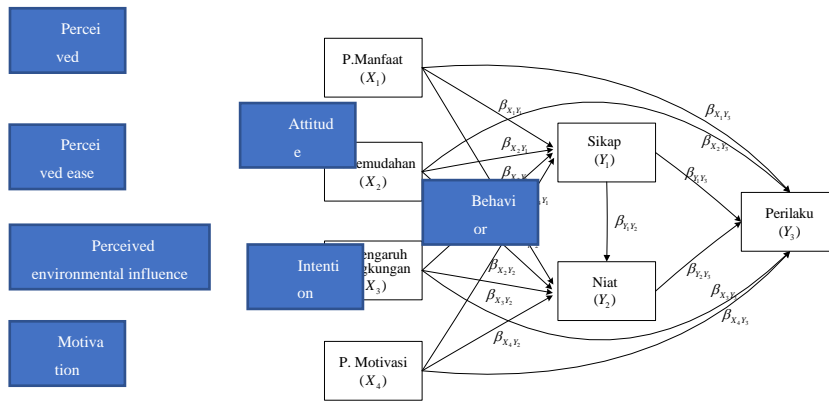


Figure 1. Integration of TPB and TAM

Figure 1 is used to achieve the fulfillment of attitudes, intentions, and behavior in constructing a vertical garden through various perceptions in the system of life.

Path Analysis Model

Variables in a path analysis require standardization of the data obtained to equalize the average and variety, so the coefficients obtained have the same units [7]. Transforming the data into standardized data is done as follows:

$$Z_{Xl} = \frac{X_l - \bar{X}_l}{S_{Xl}} \quad \text{and} \quad Z_{Yk} = \frac{Y_k - \bar{Y}_k}{S_{Yk}} \quad (2.2)$$

in which:

l : 1,2,3,.. q

k : 1,2,3,.. p

Z_{Xl} : the value of the exogenous variable X on the first observation that has been standardized

X_l : the value of the exogenous variable X on the l observation

\bar{X}_l : the average value of the exogenous variable X

S_{Xl} : the standard deviation the exogenous variable X on the l observation

Z_{Yk} : the value of the endogenous variable Y on the first observation that has been standardized

Y_k : the value of the endogenous variable Y on the k observation

\bar{Y}_k : the average value of the endogenous variable Y

S_{Yk} : the standard deviation the endogenous variable Y on the k observation

Based on Figure 1, a path analysis model can be formed that has been standardized for each observation as follows:

$$\begin{aligned} Z_{Y1} &= \beta_{X1Y1}Z_{X1} + \beta_{X2Y1}Z_{X2} + \beta_{X3Y1}Z_{X3} + \beta_{X4Y1}Z_{X4} + \varepsilon_{Y11} \\ Z_{Y2} &= \beta_{X1Y2}Z_{X1} + \beta_{X2Y2}Z_{X2} + \beta_{X3Y2}Z_{X3} + \beta_{X4Y2}Z_{X4} + \beta_{Y1Y2}Z_{Y1} + \varepsilon_{Y21} \\ Z_{Y3} &= \beta_{X1Y3}Z_{X1} + \beta_{X2Y3}Z_{X2} + \beta_{X3Y3}Z_{X3} + \beta_{X4Y3}Z_{X4} + \beta_{Y1Y3}Z_{Y1} + \beta_{Y2Y3}Z_{Y2} + \varepsilon_{Y31} \end{aligned} \quad (2.3)$$

Equation (2.3) is written in the form of a matrix as follows:

$$\underline{Y} = \mathbf{X}\underline{\beta} + \varepsilon \quad (2.4)$$

Estimating the Parameter Coefficient

The OLS Method

The Ordinary Least Square (OLS) method is a function of linear estimation method most often used in estimating the parameters of a regression model. According to [8] the OLS criterion is “Line of Best Fit”, or, in other words, the sum of squares of deviations between observation point and regression line is minimum. The OLS method is used if the model is linear in parameters, namely by minimizing the number of error squares. Based on equation (2.4), we obtain:

$$\underline{Y} = \mathbf{X}\underline{\beta} + \varepsilon$$

$$\varepsilon = \underline{Y} - \mathbf{X}\underline{\beta}$$

The OLS method minimizes the following functions::

$$\min \{Q\} = \min \{ \varepsilon^T \varepsilon \} \quad (2.5)$$

$$\min \{Q\} = \min \{ (\underline{Y} - \mathbf{X}\underline{\beta})^T (\underline{Y} - \mathbf{X}\underline{\beta}) \}$$

Parameter estimation with the OLS approach is done by minimizing the following Q :

$$\begin{aligned} Q &= (\varepsilon^T \varepsilon) = (\underline{Y} - \mathbf{X}\underline{\beta})^T (\underline{Y} - \mathbf{X}\underline{\beta}) \\ &= (\underline{Y}^T - \mathbf{X}^T \underline{\beta}^T) (\underline{Y} - \mathbf{X}\underline{\beta}) \\ &= (\underline{Y}^T \underline{Y} - \underline{Y}^T \mathbf{X}\underline{\beta} - \underline{\beta}^T \mathbf{X}^T \underline{Y} + \underline{\beta}^T \mathbf{X}^T \mathbf{X}\underline{\beta}) \\ &= (\underline{Y}^T \underline{Y} - 2\underline{\beta}^T \mathbf{X}^T \underline{Y} + \underline{\beta}^T \mathbf{X}^T \mathbf{X}\underline{\beta}) \end{aligned} \quad (2.6)$$

The solution to the optimization of equation (2.6) is done by getting a derivative of Q to get the estimator value for $\hat{\underline{b}}$:

$$\begin{aligned} \frac{\partial(Q)}{\partial(\underline{\beta})} &= 0 \\ -2\mathbf{X}^T \underline{Y} + 2\mathbf{X}^T \mathbf{X}\hat{\underline{\beta}} &= 0 \\ -\mathbf{X}^T \underline{Y} + \mathbf{X}^T \mathbf{X}\hat{\underline{\beta}} &= 0 \\ \mathbf{X}^T \mathbf{X}\hat{\underline{\beta}} &= \mathbf{X}^T \underline{Y} \\ \hat{\underline{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{Y} \end{aligned} \quad (2.7)$$

Based on equation (2.3), parameter estimation is done using OLS method for each path analysis model. The following is a parameter estimate for each path analysis model:

1. Equation (1)

$$\begin{aligned} Z_{Y1} &= \beta_{X1Y1} Z_{X1} + \beta_{X2Y1} Z_{X2} + \beta_{X3Y1} Z_{X3} + \\ &\beta_{X4Y1} Z_{X4} + \varepsilon_{Y11} \end{aligned}$$

In the path analysis model above, the parameter estimation results based on equation (2.7) are:

$$\hat{\beta}_{\sim(1)} = \begin{bmatrix} \hat{\beta}_{X1Y1} \\ \hat{\beta}_{X2Y1} \\ \hat{\beta}_{X3Y1} \\ \hat{\beta}_{X4Y1} \end{bmatrix}$$

2. Equation (2)

$$Z_{Y2} = \beta_{X1Y2}Z_{X1} + \beta_{X2Y2}Z_{X2} + \beta_{X3Y2}Z_{X3} + \beta_{X4Y2}Z_{X4} + \beta_{Y1Y2}Z_{Y1} + \varepsilon_{Y21}$$

In the path analysis model above, the parameter estimation results based on equation (2.7):

$$\hat{\beta}_{\sim(2)} = \begin{bmatrix} \hat{\beta}_{X1Y1} \\ \hat{\beta}_{X2Y1} \\ \hat{\beta}_{X3Y1} \\ \hat{\beta}_{X4Y1} \\ \hat{\beta}_{Y1Y2} \end{bmatrix}$$

3. Equation (3)

$$Z_{Y3} = \beta_{X1Y3}Z_{X1} + \beta_{X2Y3}Z_{X2} + \beta_{X3Y3}Z_{X3} + \beta_{X4Y3}Z_{X4} + \beta_{Y1Y3}Z_{Y1} + \beta_{Y2Y3}Z_{Y2} + \varepsilon_{Y31}$$

In the path analysis model above, the parameter estimation results based on equation (2.7):

$$\hat{\beta}_{\sim(3)} = \begin{bmatrix} \hat{\beta}_{X1Y3} \\ \hat{\beta}_{X2Y3} \\ \hat{\beta}_{X3Y3} \\ \hat{\beta}_{X4Y3} \\ \hat{\beta}_{Y1Y3} \\ \hat{\beta}_{Y2Y3} \end{bmatrix}$$

Based on the three equations above, the total parameter estimates are as follows:

$$\hat{\beta} = \begin{bmatrix} \beta_{X1Y1} \\ \beta_{X2Y1} \\ \beta_{X3Y1} \\ \beta_{X4Y1} \\ \beta_{X1Y2} \\ \beta_{Y2Y2} \\ \beta_{Y3Y2} \\ \beta_{Y4Y2} \\ \beta_{Y1Y2} \\ \beta_{Y1Y3} \\ \beta_{Y2Y3} \\ \beta_{Y3Y3} \\ \beta_{Y4Y3} \\ \beta_{Y1Y3} \\ \beta_{Y2Y3} \end{bmatrix} \quad (2.8)$$

The WLS Method

The OLS method has not accommodated the correlation between equations in the three equations [9]. According to [10], the estimation is carried out with the Weighted Least Square (WLS) optimization, which is able to accommodate the correlation between weighted equations in the form of inverses from the error variance-covariance matrix. Estimation of the error variance-covariance matrix for the path analysis model is as follows:

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \cdots & \hat{\Sigma}_{1q} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & \cdots & \hat{\Sigma}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Sigma}_{q1} & \hat{\Sigma}_{q2} & \cdots & \hat{\Sigma}_{qq} \end{bmatrix}_{q \times q} \quad (2.9)$$

The solution to this equation uses weighting with the value of X and β as follows: (2.10)

$$\begin{aligned} \min\{Q\} &= \min\{\varepsilon^T \Sigma^{-1} \varepsilon\} \\ \min\{Q\} &= \min\{(Y - X\beta)^T \Sigma^{-1} (Y - X\beta)\} \end{aligned} \quad (2.10)$$

To complete the optimization in equation (2.10), the partial derivative will be carried out as follows:

$$\begin{aligned} Q(\beta) &= (Y - X\beta)^T \Sigma^{-1} (Y - X\beta) = (Y^T - \beta^T X^T) \Sigma^{-1} (Y - X\beta) = Y^T \Sigma^{-1} Y - 2\beta^T X^T \Sigma^{-1} Y + \beta^T X^T \Sigma^{-1} X \beta \\ &= Y^T \Sigma^{-1} Y - \beta^T X^T \Sigma^{-1} Y - Y^T \Sigma^{-1} X \beta + \beta^T X^T \Sigma^{-1} X \beta \end{aligned} \quad (2.11)$$

Note that $Y^T X \beta = \beta^T X^T Y$ because both results are scalar. The next process is to reduce the above equation toward β to produce the above equation. After that, it is equated with zero; and the result can be seen in the equation below [11]. Thus, the β estimator is obtained in the equation below:

$$\frac{\partial Q(\beta)}{\partial \beta} = -2\mathbf{X}^T \Sigma^{-1} \underline{Y} + 2\mathbf{X}^T \Sigma^{-1} \mathbf{X} \underline{\beta} \qquad \mathbf{X}^T \Sigma^{-1} \underline{Y} = \mathbf{X}^T \Sigma^{-1} \mathbf{X} \underline{\tilde{\beta}}$$

$$\underline{\hat{\beta}} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \underline{Y} \qquad (2.12)$$

In WLS at iteration = 1, then $\Sigma = \mathbf{I}$ so the WLS equation becomes the original form like OLS, namely:

$$t = 1$$

$$\underline{\tilde{\beta}}_{(t)} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \underline{Y}$$

$$\underline{\tilde{\beta}}_{(1)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{Y} \qquad (2.17)$$

$$e_{i(1)} = y_{i(1)} - \hat{y}_{i(1)}$$

When the iteration of the Weighted Least Square method is 1, it is proportional to the Ordinary Least Square method. Whereas when WLS uses iteration > 1 , it will get a different parameter estimating function. In equation 2.18, the estimation of WLS parameters will be shown when iterations = 2, as follows:

$$t = 2, \text{ so } \omega_{i(t)} = \omega_{i(2)} = \frac{1}{\hat{e}_{i(t-1)}^2}$$

$$= \frac{1}{\hat{e}_{i(1)}^2} \qquad (2.18)$$

The estimated error variance-covariance matrix will change to:

$$\hat{\Sigma}^{-1} = \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{bmatrix}_{n \times n} \qquad (2.19)$$

And the WLS equation changes to:

$$\underline{\tilde{\beta}}_{(t)} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \underline{Y} \quad \underline{\tilde{\beta}}_{(2)} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \underline{Y} \qquad (2.20)$$

The Assumption Test in the Path Analysis

According to [12], a regression analysis requires several assumptions underlying the estimation of parameters using the Ordinary Least Square approach. This is the same for the path analysis. These assumptions are:

Linearity

The linearity assumption is needed to determine the shape of the regression curve correctly. According to [13], if the linearity assumption is not fulfilled, then relationship of the predictor variable and the response variable are not linear. One method to test the linearity of the relationship between exogenous and endogenous variables is the Regression Specification Error Test as follows:

a) The first regression equation is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.21)$$

Parameter estimation with the OLS approach obtains the following.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad (2.22)$$

R_1^2 calculation is done as follows:

$$R_1^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.23)$$

b) The second regression equation is:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 \hat{Y}_i^2 + \alpha_3 \hat{Y}_i^3 + \varepsilon_i$$

R_2^2 calculation is done as follows:

$$R_2^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i^*)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.24)$$

c) The test on the relationship of the predictor variable and the response variable to know whether it is linear or not:

I. Hypothesis

$$H_0 : \alpha_2 = \alpha_3 = 0, \text{ vs}$$

$$H_1 : \text{minimum of one } \alpha_j \neq 0, j = 2, 3$$

II. Statistical t-test follow the distribution of F as follows.

$$F = \frac{(R_2^2 - R_1^2) / 2}{(1 - R_2^2) / (n - (k + 2))} \sim F_{(k-1, n-k-2)} \quad (2.25)$$

III. H_0 is rejected if the F test result < the F critical value $F_{\alpha(k-1, n-k-2)}$; this means that the relationship of the endogenous and exogenous variable is non-linear.

Error Normality

According to [14], errors will be normally distributed in estimating regression parameters as needed because it is expected that the mean error value is zero. The method for testing error normality is Kolmogorov-Smirnov, which is one part of the goodness of fit. The Kolmogorov-Smirnov test procedure is:

a. Hypothesis:

$$H_0: F_N(X) = F_0(X) \text{ (Error normally distributed), vs}$$

$$H_1: F_N(X) \neq F_0(X) \text{ (Error not normally distributed)}$$

b. Kolmogorov-Smirnov statistics

$$D_N = \sup [|F_N(X) - F_0(X)|] \quad (2.26)$$

in which:

D_N : maximum absolute difference between empirical distribution functions and normal distribution functions

$F_N(X)$: cumulative probability functions of observation

$F_0(X)$: cumulative probability functions of normal distribution.

- c. H_0 is accepted if the test statistic is $D_N \leq D_{tabel}$ then error is normally distributed.

3. Homoscedasticity

Homoscedasticity is a condition in which a variety of homogeneous errors or a variety of errors is constant. To test the homogeneity of various errors we can use the Bseusch-Pagan test [15]. This test is in a large sample assumed to be mutually free and to have a normal distribution, and various errors σ_i^2 are influenced by the level of the predictor variable (variable X) like the following equation:

$$\text{Log}_e \sigma_i^2 = \gamma_0 + \gamma_1 X_i \quad (2.27)$$

Based on equation (2.27), it is known that $\sigma_i^2 \sigma$ will increase or decrease depending on γ_1 on X_i . The error range will be constant if $\gamma_1 = 0$ so the Breusch-Pagan test hypothesis is:

$H_0: \gamma_1 = 0$ (Variety of errors is constant), vs

$H_1: \gamma_1 \neq 0$ (Variety of errors is not constant)

The test statistics used are χ_{BP}^2 as follows:

$$\chi_{BP}^2 = \frac{\frac{JKR^*}{2}}{\left(\frac{JKS}{n} \right)} \sim \chi_1^2 \quad (2.28)$$

in which:

JKR^* : number of squares of regression between residual (e_2) and the exogenous variable (X)

JKS : number of squares of regression between the endogenous variable (Y) and the exogenous variable (X).

If $H_0: \gamma_1 = 0$ in an observation with a large n then χ_{BP}^2 follow the distribution of chi-square with one free degree. H_0 if accepted if the test statistic χ_{BP}^2 is smaller than $\chi_{\alpha(1)}^2$ so the assumption that the homogeneity of variety has been fulfilled.

4. Multicollinearity

According to [16], multicollinearity occurs because of a linear combination between two or more independent variables; if there are no correlations between the independent variables used, then it can be interpreted that multicollinearity does not occur. If the assumption of multicollinearity is not fulfilled, it will be very difficult to separate the influence of each independent variable on the dependent variable.

One method that can be used to detect multicollinearity is to look at the value of the Variance Inflation Factor (VIF) of each independent variable. VIF is a measure of the total variety that the variables can be explained by the variety of other independent variables. The VIF value can be defined as follows [7]:

$$VIF = \frac{1}{1 - R_j^2} \quad (2.29)$$

$$R_j^2 = \frac{JK_{reg}}{JK_{total}} \quad (2.30)$$

in which:

$$j= 1, 2, 3, \dots, p$$

p = the number of independent variable

R_j^2 = coefficient of determination between independent variables X_j with other independent variables

$$JK_{reg} = \text{number of squares of regression} = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

$$JK_{total} = \text{number of total regression} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

The coefficient of determination or R_j^2 can be determined by regressing the independent variable X_j with all other independent variables. If the obtained value of VIF is greater, there is an increasingly large correlation among the independent variables. If $VIF > 10$, the correlation between independent variables is very high and the vice versa [17].

Hypothesis Testing

According to [18], hypothesis testing is done to determine the significance of the path analysis model formed by comparing the values of $|t_{count}|$ with $|t_{table}|$ at 5% real level. To calculate hypothesis testing using the t test, it is necessary to know the calculation of variance analysis for linear regression with the following calculations:

Table 1. Analysis of Linear Regression Variants

Source of Variety	Degree of Freedom	JK	KT
Regression	p	$JKR = (S_{xy})^2 / S_{xx}$	$MSR = JKR / p$
Errors	$n - p - 1$	$JKG = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$	$MSE = \frac{JKG}{n - p - 1}$
Total	$n - 1$	S_{yy}	

Note:

$$S_{xy} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - \frac{(\sum X_i)^2 (\sum Y_i)^2}{n}$$

$$S_{xx} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$S_{yy} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

The hypothesis used for hypothesis testing is:

$$H_0 : \beta_{YX_j} = 0, \text{ vs}$$

$$H_1 : \beta_{YX_j} \neq 0$$

The statistic for t-test:

$$|t_{(n-2)}| = \left| \frac{\hat{\beta} - \beta_{YX_j}}{Se(\hat{\beta})} \right| = \left| \frac{\hat{\beta} - 0}{\sqrt{\sigma^2 / S_{XX}}} \right| = \left| \frac{\hat{\beta} - 0}{\sqrt{MSE / S_{XX}}} \right| \quad (2.31)$$

If $|t_{count}| > |t_{table}|$, then H_0 is rejected so it can be concluded that there is an influence on the path coefficient. In addition to using t-test, we can also use the p value. If the p value is less than α (0.05) then H_0 is rejected, meaning that there is an influence on the path coefficient.

Model Validity

A model is said to be valid if the underlying assumptions have been fulfilled. In addition, according to [19], there are two indicators of model validity in path analysis, namely:

- 1) Total Coefficient of Determination (TCD)

The total variety of data that can be explained by the model can be shown by the total determination coefficient with the following formula:

$$R_t^2 = 1 - P_{e1}^2 P_{e2}^2 \dots P_{ep}^2 \quad (2.32)$$

As for the calculation of the residual effect (error) can be done using the following formula:

$$P_{ei} = \sqrt{1 - R_i^2} \quad (2.33)$$

in which:

$$R_i^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.34)$$

Note:

R_t^2 = total coefficient of determination

R_i^2 = coefficient of determination in each equation

p_{ei} = residual effects on each equation

The interpretation for R_t^2 is the same as the one for the coefficient of determination (R^2) in a regression analysis. Total coefficient of determination has a value range from 0% to 100%. According to Hair and Ringle (2011), if $R^2 > 0.75$, then the regression model is substantial. If $0.05 < R^2 < 0.75$ then the regression model is moderate. If $R^2 < 0.50$ then the regression model is weak in representing the model.

- 2) Trimming Theory

Path validation test on each path for direct influence on the trimming method adopts the same strategy with the regression analysis, using the t-test probability value, by testing the regression coefficient which is partially standardized. According to [1], this method is an approach to improve the path analysis model done by removing path coefficients that do not meet the significance criteria.

III. Research methodology

The data used in this study are primary data obtained through a questionnaire distributed to the community of Bendosari Village, Malang Regency. The data obtained through the questionnaire is used at modeling the potential development of Bendosari Village with several variables influencing attitudes, behavioral intentions, and community behavior in building a vertical garden by considering perceived benefits, perceived ease, perceived environmental influences, and motivation.

The research stages are as follows:

1. Determine the variables used to solve the problem in this study (perceived benefits, perceived ease, perceived environmental influence, and motivation) that influence attitudes, behavioral intentions, and behavior of the community in building a vertical garden.
2. Review and determine the theory according to experts, i.e. the TPB and TAM integration model.
3. Determine the scale to be used.
4. Make a research instrument grid.
5. Make a research questionnaire.
6. Determine the population and sample.
7. Try out the research instruments.
8. Examination of validity and reliability on research instruments.
9. Collect data by distributing questionnaires to respondents.
10. Change the score to scale using MSI as input for further analysis using the average raw score.
11. Perform path analysis.

The steps in the path analysis are as follows:

1. Make a path diagram in accordance with the predetermined theory.
2. Create a path analysis model according to the standardized path diagram.
3. Estimate path coefficients using the Weighted Least Square (WLS) method.
4. Check path analysis assumptions.
5. Test the path analysis hypothesis.
6. Test the validity of the model.
7. Calculate relative efficiency.
8. Calculate direct, indirect and total influence.
9. Interpret the size of the path analysis coefficient and the effect in the formed path analysis model.

IV. FINDINGS AND DISCUSSION

Testing Path Assumptions

The following are the results of testing and checking assumptions that must be met in path analysis:

1. Assumption of Error Normality

Testing is done using the SPSS trial version software. The summary of test results is presented in Table 1.

Table 1. Test results on error normality assumptions

Endogenous variable	p-value	Conclusion
Attitude	0,200	Normal
Behavioral intention	0,200	Normal
Behavior	0,089	Normal

Based on Table 1, all errors on each residual produce an probability value $> \alpha$ (0.05) that H_0 is accepted. Thus, it can be concluded that errors are normally distributed.

2. Homoscedasticity Assumptions

Testing is done using the SPSS trial version software. The summary of test results is presented in Table 2.

Table 2. Test results on homoscedasticity assumptions

Exogenous variable	p-value	Conclusion
Equation I		
Perceived Benefits	0.023	Error Not Constant
Perceived Ease	0.004	Error Not Constant
Perceived Environmental Influence	0.136	Error Constant
Motivation	0.149	Error Constant
Equation II		
Perceived Benefits	0.059	Error Constant
Perceived Ease	0.730	Error Constant
Perceived Environmental Influence	0.142	Error Constant
Motivation	0.515	Error Constant
Attitude	0.310	Error Constant
Equation III		

Perceived Benefits	0.220	Error Constant
Perceived Ease	0.587	Error Constant
Perceived Environmental Influence	0.196	Error Constant
Motivation	0.118	Error Constant
Attitude	0.945	Error Constant
Behavioral Intention	0.086	Error Constant

Table 2 shows if the probability value $> \alpha (0.05)$, then the range of errors is constant. There are errors that are not constant in the first equation of perceived benefits and perceived ease; thus, it must be overcome by using the Weighted Least Square method.

3. Linearity Assumptions

Testing is done using the R software. The summary of test results is presented in Table 1.

Table 3. Test results on linearity assumption

Variable	P-value	Relationship
X_1 with Y_1	0.971	Linear
X_2 with Y_1	0.706	Linear
X_3 with Y_1	0.362	Linear
X_4 with Y_1	0.819	Linear
X_1 with Y_2	0.718	Linear
X_2 with Y_2	0.156	Linear
X_3 with Y_2	0.816	Linear

Variable	P-value	Relationship
X_4 with Y_2	0.884	Linear
Y_1 with Y_2	0.175	Linear
X_1 with Y_3	0.768	Linear
X_2 with Y_3	0.414	Linear
X_3 with Y_3	0.309	Linear
X_4 with Y_3	0.555	Linear
Y_1 with Y_3	0.216	Linear
Y_2 with Y_3	0.424	Linear

Table 3 shows that the relationship between exogenous and endogenous variables produces an probability value $> \alpha$ (0.05), so H_0 is accepted, which means that the relationship is linear.

Parameter Estimation

In this study, parameter estimation in the path analysis is done by estimating the standardized Weighted Least Square (WLS) coefficient. The WLS method is chosen because there are errors that are not constant in equation I, so data is heterogeneous. The following is a diagram and path coefficient that has been predicted with the help of R software.

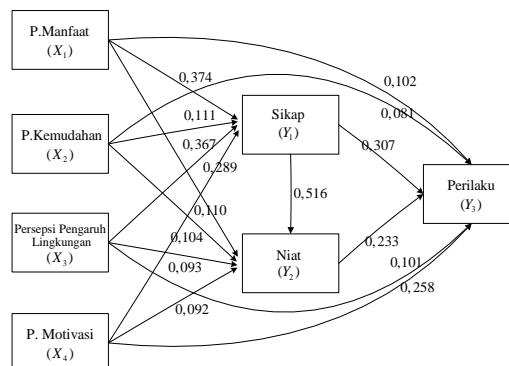


Figure 1. Diagram and path coefficient

Based on the estimation of the path coefficients above, the path analysis equation obtained from the WLS method can be formed as follows:

$$Z_{Y1} = 0,374Z_{X1} + 0,110Z_{X2} + 0,367Z_{X3} + 0,289Z_{X4} + \epsilon_{Y11}$$

$$Z_{Y2} = 0,110Z_{X1} + 0,104Z_{X2} + 0,934Z_{X3} + 0,920Z_{X4} + 0,516Z_{Y1} + \epsilon_{Y21}$$

$$Z_{Y3} = 0,102Z_{X1} + 0,081Z_{X2} + 0,101Z_{X3} + 0,258Z_{X4} + 0,307Z_{Y1} + 0,233Z_{Y2} + \epsilon_{Y31}$$

Hypothesis Testing

Hypothesis testing is done to determine the significance of the model formed. The hypothesis testing refers to equation (2.21). The results are presented in Table 4.5.

Table 4. Hypothesis testing results

Variable	P-value	Relationship
X ₁ with Y ₁	<0.001	Significant
X ₂ with Y ₁	0.1686	Not Significant
X ₃ with Y ₁	<0.001	Significant
X ₄ with Y ₁	<0.001	Significant
X ₁ with Y ₂	<0.001	Significant
X ₂ with Y ₂	0.1899	Not Significant
X ₃ with Y ₂	0.2146	Not Significant
X ₄ with Y ₂	0.1872	Not Significant
Y ₁ with Y ₂	<0.001	Significant
X ₁ with Y ₃	0.1968	Not Significant
X ₂ with Y ₃	0.2555	Not Significant
X ₃ with Y ₃	0.1950	Not Significant
X ₄ with Y ₃	<0.001	Significant
Y ₁ with Y ₃	0.0106	Significant

Y_2 with Y_3	0.0250	Significant
---------------------	--------	-------------

Table 4 shows that the probability value $< \alpha$ (0.05) which results in the rejection of H_0 ; thus, it can be concluded that there is an influence on the path. The not significant path occurs in the direct influence between:

1. Perceived ease with attitude
2. Perceived ease with behavioral intention
3. Perceived environmental influence with behavioral intention
4. Motivation with behavioral intention
5. Perceived benefits with behavior
6. Perceived ease with behavior
7. Perceived environmental influence with behavior

Model Validity

The results of Weighted Least Square as a parameter estimator in path analysis with heterogeneous data are:

1) Total Coefficient of Determination

The total coefficient of determination is calculated, so the variety of data that can be explained by the model is known. The total coefficient of determination for the WLS method is 0.9305, so it can be concluded that the variety of data that can be explained by the model is 93.05%, while the remaining 6.95% is explained by other variables that are not in the model.

2) Trimming Theory

The trimming method is an approach to improve the path analysis model by removing insignificant path coefficients. The path is then removed, so the formed trimming model is as follows:

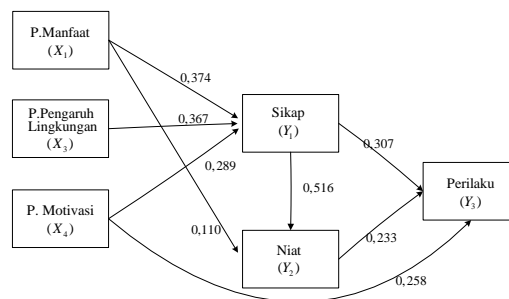


Figure 2. Final Model After Trimming

The construction of a vertical garden as one of the facilities of green building is an important factor in realizing one of the programs of the Indonesian Government, namely Indonesia Energy-Independent Based on Eco-Friendly Concept. The application in Bendosari Village has attracted the researchers to find out factors influenced the interest of the Bendosari Village community in building vertical gardens.

Perceived benefits, perceived environmental influence, and motivation are three variables that have a significant influence in determining the attitude of the Bendosari Village community in constructing a vertical garden. However, we can see that perceived benefits have the greatest role in determining the attitude of the Bendosari Village Community in

constructing vertical garden with a value of 0.374, followed by perceived environmental influence with a value of 0.367, and motivation with a value of 0.289. The rest is influenced by other variables that have not been explained in the model.

Behavioral intentions are a person's desire to carry out certain behaviors. The intention in building a vertical garden depends on the attitude of the community. In other words, the better the attitude of the Bendosari Village community in building a vertical garden, the better their intentions in building a vertical garden will be. Behavior is an action of a person. In this study, people's behavior depends on the attitude of the community in building a vertical garden; this results in a value of 0.427. In other words, the better the attitude of the community in accepting ideas, the better the behavior of the community in building a vertical garden will be. Perceived ease of development is a factor that does not influence attitudes, intention, and behavior of the Bendosari Village community in constructing a vertical garden. This shows that the community is a reliable and competent community because they do not think much about the ease or difficulty in developing something—they, instead, focus more on the benefits and use. They are motivated to realize the idea.

V. CONCLUSION

Based on findings and discussion, the following conclusions can be drawn:

1. Based on the path analysis model, it can be seen that perceived benefits, perceived ease, perceived environmental influence, and motivation have a significant effect on attitudes, intentions, and behavior of the Bendosari Village community in the construction of vertical gardens, where the coefficient of determination is 93.05%. This means that the variety of data that can be explained by the model formed is 93.05%, while the remaining 6.95% is influenced by other variables.

2. The parameter estimation in this study should use the Weighted Least Square method because the data obtained does not meet the assumptions of homoscedasticity or the data has a non-constant error.

REFERENCE

- [1] Dillon, W. R. dan Goldstein, M. (1984). *Multivariate Analysis Methods and Application*. New York: John Wiley Sons. Inc.
- [2] Fernandes, A.A.R., Nyoman Budiantara, I., Otok, B.W., Suhartono, (2014), "Reproducing Kernel Hilbert space and penalized weighted least square in nonparametric regression", *Applied Mathematical Sciences*, Vol 8 No 146, pp 7289-7300
- [3] Mulyaningsih, Y. (2015). *Makalah Arsitektur Lanskap Vertical Garden*. Bogor: Universitas Djuanda.
- [4] Fernandes, A.A.R., Budiantara, I.N, Otok, B.W., and Suhartono. (2015). "Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data", *Journal of Mathematics and Statistics*, Vol 11, No 2, pp. 61-69.
- [5] Wang, Y. (2011). *Smoothing Spline Methods Applications*. New York: CRC Press.
- [6] Gudono. (2012). *Analisis Data Multivariat*. Edisi Kedua. Yogyakarta: BPF.
- [7] Solimun, Fernandes, A.A.R. (2017). "Investigation the mediating variable: What Necessary? (Case Study in Management Research)", *International Journal of Law and Management*, Vol 59 No 6, pp 101-110.
- [8] Mendenhall, S. (1996). *A Second Course In Statistics. Regression Analysis. Fifth Edition. Prentice Hall International Edition*.
- [9] Fernandes, A.A.R., Hutahayan, B., Solimun, Arisoelaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, & Amaliana, L, (2019), "Comparison of Curve Estimation of the Smoothing Spline Nonparametric Function Path Based on PLS and PWLS In Various Levels of Heteroscedasticity", *IOP Conference Series: Materials Science and Engineering*, Forthcoming Issue.
- [10] Susanti, Yuliana. (2014). *M Estimation, S Estimation, and MM Estimation In Robust Regression*. Surakarta: Universitas Sebelas Maret.
- [11] Hutahayan, B., Solimun, Fernandes, A.A.R., Arisoelaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, & Amaliana, L, (2019), "Mixed Second Order Indicator Model: The First Order Using Principal Component Analysis and The

- Second Order Using Factor Analysis”, *IOP Conference Series: Materials Science and Engineering*, Forthcoming Issue.
- [12] Gujarati, D. (2004). *Basic Econometrics*. Fourth Edition. New York: McGraw Hill.
- [13] Wu, H. dan Zhang, J.T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. New Jersey: John Wiley and Sons. Inc.
- [14] Sheskin, D.J. (2000). *Parametric and Nonparametric Statistical Procedures*. Second Edition. New York: CRC Press
- [15] Kutner, M.H., Nachtsheim, C.J., Neter, J. dan Li W. (2005). *Applied Linear Statistical Models*. Fifth Edition. Boston: McGraw-Hill International.
- [16] Djarwanto PS dan Pangestu Subagyo. (2011). *Statistik Induktif*. Edisi 4. Yogyakarta: Penerbit BPFE.
- [17] Fernandes, A.A.R, Jansen, P., Sa’adah, U, Solimun, Nurdjannah, Amaliana, L., and Efendi, A. (2018). “Comparison of Spline Estimator at Various Levels of Autocorrelation in Smoothing Spline Nonparametric Regression For Longitudinal Data”, *Communications in Statistics – Theory and Methods*, Vol 46, No 24, pp 12401-12424.
- [18] Mendenhall, W., Beaver, R.J. dan Beaver, B.M. (2009). *Introduction to Probability and Statistics, Thirteenth edition*. USA. Brooks/Cole Cengage Learning.
- [19] Solimun. (2010). *Analisis Multivariat Pemodelan Struktural Metode Partial Least Square-PLS*. Malang: CV. Citra Malang.