Exploring Heterogeneity in Food-Security, Socioeconomic, and Health-Service Determinants of Toddler Undernutrition across Subdistricts of Banyumas Regency Using Prediction-Oriented PLS-SEM Segmentation

Felinda Arumningtyas[1*], Melda Juliza[1], Isnu Aji Saputro[2], Abdurrahman Yusuf[1], Al Fath Miftahurrizky Hikmawan[1]
[1]Department of Statistics, Universitas Jenderal Soedirman, Indonesia
[2]Department of Mathematics, Universitas Jenderal Soedirman, Indonesia

Corresponding email: felinda.arumningtyas@unsoed.ac.id

Abstract

Stunting remains a public health concern in Banyumas Regency, Central Java, where the 2024 national nutrition survey placed prevalence at 19.6%, above the provincial figure of 17.1%. A single regency-wide model can hide differences in how determinants operate across places. This exploratory study models toddler undernutrition at the subdistrict level using partial least squares structural equation modelling (PLS-SEM) with four exogenous constructs (food security, socioeconomic conditions, nutrition behaviour, and health services and environment), and then applies prediction-oriented segmentation (PLS-POS) to explore unobserved heterogeneity among the 27 subdistricts. Secondary data from the district health office, the agriculture and food security office, and Statistics Indonesia were used. The global model explained 44.7% of the in-sample variance in the latent undernutrition construct (adjusted $R^2 = 0.347$), with the largest path coefficient for socioeconomic conditions (-0.497), a composite dominated by the poverty ratio. In bootstrapping with 5,000 subsamples this was the only path with one-tailed $p < 0.05$ ($p = 0.036$), and its direction runs opposite to individual-level evidence. PLSpredict yielded negative Q^2predict values (-0.348 for the construct), indicating no out-of-sample predictive relevance. PLS-POS partitioned the subdistricts into three segments of 8, 10 and 9 units with different path patterns; the near-saturated segment models (R^2 0.948–0.974 with 3–5 residual degrees of freedom) and a coefficient exceeding 1 are consistent with overfitting. The segmentation is therefore presented as hypothesis-generating evidence that a single regency-wide model may oversimplify, not as a basis for allocating interventions.

Keywords: Food Security; Health Services; Nutritional Risk; Prediction-Oriented PLS-SEM; Toddlers

I. INTRODUCTION

Undernutrition in early childhood is not a single condition. Stunting reflects chronic nutritional deprivation and repeated infection, wasting reflects acute weight loss, and underweight blends the two. Together they remain a leading contributor to child mortality in low- and middle-income countries; a Lancet series estimated that undernutrition, broadly defined, accounted for 3.1 million child deaths in 2011, or 45% of all child deaths that year (Black et al., 2013). In Indonesia, a review organised around the WHO conceptual framework found consistent evidence that non-exclusive breastfeeding, low household socioeconomic status, poor access to health care and rural residence are linked to stunting, while the contribution of agriculture and food systems and of water, sanitation and the environment had been studied far less (Beal et al., 2018).

Banyumas Regency in Central Java illustrates why these gaps matter for planning. The 2024 Indonesian Nutritional Status Survey (SSGI) put stunting prevalence in Banyumas at about 19.6% (Dinas Komunikasi dan Informatika Kabupaten Banyumas, 2025), compared with 17.1% for Central Java and 19.8% nationally (Pemerintah Provinsi Jawa Tengah, 2025). A regency-level average, however, says little about where programmes should be concentrated. Health, nutrition and food-security programmes are delivered through subdistrict (kecamatan) health centres and offices, so the subdistrict is the level at which prioritisation decisions are actually taken.

Food security is the factor most often invoked to explain undernutrition, yet the evidence is less uniform than the policy narrative suggests. Food security is multidimensional, and the metrics used to measure it differ in the domains and constructs they capture (Jones et al., 2013). A systematic review and meta-analysis of nine cohort studies covering 46,300 children found that most individual studies reported an association between household food insecurity and stunting, but the pooled hazard ratio for moderate and severe insecurity was 1.02 (95% CI 0.84–1.22) with very high heterogeneity (I^2 = 85.96%; Patriota et al., 2024). A review of Asian studies likewise described substantial variation in how household food insecurity, dietary diversity and stunting were related (Islam et al., 2025), and a narrative review for the Indian subcontinent concluded that the evidence pointed to a possible association in which dietary diversity might mediate the link (Yadav et al., 2024). Such inconsistency is compatible with an association that depends on context, although cross-study heterogeneity cannot by itself show that structural relationships differ between places within one regency.

The four exogenous constructs in this study correspond to the determinant domains that recur in this literature. Socioeconomic conditions, here captured by poverty and schooling, are repeatedly linked to child growth in Indonesia (Beal et al., 2018). Nutrition behaviour, represented by exclusive breastfeeding and attendance at community health posts (posyandu), addresses the feeding and care practices that mediate between household resources and child outcomes; suboptimal breastfeeding is one of the contributors to undernutrition-related mortality described by Black et al. (2013). Health services and environment combine health personnel, immunisation, drinking water and sanitation, which are underlying determinants because they affect exposure to infection and access to care. Food security is the domain through which the availability of and access to food reach the child. Treating these domains as latent constructs rather than as isolated variables acknowledges that each is observed through several imperfect indicators and that the indicators sometimes pull in different directions.

Most regional analyses in Indonesia nonetheless estimate one set of relationships for all units. Partial least squares structural equation modelling (PLS-SEM) allows several latent constructs and their pathways to be handled at once, and it has been applied to toddler undernutrition in Aceh (Surayya & Sirait, 2024), to stunting in Pidie combined with geographic information system mapping (Sari et al., 2025), and to toddler malnutrition in East Java (Agustina & Arumningtyas, 2024). The findings are not uniform: in Aceh, food intake and health access acted directly on undernutrition, whereas in East Java the paths from food security and from health services and environment were not statistically significant at the 10% level. Differences of this kind may reflect different constructs and data, but they also raise the possibility that a pooled model averages over subpopulations whose structural relationships differ. When such unobserved heterogeneity is ignored, parameter estimates can be biased and both Type I and Type II errors become more likely (Becker et al., 2013).

Prediction-oriented segmentation for PLS path models (PLS-POS) was proposed by Becker et al. (2013) to detect this kind of heterogeneity. It reassigns observations among segments with the aim of improving the explained variance of the endogenous constructs, and it can be applied to models containing both reflective and formative measurement.

FIMIX-PLS, a latent class approach, is considered particularly useful for deciding how many segments to extract (Hair et al., 2016). Segmentation approaches have started to appear in Indonesian health research. Nurhanifa et al. (2026) combined SEM-PLS and POS across 34 provinces to classify maternal and child health outcomes, identified three segments, and noted that better access to services did not always translate into better outcomes. Spatial work on stunting has so far relied mainly on spatial autocorrelation and cluster maps; a national provincial analysis, for example, found modest positive spatial autocorrelation in stunting (Moran's I = 0.182; Sari & Siregar, 2023). To our knowledge, within the literature we could identify, PLS-POS has not been used to segment subdistricts of a regency on toddler undernutrition with food security as a central construct.

The present study is exploratory and has three objectives. First, it describes the relationships between food security, socioeconomic conditions, nutrition behaviour, and health services and environment on the one hand and toddler undernutrition on the other, using a global PLS-SEM at the subdistrict level. Second, it explores whether segments of subdistricts differ in their estimated relationships and in their relative undernutrition level using PLS-POS. Third, it compares the in-sample explanatory power of the global and segment-specific models through R^2 and assesses the out-of-sample predictive performance of the global model with PLSpredict. Traditional PLS-SEM reporting has relied largely on explanatory metrics (Shmueli et al., 2019), and we keep the distinction between explanation and prediction explicit throughout.

II. METHOD

Study area and unit of analysis. The study covers Banyumas Regency, Central Java, Indonesia, which comprises 27 subdistricts. Each subdistrict is one observation, so the analysis uses the full set of first-level administrative units of the regency, not a sample drawn from a larger population.

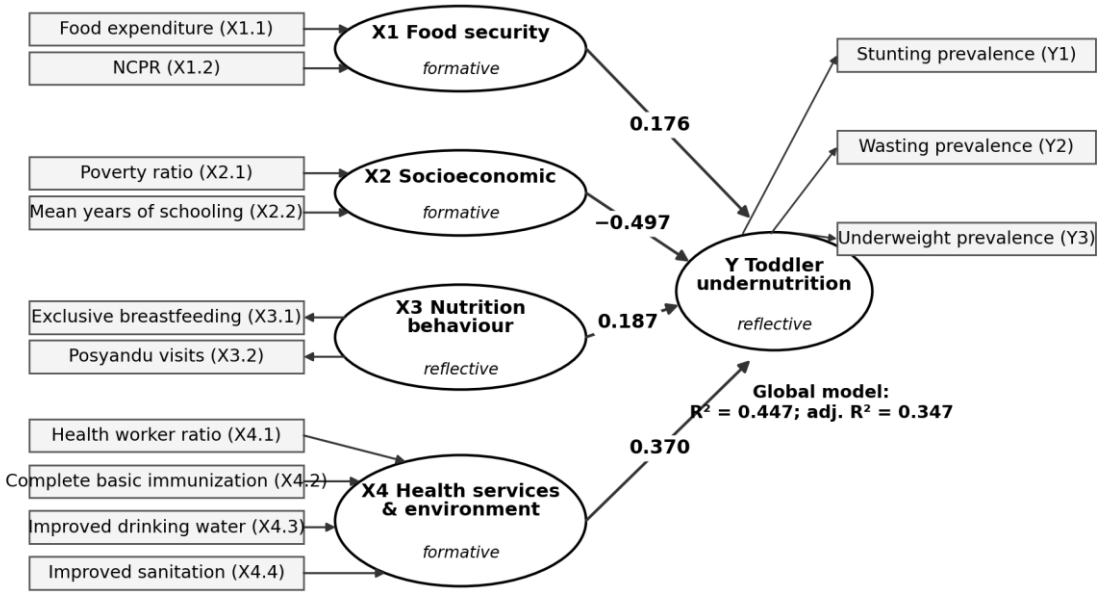
Data sources and constructs. All data are secondary. Indicators on child nutritional status, nutrition behaviour, immunisation and health personnel were obtained from the Banyumas District Health Office; food-security indicators from the Banyumas Office of Agriculture and Food Security; and poverty, schooling, drinking water and sanitation indicators from Statistics Indonesia (BPS). Data refer to 2025. Five latent constructs were specified (Table 1; Figure 1). Toddler undernutrition (Y) is measured by the prevalence of stunting, wasting and underweight. Four exogenous constructs were specified: food security (X1), socioeconomic conditions (X2), nutrition behaviour (X3), and health services and environment (X4). The Indonesian term *gizi buruk* is rendered here as undernutrition because the indicators are prevalence rates of stunting, wasting and underweight and do not denote severe acute malnutrition.

Table 1. Latent constructs, indicators, measurement mode and data sources

Construct	Indicators	Mode	Source
Toddler undernutrition (Y)	Prevalence of stunting (Y1); wasting (Y2); underweight (Y3)	Reflective	Banyumas District Health Office
Food security (X1)	Food expenditure (X1.1); normative consumption-to-production ratio, NCPR (X1.2)	Formative	Banyumas Office of Agriculture and Food Security
Socioeconomic conditions (X2)	Poverty ratio (X2.1); mean years of schooling (X2.2)	Formative	Statistics Indonesia (BPS)
Nutrition behaviour (X3)	Exclusive breastfeeding (X3.1); posyandu visit coverage (X3.2)	Reflective	Banyumas District Health Office
Health services and environment (X4)	Health worker ratio (X4.1); complete basic immunisation of toddlers (X4.2); households with improved drinking water (X4.3); households with improved sanitation (X4.4)	Formative	District Health Office (X4.1, X4.2); BPS (X4.3, X4.4)

Notes: NCPR = normative consumption per capita to net production per capita ratio, Posyandu = integrated community health post. All indicators entered the model in their original orientation, without reverse coding. Higher values denote worse conditions for Y1–Y3, the poverty ratio (X2.1) and NCPR (X1.2; a larger gap between normative consumption and net production), and better conditions for mean years of schooling (X2.2), complete basic immunisation (X4.2), improved drinking water (X4.3) and improved sanitation (X4.4). Food expenditure (X1.1) is total household spending on food, so higher values indicate greater food purchasing and, in that sense, better food access. The health worker indicator (X4.1) reflects the number of health workers, so higher values denote better staffing. Exclusive breastfeeding (X3.1) and posyandu visits (X3.2) are on the scale recorded by the District Health Office (Table 6).

Measurement specification. Food security, socioeconomic conditions, and health services and environment were treated as formative constructs, because their indicators describe different, non-interchangeable facets (for instance, health personnel, immunisation, water and sanitation) that jointly define the construct. Nutrition behaviour and undernutrition were treated as reflective constructs. The structural model contains four direct paths from X1–X4 to Y. Nutrition behaviour is measured by only two indicators, which is thin for a reflective construct.



Arrows pointing into a construct: formative measurement; arrows pointing out of a construct: reflective measurement.

Figure 1. Structural model with global PLS-SEM path coefficients (standardised).

Pre-processing. Datasets from the three agencies were merged by subdistrict, checked for completeness and cleaned, and all indicators were standardised before estimation. No missing values remained after consolidation.

Global PLS-SEM. The global model was estimated on all 27 subdistricts as the baseline. PLS-SEM estimates construct scores iteratively and then obtains path coefficients by regression among the scores, which suits prediction-oriented work with composite-based models (Hair et al., 2019). Measurement quality is conventionally judged by indicator loadings, composite reliability and average variance extracted for reflective constructs, and by collinearity (variance inflation factors) and the size and significance of outer weights for formative constructs; structural quality is judged by the sign and magnitude of path coefficients, R^2, and predictive relevance (Hair et al., 2019). The research protocol set minimum targets of loadings ≥ 0.5, AVE ≥ 0.5 and R^2 ≥ 0.25. Under the commonly used benchmarks of 0.25, 0.50 and 0.75 for weak, moderate and substantial R^2, a value such as 0.447 lies between the weak and moderate benchmarks; these are rules of thumb that depend on the research context (J. F. Hair et al., 2019).

Prediction-oriented segmentation. Unobserved heterogeneity was explored with PLS-POS (Becker et al., 2013). Starting from an initial partition, the procedure uses a distance measure derived from the PLS path model to move observations between segments and retains reassignments that improve the explained variance of the endogenous construct, repeating the process until no improvement remains. Because the criterion is built on the outcome construct, the resulting segments are by design the partition under which the model fits best within segments. The segmentation produced three segments.

Segment-specific models and profiling. The structural model was re-estimated separately within each segment. Segments were then profiled with (i) segment-level mean scores of the standardised latent undernutrition variable, which express a subdistrict's position relative to the average of all 27 subdistricts and are not probabilities, and (ii) segment means of the raw indicators. Because adjusted R^2 penalises the number of predictors relative to sample size, we additionally derived approximate adjusted values for the segment models from the reported R^2, segment size and the four predictors, using $1 - (1 - R^2)(n - 1)/(n - k - 1)$. In this paper, "nutritional risk" is used descriptively for the relative position of a subdistrict or segment on the latent undernutrition score.

Rationale for a composite-based model with few units. PLS-SEM was chosen over covariance-based SEM for three reasons that follow from the data and the aim. The constructs include formative measurement, which is easier to accommodate in PLS-SEM; the analysis is oriented toward explaining and eventually predicting an outcome construct; and only 27 units are available, which would make covariance-based estimation of a five-construct model unreliable. These are reasons for choosing the method, not guarantees of accurate estimates: small samples limit the precision of any estimator, and PLS-SEM does not remove the need for adequate sample size (Kock & Hadaya, 2018). The same reasoning applies with more force to the segment models, which are estimated on 8–10 units.

Model-fit evaluation. Model performance was described using the coefficient of determination (R^2) for the global model and each segment-specific model, which characterises in-sample explanatory fit only. Because PLS-POS optimises explained variance during segmentation, higher segment-specific R^2 values were not interpreted as evidence of superior predictive performance. The out-of-sample predictive performance of the global model was assessed with PLSpredict (10 folds, 10 repetitions), using Q^2predict and the cross-validated predictive ability test (CVPAT) against the indicator-average and linear-model benchmarks (Shmueli et al., 2019 ;Liengaard et al., 2021). Inference for the global model used bootstrapping with 5,000 subsamples, reporting percentile and bias-corrected confidence intervals and one-tailed p-values as produced by SmartPLS 4; two-tailed equivalents are also given. Bootstrapping was not applied to the segment models, whose 8–10 units are too few for stable resampling, so segment coefficients are reported descriptively.

III. RESULTS AND DISCUSSIONS

A. Result

Results are organised around the global model and its bootstrap and predictive assessment, the segmentation, the segment-specific models, and segment profiles. All coefficients are standardised. Bootstrap inference is reported for the global model only; statements about the segment models concern relative size or direction and should not be read as statements of statistical significance.

Global model. The global PLS-SEM explained 44.7% of the in-sample variance in undernutrition ($R^2 = 0.447$; adjusted $R^2 = 0.347$). Socioeconomic conditions had the largest path coefficient in absolute terms (−0.497), followed by health services and environment (0.370), nutrition behaviour (0.187) and food security (0.176) (Figure 1). Bootstrapping with 5,000 subsamples gave little support for any individual path. Only the socioeconomic path

reached $p < 0.05$ in a one-tailed test ($t = 1.79$; $p = 0.036$), which corresponds to $p \approx 0.07$ two-tailed; its 90% percentile interval included zero (−0.717 to 0.208), whereas the bias-corrected interval did not (−0.875 to −0.219). The paths from health services and environment ($p = 0.179$), nutrition behaviour ($p = 0.135$) and food security ($p = 0.228$) were not significant. Among the formative outer weights, the poverty ratio dominated the socioeconomic composite (1.081, $p < 0.001$, against 0.212 for mean years of schooling, $p = 0.314$). Both food-security indicators carried positive weights (food expenditure 0.849, $p = 0.035$; NCPR 0.819, $p = 0.039$). The health-services-and-environment composite combined positive weights for improved sanitation (0.543), improved drinking water (0.321) and the health worker ratio (0.152) with a negative weight for complete basic immunisation (−0.607), none of them significant at the 5% level. Because higher undernutrition scores denote worse outcomes and no indicator was reverse-coded (Table 1), these weights determine how the signs of the paths are read; they are interpreted in the Discussion.

Measurement model assessment. The reflective constructs showed acceptable internal consistency and convergent validity. Nutrition behaviour had a Cronbach's alpha of 0.966, composite reliability of 0.983, and AVE of 0.967, while undernutrition status had corresponding values of 0.732, 0.848, and 0.657. The formative indicators showed VIF values ranging from 1.081 to 1.274, indicating no substantial collinearity. Stunting had a lower outer loading (0.607) than wasting (0.856) and underweight (0.934), while the rho_A estimate for nutrition behaviour (1.035) warrants further verification.

Predictive assessment. PLSpredict did not support out-of-sample predictive relevance for the global model. Q^2predict was negative for all three outcome indicators (stunting −0.204, wasting −0.206, underweight −0.333) and for the latent undernutrition construct (−0.348), meaning that the cross-validated predictions were less accurate than the naive benchmark of the training-sample mean. In the CVPAT, the average loss of PLS-SEM (4.212) was higher than that of the indicator-average benchmark (3.399; $p = 0.134$) and lower than that of the linear-model benchmark (6.167; $p = 0.048$). The global model should therefore be read as a description of in-sample associations among the 27 subdistricts, not as a predictive tool.

Segmentation. PLS-POS divided the 27 subdistricts into three segments of 8, 10 and 9 subdistricts (29.63%, 37.04% and 33.33%; Table 2). Segment sizes are fairly balanced. The map in Figure 2 shows the distribution; the segments are labelled by their mean latent undernutrition score (relatively low, relatively high and relatively moderate for Segments 1, 2 and 3). The four urban subdistricts of Purwokerto fall into three different segments (Purwokerto Barat in Segment 1, Purwokerto Timur in Segment 2, and Purwokerto Utara and Selatan in Segment 3). Visually, Segment 1 includes a contiguous block in the north-west (Gumelar, Pekuncen, Cilongok, Kedungbanteng) plus Sokaraja, Banyumas and Kemranjen to the east and south-east; Segment 2 forms a central band (Ajibarang, Wangon, Purwojati, Rawalo, Patikraja), a north-eastern group (Sumbang, Kembaran, Karanglewas) and Tambak in the far south-east; and Segment 3 covers the far west and south-west (Lumbir, Jatilawang), the south (Kebasen, Sumpiuh), and east-central and northern units. No geographic information entered the segmentation and no spatial statistic was computed, so this partial spatial coherence is descriptive only.

Table 2. Segment composition of the 27 subdistricts of Banyumas Regency

Segment	n (%)	Subdistricts
1 (Relatively lower latent undernutrition score)	8 (29.63)	Gumelar, Pekuncen, Cilongok, Kedungbanteng, Purwokerto Barat, Sokaraja, Banyumas, Kemranjen
2 (Relatively higher latent undernutrition score)	10 (37.04)	Ajibarang, Wangon, Purwojati, Rawalo, Patikraja, Karanglewas, Sumbang, Kembaran, Purwokerto Timur, Tambak
3 (Close to the overall mean latent undernutrition score)	9 (33.33)	Lumbir, Jatilawang, Kebasen, Baturraden, Purwokerto Utara, Purwokerto Selatan, Kalibagor, Somagede, Sumpiuh

Segment	n (%)	Subdistricts
Total	**27 (100.00)**	

Notes: Membership is based on the PLS-POS segment assignment. Segment labels refer to the mean standardised latent undernutrition score of each segment (Table 5).

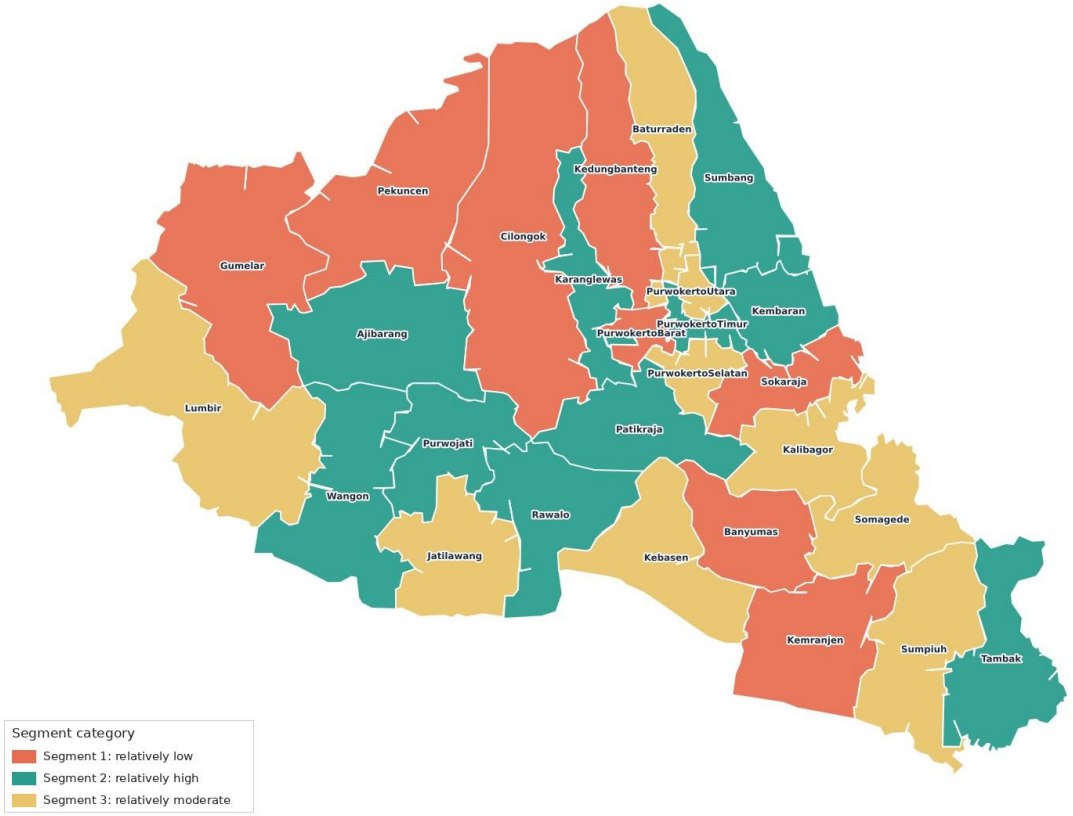


Figure 2. Map of PLS-POS segments of toddler undernutrition across subdistricts of Banyumas Regency. Segment labels refer to the mean standardised latent undernutrition score of each segment.

Segment-specific models. Re-estimating the model within segments yielded R^2 values of 0.969, 0.974 and 0.948 for Segments 1, 2 and 3, respectively, compared with 0.447 for the global model (Table 3); the size-weighted average segment R^2 was 0.964. These values should not be read as better explanation. With four predictors and 8–10 subdistricts per segment, only 3–5 residual degrees of freedom remain, so the segment models are close to saturated, and PLS-POS assigns units precisely so as to maximise within-segment R^2. Near-perfect R^2 under these conditions is a signature of overfitting. The segment path coefficients (Table 4) point the same way. The socioeconomic coefficient in Segment 3 (−1.471) exceeds 1 in absolute value, which for standardised coefficients signals collinearity among construct scores within a small group, and the health-services-and-environment coefficient in Segment 2 (0.983) approaches 1. The socioeconomic path also changes sign between segments (−0.313 and −1.471 in Segments 1 and 3; +0.833 in Segment 2), and the food-security path ranges from −0.341 to 0.134. We report these coefficients only to show that the pattern of estimates is not homogeneous across subdistricts; their individual values are unstable and are not interpreted substantively.

Table 3. Explanatory power of the global and segment-specific models

Model	n	R^2	Approximate adjusted R^2	Residual df
Global	27	0.447	0.347	22

Model	n	R^2	Approximate adjusted R^2	Residual df
Segment 1	8	0.969	0.928	3
Segment 2	10	0.974	0.953	5
Segment 3	9	0.948	0.896	4
Size-weighted mean of segments	27	0.964	–	–

Notes: Global R^2 and adjusted R^2 are as reported by the analysis. Segment adjusted R^2 values were derived from the reported R^2 as $1 - (1 - R^2)(n - 1)/(n - k - 1)$ with $k = 4$ predictors, and are approximate. The size-weighted mean is not equivalent to a pooled R^2 because segment R^2 refers to variance within segments.

Table 4. Standardised path coefficients to toddler undernutrition (Y)

Path	Global	Segment 1	Segment 2	Segment 3
X1 Food security → Y	0.176	−0.033	0.134	−0.341
X2 Socioeconomic → Y	−0.497	−0.313	0.833	−1.471
X3 Nutrition behaviour → Y	0.187	0.337	0.207	0.698
X4 Health services and environment → Y	0.370	0.618	0.983	0.645

Notes: One-tailed bootstrap p-values for the global paths (5,000 subsamples): X1 0.228, X2 0.036, X3 0.135, X4 0.179. Segment coefficients are descriptive and were not bootstrapped. Positive values mean that higher construct scores accompany higher undernutrition scores. Absolute values near or above 1 in the segment models indicate overfitting (see text).

Segment profiles. Mean standardised latent undernutrition scores were −0.116 in Segment 1, 0.099 in Segment 2 and 0.010 in Segment 3 (Table 5). Segment 1 therefore sits below the average of the 27 subdistricts, Segment 2 above it, and Segment 3 close to it. The differences are small (about 0.2 standardised units between the extreme segments), segments contain only 8–10 units, and no inferential test was performed; the labels should therefore be read as descriptive. At the indicator level (Table 6), the segments are similar. Stunting prevalence averaged 14.28, 13.96 and 13.79 in Segments 1–3, wasting 4.70, 4.91 and 4.86, and underweight 11.73, 12.16 and 12.03. Segment 1 thus had the highest mean stunting but the lowest wasting and underweight, and the lowest latent score, indicating that the latent score does not follow stunting alone. Segment 2 had the highest wasting and underweight averages together with comparatively high exclusive breastfeeding and posyandu attendance and a lower poverty ratio, and Segment 3 combined the lowest stunting with the lowest NCPR and schooling values among the three. Membership in a segment does not identify the level of any single subdistrict, because a segment is defined by structural pattern and a subdistrict's own latent score can lie above or below its segment mean.

Table 5. Mean standardised latent undernutrition score by segment

Segment	Mean latent score (Y)	Position relative to the overall mean
1	−0.116	Relatively lower
2	0.099	Relatively higher
3	0.010	Close to the mean

Notes: Scores are standardised, so values are relative positions and not probabilities of undernutrition.

Table 6. Segment means of the raw indicators

Construct	Indicator	Segment 1	Segment 2	Segment 3
Undernutrition	Stunting	14.28	13.96	13.79
	Wasting	4.70	4.91	4.86
	Underweight	11.73	12.16	12.03
Food security	Food expenditure	19.00	19.00	19.43
	NCPR	1.89	1.80	1.74
Socioeconomic	Poverty ratio	13.98	13.21	13.69

Construct	Indicator	Segment 1	Segment 2	Segment 3
	Mean years of schooling	8.95	8.87	8.82
Nutrition behaviour	Exclusive breastfeeding	295.75	311.54	309.08
	Posyandu visits	2,246.84	2,363.35	2,323.77
Health services and environment	Health worker ratio	30.73	30.96	30.92
	Complete basic immunisation	95.63	95.29	95.10
	Improved drinking water	27.95	25.07	26.99
	Improved sanitation	78.93	78.12	78.71

Notes: Values are segment means of the raw indicators. Exclusive breastfeeding and posyandu visits are on the scale recorded by the District Health Office. No inferential tests were applied.

Read together, Tables 4 and 6 show that the segments differ little in the average levels of the indicators; for example, complete basic immunisation ranged from 95.10 to 95.63 and the poverty ratio from 13.21 to 13.98. Whatever distinguishes the segments therefore lies in how the constructs co-vary with the outcome within each group rather than in large level differences, and this is precisely the quantity most affected by small-sample instability.

B. Discussion

Global relationships. The global model explains a little under half of the in-sample variance in the latent undernutrition construct, has no out-of-sample predictive relevance (Q^2predict < 0), and receives little bootstrap support, so its coefficients describe associations among these 27 subdistricts rather than generalisable effects. Their signs can nonetheless be read from the indicator orientation (Table 1) and the outer weights. Because no indicator was reverse-coded and higher undernutrition scores denote worse outcomes, a positive path means that subdistricts scoring higher on a composite tend to have higher undernutrition, and each formative composite rises with its positively weighted indicators. The socioeconomic composite is dominated by the poverty ratio (weight 1.081), so the negative path (−0.497) means that subdistricts with higher poverty ratios tend to have lower latent undernutrition scores. The raw data show the same pattern, since the poverty ratio is negatively correlated with wasting ($r = -0.64$) and underweight ($r = -0.32$). This runs opposite to the Indonesian evidence that low household socioeconomic status is a consistent determinant of stunting (Beal et al., 2018), and we do not read it as a protective effect of poverty. A weight above 1 alongside a positive weight for schooling, an indicator negatively correlated with poverty ($r = -0.46$), points to suppression within the composite. Aggregation can also reverse relationships that hold for individuals (Robinson, 1950), and differences in the completeness of growth-monitoring records between subdistricts could produce the same pattern, although we have no data to test this. The finding is also fragile, being significant only one-tailed. The food-security composite rises with both NCPR and food expenditure, but the two indicators point in opposite directions. A higher NCPR denotes a larger gap between normative consumption and net production, that is, weaker local food availability, whereas higher household food expenditure indicates greater spending on food and, in that sense, better access. The composite therefore mixes a supply-side deficit with a demand-side access indicator, and the small positive path (0.176), which was not significant ($p = 0.228$), cannot be read as either better or worse food security. Its small size is in line with other evidence: in East Java the path from food security to malnutrition was not significant (Agustina & Arumningtyas, 2024), the pooled estimate across cohort studies was not distinguishable from no association (Patriota et al., 2024), and in Aceh food intake rather than food availability indicators was the principal determinant (Surayya & Sirait, 2024). Aggregate indicators of regional food supply and expenditure capture access and availability

at most, whereas household utilisation and dietary quality, which are not in our model, may matter more for toddlers (Jones et al., 2013; Yadav et al., 2024).

The positive path from health services and environment. The positive coefficient for health services and environment (0.370 globally and positive in all three segments) is the least intuitive finding. The composite rises with improved sanitation, improved drinking water and the health worker ratio and falls with complete basic immunisation, so the positive path means that subdistricts with better sanitation and water access but lower immunisation coverage tend to have somewhat higher undernutrition scores. Because the weights have mixed signs and none is significant, and the path itself was not significant ($p = 0.179$), the composite does not represent "better services" in a simple sense and the path should not be read as harm from services. Two further explanations are plausible: services and infrastructure may be concentrated where need is greatest, which produces a positive cross-sectional association without implying harm, and correlation between this construct and socioeconomic conditions may distort the individual coefficients in a model estimated on 27 units. A comparable pattern appears in a provincial PLS-POS study, where the health-services construct had a negative path to a health-degree construct, and the authors concluded that improved access alone does not always translate into better outcomes (Nurhanifa et al., 2026). That parallel is suggestive and no more, since the outcome constructs differ. The policy reading of the X4 path is therefore neither "more services produce more undernutrition" nor "strengthen services to reduce it"; it is that the composition and orientation of service and environment variables need to be examined before they guide allocation.

What the segmentation does and does not show. The jump in R^2 from 0.447 to values above 0.94 is the most visible result of the segmentation analysis, and it needs careful reading. Part of it is built into the method: PLS-POS reassigns units to maximise explained variance, so some improvement in fit is expected even when the data contain no true subpopulations (Becker et al., 2013). A second part is a consequence of sample size. With four predictors and 8–10 observations per segment, the residual degrees of freedom are only about 3, 5 and 4, the approximate adjusted R^2 values drop to roughly 0.93, 0.95 and 0.90 (Table 3), and the models are close to saturated. The inverse square root method suggests that detecting a path coefficient of 0.30 with 80% power at the 5% level requires about 69 observations (approximately $(2.486/0.30)^2$; Kock & Hadaya, 2018), far more than any single segment contains; the older 10-times rule has also been criticised as imprecise (Kock & Hadaya, 2018). A third consideration is that segment R^2 values describe variance within a segment, so the size-weighted mean of 0.964 is not comparable on a like-for-like basis with the global R^2, which also contains between-segment variance. The coefficient above 1, the near-1 value in Segment 2 and the sign reversals of the socioeconomic path (-1.471 in Segment 3 and $+0.833$ in Segment 2) point in the same direction: they are more consistent with overfitting and unstable estimation under collinearity and small n than with a stable economic mechanism that reverses sign across places. For these reasons we regard the segmentation as hypothesis-generating. It indicates that a single set of paths is probably an oversimplification, but the specific segment-level coefficients should not be used for causal or quantitative inference.

Segments, geography and relative level. The labels attached to the segments are based on mean latent scores that differ by about 0.2 standardised units; indicator means differ by less than one percentage point for stunting, wasting and underweight. Segment 1 combines the lowest latent score with the highest mean stunting, which shows that the latent construct weights indicators unequally and that rankings based on the construct need not agree with rankings based on stunting alone. The map shows that segments are partly spatially coherent, although the algorithm used no spatial information; whether this reflects shared geography, administrative practice or chance would require a spatial statistic, and

previous work indicates only modest spatial autocorrelation in stunting at the provincial scale (Sari & Siregar, 2023). The placement of the four urban Purwokerto subdistricts in three different segments suggests that urban versus rural status alone does not define segment membership, but this observation is also descriptive.

Implications for programming. Because the segment-specific coefficients are unstable and the global model lacks predictive relevance, the results do not identify which intervention should be prioritised in which subdistrict, and we do not derive segment-specific programme recommendations from them. What the analysis offers is narrower: evidence that a single regency-wide set of paths may oversimplify how food-security, socioeconomic and service conditions relate to toddler undernutrition, and a set of questions for local assessment, such as why neighbouring subdistricts fall into different segments and whether service coverage follows need. Any targeting decision would require validated estimates from larger or multi-year data together with information on programme coverage, local service capacity and household-level nutrition. Because the analysis uses aggregated subdistrict-level data, the findings describe patterns across subdistricts and should not be interpreted as conclusions about individual children or households (Robinson, 1950).

Measurement considerations. Several measurement considerations constrain the interpretation of the results. First, the formative constructs are defined by their indicators, and their interpretation depends on indicator selection, coding, and the estimated weights; the food-security composite, for example, combines two positively weighted indicators of opposite orientation (NCPR and food expenditure). Although the formative indicators showed no substantial collinearity in the global model, the health services and environment construct combines service-delivery indicators (health personnel and immunisation) with environmental indicators (drinking water and sanitation). These domains may represent distinct mechanisms that a single composite cannot fully distinguish. Second, nutrition behaviour is measured by only two reflective indicators, which limits the breadth of the construct and its ability to capture other dimensions of nutritional practices. Third, the undernutrition construct combines stunting, wasting, and underweight, which represent related but distinct nutritional outcomes. Stunting generally reflects chronic nutritional deprivation, whereas wasting may respond more rapidly to changes in nutritional and health conditions. Consequently, the latent undernutrition score need not rank subdistricts in the same order as stunting prevalence alone, as illustrated by the contrasting patterns observed in Segment 1. Further analyses could examine the three outcomes separately and assess the robustness of the results to alternative indicator specifications or fixed composite weights. In addition, exclusive breastfeeding and posyandu visits enter on the scale recorded by the health office (Table 6); if these are counts rather than coverage rates, they partly reflect subdistrict population size, and coverage rates would be preferable in future specifications.

Using the segment map responsibly. The map is likely to attract more attention than the coefficients behind it, and it can be misread as a ranking of risk. Its colours show which segment a subdistrict was assigned to by the algorithm, not how many children are undernourished there. Two subdistricts in the same colour may have very different latent scores, and a subdistrict with a relatively low score may still contain villages with serious problems. The map is therefore best used as a screening device that prompts questions for local health and agriculture offices, for example why neighbouring subdistricts fall in different segments or which indicators drive the assignment of a particular subdistrict, and not as a basis for deciding that some subdistricts need no attention. The segmentation should also be updated when new annual data become available, because membership may change as the underlying indicators move.

Contribution relative to earlier work. Previous Indonesian PLS-SEM studies of toddler undernutrition have mostly produced a single set of path estimates for a province or district set (Agustina & Arumningtyas, 2024 ;A. Sari & Siregar, 2023 ;Surayya & Sirait, 2024).

Segmentation of Indonesian provinces with SEM-PLS and POS has been applied to maternal and child health degree (Nurhanifa et al., 2026) . The present analysis moves the same idea down to subdistricts of one regency, places food security alongside socioeconomic, behavioural and service constructs, and keeps explanatory and predictive evaluation separate. Its contribution is therefore methodological and applied, an illustration of how heterogeneity can be explored and where the evidence is thin, and not an estimate of effects.

Limitations and next steps. This study has several limitations. First, the analysis included only 27 subdistricts, and the segment-specific models were estimated using 8–10 units each. The resulting path coefficients may therefore be unstable and should be interpreted as exploratory rather than definitive evidence of structural heterogeneity. Second, the data were aggregated, cross-sectional, and obtained from administrative sources. Measurement error, reverse causality, and ecological bias (Robinson, 1950) cannot be ruled out, and the findings cannot be directly generalised to individual children or households. Third, the food-security construct was represented by only two aggregate indicators and did not capture all dimensions of household food security, such as access, utilisation, and stability. Fourth, although the global measurement model was assessed, further checks are warranted, including verification of the rho_A estimate for nutrition behaviour and sensitivity analyses of alternative measurement specifications. No inferential claims are made about the segment-specific path coefficients, and the analysis does not establish causal effects. Fifth, bootstrapping supported none of the global paths at the 5% level in a two-tailed test, the segment models could not be bootstrapped, and PLSpredict indicated no out-of-sample predictive relevance for the global model; none of the coefficients should therefore be read as an established effect. Future work should evaluate predictive performance using a validation procedure that repeats segmentation within each training sample, assess robustness to alternative numbers of segments and indicator specifications, and extend the analysis using village-level or multi-year data.

IV. CONCLUSION

This exploratory study found that a single regency-wide PLS-SEM describes toddler undernutrition across the 27 subdistricts of Banyumas Regency only partially: the global model explained 44.7% of the in-sample variance, socioeconomic conditions, a composite dominated by the poverty ratio, had the largest coefficient but only weak bootstrap support and a direction opposite to individual-level evidence, and the model showed no out-of-sample predictive relevance. PLS-POS identified three segments of 8–10 subdistricts with differing estimated path patterns, but the near-saturated segment models, a coefficient exceeding 1 and sign reversals indicate overfitting, so the segmentation is evidence of possible heterogeneity rather than a validated typology. The results are therefore hypothesis-generating and are not yet suitable for targeting interventions; validation with larger samples, village-level or multi-year data, and resampling-based inference is needed before segment-specific findings can inform programme decisions.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Institute for Research and Community Service (Lembaga Penelitian dan Pengabdian kepada Masyarakat, LPPM), Universitas Jenderal Soedirman, for funding this research through the Research Competency Enhancement scheme (Riset Peningkatan Kompetensi) under contract number 20.87/UN23.34/PT.01.00/IV/2026. The authors also thank the Banyumas District Health Office, the Banyumas Office of Agriculture and Food Security, and Statistics Indonesia (BPS) for providing the data used in this study.

REFERENCES

Agustina, F. R., & Arumningtyas, F. (2024). Analisis SEM-PLS pada kasus gizi buruk balita di Provinsi Jawa Timur. *Skewness: Jurnal Statistika, Aktuaria Dan Sains Data*, *1*(2), 63–74. http://jos.unsoed.ac.id/index.php/skewness/

Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. *Maternal & Child Nutrition*, *14*(4). https://doi.org/10.1111/mcn.12617

Becker, J.-M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats1. *MIS Quarterly*, *37*(3), 665–694. https://doi.org/10.25300/MISQ/2013/37.3.01

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, *382*(9890), 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X

Dinas Komunikasi dan Informatika Kabupaten Banyumas. (2025). *Rembuk Stunting 2025 di Banyumas: Kolaborasi untuk generasi sehat*. https://dinkominfo.banyumaskab.go.id/news/52701/rembuk-stunting-2025-di-banyumas-kolaborasi-untuk-generasi-sehat.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. In *European Business Review* (Vol. 31, Number 1, pp. 2–24). Emerald Group Publishing Ltd. https://doi.org/10.1108/EBR-11-2018-0203

Hair, Jr. , J. F., Sarstedt, M., Matthews, L. M., & Ringle, C. M. (2016). Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method. *European Business Review*, *28*(1), 63–76. https://doi.org/10.1108/EBR-09-2015-0094

Islam, B., Ibrahim, T. I., Wang, T., Wu, M., & Qin, J. (2025). Current trends in household food insecurity, dietary diversity, and stunting among children under five in Asia: a systematic review. *Journal of Global Health*, *15*, 04049. https://doi.org/10.7189/jogh.15.04049

Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What Are We Assessing When We Measure Food Security? A Compendium and Review of Current Metrics. *Advances in Nutrition*, *4*(5), 481–505. https://doi.org/10.3945/an.113.004119

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, *28*(1), 227–261. https://doi.org/10.1111/isj.12131

Liengaard, B. D., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2021). Prediction: Coveted, yet forsaken? Introducing a cross-validated

predictive ability test in partial least squares path modeling. *Decision Sciences*, *52*(2), 362–392. https://doi.org/10.1111/deci.12445

Nurhanifa, N., Perdana, H., & Satyahadewi, N. (2026). SEGMENTING MATERNAL AND CHILD HEALTH DEGREE IN INDONESIA USING SEM-PLS POS. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *20*(2), 0897–0912. https://doi.org/10.30598/barekengvol20iss2pp0897-0912

Patriota, É. S. O., Abrantes, L. C. S., Figueiredo, A. C. M. G., Pizato, N., Buccini, G., & Gonçalves, V. S. S. (2024). Association between household food insecurity and stunting in children aged 0−59 months: Systematic review and meta-analysis of cohort studies. *Maternal & Child Nutrition*, *20*(2). https://doi.org/10.1111/mcn.13609

Pemerintah Provinsi Jawa Tengah. (2025, November 12). *Sukses intervensi penurunan stunting, Gubernur Ahmad Luthfi terima penghargaan Kemenkes*. https://jatengprov.go.id/publik/sukses-intervensi-penurunan-stunting-gubernur-ahmad-luthfi-terima-penghargaan-kemenkes/

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357. https://doi.org/10.2307/2087176

Sari, A., & Siregar, R. (2023). A Spatial Analysis of Stunting and Its Determinants in Indonesia. *Proceedings of the 5th International Conference on Social Determinants of Health*, 198–202. https://doi.org/10.5220/0012903400004564

Sari, P. I., Dimiati, H., Sofia, S., & Subianto, M. (2025). Predicting stunting incidence in Pidie, Indonesia using SEM-PLS and GIS: A multilevel analysis of determinants. *International Journal of ADVANCED AND APPLIED SCIENCES*, *12*(8), 273–286. https://doi.org/10.21833/ijaas.2025.08.025

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Surayya, N. N., & Sirait, T. (2024). DETERMINANTS MODELING OF UNDERNUTRITION IN TODDLERS IN ACEH PROVINCE: A PLS-SEM APPROACH. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(3), 1853–1864. https://doi.org/10.30598/barekengvol18iss3pp1853-1864

Yadav, S. S., Matela, H., Panchal, P., & Menon, K. (2024). Household food insecurity, dietary diversity with undernutrition among children younger than five years in Indian subcontinent–a narrative review. *The Lancet Regional Health - Southeast Asia*, *26*, 100426. https://doi.org/10.1016/j.lansea.2024.100426