CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 2494–2514
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Identifying Regional Heterogeneity and Priority Factors for Infrastructure in Indonesia Using FIMIX-PLS and Importance-Performance Map Analysis

Arum Nanda Putri Aprilia and Rahmawati Erma Standsyah*

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Indonesia

Abstract

Infrastructure development is important for reducing regional disparities, but Indonesian provinces differ in economic conditions, human resources, and regional characteristics. This study examines the relationships between the economy, Human Resources, and Regional Characteristics and Infrastructure in the 38 provinces of Indonesia, based on secondary data from Statistics Indonesia (BPS) for 2024 and 2025. To examine the structural relationships, Partial Least Squares Structural Equation Modeling (PLS-SEM) was employed, and to investigate unobserved heterogeneity, Finite Mixture PLS (FIMIX-PLS) was used together with Importance–Performance Map Analysis (IPMA) to determine the priorities within the segments. The model accounted for 82.3% of the variance in Infrastructure ($R^2 = 0.823$). Yet none of the structural paths were found to be statistically significant at the 5% level: for Economy ($t = 1.195$, $p = 0.232$), for Human Resources ($t = 0.895$, $p = 0.371$), and for Regional Characteristics ($t = 1.928$, $p = 0.054$). FIMIX-PLS identified a two-segment solution, with the final partition assigning 27 provinces to Segment 1 and 11 provinces to Segment 2. IPMA showed different patterns across segments: Regional Characteristics had the highest importance in Segment 1, while Economy ranked highest in Segment 2. Regional Characteristics also had the lowest performance in both segments. As this construct represents disaster exposure rather than development quality, the results highlight the need to consider disaster risks and economic differences when setting infrastructure priorities. However, the findings remain exploratory because none of the overall structural paths were statistically significant, and the segmentation is exploratory.

Keywords: FIMIX-PLS; Infrastructure; IPMA; PLS-SEM.

Copyright © 2026 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1. Introduction

Sustainable development through the Sustainable Development Goals (SDGs) ranks infrastructure as one of the key foundations for stimulating economic growth, improving the well-being of the population, and expanding access to various basic services [1, 2]. The availability of adequate infrastructure is an important factor in strengthening competitiveness, promoting more inclusive development, and enhancing interregional connectivity, in addition to facilitating economic activities [3, 4]. That is why infrastructure development continues to be a strategic priority of many countries, including Indonesia, as part of its efforts to support sustainable development.

*Corresponding author. E-mail: rahmawatistandsyah@unesa.ac.id

Received: August 12, 2026 Revised: October 14, 2026 Accepted: October 22, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.45621

Although Indonesia has made infrastructure development a national priority, equitable development continues to face various challenges [5, 6]. Disparities in geographical conditions, regional characteristics, economic capacities, and development levels among provinces result in uneven development of the quality and availability of infrastructure [7]. This situation leads to inequalities in access to essential services, economic activities, and development opportunities, particularly in regions outside of Java [4]. Therefore, infrastructure development requires a region-specific approach to ensure that implemented policies are more effective and promote inclusive and sustainable development [4, 8].

In recent years, the use of finite mixture partial least squares (FIMIX-PLS) has expanded considerably for identifying unobserved heterogeneity across various research fields [9–11]. FIMIX-PLS allows data to be grouped into segments with different relational structures, so that the resulting model better represents the characteristics of each group. Simultaneously, FIMIX-PLS has begun to be combined with importance-performance map analysis (IPMA) to determine improvement priorities based on the characteristics of each segment [12–14]. This approach has been widely applied in various research areas. Previous studies have shown that combining FIMIX-PLS and IPMA not only reveals hidden heterogeneity but also provides more accurate information for strategy development compared to the use of conventional PLS-SEM.

Research on underdeveloped regions and infrastructure development in Indonesia has generally relied on a single model to represent the entire study area, as in studies focusing on Java Island [15, 16]. This approach provides a global view of the relationships between variables at the population level but overlooks unobserved heterogeneity across regions. Such heterogeneity is plausible in Indonesia, where regions differ markedly in economic structure and development levels [8], and where infrastructure is exposed to different hazards [17], so the factors driving infrastructure development may differ from one region to another.

Although FIMIX-PLS and IPMA have been combined in previous studies, for instance in the context of social media marketing [14], their application to provincial infrastructure development in Indonesia remains limited. This study therefore integrates partial least squares structural equation modeling (PLS-SEM) [18], finite mixture partial least squares (FIMIX-PLS) [19], and importance-performance map analysis (IPMA) [20], and contributes in three ways. First, it examines the effects of the economy, human resources, and regional characteristics on infrastructure across Indonesian provinces. Second, it uncovers latent regional heterogeneity by identifying segments of provinces in which these relationships differ. Third, it derives segment-specific improvement priorities through IPMA. These contributions support more selective, differentiated infrastructure policies rather than uniform national approaches. Accordingly, this study aims to analyze the effects of the economy, human resources, and regional characteristics on infrastructure using PLS-SEM, identify unobserved heterogeneity among Indonesian provinces using FIMIX-PLS, and determine segment-specific infrastructure improvement priorities using IPMA. The results are expected to inform infrastructure development policies tailored to the characteristics of each group of provinces, so that the adopted policies can be more selective.

2. Methods

This study uses a quantitative method to analyze 38 provinces in Indonesia using Partial Least Squares-Structural Equation Modeling (SEM-PLS). After checking the validity and reliability criteria, Finite Mixture Partial Least Squares (FIMIX-PLS) analysis was used to detect the existence of unobserved heterogeneity and assign the provinces to segments with homogeneous characteristics. Then, an Importance-Performance Map Analysis (IPMA) was carried out for each segment to determine the priority indicators of infrastructure development in Indonesia.

2.1. Data Source

The research used secondary data from official publications of Statistics Indonesia (BPS). The unit of analysis is 38 provinces of Indonesia. The data used in this study were obtained for

2024 and 2025 from several publications of BPS, including Potensi Desa (PODES) and other statistical reports, which provide indicators that are related to the latent variables used in this research. These data from different sources provided full information on the situation of the infrastructure, economy, human resources, and regional characteristics of each province.

Data were available for all 38 provinces in both 2024 and 2025. Because the study aims to represent the general condition of each province during the 2024–2025 period rather than examine temporal dynamics and therefore does not use time-series or panel methods, the values for the two years were combined into a single observation for each province. For indicators reported as the number of villages or urban villages with a particular characteristic, a scale adjustment was performed separately for each province and year before the two-year values were averaged. The scale-adjusted value was calculated as:

$$X_{it}^{*} = \frac{X_{it}}{V_{it}} \times 100 \tag{1}$$

In Eq. (1), X_{it} denotes the number of villages or urban villages with the relevant characteristic in province i and year t, V_{it} denotes the total number of villages or urban villages in the corresponding province and year, and X_{it}^{*} denotes the scale-adjusted value expressed per 100 villages or urban villages. The final value for the 2024–2025 period was then calculated as:

$$X_{i} = \frac{X_{i,2024}^{*} + X_{i,2025}^{*}}{2} \tag{2}$$

In Eq. (2), X_i represents the final scale-adjusted indicator value for province i. This procedure was applied to reduce the potential influence of differences in provincial administrative size on indicators expressed as village or urban-village counts. Other indicators were retained in their respective measurement units as reported by BPS, including monetary, percentage, and year-based measures. Data preprocessing and calculation of the scale-adjusted indicators were performed using Microsoft Excel.

The operationalization of the constructs was re-examined to ensure stronger alignment with the conceptual framework underlying the identification of underdeveloped areas. The indicator selection was primarily guided by the *Petunjuk Teknis Penentuan Indikator dalam Penetapan Daerah Tertinggal secara Nasional* (Technical Guidelines for Determining Indicators in the National Designation of Underdeveloped Regions) issued in 2016 by the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration of the Republic of Indonesia [21], which identifies economic, human resources, infrastructure, and regional characteristics as dimensions relevant to regional underdevelopment. Previous empirical studies that operationalized this framework [15, 16] were also considered as supporting references for indicator selection. The revised indicator specification was therefore theoretically and conceptually motivated rather than determined by the statistical performance of the indicators in the PLS-SEM model.

The study comprises four constructs: Economy, Human Resources, Infrastructure, and Regional Characteristics. Infrastructure is specified as the endogenous construct, whereas Economy, Human Resources, and Regional Characteristics are specified as exogenous constructs. Based on the final operationalization, Economy is represented by per-capita expenditure and the percentage of poor population. Human Resources is represented by life expectancy, average years of schooling, and literacy rate. Infrastructure is represented by four indicators reflecting different aspects of infrastructure provision, including road surface conditions, the availability of markets without buildings, health facilities, and basic education facilities. Regional Characteristics is represented by the selected indicators of villages or urban villages affected by different types of natural disasters, including earthquakes, floods, whirlwinds, and drought. Table 1 presents the constructs and indicators used in the study.

Table 1: Latent Variables and Measurement Indicators

Latent Variable	Indicator	Code	Data Source
Infrastructure	Asphalt / Concrete Road	I1	[22, 23]
	Market Without Buildings	I2	
	Health Facility	I3	
	Basic Education Facilities	I4	
Economy	Per capita Expenditure	E1	[24]
	Percentage of Poor Population	E2	[25]
Human Resources	Life Expectancy	H1	[26]
	Average years of schooling	H2	[27]
	Literacy Rate	H3	[28]
Regional Characteristics	Earthquake	R1	[22, 23]
	Flood	R2	
	Whirlwind / Puting Beliung / Typhoon	R3	
	Drought	R4	

2.2. SEM-PLS

Multivariate analysis is a statistical approach that allows for the simultaneous analysis of several variables, thus revealing more complex relationships than univariate or bivariate analyses. Among the most widely used multivariate analysis methods is structural equation modeling (SEM), which examines how a latent construct is measured by several observed indicators [18]. This research uses partial least squares-structural equation modeling (PLS-SEM) to analyze the relationship between economic factors, human resources, regional characteristics, and infrastructure development in Indonesia. PLS-SEM allows the measurement and structural models to be estimated within the same analytical framework and can accommodate models containing formative constructs [18]. The analysis was conducted using SmartPLS version 4.1.1.8.

All four constructs—Economy, Human Resources, Infrastructure, and Regional Characteristics—were specified as formative. This decision was based on the conceptual role of the indicators, which represent distinct aspects that jointly define each construct rather than interchangeable manifestations of an underlying latent variable. The distinction between formative and reflective measurement was considered in accordance with the conceptual criteria discussed by [29, 30]. The analysis comprised two stages: evaluation of the formative measurement models and evaluation of the structural model.

2.2.1. Evaluation of the Measurement Model (Outer Model)

The measurement model was evaluated using criteria appropriate for formative constructs. The assessment focused on indicator collinearity. Outer weights and outer loadings were also examined [18].

a. **Indicator Collinearity**
Indicator collinearity was assessed using the variance inflation factor (VIF). This assessment was used to identify potential collinearity among indicators within each formative construct. Excessive collinearity can make it difficult to distinguish the individual contribution of each indicator [18].

b. **Outer Weight**
Outer weights were examined to assess the relative contribution of each indicator to its formative construct. Their statistical significance was evaluated using bootstrapping. A significant outer weight indicates that an indicator makes a statistically significant unique contribution to the construct [18].
However, an indicator was not automatically removed when its outer weight was nonsignificant. Its outer loading and theoretical relevance were also considered, particularly because

the indicators represent distinct dimensions of regional underdevelopment. Removing an indicator solely because its outer weight was nonsignificant could reduce the content coverage of the construct [18, 29].

c. **Outer Loadings**
Outer loadings were examined to assess the absolute contribution of each indicator to its associated construct, particularly when its outer weight was nonsignificant. The magnitude and statistical significance of the outer loadings were considered alongside the conceptual relevance of the indicators when making retention decisions [18]. The final indicator specification was based on the conceptual framework and operationalization described in Section 2.1. Statistical results were used to evaluate the specified formative measurement models rather than to determine the conceptual composition of the constructs.

2.2.2. Evaluation of the Structural Model (Inner Model)

After the measurement models had been evaluated, the structural model was assessed to examine the relationships among the constructs [18]. Economy, Human Resources, and Regional Characteristics were specified as exogenous constructs, while Infrastructure was specified as the endogenous construct. The structural model was evaluated using the path coefficients, the coefficient of determination (R^2), the effect size (f^2), and bootstrapping to assess the statistical significance of the estimated relationships [18].

a. **Coefficient of Determination (R^2)**
The coefficient of determination (R^2) shows how well a structural model explains the variance of an endogenous variable using the exogenous variables that affect it. A higher R^2 means the model explains the endogenous variable more effectively [18]. The interpretation of R^2 values, which vary from 0 to 1, depends on the research field and model complexity. In marketing and social sciences, R^2 values of 0.75, 0.50, and 0.25 are regarded as high, moderate, and low, respectively [18]. However, to make sure the model has both strong predictive power and high parsimony, R^2 should be interpreted alongside other structural model evaluation criteria [18].

b. **Effect Size (f^2)**
The effect size known as f^2 is a measure of the amount that each exogenous variable contributes to an endogenous variable in a structural model, whereas R^2 examines the model's overall predictive power, f^2 is concerned with the strength of the influence of an exogenous variable on a particular endogenous variable [18]. The f^2 value is obtained by comparing the R^2 value of the endogenous variable before and after removing an exogenous variable from the model. This comparison indicates the exogenous variable's contribution to explaining the variance of the endogenous variable. The f^2 value is calculated using the following equation [18].

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \tag{3}$$

In Eq. (3), $R^2_{included}$ represents the coefficient of determination when the exogenous variable is included in the structural model, while $R^2_{excluded}$ represents the coefficient of determination after its exclusion from the model. The difference between these two values indicates the significance of the exogenous variable's contribution to the predictive power of the endogenous variable [18]. The f^2 value allows us to assess the relative contribution of an exogenous variable to explaining the endogenous variable. f^2 values of 0.02, 0.15, and 0.35 indicate small, medium, and large effects, respectively, while values below 0.02 indicate a negligible contribution from the exogenous variable [18].

c. **Bootstrapping**
The statistical significance of the estimated path coefficients was assessed by means of

bootstrapping. For this purpose, 5,000 subsamples were used, a two-tailed test was carried out, and the significance level was set at 5%. The standard errors, t-statistics, and p-values obtained through bootstrapping were employed in the evaluation of the statistical significance of the relationships between the constructs [18].

The study included all 38 provinces of Indonesia within its defined geographical scope. Accordingly, the observed data covered the complete set of provincial units examined in this study. Bootstrapping was used to assess the uncertainty of the estimated model parameters under resampling. The resulting p-values were interpreted as evidence from the bootstrap procedure, rather than as evidence that the 38 provinces constituted a random sample from a larger population of provinces. Statistical significance was assessed using a two-tailed test at the 5% level. A relationship was considered statistically significant when its p-value was below 0.05; a t-statistic greater than 1.96 was used as the corresponding conventional large-sample reference for a two-tailed test [18].

2.3. FIMIX-PLS

Finite Mixture Partial Least Squares (FIMIX-PLS) is a segmentation technique used to examine unobserved heterogeneity in Partial Least Squares Structural Equation Modeling (PLS-SEM). Unlike conventional PLS-SEM, which estimates structural relationships for the overall sample, FIMIX-PLS accounts for the possibility that observations belong to latent segments with different structural relationships. Segment membership is represented by probabilities, allowing each observation to have a probability of belonging to each segment [19].

In this study, FIMIX-PLS was applied to the data from 38 Indonesian provinces to examine potential heterogeneity in the structural relationships between Economy, Human Resources, Regional Characteristics, and Infrastructure. Candidate solutions with two to five segments (K = 2, 3, 4, 5) were evaluated. Each candidate solution was estimated using 10 repeated runs to reduce dependence on a single initialization and to assess convergence consistency.

The candidate solutions were compared using multiple model-selection and classification criteria, including the Akaike information criterion (AIC), modified Akaike information criteria (AIC3 and AIC4), Bayesian information criterion (BIC), consistent Akaike information criterion (CAIC), Hannan–Quinn criterion (HQ), minimum description length criterion with factor 5 (MDL5), log-likelihood (LnL), normed entropy statistic (EN), non-fuzzy index (NFI), and normalized entropy criterion (NEC). Lower values are generally preferred for the information criteria and NEC, whereas higher LnL, EN, and NFI values generally indicate better model fit or classification quality, depending on the criterion. The information criteria provide complementary assessments of model fit and complexity, while EN and NFI provide information on the clarity of segment classification. Because the criteria may favor different numbers of segments, the final solution was selected by considering their overall pattern, classification quality, interpretability, and model parsimony rather than relying on a single criterion.

After selecting the solution for further analysis, each province was assigned to the segment with the highest estimated membership probability. The estimated segment proportions and the final province assignments were examined separately because model-estimated membership proportions and counts based on maximum-probability classification represent different aspects of the segmentation. Finally, Importance–Performance Map Analysis (IPMA) was performed separately for each segment to examine the relative importance and performance of the antecedent constructs in relation to Infrastructure.

2.4. IPMA

Importance-Performance Map Analysis (IPMA) is an evolution of PLS-SEM analysis that provides additional information on improvement priorities for a construct or indicator. Unlike PLS-SEM results, which only show the magnitude of influence between constructs, IPMA combines two dimensions: importance and performance. It thus helps identify factors with a strong influence

but a relatively low level of performance. IPMA can therefore serve as a basis for determining improvement priorities for constructs and indicators contributing to the target construct [20].

In IPMA, the importance value is calculated from the total effects generated by the PLS-SEM model. This value indicates the magnitude of a construct's contribution to the target construct. Concurrently, the performance value is derived from the latent variable score and converted to a 0 to 100 scale to facilitate interpretation. Integrating these two measures enables visualization of each construct's position based on its importance and performance levels [20].

In this study, importance-performance map analysis (IPMA) was applied separately to each segment generated by the FIMIX-PLS analysis. This approach allows for the evaluation of the importance and performance of each variable based on the characteristics of each segment. Therefore, the resulting improvement priorities are not based solely on overall data but also take into account the differences in characteristics between segments, thus enabling the formulation of more specific and targeted recommendations.

3. Results and Discussion

The research results are presented in stages, including the evaluation of the PLS-SEM, FIMIX-PLS analysis, and importance-performance map analysis (IPMA). The discussion at each stage is conducted in a way that explains the main conclusions of this study. Fig. 1 presents the formative measurement model used in this study to examine provincial infrastructure in Indonesia. The model comprises four constructs: Economy, Human Resources, Infrastructure, and Regional Characteristics. Economy, Human Resources, and Regional Characteristics are specified as exogenous constructs, while Infrastructure is the endogenous construct. The model is designed to examine how these complementary dimensions relate to infrastructure and to provide a basis for investigating heterogeneity in these relationships across provinces.

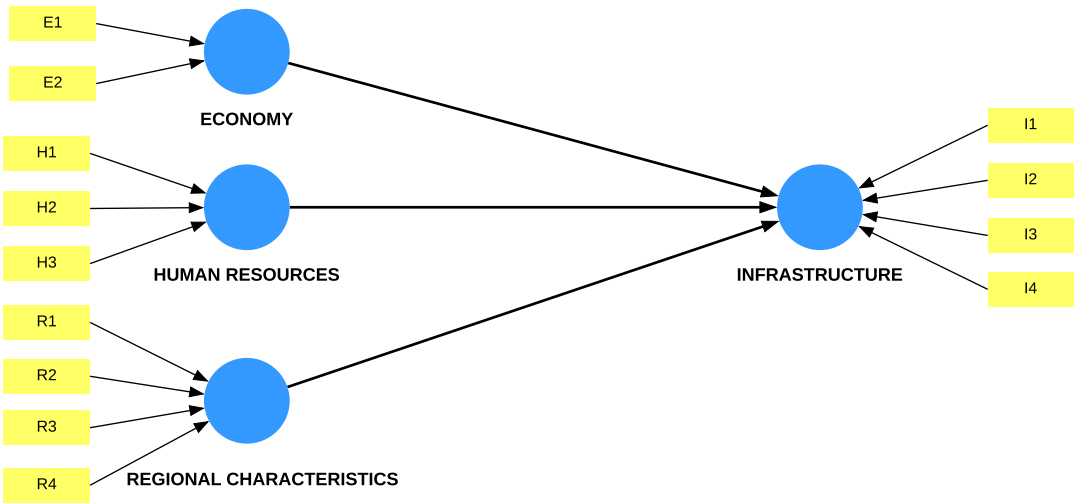


Fig. 1: Conceptual Research Model

The selection of the four constructs and their indicators was primarily guided by the *Petunjuk Teknis Penentuan Indikator dalam Penetapan Daerah Tertinggal secara Nasional* (Technical Guidelines for Determining Indicators in the National Designation of Underdeveloped Regions) issued in 2016 by the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration of the Republic of Indonesia [21]. The guidelines provide a conceptual framework that groups regional underdevelopment indicators into several dimensions, including economy, human resources, infrastructure, and regional characteristics [21]. This study adapts that framework to the provincial level, taking into account the research objectives and the availability of comparable provincial data. Two relevant empirical studies have examined regional underdevelopment indica-

tors using the 2016 guidelines, providing a basis for comparison with this study [15, 16]. However, their use of confirmatory factor analysis differs from the formative measurement specification adopted in this study. Their work is therefore cited as a reference for indicator operationalization rather than as methodological justification for the formative model.

The formative specification was determined by the conceptual meaning of each construct rather than by statistical results alone. In a formative measurement model, indicators represent distinct components that jointly define a construct and are not assumed to be interchangeable manifestations of a single underlying characteristic. Removing an indicator may alter the conceptual coverage of the construct, even when its estimated statistical contribution is relatively weak. This distinction is consistent with the guidance on construct specification provided by [29, 30].

The Economy construct is represented by per-capita expenditure and the percentage of the population living in poverty. These indicators capture different aspects of provincial economic conditions: economic welfare and the prevalence of poverty. Higher per-capita expenditure and a higher poverty rate do not have the same substantive meaning; therefore, both indicators are retained to represent complementary aspects of the economic dimension considered in this study.

Human Resources is represented by life expectancy, average years of schooling, and literacy rate. These indicators capture distinct dimensions of human development, including health and longevity, educational attainment, and basic literacy. Although related, these dimensions are not identical and may vary independently across provinces. They are therefore combined to represent the broader human-resource dimension rather than treated as interchangeable measures of a single outcome.

Infrastructure is represented by road surface conditions, the availability of markets without buildings, health facilities, and basic education facilities. These indicators describe different types of physical and service infrastructure. A province may have relatively good road conditions but limited access to health or educational facilities, or the reverse. The indicators consequently represent complementary components of provincial infrastructure provision rather than equivalent manifestations of a single underlying facility condition. Infrastructure is the central endogenous construct in the model because the study focuses on differences in infrastructure conditions and on the relationships between infrastructure and the other provincial dimensions.

Regional Characteristics are represented by the prevalence of villages and urban villages experiencing different types of natural disasters. These indicators capture distinct aspects of provincial exposure to natural hazards. Exposure to one type of disaster does not necessarily imply exposure to another; therefore, the indicators jointly describe the regional hazard context relevant to infrastructure planning. This construct represents hazard exposure rather than a measure of development performance, so its interpretation should account for the potential implications of natural hazards for infrastructure provision.

Taken together, the four constructs provide a multidimensional representation of provincial conditions relevant to infrastructure. Their formative specification allows each indicator to contribute a distinct element to its corresponding construct. The subsequent measurement-model assessment therefore examines indicator collinearity, the magnitude and significance of outer weights, and outer loadings when appropriate. These statistical assessments are interpreted alongside the conceptual relevance of the indicators and the framework established during operationalisation, rather than being used as the sole basis for determining the composition of the constructs.

3.1. Evaluation of the Measurement Model (Outer Model)

The formative measurement model was evaluated by examining indicator collinearity, the significance of outer weights, and the magnitude of outer loadings. Following the formative-indicator assessment procedure described by [18], outer weights were first assessed to determine the relative contribution of each indicator. When an outer weight was not statistically significant, the

corresponding outer loading was considered to assess the indicator's absolute contribution. These statistical results were interpreted alongside the conceptual relevance of each indicator and its role in representing the corresponding construct.

Table 2: Collinearity Assessment of the Formative Indicators

Construct	Indicator	Outer VIF
Infrastructure	I1	2.106
	I2	1.272
	I3	1.640
	I4	2.684
Economy	E1	2.019
	E2	2.019
Human Resource	H1	1.673
	H2	2.382
	H3	2.435
Regional Characteristics	R1	1.155
	R2	1.701
	R3	2.296
	R4	2.720

Table 2 shows that the outer VIF values range from 1.155 to 2.720. All values are below the commonly used threshold of 3.3, indicating that severe collinearity among the formative indicators is not apparent. R4 has the highest VIF (2.720), followed by I4 (2.684), whereas I2 has the lowest value (1.272). Overall, the results suggest that indicator collinearity is unlikely to be a major concern in the formative measurement model.

Table 3: Outer Weight and Outer Loadings of the Formative Indicators

Construct	Indicator	Outer Weight	t-value	p-value	Description	Outer Loading
Infrastructure	I1	0.442	2.288	0.022	Significant	0.865
	I2	0.241	1.213	0.225	Not significant	0.637
	I3	0.318	1.207	0.228	Not significant	0.768
	I4	0.253	0.912	0.362	Not significant	0.871
Economy	E1	0.636	1.316	0.188	Not significant	0.950
	E2	-0.442	0.923	0.356	Not significant	-0.894
Human Resource	H1	0.769	4.717	<0.001	Significant	0.969
	H2	0.123	0.456	0.642	Not Significant	0.732
	H3	0.214	0.734	0.463	Not Significant	0.765
Regional Characteristics	R1	0.677	3.329	0.001	Significant	0.397
	R2	0.980	5.365	<0.001	Significant	0.738
	R3	0.463	2.090	0.037	Significant	0.207
	R4	-0.175	0.737	0.461	Not Significant	0.501

As shown in Table 3, the outer-weight results indicate that H1, I1, R1, R2, and R3 have statistically significant relative contributions at the 5% significance level, whereas the outer weights of E1, E2, H2, H3, I2, I3, I4, and R4 are not statistically significant. For indicators with nonsignificant outer weights, the magnitude of the corresponding outer loadings was examined to assess their absolute contributions, following the formative-indicator assessment procedure described by [18]. For the Economy construct, neither E1 nor E2 shows a statistically significant outer weight. Still, their absolute outer loadings are 0.950 and 0.894, both above 0.50. Following the indicator-retention procedure in [18], both indicators are kept. This decision also makes sense conceptually, since per-capita expenditure and the poverty rate reflect different sides of provincial economic conditions. The negative sign of E2 shows the opposite relationship between poverty and economic welfare.

For Human Resources, H1 has a statistically significant outer weight and an outer loading of 0.969. H2 and H3 do not have significant outer weights, but their outer loadings are 0.732 and

0.765. Since both values are above 0.50, these indicators are kept according to the formative measurement assessment procedure in [18]. Including them also keeps the separate aspects of educational attainment and literacy in the construct.

For Infrastructure, I1 has a statistically significant outer weight and an outer loading of 0.865. The outer weights of I2, I3, and I4 are not statistically significant, but their outer loadings are 0.637, 0.768, and 0.871, respectively. As all three values exceed 0.50, the indicators are retained following the same assessment procedure [18]. This decision preserves the coverage of markets without buildings, health facilities, and basic education facilities as distinct components of provincial infrastructure.

For Regional Characteristics, R1, R2, and R3 have statistically significant outer weights, whereas R4 does not. The outer loadings of R2 and R4 are 0.738 and 0.501, respectively. R4 therefore meets the 0.50 threshold, although only narrowly. In contrast, R1 and R3 have outer loadings of 0.397 and 0.207 despite their statistically significant outer weights. Their relative contributions are statistically supported, but their absolute loadings are comparatively low. Under the formative-indicator assessment procedure, these results should be interpreted by considering the outer weights alongside the conceptual relevance of the indicators [18]. In this study, the indicators represent different types of natural-hazard exposure and are not necessarily interchangeable.

The collinearity assessment also showed that all outer VIF values were below 3.3, suggesting that severe collinearity among the formative indicators was not apparent. VIF was examined as part of the assessment of potential collinearity in the formative measurement model [18]. Overall, the outer-weight and outer-loading assessments support retaining the indicators under the specified decision procedure, while also revealing differences in the strength of their relative and absolute contributions. The indicators were therefore evaluated alongside their conceptual relevance and the framework established during operationalisation. In particular, Regional Characteristics is interpreted as a measure of hazard exposure rather than as a direct measure of desirable or undesirable development performance.

3.2. Evaluation of the Structural Model (Inner Model)

The structural model was evaluated by examining predictor collinearity, the coefficient of determination (R^2), effect sizes (f^2), and the significance of the structural path coefficients. Together, these criteria provide information about the model's ability to explain the data, the relative amount that each predictor contributes, and the uncertainty surrounding the estimated relationships. Since the constructs were based on the study's conceptual framework, the results are interpreted in light of their theoretical roles rather than merely on the basis of statistical significance.

3.2.1. Collinearity Assessment

Collinearity among the predictor constructs was assessed using the inner variance inflation factor (VIF). The results are presented in Table 4. The inner VIF values for Economy and Human Resources were 6.030 and 6.010, which are both above the common threshold of 5. This suggests there may be collinearity, so it is important to be careful when interpreting how each of these factors contributes to Infrastructure. In contrast, the inner VIF for Regional Characteristics was 1.809, showing no similar collinearity issue for this predictor. Because of this, the coefficients for Economy and Human Resources should be interpreted with caution, as it may be hard to separate their individual effects when the predictors overlap. However, these results alone do not justify removing either factor, especially since Economy and Human Resources are considered separate parts of the study's conceptual framework.

Table 4: Inner VIF Value

Structural Path	**VIF**
Economy → Infrastructure	6.030
Regional Characteristics → Infrastructure	1.809
Human Resources → Infrastructure	6.010

3.2.2. Coefficient of Determination

The explanatory power of the structural model was assessed using the coefficient of determination (R^2) and its adjusted value. The results are reported in Table 5. The Infrastructure construct had an R^2 value of 0.823 and an adjusted R^2 value of 0.808. Thus, Economy, Human Resources, and Regional Characteristics jointly accounted for 82.3% of the variance in Infrastructure, while the adjusted value was 80.8%. These results indicate substantial explanatory power within the specified model. However, the R^2 value reflects the predictors' joint explanatory contribution and should not be interpreted as evidence that every individual structural path is statistically significant.

Table 5: Coefficient of Determination (R^2) Results

Construct	R^2	**Adjusted** R^2
Infrastructure	0.823	0.808

3.2.3. Effect Size Assessment

The relative contribution of each predictor was further examined using the effect size (f^2). The results are presented in Table 6. Regional Characteristics had the largest effect size on Infrastructure ($f^2 = 0.402$), followed by Economy ($f^2 = 0.158$) and Human Resources ($f^2 = 0.051$). According to conventional guidelines, these values correspond to large, medium, and small effects, respectively. The f^2 values describe the relative change in the model's explanatory power associated with each predictor. They should not, however, be treated as substitutes for the bootstrapping results. In particular, the comparatively large effect size of Regional Characteristics does not establish a statistically significant path or imply that greater exposure to natural hazards is a desirable development outcome. This construct represents regional hazard exposure, and its substantive interpretation must remain consistent with that meaning.

Table 6: Effect Size (f^2) of the Structural Model

Structural Path	f^2
Economy → Infrastructure	0.158
Human Resources → Infrastructure	0.051
Regional Characteristics → Infrastructure	0.402

3.2.4. Structural Path Coefficients and Bootstrapping

The structural relationships were assessed using bootstrapping with 5,000 subsamples and a two-tailed significance level of 5%. The path coefficients and corresponding inference statistics are reported in Table 7. The estimated path coefficients were positive for all three predictor constructs. However, none of the paths reached statistical significance at the 5% level. The path from Economy to Infrastructure had a coefficient of 0.410 ($t = 1.195,\ p = 0.232$), while the path from Human Resources to Infrastructure was 0.232 ($t = 0.895,\ p = 0.371$). The path from Regional Characteristics to Infrastructure had a coefficient of 0.358 ($t = 1.928,\ p = 0.054$).

Although this last result was close to the conventional significance threshold, it did not meet the specified criterion and is therefore treated as statistically nonsignificant.

Table 7: Structural Path Coefficient and Bootstrapping Results

Construct	Original sample (O)	T statistics	P values	Description
Economy → Infrastructure	0.410	1.195	0.232	Not significant
Human Resources → Infrastructure	0.232	0.895	0.371	Not significant
Regional Characteristics → Infrastructure	0.358	1.928	0.054	Not significant

It is important to note that the fact that the path tests are not significant does not contradict the model's high R^2, nor does it prove that the respective relationships are absent. Instead, the present analysis does not offer enough statistical evidence to verify the individual paths at the 5% level of significance. The results thus lend support to a careful interpretation: the three constructs together explain a large portion of the variance in Infrastructure, while the evidence regarding their separate effects is still inconclusive within the present model specification. Economy and Human Resources are kept as separate constructs in line with the conceptual framework, and their high collinearity is recognised as a limitation when interpreting their individual path coefficients.

3.3. Finite Mixture Partial Least Squares (FIMIX-PLS)

Finite Mixture Partial Least Squares (FIMIX-PLS) was employed to explore potential unobserved heterogeneity among the 38 Indonesian provinces. FIMIX-PLS identifies latent segments that may differ in the relationships between constructs in the structural model. In this study, the method was used to investigate whether the relationships between Economy, Human Resources, Regional Characteristics, and Infrastructure varied across provinces. Because the appropriate number of segments may depend on the model-selection criterion and the characteristics of the data, the segmentation results were treated as exploratory rather than as a definitive classification [19].

Candidate solutions with two to five segments, K = 2 to K = 5, were examined. Model selection was evaluated using the information criteria and segmentation measures reported by the FIMIX-PLS procedure, including AIC, AIC3, AIC4, BIC, CAIC, HQ, MDL5, entropy (EN), non-fuzzy index (NFI), and normalized entropy criterion (NEC). These criteria were considered jointly because different criteria may indicate different segment solutions [19]. Each candidate solution was estimated using ten repetitions to examine whether the estimation procedure consistently converged from different starting values.

Table 8: FIMIX-PLS Model Selection Criteria

	$K = 2$	$K = 3$	$K = 4$	$K = 5$
AIC	47.198	49.196	46.102	31.177
AIC3	56.198	63.196	65.102	55.177
AIC4	65.198	77.196	84.120	79.177
BIC	61.936	72.122	77.216	70.479
CAIC	70.936	86.122	96.216	94.749
HQ	52.442	57.353	57.172	45.160
MDL5	192.889	275.827	353.673	419.687
LnL	-14.599	-10.598	-4.051	8.412
EN	0.755	0.547	0.665	0.867
NFI	0.778	0.535	0.622	0.826
NEC	9.315	17.210	12.746	5.039

Based on Table 8, the selection of the number of segments was treated as an exploratory

analysis because the model-selection criteria yielded mixed results. As shown in Table 8, AIC4 (65.198), BIC (61.936), CAIC (70.936), and MDL5 (192.889) attained their lowest values at (K=2). In contrast, AIC (31.177), AIC3 (55.177), and HQ (45.160) reached their lowest values at (K=5). The log-likelihood, entropy (EN), and non-fuzzy index (NFI) were also highest at (K=5), with values of 8.412, 0.867, and 0.826, respectively. NEC was lowest at (K=5) (5.039). Thus, seven of the eleven criteria favor (K=5) when assessed according to their respective optimization directions, whereas four favor (K=2). This disagreement indicates that the criteria do not provide a uniform basis for selecting the number of segments.

The model-selection criteria did not consistently identify the same number of segments. Therefore, the results do not support the claim that one solution is unequivocally optimal across all criteria. The two-segment solution was retained as a parsimonious exploratory specification to facilitate the interpretation of potential differences among provinces. This decision should not be interpreted as evidence that (K=2) outperforms the alternative solutions on most criteria or that provincial heterogeneity can be represented definitively by exactly two segments. Consistent with the discussion of FIMIX-PLS model-selection criteria by [19], the selected solution is interpreted cautiously because the available criteria do not provide a conclusive basis for identifying a unique number of segments.

The two-segment solution was further examined in terms of computational consistency, segment size, and membership probabilities. Across ten repeated estimations, all runs converged, with the same log-likelihood value of -14.598923 and convergence reached within 103–116 iterations. These results indicate computational consistency across the repeated estimations under the tested specification. However, consistent convergence and log-likelihood values do not establish that the selected number of segments is optimal, nor do they demonstrate that every province is classified unambiguously. The distinction between computational consistency and the substantive adequacy of the segmentation is therefore maintained throughout the interpretation.

Table 9: Computational Consistency and Segment Characteristics of the Exploratory $K = 2$ Solution

Aspect	**Result**
Number of repetitions	10
Converged repetitions	10/10
Log-likelihood across repetitions	-14.598923
Number of iterations	103–116
Estimated proportion of Segment 1	76.6%
Estimated proportion of Segment 2	23.4%
Final partition Segment 1	27 provinces
Final partition Segment 2	11 provinces
Minimum maximum membership probability	0.6153

Table 9 reports the estimated segment proportions for the 38 provincial observations. Segment 1 accounted for 76.6% of the estimated membership, while Segment 2 accounted for 23.4%, corresponding to approximately 29.1 and 8.9 observations, respectively. However, the final partition assigned 27 provinces to Segment 1 and 11 provinces to Segment 2. This difference reflects the distinction between the estimated relative segment sizes and the final classification of individual observations in FIMIX-PLS. The estimated proportions represent the mixing proportions of the model, whereas the final partition assigns each observation to the segment with the highest membership probability [31]. Consequently, the estimated proportions do not necessarily correspond exactly to the observed counts in the final partition. In this study, the estimated segment proportions and the final classification are therefore reported separately. The final partition was used to identify provincial membership and interpret the characteristics of each segment, particularly when examining the smaller group.

Membership probabilities also give a deeper understanding of how the provinces have been

assigned to the estimated segments. The lowest value of the maximum membership probability was 0.6153, which shows that the clarity of the segment assignment differed from observation to observation. When taken together with the convergence achieved in all ten repeated estimations, these probabilities offer additional information regarding both the computational consistency and the classification features of the two-segment solution. Thus, the segmentation provides a useful exploratory approach for looking at possible differences in the structural relationships among the provinces, and the membership probabilities aid in interpreting the individual assignments of the provinces.

Table 10: Final Partition Based on FIMIX-PLS Segmentation

Segment 1		**Segment 2**
Aceh	Kalimantan Tengah	Bengkulu
Sumatera Utara	Kalimantan Timur	Lampung
Sumatera Barat	Kalimantan Utara	DKI Jakarta
Riau	Sulawesi Utara	Jawa Tengah
Jambi	Sulawesi Tengah	DI Yogyakarta
Sumatera Selatan	Sulawesi Selatan	Jawa Timur
Kepulauan Bangka Belitung	Gorontalo	Kalimantan Selatan
Kepulauan Riau	Sulawesi Barat	Sulawesi Tenggara
Jawa Barat	Maluku	Papua Selatan
Banten	Maluku Utara	Papua Tengah
Bali	Papua Barat	Papua Pegunungan
Nusa Tenggara Barat	Papua Barat Daya	
Nusa Tenggara Timur	Papua	
Kalimantan Barat		



Fig. 2: Geographical Distribution of the Two FIMIX-PLS Segments

Table 10 and Fig. 2 show the results of the FIMIX-PLS classification. Segment-specific path coefficients provide an initial indication of how the relationships between the constructs may differ across the two estimated segments. For Segment 1, the estimated path coefficients from Economy, Human Resources, and Regional Characteristics to Infrastructure were 0.415, 0.112, and 0.410, respectively. For Segment 2, the corresponding coefficients were 0.346, 0.462, and 0.260. These estimates suggest different relative patterns of structural relationships across the two groups. In particular, the coefficient for Human Resources was larger in Segment 2, whereas

the coefficients for Economy and Regional Characteristics were larger in Segment 1. Given the exploratory nature of the segmentation, these differences are interpreted descriptively and are not treated as conclusive evidence of statistically significant differences between segments.

Table 11: Structural Path Coefficients for Each FIMIX-PLS Segment

Construct	Aggregate Model	Segment 1	Segment 2
Economy → Infrastructure	0.410	0.415	0.346
Human Resources → Infrastructure	0.232	0.112	0.462
Regional Characteristics → Infrastructure	0.358	0.410	0.260

Table 11, a comparison of the aggregate and segment-specific estimates, provides a more differentiated view of the structural relationships than the aggregate model alone. In Segment 1, the coefficients for Economy and Regional Characteristics were relatively larger than those for Human Resources. In Segment 2, Human Resources had the largest coefficient among the three predictors, while the coefficients for Economy and Regional Characteristics were lower than their corresponding estimates in Segment 1. These patterns offer a basis for further investigation into whether the relationships associated with provincial infrastructure differ across groups of provinces. It cannot be concluded from the magnitude of the coefficients that there are statistically significant differences between the segments, and the estimates must therefore be regarded as describing certain patterns rather than as confirming effects that are specific to the segments.

The interpretation of Regional Characteristics requires particular care because this construct represents exposure to natural hazards, operationalized through normalized indicators of villages and urban villages experiencing different types of disasters. The fact that the coefficient for Regional Characteristics is positive does not mean that greater hazard exposure is beneficial or that disasters have a positive effect on infrastructure. Rather, the coefficient shows the estimated association between the hazard-exposure construct and infrastructure within the given model. It could be due to differences in the context regarding hazard exposure, in the provision of infrastructure, or in the way disasters are responded to; yet the path coefficient by itself does not show which of these explanations is the most accurate. For this reason, the policy recommendations should focus on hazard-adaptive and resilient infrastructure planning, not on seeing an increased hazard exposure as a desirable result.

In general, the FIMIX-PLS analysis provides an exploratory view of possible heterogeneity in the relationships between Economy, Human Resources, Regional Characteristics, and Infrastructure among the provinces of Indonesia. The two-segment solution showed consistent convergence in the course of repeated estimations and generated coefficient patterns specific to each segment, which can be examined in relation to the provincial characteristics. Moreover, when interpreting the resulting classification, the use of mixed model-selection criteria should be considered, as well as the unequal proportions of the segments and the variation in membership probabilities. The two-segment solution is therefore used to organize and discuss potential patterns of heterogeneity, rather than to claim a definitive or uniquely optimal classification of the provinces. Given these differences, the Importance-Performance Map Analysis (IPMA) method is applied in this study to identify priority factors for infrastructure development within each segment.

3.4. Importance–Performance Map Analysis (IPMA)

Importance–Performance Map Analysis (IPMA) was conducted to complement the PLS-SEM and FIMIX-PLS results by examining the relative importance and performance of Economy, Human Resources, and Regional Characteristics in explaining Infrastructure. Following the FIMIX-PLS segmentation, IPMA was performed separately for the two groups of provinces to explore whether their relative patterns of importance and performance differed. For each segment, the PLS-SEM model was estimated using the provinces assigned to that segment, with

Infrastructure specified as the target construct. The importance was evaluated in terms of the total impact of the earlier constructs on Infrastructure, while performance was given on a scale from 0 to 100. The reference lines for the quadrants were established on the basis of the average importance and average performance values for each segment. Thus, the quadrant classifications indicate the relative positions within each group and not absolute development thresholds as stated in Table 12.

Table 12: Importance and Performance Values for Each FIMIX-PLS Segment

Construct	Segment 1		Segment 2	
	Importance	**Performance**	**Importance**	**Performance**
Economy	0.310	34.008	0.985	38.016
Human Resources	0.076	68.617	-0.378	52.344
Regional Characteristics	0.601	24.031	0.382	20.749
Mean (quadrant reference line)	0.329	42.218	*0.329*	37,036

3.4.1. IPMA Analysis for Segment 1

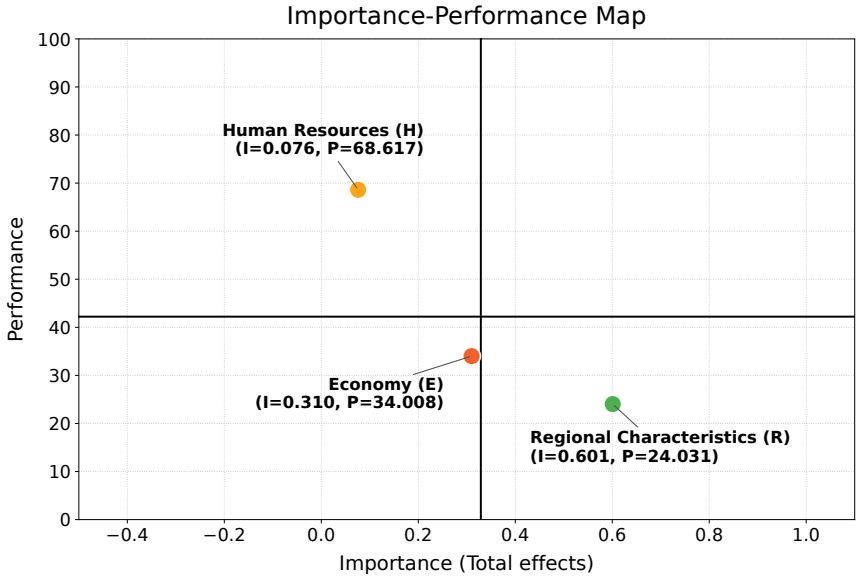


Fig. 3: Importance–Performance Map Analysis (IPMA) for Segment 1

The IPMA results for Segment 1 are shown in Fig. 3. Among the three constructs, Regional Characteristics had the highest importance score of 0.601 and the lowest performance score of 24.031. The fact that it lies in the quadrant of high importance and low performance shows that infrastructure in this segment is comparatively closely linked with Regional Characteristics, even though its performance score is below the reference value specific to the segment. This demonstrates the need to take regional characteristics into account when planning infrastructure in the provinces that belong to Segment 1.

The economy had an importance value of 0.310 and a performance score of 34.008 and, according to the reference lines indicated on the map, lies in the quadrant representing low importance and low performance. Human Resources had a lower importance value of 0.076 but a higher performance score of 68.617, so it is in the quadrant for low importance and high performance. Therefore, of the three constructs, Human Resources showed the highest level of performance in Segment 1, even though its estimated total effect on Infrastructure was relatively small.

3.4.2. IPMA Analysis for Segment 2

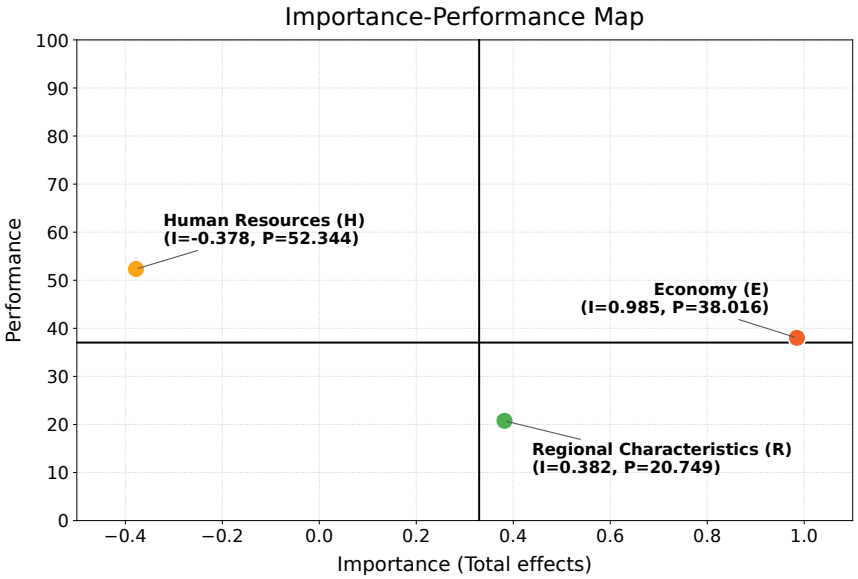


Fig. 4: Importance–Performance Map Analysis (IPMA) for Segment 2

Fig. 4 presents the IPMA results for Segment 2. Economy had the highest importance value of 0.985, with a performance score of 38.016. Relative to the quadrant reference lines in the map, Economy occupied the high-importance, slightly above-average-performance quadrant. Compared with Segment 1, its estimated importance was substantially higher, suggesting that economic conditions may be particularly relevant to understanding Infrastructure in this group. However, this difference is descriptive and should not be interpreted as evidence of a statistically significant difference between segments.

Regional Characteristics had an importance value of 0.382 and the lowest performance score of 20.749; its position in the high-importance, low-performance quadrant shows that the construct is still relevant to the Infrastructure model, although its performance is comparatively low. Human Resources had a negative importance value of -0.378 and a performance score of 52.344, which places it in the low-importance, high-performance quadrant. The negative importance value shows that there is a negative estimated total effect on Infrastructure in this segment; it does not imply that the performance score is negative.

3.4.3. Comparative Interpretation and Policy Implications

The IPMA maps reveal different patterns of relative importance across the two estimated segments. In Segment 1, Regional Characteristics had the highest importance, whereas in Segment 2, Economy had the highest importance. Regional Characteristics recorded the lowest performance score in both groups (24.031 in Segment 1 and 20.749 in Segment 2). Human Resources had the highest performance score in Segment 1 and remained above the performance reference line in Segment 2, although its estimated importance was negative in the latter group. These patterns suggest that the relative importance and performance of the constructs vary across the two groups of provinces.

The results for Regional Characteristics require particular care in substantive interpretation. Its indicators represent the proportion of villages or urban villages experiencing natural disasters, rather than a conventional development resource for which a higher value is necessarily desirable. Therefore, a low IPMA performance score should not be interpreted as a deficiency that can be addressed by increasing or decreasing disaster exposure. Instead, the combination of comparatively high importance and low performance highlights the need to account for hazard exposure when planning and prioritizing infrastructure.

This interpretation is consistent with research reviewing the fragility of transport assets exposed to multiple hazards and emphasizing the importance of infrastructure resilience [17]. Research on flooding in Greater Jakarta has also examined disruptions in access to critical services using social, geospatial, and network perspectives [32]. These studies provide relevant context for interpreting the present findings, although they do not independently validate the relationships estimated in this study.

From a policy perspective, the findings suggest that hazard-adaptive infrastructure planning should be considered, including the assessment of disaster exposure, the continuity of transport and access to services, and the resilience of infrastructure. Since Economy also had the highest estimated importance within that group, it too may deserve attention in Segment 2. However, the IPMA results must be seen as preliminary evidence of prioritization rather than as proof of causal effects or as indicating statistically significant differences between the segments.

4. Conclusion

This study examined the relationships between Economy, Human Resources, and Regional Characteristics with Infrastructure across 38 Indonesian provinces using PLS-SEM, followed by FIMIX-PLS and segment-specific IPMA. The overall structural model explained 82.3% of the variance in Infrastructure ($R^2 = 0.823$). However, none of the three structural relationships was statistically significant at the 5% level. The relationships of Economy ($t = 1.195, \quad p = 0.232$) and Human Resources ($t = 0.895, \quad p = 0.371$) were not statistically significant, while the relationship of Regional Characteristics was close to, but did not reach, the significance threshold ($t = 1.928, \quad p = 0.054$). Therefore, the results do not provide sufficient statistical evidence to confirm the proposed relationships, despite the model's relatively high explanatory power.

FIMIX-PLS identified a two-segment solution that was retained as a parsimonious exploratory representation of potential provincial heterogeneity. The segment-specific path coefficients showed different descriptive patterns: Regional Characteristics had the largest coefficient in Segment 1, whereas Human Resources had the largest coefficient in Segment 2. These differences should not be interpreted as statistically established differences between segments because no formal between-segment test was conducted. The IPMA results also indicated different relative-importance patterns: Regional Characteristics had the highest importance in Segment 1, whereas Economy had the highest importance in Segment 2. Regional Characteristics showed the lowest performance in both segments.

Because Regional Characteristics represents exposure to natural disasters, its low IPMA performance should not be interpreted as a deficiency that can be addressed by changing disaster exposure itself. Instead, the finding points to the relevance of considering disaster exposure when planning infrastructure that is resilient to hazards and supports access to essential services. These implications remain exploratory and should not be interpreted as evidence of causal effects or definitive segment-specific priorities.

Several limitations should be considered when interpreting the findings. The analysis used provincial-level secondary data averaged across 2024 and 2025, which limits the ability to examine changes over time. The analysis also covered 38 provinces, with the smaller FIMIX segment comprising 11 provinces, and the overall structural paths were not statistically significant at the 5% level. In addition, the high correlation between Economy and Human Resources and the corresponding inner VIF values indicate potential structural collinearity. Future research could use longitudinal data, a larger number of regional units, additional indicators of infrastructure and regional conditions, and formal tests of between-segment differences to assess whether the exploratory patterns identified in this study are reproducible.

CRediT Authorship Contribution Statement

Arum Nanda Putri Aprilia: Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Visualization, Writing – Original Draft. **Rahmawati Erma Standsyah:** Conceptualization, Methodology, Validation, Writing – Review & Editing, Supervision.

Declaration of Generative AI and AI-assisted technologies

During the preparation of this study, the authors used ChatGPT for translation and language editing purposes. After using this tool, the authors reviewed, verified, and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interests

The authors declare no competing interests.

Funding and Acknowledgments

This research was not supported by any specific grant from public, commercial, or non-profit funding agencies.

We would like to thank Statistics Indonesia (Badan Pusat Statistik/BPS) for supplying the official statistical data used in this study, including the Village Potential Statistics (PODES) and other key publications that served as our main data sources.

Data and Code Availability

The data for this study come from official publications by Statistics Indonesia (Badan Pusat Statistik/BPS), such as the Village Potential Statistics (PODES) and other official reports. If you would like access to the processed dataset used in this research, please contact the corresponding author.

References

[1] M. A. Berawi, P. Miraj, and M. Sari. "Fostering Infrastructure Development Aligned With Sustainable Development Goals". In: *CSID Journal of Infrastructure Development* 4.1 (2021), pp. 1–3. DOI: 10.32783/csid-jid.v4i1.236.

[2] I. Kustova, J. Hudenko, and N. Lace. "A Systematic Review of Sustainability Criteria in Infrastructure Development". In: *Sustainability* 16.11 (2024), p. 4564. DOI: 10.3390/su16114564.

[3] A. Jatayu et al. "Measuring Levels of Infrastructure Development and Its Impact on Regional Growth: Insights from Indonesia". In: *IOP Conference Series: Earth and Environmental Science* 1353.1 (2024), p. 012011. DOI: 10.1088/1755-1315/1353/1/012011.

[4] R. D. Ramadhan and F. Z. Fauzi. "Driving Progress Together: Embracing Inclusive Development for Infrastructure in East Indonesia". In: *J-3P (Jurnal Pembangunan Pemberdayaan Pemerintahan)* 8.2 (2023), pp. 93–112. DOI: 10.33701/j-3p.v8i2.3691.

[5] W. Nurdina. "Infrastructure and Income Inequality in Indonesia: 2009–2017". In: *The Journal of Indonesia Sustainable Development Planning* 2.2 (2021), pp. 129–144. DOI: 10.46456/jisdep.v2i2.132.

[6] K. Ariansyah et al. "Unleashing the Potential of Mobile Broadband: Evidence from Indonesia's Underdeveloped Regions on Its Role in Reducing Income Inequality". In: *Telematics and Informatics* 82 (2023), p. 102012. DOI: 10.1016/j.tele.2023.102012.

[7] R. N. T. Wahyuni, M. Ikhsan, A. Damayanti, and K. Khoirunurrofik. "Inter-District Road Infrastructure and Spatial Inequality in Rural Indonesia". In: *Economies* 10.9 (2022), p. 229. DOI: 10.3390/economies10090229.

[8] P. R. Halim, A. A. Yusuf, A. Sumner, and R. M. Purnagunawan. "Structural Transformation and Regional Inequality Differential in Indonesia". In: *Regional Studies, Regional Science* 12.1 (2025), pp. 790–815. DOI: 10.1080/21681376.2025.2565318.

[9] J. C. Londono, B. Wilson, and F. Osorio-Tinoco. "Understanding the Entrepreneurial Intentions of Youth: A PLS Multi-Group and FIMIX Analysis Using the Model of Goal-Directed Behavior". In: *Journal of Entrepreneurship in Emerging Economies* 13.3 (2021), pp. 301–326. DOI: 10.1108/JEEE-07-2019-0090.

[10] B. W. Otok, Purhadi, R. Sriningsih, and D. S. Dila. "Segmentation of Toddler Nutritional Status Using REBUS and FIMIX Partial Least Square in Southeast Sulawesi". In: *MethodsX* 12 (2024), p. 102515. DOI: 10.1016/j.mex.2023.102515.

[11] A. S. H. Wahyuningtyas et al. "Consumer Segmentation Analysis of Microalgae-Based Foods in Indonesia: A Finite Mixture Partial Least Squares (FIMIX-PLS) Approach". In: *Sustainable Futures* 10 (2025), p. 101555. DOI: 10.1016/j.sftr.2025.101555.

[12] Y. Yuan, M. Yang, T. Feng, S. Rasouli, D. Li, and X. Ruan. "Heterogeneity in Passenger Satisfaction with Air-Rail Integration Services: Results of a Finite Mixture Partial Least Squares Model". In: *Transportation Research Part A: Policy and Practice* 147 (2021), pp. 133–158. DOI: 10.1016/j.tra.2021.03.003.

[13] H. Joshi and D. Chawla. "Identifying Unobserved Heterogeneity in Mobile Wallet Adoption: A FIMIX-PLS Approach for User Segmentation". In: *International Journal of Bank Marketing* 41.1 (2023), pp. 210–236. DOI: 10.1108/IJBM-03-2022-0132.

[14] A. Jo and S. An. "Customer Segmentation in Social Media Marketing: Analyzing Hidden Customer Heterogeneity through Finite Mixture Partial Least Squares (FIMIX-PLS) and Importance-Performance Map Analysis (IPMA)". In: *Business Communication Research and Practice* 7.1 (2024), pp. 35–47. DOI: 10.22682/bcrp.2024.7.1.35.

[15] D. A. Fadhila and B. W. Otok. "Pemodelan Faktor-Faktor yang Mempengaruhi Daerah Tertinggal di Pulau Jawa Menggunakan Second-Order Confirmatory Factor Analysis (CFA)". In: *Jurnal Sains dan Seni ITS* 8.2 (2020), pp. D153–D160. DOI: 10.12962/j23373520.v8i2.43216.

[16] B. W. Otok, A. Suharsono, Purhadi, R. E. Standsyah, and H. Al Azies. "MASEM Infrastructure in Underdeveloped Areas of Java Island". In: *Journal of Southwest Jiaotong University* 56.1 (2021). DOI: 10.35741/issn.0258-2724.56.1.10.

[17] S. A. Argyroudis, S. A. Mitoulis, M. G. Winter, and A. M. Kaynia. "Fragility of Transport Assets Exposed to Multiple Hazards: State-of-the-Art Review toward Infrastructural Resilience". In: *Reliability Engineering & System Safety* 191 (2019), p. 106567. DOI: 10.1016/j.ress.2019.106567.

[18] Joseph F. Hair, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd ed. Los Angeles: SAGE Publications, 2017.

[19] M. Sarstedt, J.-M. Becker, C. M. Ringle, and M. Schwaiger. "Uncovering and Treating Unobserved Heterogeneity with FIMIX-PLS: Which Model Selection Criterion Provides an Appropriate Number of Segments?" In: *Schmalenbach Business Review* 63.1 (2011), pp. 34–62. DOI: 10.1007/BF03396886.

[20] C. M. Ringle and M. Sarstedt. "Gain More Insight from Your PLS-SEM Results: The Importance-Performance Map Analysis". In: *Industrial Management & Data Systems* 116.9 (2016), pp. 1865–1886. DOI: 10.1108/IMDS-10-2015-0449.

[21] Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia. *Peraturan Menteri Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Nomor 3 Tahun 2016 tentang Petunjuk Teknis Penentuan Indikator dalam Penetapan Daerah Tertinggal Secara Nasional*. Peraturan Menteri Nomor 3 Tahun 2016. 2016. https://peraturan.bpk.go.id/Details/150586/permendes-pdtt-no-3-tahun-2016 (visited on 05/10/2026).

[22] Badan Pusat Statistik. *Statistik Potensi Desa Indonesia 2024*. 04300.24002. Jakarta: Badan Pusat Statistik, 2024. https://www.bps.go.id/id/publication/2024/12/10/2f5217e2d6a695a0830290a7/statistik-potensi-desa-indonesia-2024.html (visited on 05/05/2026).

[23] Badan Pusat Statistik. *Statistik Potensi Desa Indonesia 2025*. 04300.25012. Jakarta: Badan Pusat Statistik, 2025. https://www.bps.go.id/id/publication/2025/12/11/df409d461277ba686888b7f6/statistik-potensi-desa-indonesia-2025.html (visited on 05/05/2026).

[24] Badan Pusat Statistik. *[Metode Baru] Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang/Tahun)*. 2026. https://www.bps.go.id/indicator/26/416/1/-metode-baru-pengeluaran-per-kapitadisesuaikan.html (visited on 06/29/2026).

[25] Badan Pusat Statistik. *Jumlah dan Persentase Penduduk Miskin Menurut Provinsi*. Available at: BPS Statistics Table (accessed 5 August 2026). 2026.

[26] Badan Pusat Statistik. *Umur Harapan Hidup Saat Lahir Menurut Provinsi dan Jenis Kelamin (Menggunakan UHH Hasil SP2020 LF)*. 2026. https://www.bps.go.id/id/statistics-table/2/MjI3MyMy/angka-harapan-hidup-/-ahh-/-menurut-provinsi-dan-jenis-kelamin-/-menggunakan-uhh-hasilsp2020lf-.html (visited on 06/29/2026).

[27] Badan Pusat Statistik. *Rata-Rata Lama Sekolah (RLS) Menurut Jenis Kelamin*. 2026. https://www.bps.go.id/id/statistics-table/2/NDU5IzI%5C%3D/rata-rata-lama-sekolah-rls-menurut-jenis-kelamin-tahun.html (visited on 06/29/2026).

[28] Badan Pusat Statistik. *Angka Melek Aksara Penduduk 15–24 Tahun Menurut Provinsi*. 2026. https://www.bps.go.id/id/statistics-table/2/MTQ2MiMy/angka-melek-aksara-penduduk-15-24-tahun-menurut-provinsi.html (visited on 06/29/2026).

[29] C. B. Jarvis, S. B. MacKenzie, and P. M. Podsakoff. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research". In: *Journal of Consumer Research* 30.2 (2003), pp. 199–218. DOI: 10.1086/376806.

[30] S. B. MacKenzie, P. M. Podsakoff, and C. B. Jarvis. "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions". In: *Journal of Applied Psychology* 90.4 (2005), pp. 710–730. DOI: 10.1037/0021-9010.90.4.710.

[31] L. M. Matthews, M. Sarstedt, J. F. Hair, and C. M. Ringle. "Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part II – A Case Study". In: *European Business Review* 28.2 (2016), pp. 208–224. DOI: 10.1108/EBR-09-2015-0095.

[32] P. Kiparisov, V. Lagutov, and G. Pflug. "Quantification of Loss of Access to Critical Services during Floods in Greater Jakarta: Integrating Social, Geospatial, and Network Perspectives". In: *Remote Sensing* 15.21 (2023), p. 5250. DOI: 10.3390/rs15215250.