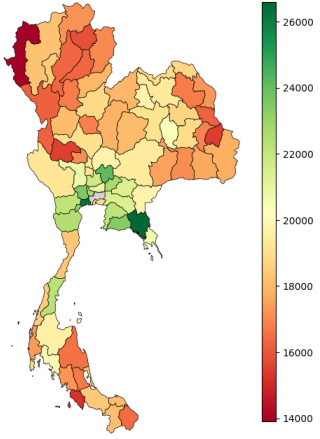
Socioeconomic Regionalization and Bayesian Hierarchical Modeling of Thailand

Irving Gómez Méndez

March, 2026

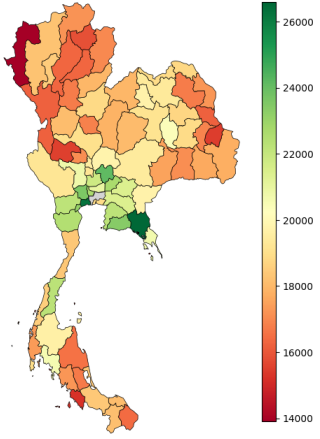
Seminar Talk - AI and Computer Engineering Program

Spatial Patterns of Income in Thailand



Monthly Income by Province (y)

Spatial Patterns of Income in Thailand

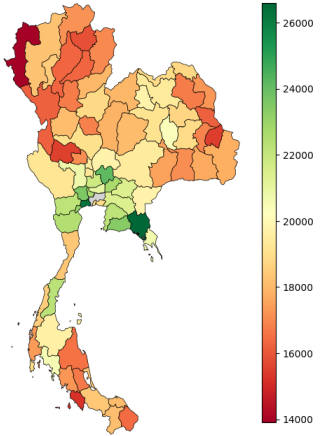


Monthly Income by Province (y)

Visual Inspection:

➔ High-income cluster centered on the Bangkok Metropolitan Region.

Spatial Patterns of Income in Thailand

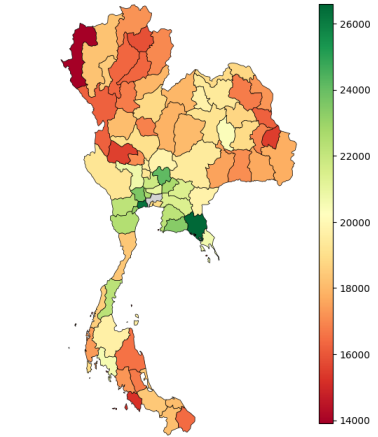


Monthly Income by Province (y)

Visual Inspection:

- ➔ High-income cluster centered on the Bangkok Metropolitan Region.
- ➔ Persistent poverty ”pockets” in the North, Northeast (Isan) and Deep South.

Spatial Patterns of Income in Thailand



Monthly Income by Province (y)

Visual Inspection:

- ➔ High-income cluster centered on the Bangkok Metropolitan Region.
- ➔ Persistent poverty ”pockets” in the North, Northeast (Isan) and Deep South.
- ➔ **Is this geographic pattern statistically significant, or is it merely a random realization?**

Spatial Autocorrelation: The Concept

To determine if the observed patterns are significant, we compare them against three theoretical states:

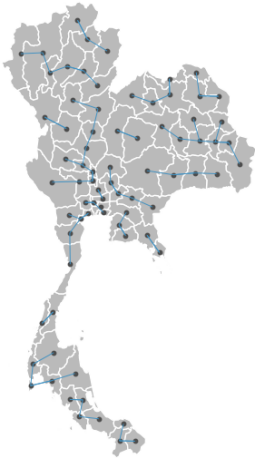
a) Positive **b) Negative** **c) Zero (Random)**

- **Positive:** Clusters of similar values (Wealthy neighbors wealthy).
- **Negative:** A "checkerboard" of dissimilar values.
- **Zero:** The **Null Hypothesis** — no discernible spatial structure.

Tobler's First Law (1970) posits that geographic data naturally leans toward state (a).

Quantifying "Nearness": The k-NN Network

We must first define the spatial topology of Thailand:



$k = 1$

We define the weight matrix W as:

$$w_{ij} = \begin{cases} 1/k & \text{if province } j \in \text{knn}(i), \\ 0 & \text{otherwise.} \end{cases}$$

This allows us to calculate the **Spatial Lag** (the neighborhood average):

$$\text{lag-}y_i = \frac{1}{k} \sum_{j \in \text{knn}(i)} y_j.$$

Quantifying "Nearness": The k-NN Network

We must first define the spatial topology of Thailand:

$k = 10$

We define the weight matrix W as:

$$w_{ij} = \begin{cases} 1/k & \text{if province } j \in \text{knn}(i), \\ 0 & \text{otherwise.} \end{cases}$$

This allows us to calculate the **Spatial Lag** (the neighborhood average):

$$\text{lag-}y_i = \frac{1}{k} \sum_{j \in \text{knn}(i)} y_j.$$

Quantifying "Nearness": The k-NN Network

We must first define the spatial topology of Thailand:

$k = 5$ (Optimal)

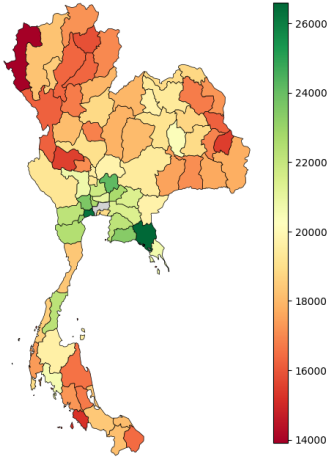
We define the weight matrix W as:

$$w_{ij} = \begin{cases} 1/k & \text{if province } j \in \text{knn}(i), \\ 0 & \text{otherwise.} \end{cases}$$

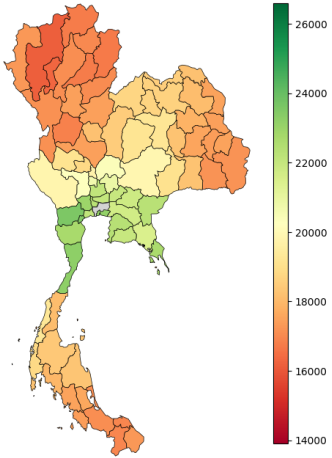
This allows us to calculate the **Spatial Lag** (the neighborhood average):

$$\text{lag-}y_i = \frac{1}{k} \sum_{j \in \text{knn}(i)} y_j.$$

The Spatial Lag: Filtering the Map



(1) Raw Income (y)



(2) Spatial Lag (Wy)

By correlating map (1) with map (2), we obtain a single metric of clustering: Moran's I.

Global Moran's I: A Weighted Correlation

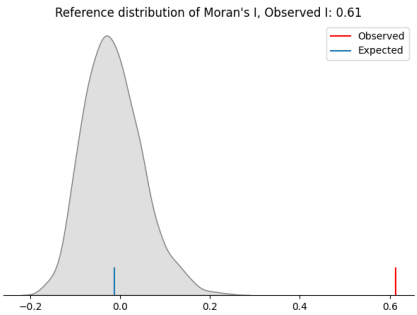
Moran's I is essentially the **Pearson Correlation** between a province and its neighborhood. It is a weighted correlation that measures how much a province "agrees" with its neighbors.

$$I = \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y}) / \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}{\sum_i (y_i - \bar{y})^2 / n}$$

- **The Core Mechanism:** It sums the product of deviations $(y_i - \bar{y}) \cdot (y_j - \bar{y})$ *only* for pairs that are neighbors ($w_{ij} > 0$).
- **Interpretation:**
  - If wealthy provinces are near wealthy ones: $(+) \cdot (+) =$ Positive I.
  - If poor provinces are near poor ones: $(-) \cdot (-) =$ Positive I.

Permutation Test

To test if our observed I is "real," we compare it against a **Null Hypothesis** where geography does not matter.



The "Shuffle" Procedure:

1. Take the 76 income values of Thailand.
2. Randomly shuffle them across the map (breaking any spatial link).
3. Calculate I for this random map.
4. Repeat 1,000 times to build the Reference Distribution.

Result: Our observed $I = 0.61$ is far outside the random range. $p < 0.001$

Conclusion: The spatial clustering of income in Thailand is *not* a coincidence.

Global Moran's I: The Regression Interpretation

Because $\sum_i \sum_j w_{ij} = n$, then the I statistic simplifies to:

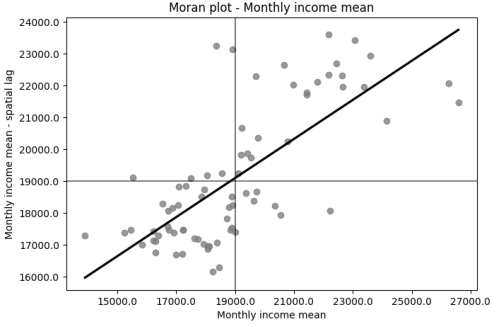
$$I = \frac{\sum_{i=1}^{n} z_i \left(\sum_{j=1}^{n} w_{ij} z_j \right)}{\sum_{\ell=1}^{n} z_\ell^2} = \frac{\sum_{i=1}^{n} z_i \cdot \text{lag-}z_i}{\sum_{\ell=1}^{n} z_\ell^2},$$

where $z_i = y_i - \bar{y}$.

This is exactly the OLS estimator $\hat{\beta}_1$ for the model:

$$\text{lag-}z = \beta_0 + \beta_1 z + \varepsilon$$

- ➔ Since z is mean-centered, the intercept $\hat{\beta}_0 = 0$.
- ➔ **Key Insight:** Moran's I is the "spatial consistency" slope.



Local Moran's I

Global Moran's I provides a single summary for all of Thailand. However, spatial processes are rarely stationary. We can decompose I into local contributions:

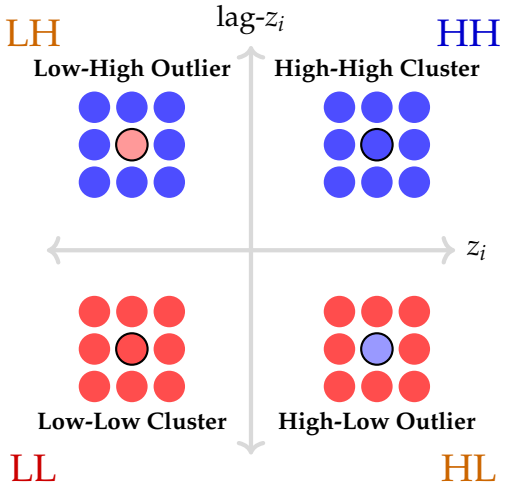
$$I = \frac{\sum_{i=1}^{n} z_i \text{lag-}z_i}{\sum_{\ell=1}^{n} z_\ell^2} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{z_i \cdot \text{lag-}z_i}{\frac{1}{n} \sum_{\ell=1}^{n} z_\ell^2} \right) = \frac{1}{n} \sum_{i=1}^{n} I_i$$

where the **Local Moran's** I_i for observation i is:

$$I_i = \frac{z_i \cdot \text{lag-}z_i}{m_2}, \quad \text{with } m_2 = \frac{1}{n} \sum z_\ell^2$$

- ➔ $I_i > 0$: Province i is similar to its neighbors (**Clustering**).
- ➔ $I_i < 0$: Province i is different from its neighbors (**Outlier**).

Local Moran's I: Quadrant Interpretation

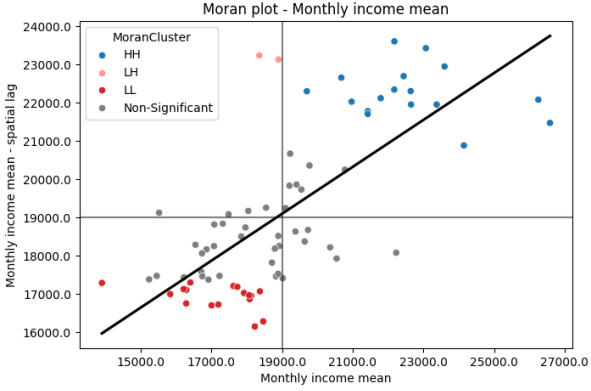


The sign of I_i identifies four distinct types of spatial association:

- ➔ **Positive I_i:** Similar values cluster together (HH or LL).
- ➔ **Negative I_i:** Spatial anomalies or "checkboards" (LH or HL).

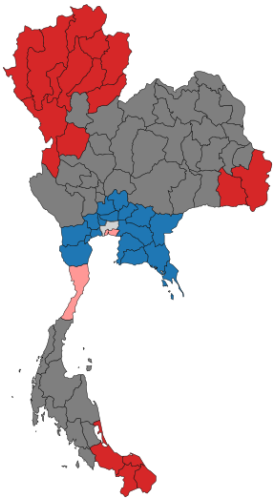
Identifying Moran's Clusters (1/2)

Not every point in the HH quadrant is a "cluster." To identify meaningful clusters, we test each I_i using a permutation test (shuffling neighbors for each province).



- **HH (Blue):** Significant wealth clusters.
- **LL (Red):** Significant poverty traps.
- **LH (Pink):** Low-income provinces surrounded by wealthy ones.
- **Gray:** No significant spatial pattern.

Identifying Moran's Clusters (2/2)



- **HH Clusters:** The economic engine of Bangkok and surrounding provinces.
- **LL Clusters:** Structural poverty in the North, Northeast and Deep South.

Feature Engineering: The 9 Dimensions of Poverty

We don't just look at income. We analyze 9 variables across 5 socio-economic aspects. Note that **all** exhibit significant spatial clustering:

Aspect	**Variable**	**Moran's I**
Education	Years of education*	0.64
Income	Monthly income	0.61
	Yearly savings	0.25
	% Households w/o savings	0.56
Inequality	Monthly income ratio 20:20	0.47
	Gini index	0.31
Debt	% Households w/ formal debt	0.72
Living	Alcohol consumption	0.53
	Smoking	0.68

*Estimated due to lack of primary data.

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").
- ✗ **Fragmented:** Often results in non-contiguous "islands."

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").
- ✗ **Fragmented:** Often results in non-contiguous "islands."

Hierarchical Clustering Benefits:

- ✓ **Multivariate:** Considers all 9 socio-economic variables.

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").
- ✗ **Fragmented:** Often results in non-contiguous "islands."

Hierarchical Clustering Benefits:

- ✓ **Multivariate:** Considers all 9 socio-economic variables.
- ✓ **Non-linear:** Captures complex relations missed by linear I.

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").
- ✗ **Fragmented:** Often results in non-contiguous "islands."

Hierarchical Clustering Benefits:

- ✓ **Multivariate:** Considers all 9 socio-economic variables.
- ✓ **Non-linear:** Captures complex relations missed by linear I.
- ✓ **Exhaustive:** Every province belongs to a region.

From Clusters to Regions

Local Moran's I identifies spatial "hotspots," but it is insufficient for a national structural model due to three main drawbacks:

Moran's I Drawbacks:

- ✗ **Univariate:** Only considers one variable at a time.
- ✗ **Incomplete:** Leaves many provinces unassigned ("holes").
- ✗ **Fragmented:** Often results in non-contiguous "islands."

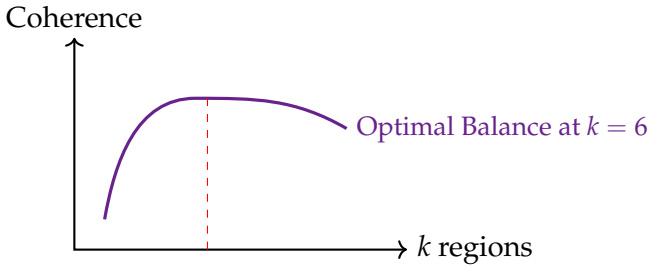
Hierarchical Clustering Benefits:

- ✓ **Multivariate:** Considers all 9 socio-economic variables.
- ✓ **Non-linear:** Captures complex relations missed by linear I.
- ✓ **Exhaustive:** Every province belongs to a region.
- ✓ **Cohesion:** Produces spatially and structurally contiguous regions.

Optimization: The k-Region Trade-off

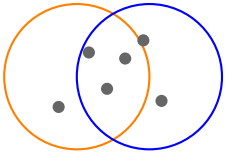
Determining the number of regions (k) is a multi-objective optimization problem. We seek a value for k that balances:

1. **Feature Coherence:** How similar are provinces within a region based on the 9 socio-economic variables?
2. **Geographical Coherence:** How compact and contiguous are the resulting shapes?

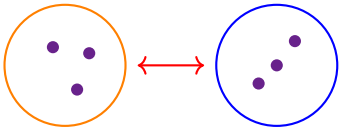


Metric 1: Feature Coherence

We use the **Silhouette Score** and the **Calinski-Harabasz Score** to measure how well-defined the regions are in the 9-dimensional variable space.



Low Score
Clusters overlap/mix



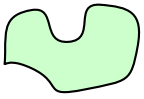
High Score
Well-separated regions

Goal: Maximize the "feature" distance between different regions while minimizing the distance between provinces in the same region.

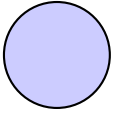
Metric 2: Geographic Coherence (IPQ)

To avoid "gerrymandered" or fragmented regions, we use the **Isoperimetric Quotient (IPQ)**.

$$Q = \frac{4\pi A}{L^2}$$



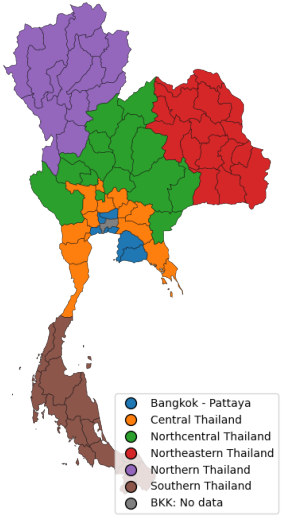
$Q \ll 1$ (Elongated)



$Q = 1$ (Perfectly Compact)

- **Area (A):** The size of the region.
- **Perimeter (L):** The length of the boundary.
- **Trade-off:** A region should be a "solid block" rather than a thin string of provinces.

The 6 Resulting Regions



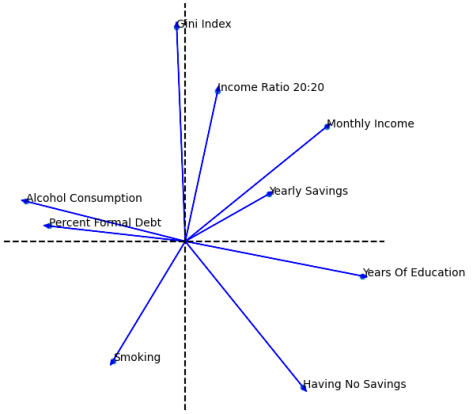
The optimization yields 6 functional regions:

1. **Bangkok-Pattaya**
2. **Central Thailand**
3. **Northcentral Thailand**
4. **Northeastern Thailand**
5. **Northern Thailand**
6. **Southern Thailand**

Note: These regions are spatially contiguous and group provinces with similar multivariate poverty profiles.

Multivariate Characterization: PCA

To interpret the 6 regions, we perform a Principal Component Analysis (PCA) on the 9 variables. This reduces the dimensionality while preserving the socio-economic ”signal.”



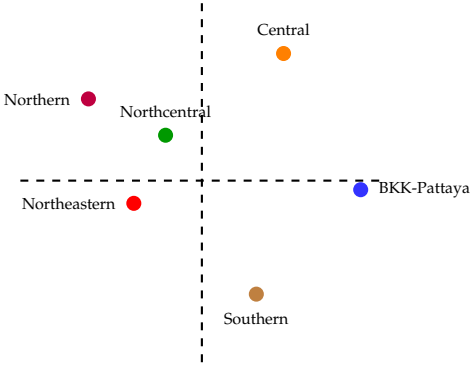
PC1 (38% Variance):

- ➔ Positive: Education, Income.
- ➔ Negative: Formal Debt, Alcohol.
- ➔ *Interpretation:* A ”Development/Wealth” axis.

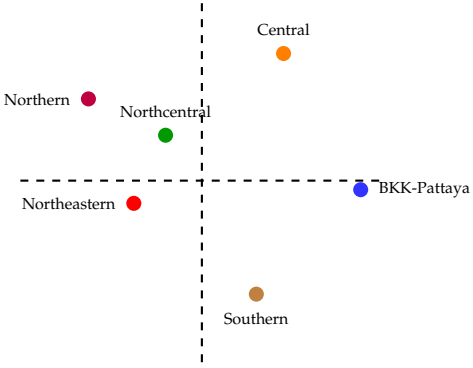
PC2 (23% Variance):

- ➔ High loading on Gini and Income Ratio.
- ➔ *Interpretation:* An ”Inequality” axis.

Characterizing the 6 Regions

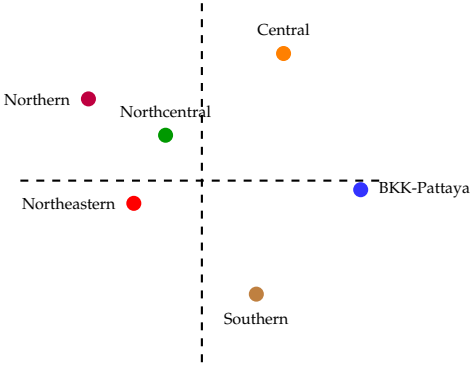


Characterizing the 6 Regions



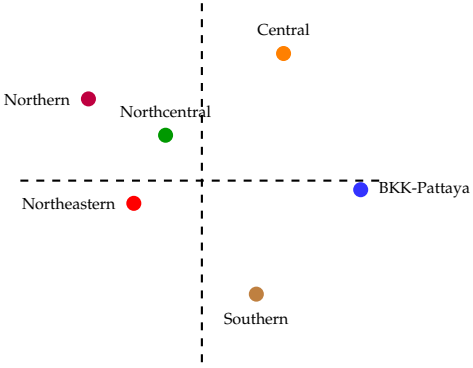
➔ **BKK-Pattaya:** Maximum PC1. Extreme wealth and education.

Characterizing the 6 Regions



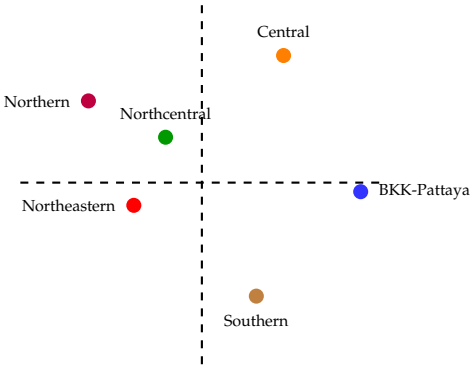
- **BKK-Pattaya:** Maximum PC1. Extreme wealth and education.
- **Central (Buffer):** High PC1 (Wealth) but also **Maximum PC2 (Inequality)**. This confirms it as a transition zone.

Characterizing the 6 Regions



- **BKK-Pattaya:** Maximum PC1. Extreme wealth and education.
- **Central (Buffer):** High PC1 (Wealth) but also **Maximum PC2 (Inequality)**. This confirms it as a transition zone.
- **North/Northeast:** Negative PC1. The ”Poverty Trap” profile—high debt and low education.

Characterizing the 6 Regions



- **BKK-Pattaya:** Maximum PC1. Extreme wealth and education.
- **Central (Buffer):** High PC1 (Wealth) but also **Maximum PC2 (Inequality)**. This confirms it as a transition zone.
- **North/Northeast:** Negative PC1. The ”Poverty Trap” profile—high debt and low education.
- **Southern:** High loadings on ”No Savings” aspect.

Quantifying Education: From Degrees to Years

We map formal educational levels to a continuous scale of "Years of Education" based on the standard Thai curriculum:

Education Level	Years
Uneducated / Kindergarten	0
Pre-elementary school	3
Elementary school	6
Junior high school	9
Senior high school	**12**
Vocational degree	14
Bachelor degree	16
Post-graduate	19

Why this mapping?

- ➔ Transforms categorical data into a **continuous interval scale**.
- ➔ Allows us to interpret the regression coefficient β as the *"Return for every additional year of schooling."*

Regional Educational Baselines

Using this metric, we calculated the average years of formal education across the 6 regions:

Region	Avg. Years
Bangkok-Pattaya	**12.52**
Southern Thailand	11.26
Central Thailand	11.16
Northeastern Thailand	10.49
Northcentral Thailand	10.42
Northern Thailand	10.32
National Average	*10.87*

Key Observations:

- ➔ Only **Bangkok-Pattaya** exceeds the 12-year threshold (Senior High School).
- ➔ The **Northern** and **Northeastern** regions lag by approximately 2 years compared to the capital.
- ➔ **The Question:** How much does this 2-year gap explain the income disparities?

Note: According to Our World in Data the average years of formal education for Thailand is 9.3 years.

The Model: Return on Education

We want to quantify the influence of education on income while accounting for regional differences.

The Research Question

Does an extra year of education yield the same income increase in the ”Buffer Zone” (Central) as it does in the ”Poverty Traps” (North)?

Why Bayesian Hierarchical Regression?

- ➔ **Varying Slopes:** Allows each region to have its own ”Return on Education” (β_j).
- ➔ **Partial Pooling:** ”Borrows strength” from the national average to stabilize estimates in regions with fewer provinces.
- ➔ **Uncertainty:** Provides a full posterior distribution for policy risk assessment.

The Regional Model: Robust Linear Regression

We model the income of province i in region j, Y_{ij} as a function of education, X_{ij}:

Likelihood (Robust)

$$Y_{ij} \sim \text{Laplace}_{\text{Truncated}}(\mu_{ij}, \sigma_j)$$

Linear Predictor

$$\mu_{ij} = \alpha_j + \beta_j(X_{ij} - \bar{X}_{\cdot j})$$

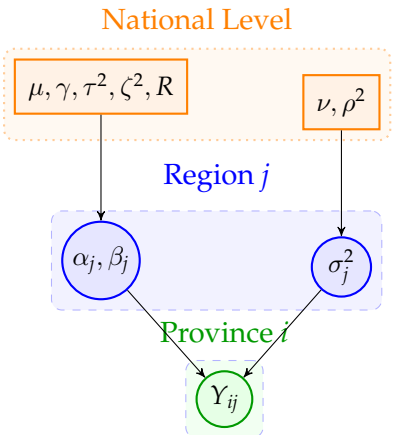
Parameter Interpretation:

- ➔ α_j: **Regional Intercept**
  Expected income at the regional mean education level.
- ➔ β_j: **Regional Slope**
  The "Return on Education" within region j.
- ➔ σ_j: **Regional Scale**
  Intra-regional income dispersion.

Note: Centering education X_{ij} around the regional mean $\bar{X}_{\cdot j}$ ensures α_j is interpretable as the average income of the region.

The National Hierarchy: Partial Pooling

To "borrow strength" across Thailand, we treat regional parameters as draws from a national distribution:



National Interpretation:

- μ: **National Average Income**
  The central tendency for regional intercepts α_j.
- γ: **National Return on Edu.**
  The central tendency for regional slopes β_j.
- R: **Correlation Matrix**
  Captures the correlation between wealth and education returns.

Partial pooling allows regions with fewer provinces to have their estimates stabilized by the national average.

The Statistical Model: Regional Level

Likelihood (Robust)

$$Y_{ij} \sim \text{Laplace}_{\text{Truncated}}(\mu_{ij}, \sigma_j)$$

Where:

$$\mu_{ij} = \alpha_j + \beta_j(X_{ij} - \bar{X}_{\cdot j})$$

Note: The Laplace distribution provides robustness against income outliers compared to Gaussian models.

Regional Parameters

$$\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu \\ \gamma \end{bmatrix}, S \right)$$

$$\sigma_j^2 \sim \text{Inv-}\chi^2(\nu, \rho^2)$$

The covariance matrix S is decomposed as:

$$S = \text{diag}(\tau, \zeta)\, R\, \text{diag}(\tau, \zeta)$$

where R is the correlation matrix.

The Statistical Model: National Level

To complete the hierarchy, we define weakly informative hyper-priors:

Location and Correlation

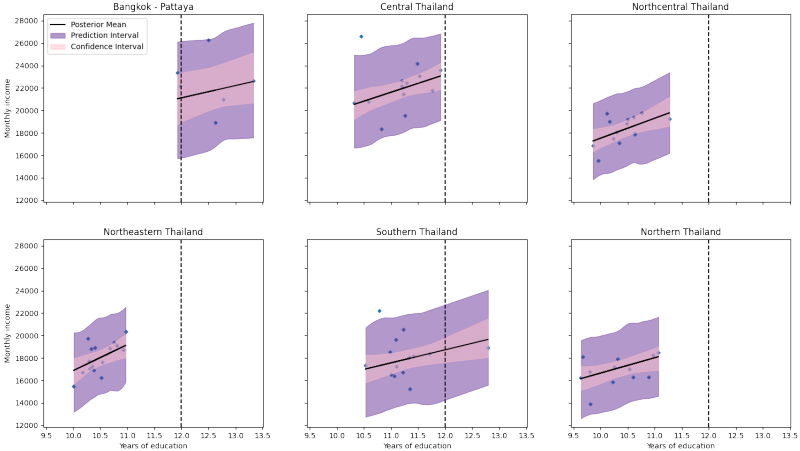
$$\mu \sim \text{Normal}(\hat{\mu}, \hat{\sigma}_{\mu}^2)$$
$$\gamma \sim \text{Normal}(\hat{\gamma}, \hat{\sigma}_{\gamma}^2)$$
$$\tau^2 \sim \text{Exponential}(1/\hat{\tau}^2)$$
$$\zeta^2 \sim \text{Exponential}(1/\hat{\zeta}^2)$$
$$R \sim \text{LKJ}(2)$$

Variance

$$\nu^2 \sim \text{Exponential}(1/\hat{\nu}^2)$$
$$p(\rho^2) \propto \frac{1}{\rho^2} \mathbb{1}_{(0,\infty)}(\rho^2)$$

We use the LKJ(2) prior to favor the identity matrix (no correlation) while allowing the data to inform $\rho_{\alpha,\beta}$.

Regional Regression Results: Education vs. Income



Model Components:

- **Solid Line:** Posterior Mean.
- **Pink Area:** 95% Credible Interval (Uncertainty in the mean).
- **Purple Area:** 95% Prediction Interval (Expected range for a new province).

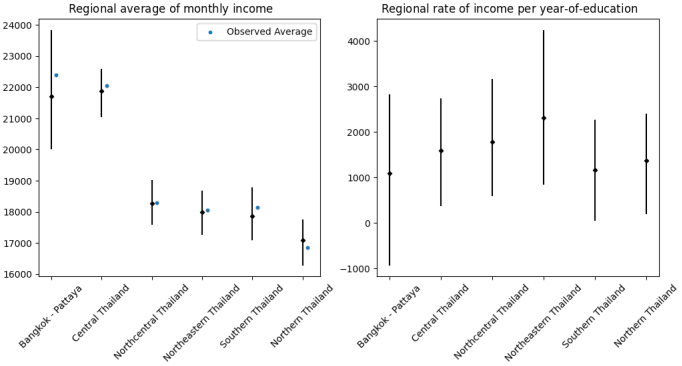
Note: The vertical dashed line represents **12 years** of education.

The 12-Year Threshold: A National Barrier

By observing the dashed line ($x = 12$) across all panels, we see a clear structural divide in Thailand:

- ➔ **The "High-Education" Regime:** Only **Bangkok-Pattaya** operates consistently to the right of the 12-year mark. Its intercept (α_j) is the highest, but the slope is comparable to other regions.
- ➔ **The "Truncated" Regimes:** Northern and Northeastern regions are "cut off" before reaching the 12-year threshold.
- ➔ **The Buffer Zone:** Central Thailand bridges the gap, showing high income variance but still struggling to push the mean education past 12 years.

Observation: Slope Homogeneity



Analysis of Posteriors:

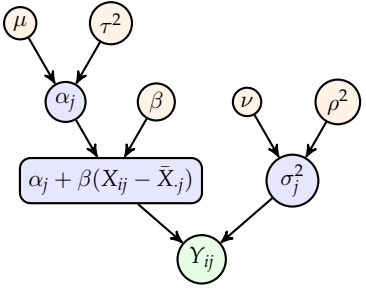
- **Intercepts (α_j):** Clearly distinct. Bangkok is significantly higher than the North.
- **Slopes (β_j):** High degree of overlap in the 95% credible intervals.

Refined Model: Unique National Slope

Hypothesis

The "Return on Education" is a **national constant**. Regional disparity is driven by the baseline (α_j), not the rate of return (β).

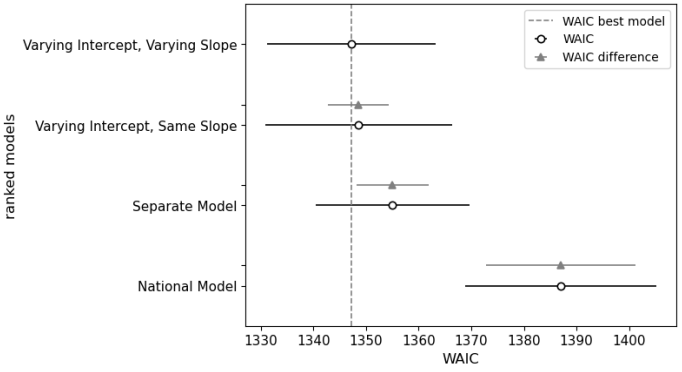
We refine the hierarchy by fixing β as a national parameter while keeping α_j regional:



Key Changes:

- ➔ β is no longer indexed by j.
- ➔ It represents the **National Average Return** on education.
- ➔ This reduces model complexity and prevents overfitting in regions with less data.

Model Comparison: Ranking



Ranking (Best to Worst):

1. **Varying Intercept, Same Slope** (○)
2. **Varying Intercept, Varying Slope**
3. **Separate Models**
4. **National Model**

Key Metric: The *difference* (△) shows the distance in standard error from the top-ranked model.

Model Selection: Why WAIC?

To choose the best architecture, we use the **Widely Applicable Information Criterion (WAIC)**.

- ➔ **Bayesian Nature:** Unlike AIC/BIC, WAIC uses the entire posterior distribution, making it ideal for hierarchical models.
- ➔ **Predictive Accuracy:** It estimates the *out-of-sample* predictive density.
- ➔ **The Penalty:** It penalizes model complexity (effective number of parameters p_{WAIC}) to prevent overfitting.

Interpretation

$$\text{WAIC} = -2 \times (\text{lppd} - p_{\text{WAIC}})$$

Lower values indicate a better balance between fit and parsimony.

The Winning Model: Parsimony Wins

The **Varying Intercept, Same Slope** model is the superior choice because:

- **Statistical Support:** It has the lowest WAIC, indicating that adding regional slopes (β_j) increases complexity without significantly improving predictive power.
- **Structural Insight:** It confirms that the "Return on Education" is a **National Constant**.
- **Policy Implication:** Regional disparity is not caused by education being "less effective" in poor areas, but by lower baseline wealth and educational ceilings.

Final Decision: Adopt the National Slope β for all 6 regions.

Results: The National Return on Education

From our winning model (Varying Intercept, Same Slope), we derive a single national estimate for the impact of schooling on income:

The Value of one Year of Education (β)

1,545 THB / Month

95% Credible Interval: (982, 2,059)

This rate is **consistent across all regions**, regardless of whether the province is in the wealthy Bangkok core or the rural North.

Results: Regional Income Baselines

While the *rate* of return is the same, the **starting baseline income** (α_j) varies significantly across the 6 regions:

Region	Monthly Income Mean (95% CI)
Central Thailand	**21,852** (20,804, 22,892)
Bangkok-Pattaya	**21,779** (19,828, 23,567)
Northcentral Thailand	18,310 (17,638, 18,982)
Southern Thailand	18,194 (17,242, 19,167)
Northeastern Thailand	18,084 (17,441, 18,754)
Northern Thailand	**16,943** (16,157, 17,722)

➔ **The Gap:** There is a $\approx$ **5,000 THB** structural difference between the wealthiest (Central/BKK) and the poorest (Northern) regions that *cannot* be explained by the education slope alone.

Future Work: Moving to District-Level Granularity

Our current model operates at the **Province level** ($N = 76$). However, socio-economic disparities often exist *within* provinces.

- ➔ **The Goal:** Extend the analysis to the **District level** to capture local pockets of poverty.
- ➔ **The Challenge:** Adding a third level to a Bayesian Hierarchical model

$$(\text{National} \rightarrow \text{Region} \rightarrow \text{Province} \rightarrow \text{District})$$

significantly increases:

1. **Computational Complexity:** High-dimensional MCMC convergence issues.
2. **Data Sparsity:** Many districts may have insufficient census samples.

*Solution: Transitioning toward **Geographically Weighted Regression (GWR)**.*

Geographically Weighted Regression (GWR)

Instead of discrete regional boundaries, GWR allows parameters to vary as a continuous function of coordinates $x = (u, v)$:

$$Y = \beta_0(x) + \sum_{j=1}^{p} \beta_j(x) X^{(j)} + \varepsilon$$

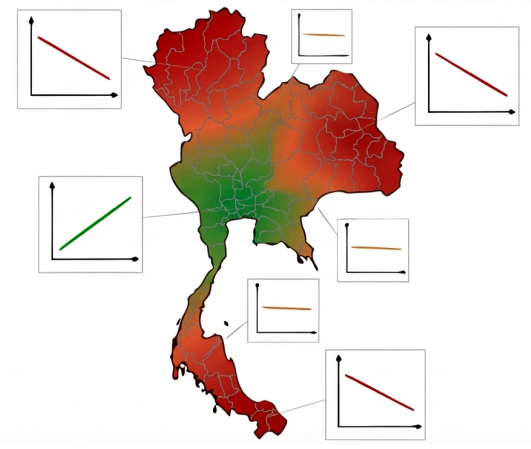
The Local Weighting Mechanism: We estimate $\hat{\beta}(x)$ using a weighted least squares approach:

$$\hat{\beta}(x) = (\mathbf{X}^T \mathbf{W}_x \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_x \mathbf{Y}$$

Where $\mathbf{W}_x$ is a diagonal matrix of weights based on proximity (e.g., k-Nearest Neighbors):

$$w_{x,j} = \begin{cases} 1/k & \text{if province } j \in \text{knn}(x) \\ 0 & \text{otherwise} \end{cases}$$

GWR: Visualizing Spatial Non-Stationarity



Key Concept:

- The relationship between education and income is **not stationary**.
- **Continuous Surface:** GWR transitions from ”Regional Blocks” to a smooth geographic gradient.

Feasibility: Testing for Multicollinearity

A common failure point for GWR is **spatial multicollinearity**, which can lead to unstable coefficient estimates and variance inflation.

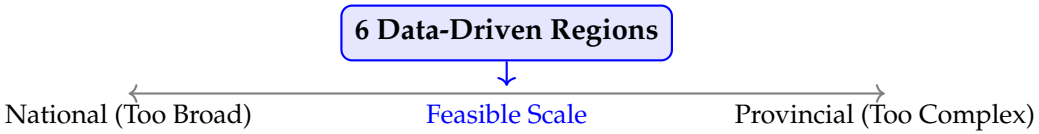
Our VIF Analysis: Before proceeding, we calculated the **Variance Inflation Factor (VIF)** for the 9 socio-economic variables:

- ➔ **Result:** All VIF values were **below 5**.
- ➔ **Significance:** This indicates no significant multicollinearity between our predictors.

Conclusion for Future Research

The lack of collinearity suggests that GWR will be numerically stable. This allows us to map the "Return on Education" as a **continuous surface** across Thailand, identifying exactly where the education-income link is strongest.

Conclusions: The 6-Region Strategic Balance



Spatial Spillovers

Positive **Moran's I** across all 9 variables confirms poverty is *clustered*.

Statistical Coherence

6 regions are geographically continuous and statistically consistent.

Targeted Regional Strategies

North & Northeast

Barrier: Low Ed / High Debt
Action: Education expansion & Debt management.

Southern Thailand

Barrier: Behavioral/Savings
Action: Financial literacy & Health programs.

Central Thailand

Barrier: High Inequality
Action: Skill upgrading & Wage redistribution.

Bangkok-Pattaya

Status: $>$ 12 yrs Education
Action: National benchmark for structural success.

*Policy effectiveness depends on the **spatial configuration** of barriers.*

Socioeconomic Regionalization and Bayesian Hierarchical Modeling of Thailand