WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Identification of Factors Affecting the Welfare Level in Central Java Province-Indonesia Using SEM-PLS POS Method.

Hilda Ayu Meylia[a], Alan Prahutama[a,b], and Sugito[a,c]*

[a]*Statistics Department, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia, 50275*
[b]*Social and Health Laboratory Statistics, Diponegoro University, Semarang, Indonesia, 50275*
[c]*Industrial Laboratory Statistics, Diponegoro University, Semarang, Indonesia, 50275*

A B S T R A C T

Welfare is a key indicator of a nation's development as it reflects individuals' quality of life in meeting basic needs and participating in society. Development disparities between regions and limited infrastructure pose significant challenges in improving public welfare. Structural Equation Modeling (SEM) with the Partial Least Squares (PLS) approach is one statistical method to analyze the complexity of dimensions affecting welfare. SEM usually assumes a homogeneous population, in practice, the population often consists of several clusters. Partial Least Squares-Prediction Oriented Segmentation (PLS-POS) is the method to detect heterogeneity. This study aims to build a structural model that describes the relationships between education, housing, health, employment, and welfare in Central Java Province based on 19 valid and reliable indicators. The results of SEM-PLS show that housing has a significant positive effect on health, education has a positive impact on employment and welfare, housing and employment have a significant negative effect on welfare, and health has a positive impact on welfare. Grouping regions using PLS-POS resulted in three segments with different characteristics. Each model local shows higher R^2 values and better Goodness of Fit than the model global, indicating that the model local has better model goodness.

Keywords: welfare; SEM; PLS; PLS-POS, Central Java-Indonesia

1. Introduction

Welfare is an important indicator in assessing the success of a country. Welfare reflects the quality of life of a population as one of the main goals of national development (Todaro and Smith, 2020). Welfare is also one of the key objectives of the Sustainable Development Goals (SDGs), particularly the third goal of healthy lives and well-being for all ages, which encompasses both physical and overall quality of life aspects (United Nations, 2015). In Indonesia, the Law Number 11 of 2009 emphasizes that welfare is a condition in which individuals' material, spiritual, and social needs are fulfilled to live a decent life, develop their potential, and play an active role in society. Welfare issues remain a challenge, especially for people unable to meet their basic needs due to economic and social constraints (Blau and Abramovitz, 2010). The Indonesian government has prioritized improving its welfare, but progress has been uneven across regions, including Central Java Province.

Central Java Province, the third most populous province in Indonesia after West Java and East Java, has a population of approximately 37 million with a high population density and covers nearly a quarter of Java Island. This makes it the fifth-largest province in Indonesia regarding land area (Statistics Indonesia, 2023). The high population, regional development disparities, and infrastructure limitations pose significant challenges in improving welfare (Rodríguez and Hardy, 2015). Data from the Central Statistics Agency (BPS) indicates that the Human Development Index (HDI) in several regencies/cities in Central Java Province in 2023 remains below 70.00 (Statistics Indonesia, 2023). Development disparities are reflected in HDI achievements and access to education, health services, employment opportunities, and inadequate basic infrastructure.

One statistical method that can be used to analyze the complexity between dimensions that influence welfare in Indonesia, particularly in Central Java Province, is Structural Equation Modeling (SEM). The variables used in this study cannot be measured directly (unobserved variables) or referred to as latent variables, thus requiring indicators capable of forming these latent variables. SEM can estimate relationships between variables with multiple relationships formed in a structural model (relationships between dependent and independent latent variables) (Yamin and Kurniawan, 2009). SEM can also describe the relationship pattern between latent variables and their indicator variables. The limitations of SEM are related to the assumption that data must be multivariate and normally distributed, indicators must be reflective, models must be based on theory, and there must be interdependence. As an alternative, variance-based SEM or Partial Least Squares (SEM-PLS) was developed to overcome these limitations (Sarstedt, 2016). SEM-PLS is a variance-based method for estimating structural equation models to maximize the variance explained by endogenous latent variables (Hair et al., 2017).

Research on Structural Equation Modeling-Partial Least Squares (SEM-PLS) has been widely conducted in various contexts. Aghili and Amirkhani (2021) used the SEM-PLS approach to analyze complex relationships and factors influencing the development and success of green buildings. Nasutian et al. (2020) evaluated the performance quality of micro, small, and medium enterprises based on management quality, innovation, and financial aspects. Ali et al. (2017) applied SEM-PLS to evaluate hotel service satisfaction. Ahmad et al. (2024) modeled community well-being and unemployment rates while considering income inequality.

These studies generally assume that the data used comes from the same population (homogeneous), whereas in practice, research data is collected from populations with different characteristics that can cause heterogeneity. According to Lubke and Muthén (2005), ignoring heterogeneity in modeling can result in biased parameter estimates and invalid conclusions. One approach that can detect unobserved heterogeneity in SEM-PLS is Partial Least Square-Prediction Oriented Segmentation (PLS-POS). This method groups observations into homogeneous segments based on the similarity of patterns of relationships between latent variables and estimates model parameters separately for each segment formed (Becker et al., 2013).

The Partial Least Square-Prediction Oriented Segmentation (PLS-POS) approach has been used in various previous studies. Stefan et al. (2024) implemented PLS-POS to examine how the application of AI affects an organization's ability to adapt to the environment. Arenas et al. (2020) identified groups of elderly people based on their internet usage, preferences, and online behavior. Ratzmann et al. (2016) applied PLS-POS in management and strategy research on alliance governance and innovation.

Previous studies have used SEM-PLS to analyze welfare, but there are still limitations in considering heterogeneity between regions that can affect the accuracy of the analysis results (Stefan et al., 2024). The PLS-POS approach was used in this study to detect and address heterogeneity, which has rarely been applied in the context of welfare in Central Java Province. Based on the above, this study aims to analyze the relationships among latent variables influencing public welfare using the Structural Equation Modeling-Partial Least Squares (SEM-PLS) method with the Partial Least Squares-Prediction Oriented Segmentation (PLS-POS) approach.

2. Literature Review

2.1 Structural Equation Modeling- Partial Least Squares (SEM-PLS)

SEM is the second generation of multivariate analysis methods that can be used to describe simultaneous linear relationships between observed variables (indicators) and variables that cannot be measured directly (latent variables), as well as the relationships between latent variables themselves (Hair et al., 2021). PLS is one of the approaches in SEM that is based on variance and is used as an alternative to the covariance-based SEM approach. The main advantage of PLS is its flexibility regarding assumptions, such as multivariate normal distribution. Additionally, PLS can be applied to various measurement scales and remains effective even with relatively small sample sizes (Kline and Sentor, 1999). Chin (1998) states that the PLS path analysis model has three sets of relationships.

(1) The inner or structural model describes the relationships between latent variables represented by the following simultaneous equations.

$$\boldsymbol{\eta} = \boldsymbol{\beta\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\varsigma} \quad (1)$$

(2) The outer or measurement model describes how each indicator block relates to its latent variable. The equation for the reflective indicator model can be written as follows.

$$\mathbf{X} = \boldsymbol{\Lambda}_{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\delta}_{\mathbf{X}} \quad (2)$$

$$\mathbf{Y} = \boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}_{\mathbf{Y}} \quad (3)$$

$\boldsymbol{\eta}$ is an endogenous latent variable, $\boldsymbol{\xi}$ is an exogenous latent variable. $\boldsymbol{\beta}$ is the coefficient of the endogenous latent variable. $\boldsymbol{\Gamma}$ is the coefficient exogenous latent variable. $\boldsymbol{\varsigma}$ is the endogenous latent variable measurement error. $\mathbf{X}$ and $\mathbf{Y}$ are indicators of exogenous and endogenous latent variables. $\boldsymbol{\Lambda}_{\mathbf{X}}$ and $\boldsymbol{\Lambda}_{\mathbf{Y}}$ are measurement coefficient (loading factor). $\boldsymbol{\delta}_{\mathbf{X}}$ and $\boldsymbol{\varepsilon}_{\mathbf{Y}}$ are measurement errors from exogenous and endogenous indicators.

(3) Weight relation can calculate latent variable scores based on the inner and outer models. Case values for each latent variable are estimated in PLS as follows.

$$\xi_j = \sum_k w_{jk} X_{jk} \quad (4)$$

$$\eta_j = \sum_k w_{jk} Y_{jk} \quad (5)$$

w_{jk} is the weight used to estimate latent variables as a linear combination of their manifest variables.

The SEM-PLS model parameter estimation was conducted in several stages, including (Samani, 2016):

Stage 1: This stage was conducted repeatedly and produced two weight estimates, namely the measurement weight (outer weight) and structural weight (inner weight), through an iterative process.

Stage 1.1 Outside Approximation

The basic idea of this stage is to obtain a set of weights to estimate a latent variable. The initial step is to initialize all indicator weights to 1 (one). The calculation is formulated in the following equation:

$$Y_j = \sum_{k=1}^{K} \tilde{w}_{jk} X_{jk} \qquad (6)$$

Stage 1.2 Inside Approximation

This stage considers the relationship between latent variables in the inner model to obtain a new approximation of each latent variable calculated in the outside approximation as a weighted aggregate of other adjacent latent variables. The estimation of Z_j from a latent variable is formulated as $Z_j = \sum_{i\leftrightarrow j} v_{ji} Y_i$, with v_{ji} is inner weight, and the sign of $\leftrightarrow$ is the latent variable Y_j correlates with Y_i. The formula of v_{ji} uses path schema as follows:

$$v_{ji} = \begin{cases} corr\left(Y_j, Y_i\right), \ for\ Y_j\ and\ Y_i\ are\ close \\ Y_j = \sum_i v_{ji} Y_i, \ for\ v_{ji}\ in\ the\ regression\ Y_j\ and\ Y_i \\ 0,\ otherwise \end{cases} \qquad (7)$$

Stage 1.3 Update the Outer Weight

When the inside approximation stage is complete, the internal estimates must be reviewed for the indicators. This is done by updating the outer weight depending on the indicator block. The reflective relationship can be formulated as follows:

$$\hat{w}_{jk} = \left(\mathbf{Z_j^T Z_j}\right)^{\mathbf{-1}} \left(\mathbf{Z_j^T X_{jk}}\right) \qquad (8)$$

Step 1.4 Checking the Convergence

For each iteration *S= 1,2,3…,* the convergence is checked by comparing the outer weight at iteration stage S with the outer weight value at stage *(S-1)*-th. According to Sanchez (2013), the following limits are recommended: $\left| w_{jk}^{S-1} - w_{jk}^{S} \right| < 10^{-5}$ as a limit of convergence.

Stage 2:

Estimation of path coefficients between latent variables using the Ordinary Least Squares (OLS) method in corresponding multiple linear regression Y_j and Y_i .

$$\hat{\boldsymbol{\beta}}_{jk} = \left(\mathbf{Y_i^T Y_i}\right)^{\mathbf{-1}} \mathbf{Y_i^T Y_j} \qquad (9)$$

Stage 3:

In the third stage, the coefficients were estimated. The loading coefficients were obtained by calculating the correlation between the latent variables and their indicators.

$$\hat{\lambda}_{jk} = cor\left(X_{jk}, Y_j\right) \qquad (10)$$

2.2 Evaluation of SEM-PLS Model

1. Evaluate the Measurement Model (*Outer Model*)

The reflective measurement model was evaluated to demonstrate that the indicators used could represent the latent variables validly and reliably. The evaluation measures used are described in Table 1 below (Samani, 2016):

Table 1. The criteria of evaluation in the measurement model

Evaluation of Measurement	Parameter	Criteria
Indicator reliability	Loading factor	A loading factor value > 0.50 indicates a valid indicator representing the measured latent variables.
Construct reliability	Cronbach's alpha, ρ_c, ρ_A	The reliability value should be higher than 0.70, but values between 0.60 and 0.70 are acceptable.
Convergent validity	Average variance extracted (AVE)	An AVE value ≥ 0.50 indicates that the latent variable can explain at least half or more than 50% of the variance of the indicators.
Discriminant validity	Cross loading	The correlation between indicators and their constructs should be higher than that between other block constructs.

2. Evaluate the Structural Model (Inner Model)

Structural model evaluation was conducted to assess the relationship between latent variables and examine the strength of the model in representing endogenous latent variables with evaluation criteria outlined in the following table:

Table 2. The criteria of evaluation in the structural model

Evaluation of Measurement	Criteria
R-Square (R^2)	The value R^2 is classified into three categories: 0.67; 0.33; and 0.19 as substantial, moderate, and weak, respectively.
Goodness of Fit (GoF)	GoF is considered low when it is 0.10, moderate when it is 0.25, and high when it is 0.36.

2.3 Hypothesis Test of SEM-PLS

Partial Least Squares (PLS) does not assume data must follow a normal distribution. Hypothesis testing to assess the significance of path coefficients depends on nonparametric bootstrap procedures (Hair et al., 2017). The hypotheses in this study are as follows:

Statistical hypotheses for the outer model:

$H_0: \lambda_{kj} = 0 \qquad H_1: \lambda_{kj} \neq 0$

For the inner model:

$H_0: \beta_{ji} = 0 \qquad H_1: \beta_{ji} \neq 0$

$H_0: \gamma_{ji} = 0 \qquad H_1: \gamma_{ji} \neq 0$

The statistical test used is the t-test, shown in Equations (11) and (12):

$$t = \frac{\hat{\lambda}_{jk}}{SE\left(\hat{\lambda}_{jk}\right)} \text{, for outer model} \tag{11}$$

$$t = \frac{\hat{\beta}_{jk}}{SE\left(\hat{\beta}_{jk}\right)} \text{, or } t = \frac{\hat{\gamma}_{jk}}{SE\left(\hat{\gamma}_{jk}\right)} \text{ for inner model} \tag{12}$$

The testing criteria is rejected H_0 for $|t| > 1.96$ or $p - value < \alpha$.

2.4 Partial Least Squares-Prediction Oriented Segmentation (PLS-POS)

PLS-POS is a segmentation method oriented toward predicting relationships between constructs and specifically developed to complement path modeling in PLS. PLS-POS follows a clustering approach that deterministically places observations into groups and uses distance measures to re-place observations into homogeneous groups, so this method does not have distribution assumptions. It aims to improve the model's predictive power as seen from the R^2 value of endogenous latent variables (Becker et al., 2013). The total distance measures used in PLS-POS are as follows.

$$D_{kig} = \sum_{j=1}^{J} \sqrt{\frac{e_{jiq}^{2}}{\sum_{i=1}^{I_k} e_{jiq}^{2}}} \quad (13)$$

J is an endogenous latent variable; I_k is the size of the sample in the initial cluster k, and e_{jiq}^2 is the square residual from i observation in the alternative cluster $g(k \neq g; k, g \varepsilon G)$.

3. METHODOLOGY

3.1 Source of Data and Variable of Research

The data used in this study are secondary data obtained from publications by the Central Java Provincial Statistics Agency (BPS). The data include information on the welfare of the people of Central Java Province in 2023, with 35 observations representing each district/city. The research variables consist of two exogenous latent variables and three endogenous latent variables.

Table 3. Latent variable and its indicators

Latent Variable	Indicator	
Education (ξ_1)	X_{11}	The average length of school
	X_{12}	The Expectation Rate for School Duration
	X_{13}	The percentage of the population aged 15 or above by highest education attainment
	X_{14}	Average monthly expenditure per capita education
Housing (ξ_2)	X_{21}	The percentage of living conditions properly
	X_{22}	The percentage of households with the main water resource is from the dwelling.
	X_{23}	Percentage of Access to Adequate Sanitation
	X_{24}	The percentage of households with sanitation is the septic tank
Health (η_1)	Y_{11}	The life expectancy
	Y_{12}	Percentage of Population with BPJS Health Insurance Coverage Who Are Not Recipients of Premium Assistance (Non-PBI)
	Y_{13}	Percentage of women aged 15–49 who have ever been married and gave birth with medical assistance
	Y_{14}	Percentage of Non-Smokers
Employment (η_2)	Y_{21}	Level of Labor Force Participation
	Y_{22}	Unemployment rate
	Y_{23}	Percentage of the workforce with secondary education
	Y_{24}	Percentage of Population Aged 15 Years and Above Working 35-48 Hours
	Y_{25}	The average Net Wages/Salaries of Formal Workers
Welfare (η_3)	Y_{31}	Human Development Index
	Y_{32}	Gender Development Index
	Y_{33}	Percentage of Non-Food Expenditure per Capita
	Y_{34}	Gini ratio

The following conceptual framework illustrates the relationship between latent variables and the indicators that represent them.

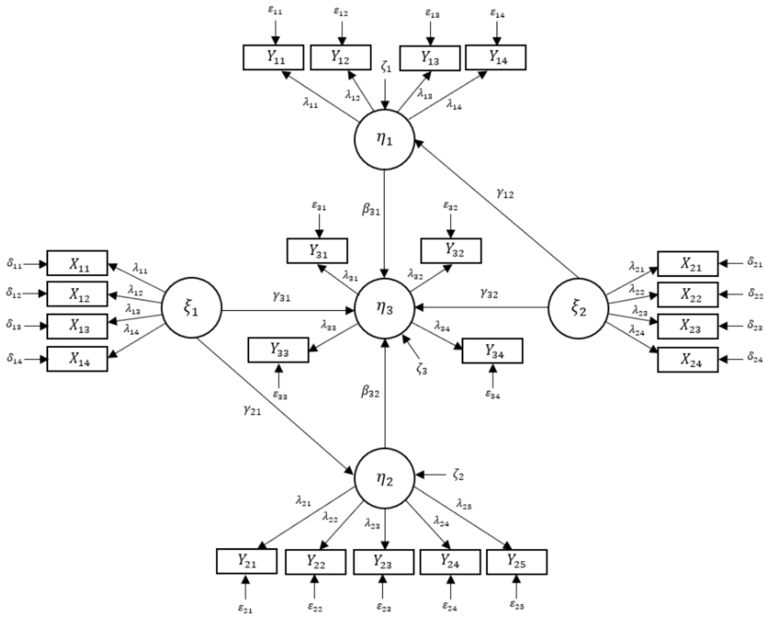

Fig. 1 Research Conceptual Framework

3.2 The Steps of Data Analysis

Data analysis in this study used the software SmartPLS4 in the following stages.

1. Conceptualize the model by designing a measurement model and a structural model.
2. Construct a path diagram that explains the relationship patterns between latent variables and their indicators and the causal relationships between exogenous and endogenous variables.
3. Convert the path diagram results into a mathematical equation model.
4. Estimate model parameters, namely the weights and scores of latent variables, path, and loading coefficients.
5. Evaluate the measurement model until the indicators are valid and reliable. If indicators do not meet the validity and reliability criteria, a dropping process is required to obtain appropriate indicators.
6. Reconstructing the path diagram by improving the structural model if significance has not been achieved or indicators are not valid and reliable.
7. Evaluating the structural model.
8. Testing hypotheses using the bootstrapping method.
9. Interpreting the analysis results. Grouping regions using the Partial Least Squares-Prediction Oriented Segmentation (PLS-POS) method.

4. RESULTS AND DISCUSSION

4.1 Evaluate the Measurement Model (Outer Model) I

The evaluation stage of the outer model in Model I shows that the indicators Y_{21} and Y_{22} do not meet the indicator reliability criteria because the loading factor value is < 0.5. The Employment latent variable also has a Composite Reliability value ($\hat{\rho}_c$), Cronbach's Alpha, and a value $\hat{\rho}_A \leq 0.7$. The AVE value for the latent variable is recorded below 0.5. In the discriminant validity evaluation, the indicators Y_{22} did not meet the criteria because the correlation values between the indicators and their constructs were lower than those of other construct blocks. Based on the results of the outer model evaluation, all criteria were not met, so dropping was performed by removing the indicators Y_{21} and Y_{22} . Reconstruction was carried out after the dropping process was completed to form Model II.

Evaluate the Measurement Model (Outer Model) I

a. Indicator Reliability

All indicators have a loading factor value > 0.5, indicating that each indicator is valid in measuring its latent variable.

Table 4. Nilai Loading Factor Value

Latent Variable	**Indicator**	***Loading Factor***	**Conclusion**
Education	X_{11}	0.973	Valid
	X_{12}	0.963	Valid
	X_{13}	0.951	Valid
	X_{14}	0.898	Valid
Housing	X_{21}	0.893	Valid
	X_{22}	0.714	Valid
	X_{23}	0.960	Valid
	X_{24}	0.911	Valid
Health	Y_{11}	0.837	Valid
	Y_{12}	0.920	Valid
	Y_{13}	0.852	Valid
	Y_{14}	0.828	Valid
Employment	Y_{23}	0.867	Valid
	Y_{24}	0.810	Valid
	Y_{25}	0.755	Valid
Welfare	Y_{31}	0.821	Valid
	Y_{32}	0.785	Valid
	Y_{33}	0.963	Valid
	Y_{34}	0.841	Valid

b. Construct Reliability

All latent variables have a composite reliability ($\hat{\rho}_c$), Cronbach's Alpha, and the value $\hat{\rho}_A$ exceeds 0.7. These results indicate that each construct meets the reliability criteria in measuring latent variables.

Table 5. Construct Reliability Model II

Latent Variable	$\hat{\boldsymbol{\rho}}_{\boldsymbol{c}}$	$\hat{\boldsymbol{\rho}}_{\boldsymbol{A}}$	***Cronbach's Alpha***
Education	0.972	0.964	0.961
Housing	0.928	0.950	0.896
Health	0.919	0.891	0.882
Employment	0.853	0.754	0.741
Welfare	0.915	0.896	0.875

c. Convergent Validity

The AVE values for each latent variable indicate that the latent variables of education, housing, health, employment, and welfare are 0.896, 0.764, 0.740, 0.660, and 0.731, respectively. All AVE values are greater than 0.5, indicating that the criterion for convergent validity is met.

d. Discriminant Validity

The cross-loading values indicate that each indicator correlates more with its construct than with other constructs, thus concluding that the discriminant validity criterion is met.

Evaluate the Structural Model (Inner Model) II

R^2 is used to assess the extent to which endogenous variables can be explained by exogenous variables in the model. The value of R^2 in the latent variable of health (η_1) is 0.418, indicating a moderate relationship between variables, while the value of R^2 or the latent variable of employment (η_2) is 0.810, indicating a substantial relationship. The value of R^2 for the latent variable of welfare (η_3) is 0.818, also indicating a substantial relationship.

The subsequent evaluation, Goodness of Fit (GoF), was used to validate the structural model as a whole. Based on the processing results, a GoF value of 0.719 was obtained, which falls into the high GoF category. It indicates that the resulting model can explain empirical data excellently.

4.2 Hypothesis Testing

Testing for each path between latent variables was conducted based on the significance level $\alpha = 5\%$, and the criteria of testing is to reject H_0 if $|t| > 1.96$ atau $p - value < 0.05$.

Table 6. Testing the Path Coefficient between Latent Variables in the Inner Model

Relationships Between Latent Variables	Path Coefficient	The estimation of Path Coefficient	t	*p-value*	Decision	Conclusion
Housing → Health	γ_{12}	0.647	9.785	< 0.001	Reject H_0	Significant
Education → Employment	γ_{21}	0.905	27.306	< 0.001	Reject H_0	Significant
Education → Welfare	γ_{31}	0.854	3.398	0.001	Reject H_0	Significant
Housing → Wealth	γ_{32}	-0.251	1.965	0.049	Reject H_0	Significant
Health → Welfare	β_{31}	0.605	2.533	0.011	Reject H_0	Significant
Employment → Welfare	β_{32}	-0.460	1.977	0.048	Reject H_0	Significant

Structural model of the influence of housing on health.

$$\eta_1 = \gamma_{12}\xi_2 + \zeta_2$$

$$\eta_1 = 0.647\xi_2 + \zeta_2$$

Structural model of the influence of education on employment.

$$\eta_2 = \gamma_{21}\xi_1 + \zeta_1$$

$$\eta_2 = 0.905\xi_1 + \zeta_1$$

Structural model of the influence of education, housing, health, and employment on welfare.

$$\eta_3 = \gamma_{31}\xi_3 + \gamma_{32}\xi_3 + \beta_{31}\xi_3 + \beta_{32}\xi_3 + \zeta_3$$

$$\eta_3 = 0.854\xi_3 - 0.251\xi_3 + 0.605\xi_3 - 0.460\xi_3 + \zeta_3$$

4.3 Clustering the Areas Use PLS-POS Approach

Clustering into three segments yielded the best results based on the highest average weighted R-squared value. Each segment exhibits different characteristics depending on the dimensions relevant to the welfare of the people in each district/city in Central Java province.

Table 7. Clustering area based on POS-PLS

Cluster	**District**	
Cluster 1	Cilacap	Jepara
	Kebumen	Pekalongan
	Purworejo	Pemalang
	Magelang	Brebes
	Grobogan	Surakarta City
	Pati	Semarang City
Cluster 2	Banyumas	
	Banjarnegara	
	Wonogiri	
	Tegal	
	Magelang City	
	Salatiga City	
Cluster 3	Purbalingga	Kudus
	Wonosobo	Demak
	Boyolali	Semarang
	Klaten	Temanggung
	Sukoharjo	Kendal
	Karanganyar	Batang
	Sragen	Pekalongan City
	Blora	Tegal City
	Rembang	

After districts/cities were grouped based on the segments formed, model estimation was performed separately for each segment using the PLS approach.

4.4 The Heterogeneity in Structural Model

Table 8. Comparison of path coefficient

The relation of Variables	***Original Sample***	**Cluster 1**	**Cluster 2**	**Cluster 3**
Housing → Health	0.647	0.678	0.851	0.863
Education → Employment	0.905	0.908	0.977	0.912
Education → Welfare	0.438	1.282	-0.128	1.441
Housing → Wealth	0.140	0.010	-0.832	-1.040
Health → Welfare	0.605	-0.206	2.409	1.382
Employment → Welfare	-0.460	-0.129	-0.687	-1.255

Based on Table 6, cluster 1 shows that education plays a dominant role in increasing employment so that improvements in welfare can be focused on improving the quality of education. In segment 2, housing contributes positively to health but has a negative impact on welfare. On the other hand, health is the primary factor driving welfare, so efforts to improve welfare should be focused on improving health through better housing quality. In cluster 3, education and health contribute positively to welfare, while employment has a negative impact. Therefore, welfare improvement strategies in this cluster can be directed toward improving the quality of education and health services.

4.5 The Heterogeneity in Measurement Model

Table 9. Comparison of Loading Factor

Latent Variable		*Original Sample*	Cluster 1	Cluster 2	Cluster 3
Education	X_{11}	0.973	0.977	0.974	0.966
	X_{12}	0.963	0.988	0.987	0.926
	X_{13}	0.951	0.981	0.966	0.955
	X_{14}	0.898	0.935	0.886	0.881
Housing	X_{21}	0.893	0.612	0.884	0.904
	X_{22}	0.714	0.878	-0.092	0.843
	X_{23}	0.960	0.782	0.903	0.975
	X_{24}	0.911	0.385	0.830	0.968
Health	Y_{11}	0.837	0.865	0.832	0.849
	Y_{12}	0.920	0.961	0.960	0.893
	Y_{13}	0.852	0.924	0.895	0.782
	Y_{14}	0.828	0.934	0.715	0.860
Employment	Y_{23}	0.867	0.840	0.893	0.872
	Y_{24}	0.810	0.854	0.889	0.832
	Y_{25}	0.755	0.732	0.925	0.562
Welfare	Y_{31}	0.821	0.957	0.863	0.491
	Y_{32}	0.785	0.818	0.846	0.818
	Y_{33}	0.963	0.979	0.970	0.952
	Y_{34}	0.841	0.860	0.904	0.882

Based on Table 7, it can be concluded that, in general, the value of the loading factor in three clusters is > 0.5, but not for X_{24} in cluster 1, X_{22} in cluster 2, and Y_{31}in cluster 3.

4.6 Model Evaluation

Table 10. The comparison of R^2

Endogenous Latent Variable	R^2			
	Original Sample	Cluster 1	Cluster 2	Cluster 3
Health (η_1)	0.418	0.459	0.725	0.745
Employment (η_2)	0.810	0.967	0.994	0.929
Welfare (η_3)	0.818	0.824	0.955	0.833
Mean	0.682	0.750	0.891	0.836

The R^2 values for each segment formed (local model) increase compared to the global model values for each endogenous latent variable. In addition, model evaluation was also carried out using Goodness of Fit (GoF) values to validate the model as a whole and to obtain a local model that is better than the global model. The GoF value for the global model is 0.719, Cluster 1 obtained a GoF value of 0.750, Cluster 2 obtained a GoF value of 0.827, and Cluster 3 obtained a GoF value of 0.784. It is concluded that the GoF values are suitable for the global model. Cluster 1, Cluster 2, and Cluster 3 are in the large GoF category. The high GoF values also indicate better structural and measurement models. Additionally, the GoF values for each segment are better than the GoF value for the global model, indicating that the model is better at the local level, thereby allowing the heterogeneity in this study to be detected effectively.

5. CONCLUSION

The conclusions drawn from this study indicate that there are 19 significant indicators in the structural model involving the latent variables of education, housing, health, employment, and welfare. The structural equation model shows significant pathways, namely the influence of housing on health, the influence of education on employment, the influence of education on welfare, the influence of housing on welfare, the influence of health on welfare, and the influence of employment on welfare. Using the PLS-POS approach, three segments with different levels of influence were identified: Cluster 1 consists of 12 districts/cities, Cluster 2 consists of 6 districts/cities, and Cluster 3 consists of 17 districts/cities. Each local model has a higher R^2 value and Goodness of Fit than the global model, thereby better explaining the endogenous variables and demonstrating that the PLS-POS segmentation approach can optimally detect heterogeneity in SEM-PLS models.

References

Todaro, M. P., & Smith, S. C. (2020). Economic Development (13th ed.). Pearson Education Limited. Harlow, United Kingdom: Pearson.

United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1). General Assembly, Seventieth session.

Blau, J., & Abramovitz, M. (2010). The dynamics of social welfare policy. Oxford University Press, USA.

Statistics Indonesia-Central Java Province. (2023). Report: Statistics Central Java Province 202. Central Java, Indonesia: Badan Pusat Statistik (BPS).

Rodríguez-Pose, A., & Hardy, D. (2015). Addressing poverty and inequality in the rural economy from a global perspective. Applied Geography, 61, 11-23.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies!. Journal of business research, 69(10), 3998-4010.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., ... & Ray, S. (2021). An introduction to structural equation modeling. Partial least squares structural equation modeling (PLS-SEM) using R: a workbook, 1-29.

Aghili, N., & Amirkhani, M. (2021). SEM-PLS approach to green building. Encyclopedia, 1(2), 472-481.

Nasution, M. I., Fahmi, M., & Prayogi, M. A. (2020, March). The quality of small and medium enterprises performance using the structural equation model-part least square (SEM-PLS). In Journal of Physics: Conference Series (Vol. 1477, No. 5, p. 052052). IOP Publishing.

Ali, F., Hussain, K., Konar, R., & Jeon, H. M. (2017). The effect of technical and functional quality on guests' perceived hotel service quality and satisfaction: A SEM-PLS analysis. Journal of Quality Assurance in Hospitality & Tourism, 18(3), 354-378.

Ahmad, A., & Armawaddin, M. (2024). SEM-PLS: Predicting Community Welfare and Unemployment Using Income Inequality. Economics Development Analysis Journal, 13(2), 275-288.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. Psychological methods, 10(1), 21.

Becker, J. M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats. MIS quarterly, 665-694.

Ștefan, S. C., Olariu, A. A., & Popa, S. C. (2024). Implications of artificial intelligence on organizational agility: A PLS-SEM and PLS-POS Approach. Amfiteatru Economic, 26(66), 403-420.

Arenas-Gaitán, J., Villarejo Ramos, A. F., & Peral-Peral, B. (2020). A posteriori segmentation of elderly internet users: applying PLS-POS. Marketing Intelligence & Planning, 38(3), 340-353.

Ratzmann, M., Gudergan, S. P., & Bouncken, R. (2016). Capturing heterogeneity and PLS-SEM prediction ability: Alliance governance and innovation. Journal of Business Research, 69(10), 4593-4603.

Kline, R. B., & Santor, D. A. (1999). Principles & practice of structural equation modelling. Canadian Psychology, 40(4), 381.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. Modern methods for business research, 295(2), 295-336.

Stein, C. M., Morris, N. J., & Nock, N. L. (2011). Structural equation modeling. In Statistical human genetics: Methods and protocols (pp. 495-512). Totowa, NJ: Humana Press.

Samani, S. A. (2016). Steps in research process (partial least square of structural equation modeling. International Journal of Social Science and Business, 1(2), 55-66.