A journey through Bayesian hierarchical models: Analyzing income and education in Thailand

Dr. Irving Gómez Méndez

April 2024

This work was done in collaboration with Dr. Chainarong Amornbunchornvej, NECTEC

Hierarchical models with one cluster: income per region

What is a Bayesian model?

Input:

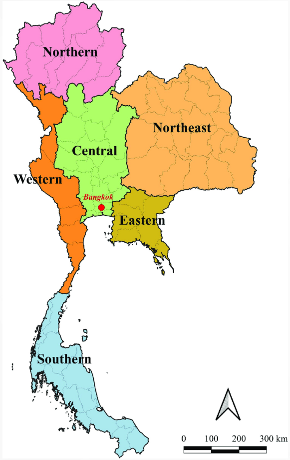
We model a random variable using the **likelihood** $Y \sim p(Y|\theta)$.

We model the uncertainty on θ using a **prior distribution** $p(\theta)$.

Output:

We update our uncertainty on θ through Bayes rule, getting the **posterior distribution** $p(\theta|\mathbf{Y})$.

We capture the total uncertainty on Y using the **posterior predictive distribution** $p(Y|\mathbf{Y})$.

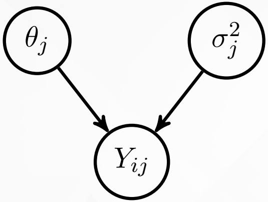


There are 76 provinces grouped in 6 different regions.

Let be Y_{ij} the income in province i belonging to region j.

We assume that $Y_{ij}|\theta_j, \sigma_j^2 \sim \mathcal{N}(\theta_j, \sigma_j^2)$ **(Likelihood)**.

No pooling model

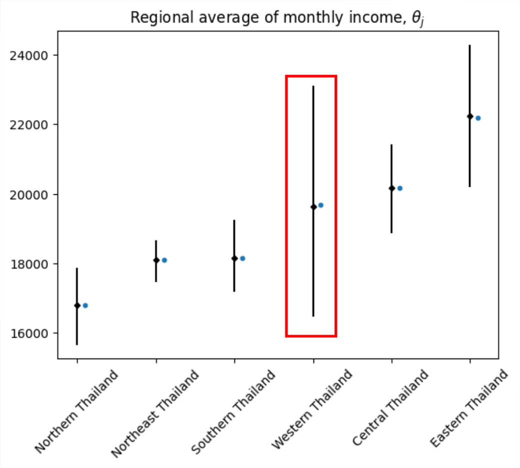


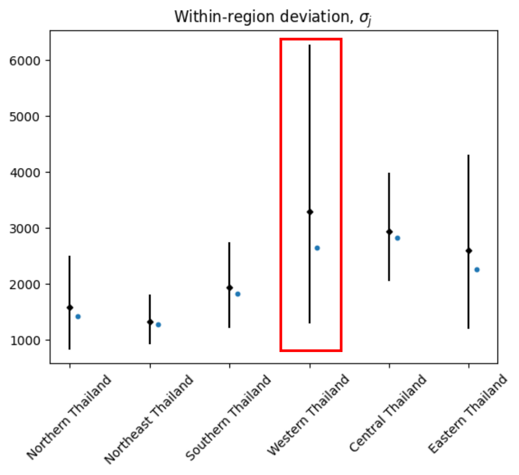
Each region has an independent model.

Prior distribution

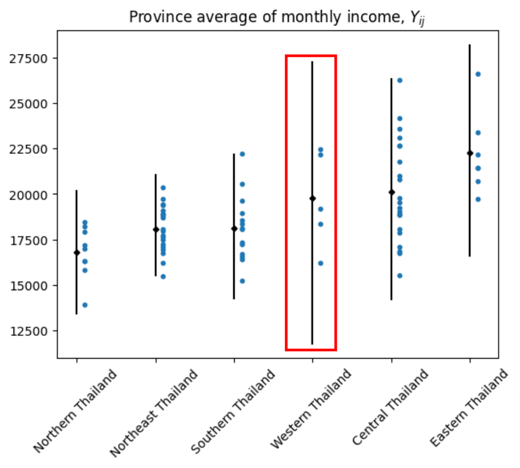
$$p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) \propto \prod_{j=1}^{J} \frac{1}{\sigma_j^2} \mathbb{1}_{\mathbb{R}}(\theta_j) \mathbb{1}_{(0,\infty)}(\sigma_j^2)$$

Posterior distributions





Posterior predictive distribution



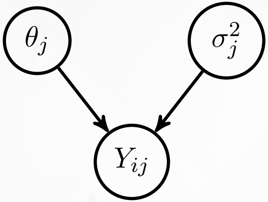
Large intervals, because regions don't share information.

Intervals overlap for most of the regions.

All the parameters might be estimating the same quantity.

It is highly unlikely that the regions are independent between them.

Complete pooling model



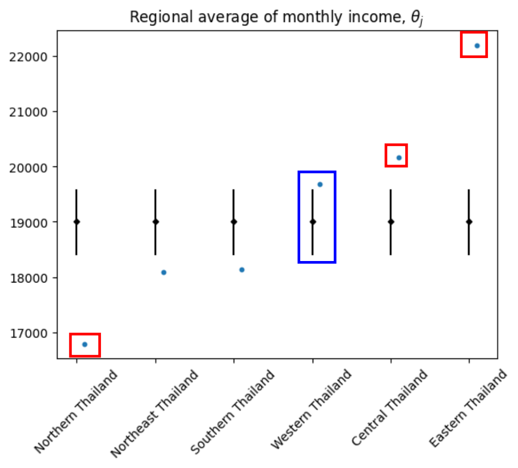
Same model for all the regions.

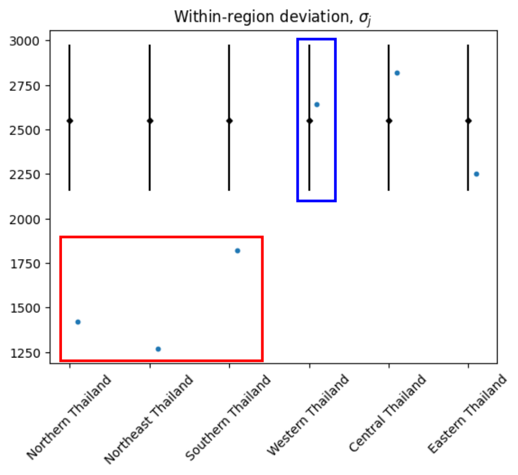
$\theta_j = \theta$, $\sigma_j = \sigma$, for all j.

Prior distribution

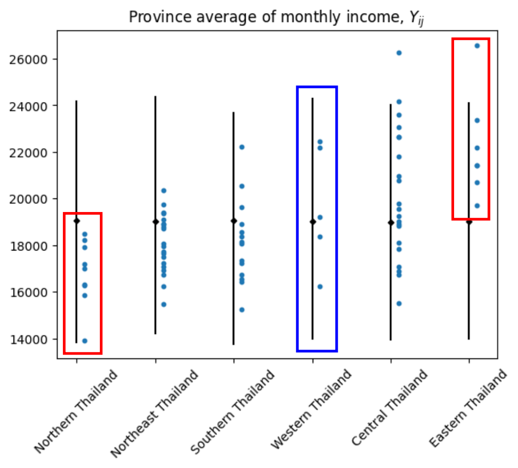
$$p(\theta, \sigma^2) \propto \frac{1}{\sigma^2} \mathbb{1}_{\mathbb{R}}(\theta) \mathbb{1}_{(0,\infty)}(\sigma^2)$$

Posterior distributions





Posterior predictive distribution



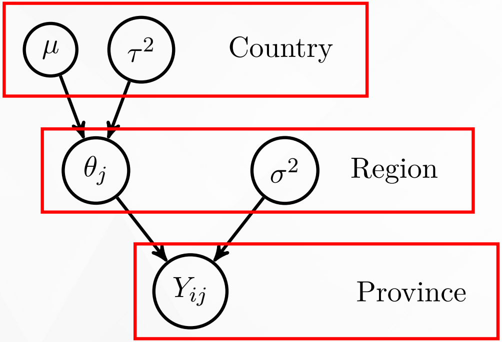
Narrow intervals.

Common mean θ can barely explain the mean of a few regions.

Overestimated σ.

Income in Northern Thailand is overestimated. Income in Eastern Thailand is underestimated.

Hierarchical model with common within-cluster variance



Likelihood

$$Y_{ij}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \sigma^2)$$

Prior distributions

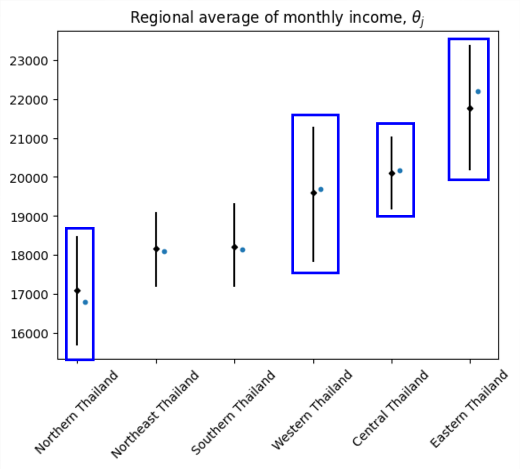
$$\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2)$$

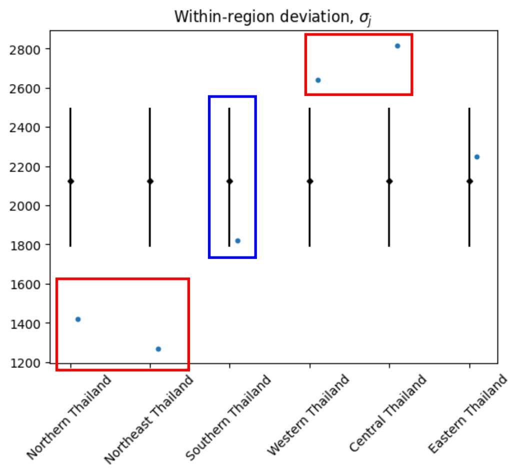
$$p(\mu) \propto \mathbb{1}_{\mathbb{R}}(\mu)$$

$$p(\tau^2) \propto \mathbb{1}_{(0,\infty)}(\tau^2)$$

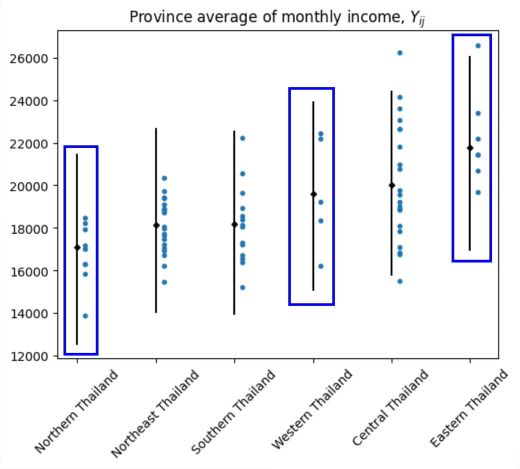
$$p(\sigma^2) \propto \frac{1}{\sigma^2} \mathbb{1}_{(0,\infty)}(\sigma^2)$$

Posterior distributions





Posterior predictive distribution



Less uncertainty than independent models.

Better performance than complete pooling model.

The common variance is not overestimated.

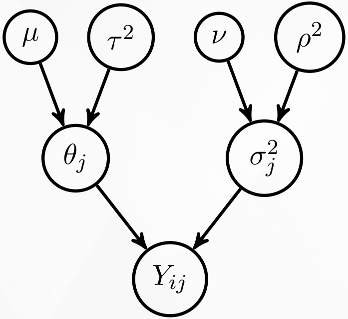
A common variance cannot explain the observed variability.

Drawbacks of Bayesian models:

We must ensure the existence of the posterior distribution. This requires that **we analyze** the mathematical properties of the model.

The model is not going to do it for us! That's the hatch!

Hierarchical model varying within-cluster variance



$$Y_{ij}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \sigma^2)$$

$$\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2)$$

$$p(\mu) \propto \mathbb{1}_{\mathbb{R}}(\mu)$$

$$p(\tau^2) \propto \mathbb{1}_{(0,\infty)}(\tau^2)$$

$$\sigma_j^2|\nu, \rho^2 \sim \textsf{Inverse-}\chi^2(\nu, \rho^2)$$

$$p(\rho^2) \propto \frac{1}{\rho^2}\mathbb{1}_{(0,\infty)}(\rho^2)$$

$$\nu \sim ?$$

It remains challenging to propose priors for the degrees of freedom of a distribution.

We consider three approaches:

- Set ν equal to some estimated value $\hat{\nu}$. For example, using the method of moments
$$\hat{\nu} = \frac{2(E_{s^2})^2}{V_{s^2}} + 4.$$
- Use a vague prior, for example $p(\nu) \propto \nu^{-h}\mathbb{1}_{(0,\infty)}(\nu)$, for some $h \in [0, \infty)$.
- Use a regularizing prior, for example $\nu \sim \mathsf{Exponential}(1/\hat{\nu})$.

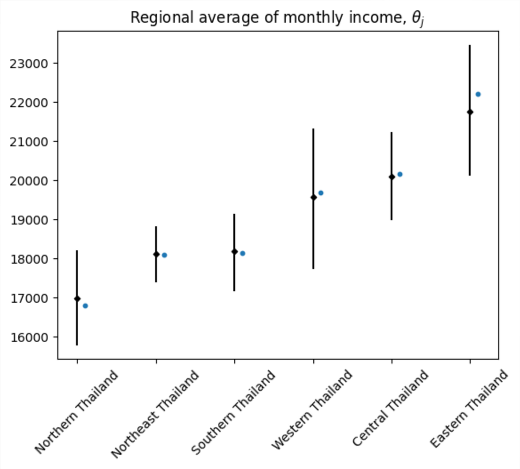
Cons:

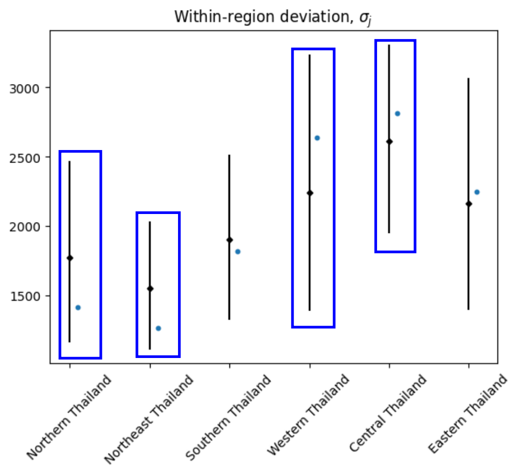
- Set $\nu = \hat{\nu}$: Eliminates the uncertainty on ν.
- Improper distribution: Does not give guarantee for the existence of the posterior distribution.
- Regularizing prior: Which distribution to use?

Pros:

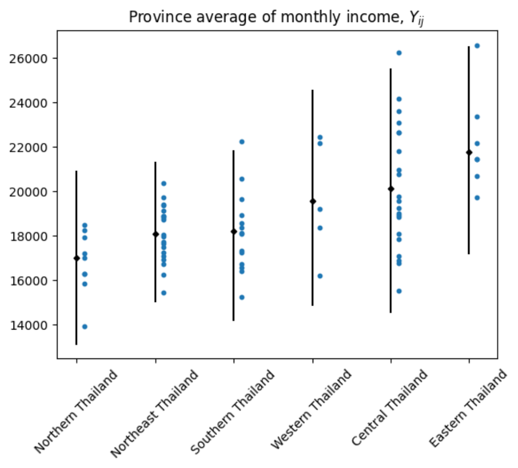
- Set $\nu = \hat{\nu}$: We don't have to worry about the posterior.
- Improper distribution: There's nothing to estimate.
- Regularizing prior: Maintain the uncertainty and guarantee the existence of the posterior distribution.

Posterior distributions

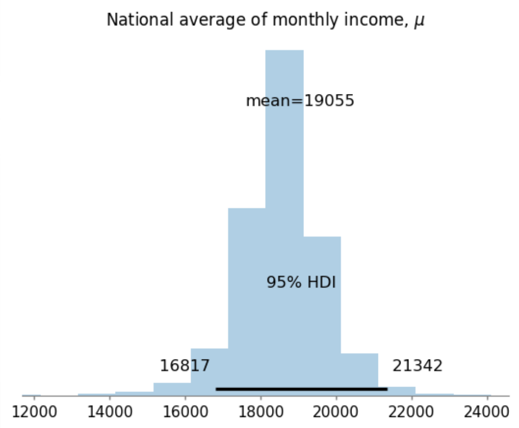


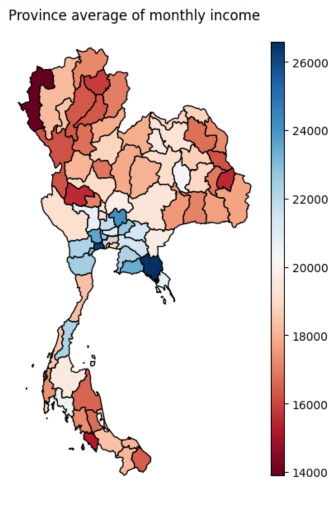


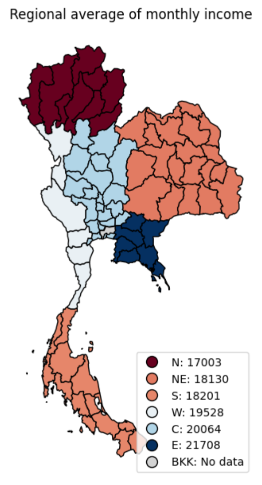
Posterior predictive distribution

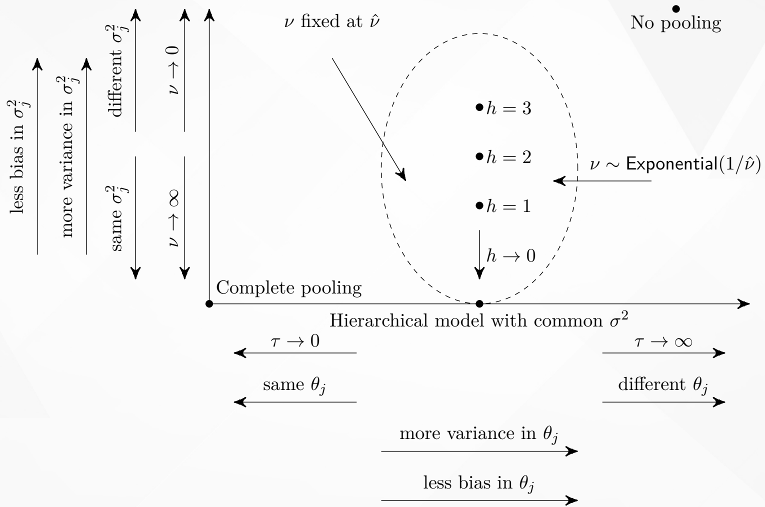


Posterior distribution




Province average of monthly income
26000
24000
22000
20000
18000
16000
14000


Regional average of monthly income
N: 17003
NE: 18130
S: 18201
W: 19528
C: 20064
E: 21708
BKK: No data

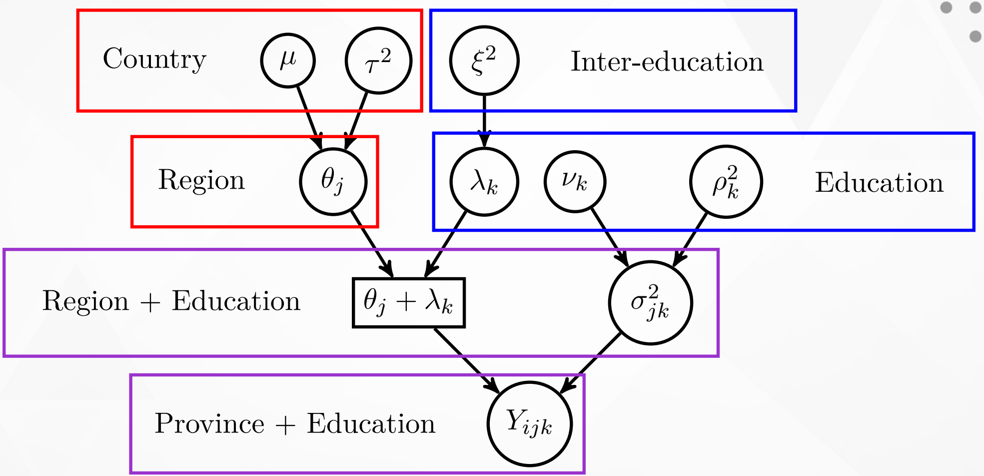

less bias in σ_j^2
more variance in σ_j^2
different σ_j^2
same σ_j^2
$\nu \to 0$
$\nu \to \infty$
ν fixed at $\hat{\nu}$
No pooling
$h = 3$
$h = 2$
$h = 1$
$h \to 0$
$\nu \sim \text{Exponential}(1/\hat{\nu})$
Complete pooling
Hierarchical model with common σ^2
$\tau \to 0$
$\tau \to \infty$
same θ_j
different θ_j
more variance in θ_j
less bias in θ_j

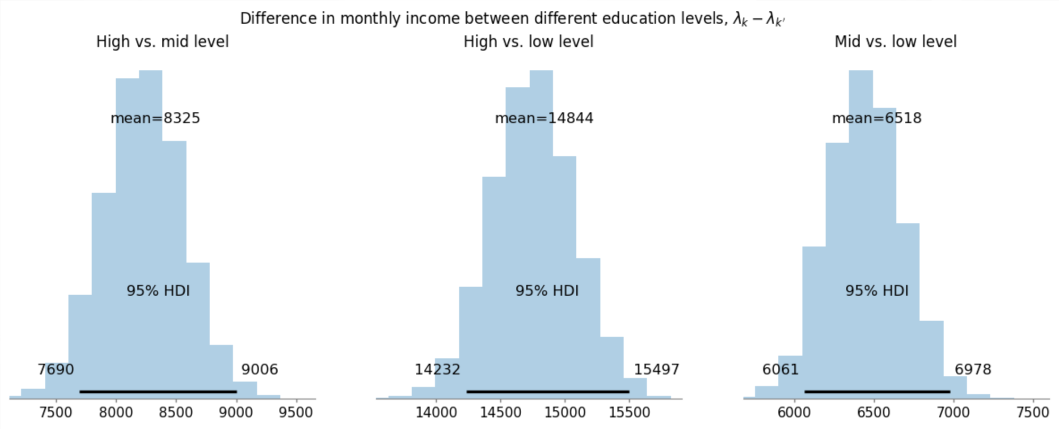
Hierarchical model with two non-nested clusters: income per region and education level

We define the next rule:

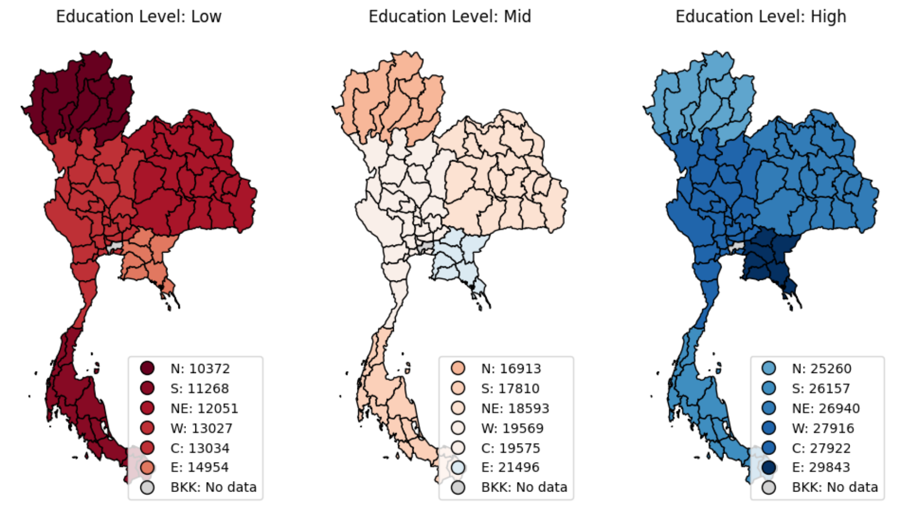
Education	Years of education	Education level
Uneducated	0	Low
Kindergarten	0	
Pre-elementary school	3	
Elementary school	6	
Junior high school	9	Mid
Senior high school	12	
Vocational degree	14	
Bachelor degree	16	High
Post-graduate	19	

$$
\begin{aligned}
Y_{ijk} | \theta_j, \lambda_k, \sigma^2_{jk} &\sim \mathcal{N}(\theta_j + \lambda_k, \sigma^2_{jk}) \\
\theta_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2) \\
\mu &\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2_{\mu}) \\
\tau^2 &\sim \textsf{Exponential}(1/\hat{\tau}^2) \\
\lambda_k | \xi^2 &\sim \mathcal{N}(0, \xi^2) \\
\xi^2 &\sim \textsf{Exponential}(1/\hat{\xi}^2) \\
\sigma^2_{jk} | \nu_k, \rho^2_k &\sim \textsf{Inverse-}\chi^2(\nu_k, \rho^2_k) \\
\nu^2_k &\sim \textsf{Exponential}(1/\hat{\nu}^2_k) \\
p(\rho^2_k) &\propto \frac{1}{\rho^2_k} \mathbb{1}_{(0,\infty)}(\rho^2_k)
\end{aligned}
$$


Country
μ
τ^2
ξ^2
Inter-education
Region
θ_j
λ_k
ν_k
ρ_k^2
Education
Region + Education
$\theta_j + \lambda_k$
σ_{jk}^2
Province + Education
Y_{ijk}



Difference in monthly income between different education levels, $\lambda_k - \lambda_{k'}$
High vs. mid level
mean=8325
95% HDI
7690
9006
7500 8000 8500 9000 9500
High vs. low level
mean=14844
95% HDI
14232
15497
14000 14500 15000 15500
Mid vs. low level
mean=6518
95% HDI
6061
6978
6000 6500 7000 7500


Education Level: Low
N: 10372
S: 11268
NE: 12051
W: 13027
C: 13034
E: 14954
BKK: No data
Education Level: Mid
N: 16913
S: 17810
NE: 18593
W: 19569
C: 19575
E: 21496
BKK: No data
Education Level: High
N: 25260
S: 26157
NE: 26940
W: 27916
C: 27922
E: 29843
BKK: No data

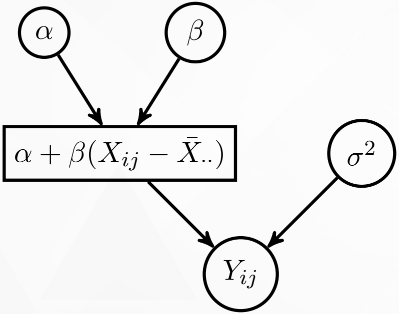
Bayesian hierarchical regression: income considering years of formal education

We define the next rule:

Education	Years of education	Education level
Uneducated	0	Low
Kindergarten	0	
Pre-elementary school	3	
Elementary school	6	
Junior high school	9	Mid
Senior high school	12	
Vocational degree	14	
Bachelor degree	16	High
Post-graduate	19	

National model

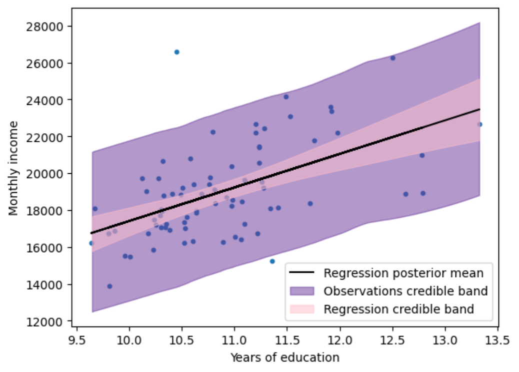
Let be X_{ij} the average years of formal education in the province i belonging to region j.



$$Y_{ij} | \alpha, \beta, \sigma^2 \sim \mathcal{N}(\alpha + \beta(X_{ij} - \bar{X}_{..}), \sigma^2)$$

$$p(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2} \mathbb{1}_{\mathbb{R}}(\alpha) \mathbb{1}_{\mathbb{R}}(\beta) \mathbb{1}_{(0,\infty)}(\sigma^2)$$


National average of monthly income, α
mean=18998
95% HDI
18532
19506
18250 18500 18750 19000 19250 19500 19750
National ratio of income per year-of-education, β
mean=1821
95% HDI
1173
2459
750 1000 1250 1500 1750 2000 2250 2500 2750

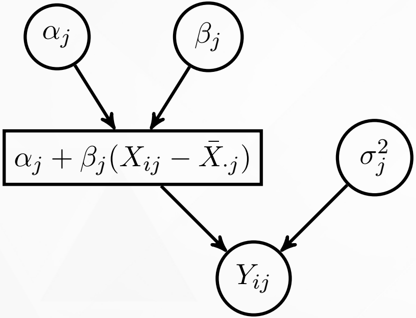


Suitable to explain national behavior.

We cannot say anything at a regional level.

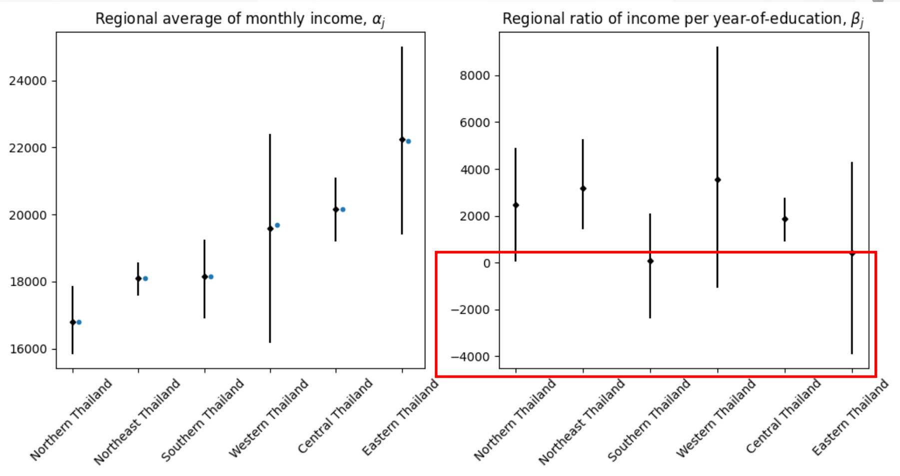
Separate models

Let be X_{ij} the average years of formal education in the province i belonging to region j.



$$Y_{ij}|\alpha, \beta, \sigma^2 \sim \mathcal{N}(\alpha_j + \beta_j(X_{ij} - \bar{X}_{\cdot j}), \sigma^2)$$

$$p(\alpha_j, \beta_j, \sigma_j^2) \propto \frac{1}{\sigma_j^2} \mathbb{1}_{\mathbb{R}}(\alpha_j) \mathbb{1}_{\mathbb{R}}(\beta_j) \mathbb{1}_{(0,\infty)}(\sigma_j^2)$$

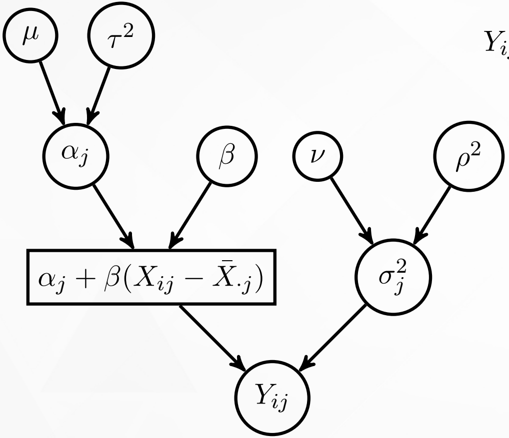

Regional average of monthly income, α_j
24000
22000
20000
18000
16000
Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand
Regional ratio of income per year-of-education, β_j
8000
6000
4000
2000
0
−2000
−4000
Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand

Large credible intervals.

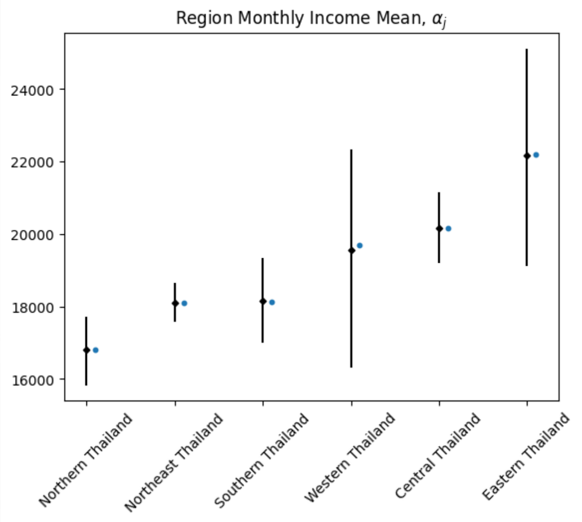
The intervals include negative values for β_j, which seems implausible.

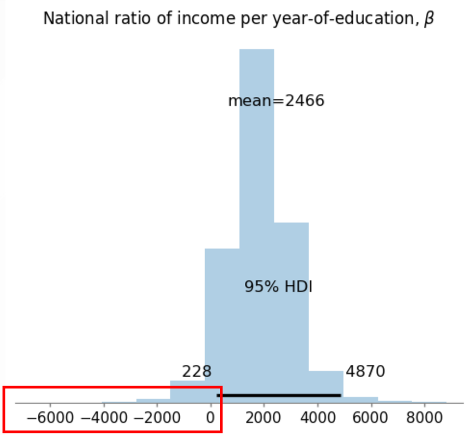
Pretending that each region is independent for the others seems unrealistic.

Bayesian hierarchical regression varying intercepts



$$
\begin{aligned}
Y_{ij} | \alpha_j, \beta_j, \sigma_j^2 &\sim \mathcal{N}(\alpha_j + \beta_j(X_{ij} - \bar{X}_{\cdot j}), \sigma_j^2) \\
\alpha_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2) \\
\mu &\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2) \\
\tau^2 &\sim \text{Exponential}(1/\hat{\tau}^2) \\
p(\beta) &\propto \mathbb{1}_{\mathbb{R}}(\beta) \\
\sigma_j^2 | \nu, \rho^2 &\sim \text{Inverse-}\chi^2(\nu, \rho^2) \\
\nu^2 &\sim \text{Exponential}(1/\hat{\nu}^2) \\
p(\rho^2) &\propto \frac{1}{\rho^2} \mathbb{1}_{(0,\infty)}(\rho^2)
\end{aligned}
$$

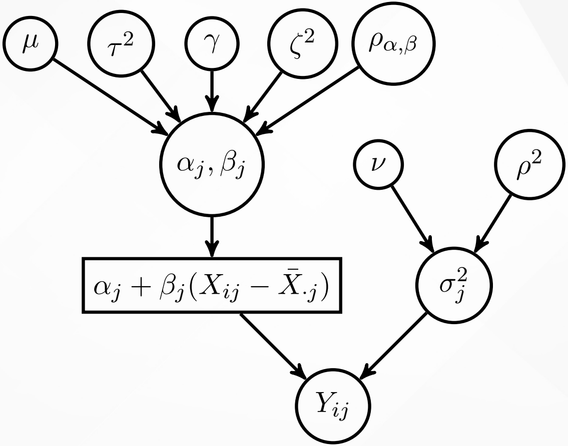

Region Monthly Income Mean, α_j
24000
22000
20000
18000
16000
Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand


National ratio of income per year-of-education, β
mean=2466
95% HDI
228
4870
−6000 −4000 −2000 0 2000 4000 6000 8000

The distribution of β includes negative values.

Different slopes could be preferred.

Bayesian hierarchical regression varying intercepts and slopes



α_j and β_j will follow a multivariate normal distribution with mean (μ, γ) and a matrix of variances and covariances

$$S = \begin{pmatrix} \tau^2 & \tau\zeta\rho_{\alpha,\beta} \\ \tau\zeta\rho_{\alpha,\beta} & \zeta^2 \end{pmatrix},$$

which can be written as

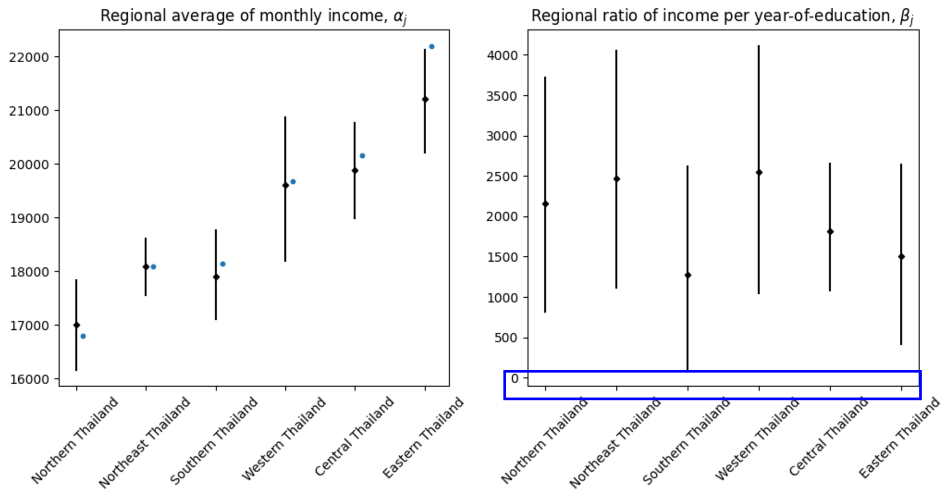
$$S = \begin{pmatrix} \tau & 0 \\ 0 & \zeta \end{pmatrix} R \begin{pmatrix} \tau & 0 \\ 0 & \zeta \end{pmatrix},$$

where

$$R = \begin{pmatrix} 1 & \rho_{\alpha,\beta} \\ \rho_{\alpha,\beta} & 1 \end{pmatrix}$$

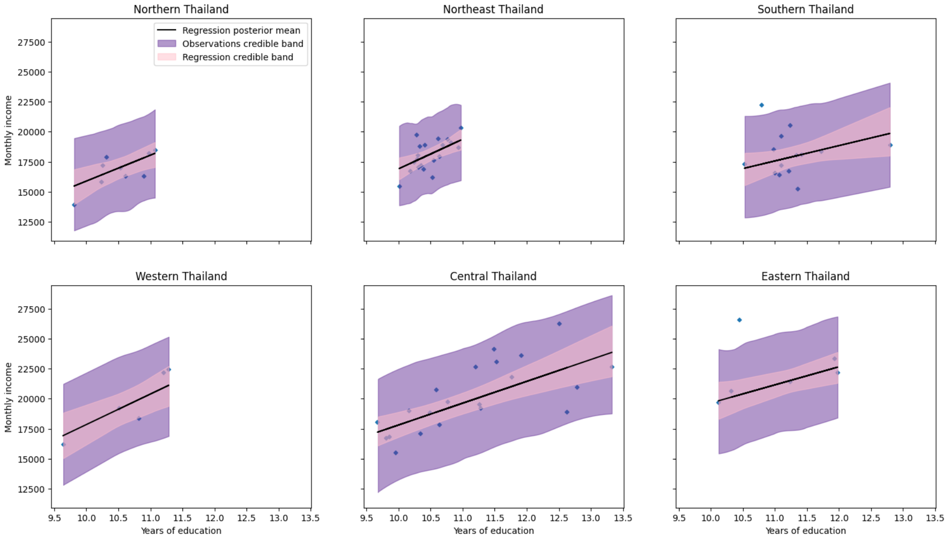
is the correlation matrix.

$$
\begin{aligned}
Y_{ij} | \alpha_j, \beta_j, \sigma_j &\sim \mathsf{Laplace}(\alpha_j + \beta_j (X_{ij} - \bar{X}_{\cdot j}), \sigma_j) \\
\alpha_j, \beta_j | \mu, \tau^2, \gamma, \zeta^2, \rho_{\alpha,\beta} &\sim MVN\left(\begin{bmatrix}\mu \\ \gamma\end{bmatrix}, S\right) \\
\mu &\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2) \\
\tau^2 &\sim \mathsf{Exponential}(1/\hat{\tau}^2) \\
\gamma &\sim \mathcal{N}(\hat{\gamma}, \hat{\sigma}_\gamma^2) \\
\zeta^2 &\sim \mathsf{Exponential}(1/\hat{\zeta}^2) \\
R &\sim \mathsf{LKJ}(2) \\
\sigma_j^2 | \nu, \rho^2 &\sim \mathsf{Inverse}\text{-}\chi^2(\nu, \rho^2) \\
\nu^2 &\sim \mathsf{Exponential}(1/\hat{\nu}^2) \\
p(\rho^2) &\propto \frac{1}{\rho^2} \mathbb{1}_{(0,\infty)}(\rho^2)
\end{aligned}
$$



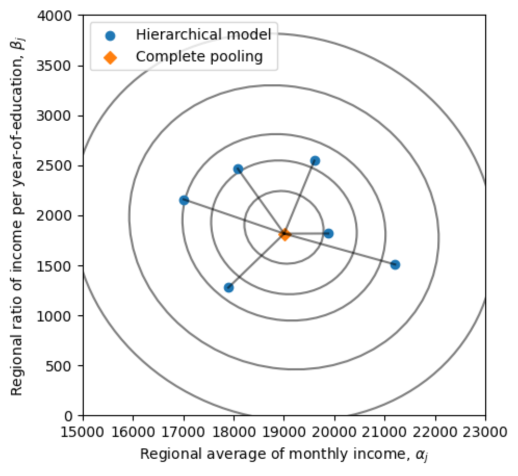
Regional average of monthly income, α_j
22000
21000
20000
19000
18000
17000
16000
Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand
Regional ratio of income per year-of-education, β_j
4000
3500
3000
2500
2000
1500
1000
500
0
Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand

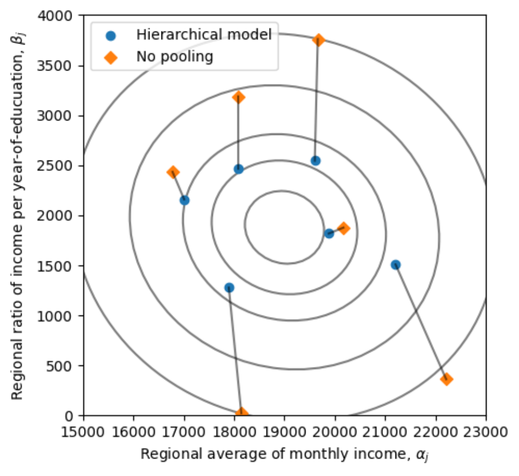

National average of monthly income, μ
mean=19001
95% HDI
18530
19424
18400
18600
18800
19000
19200
19400
19600
National ratio of income per year-of-education, γ
mean=1880
95% HDI
1322
2443
1000
1250
1500
1750
2000
2250
2500
2750


Northern Thailand
Northeast Thailand
Southern Thailand
Western Thailand
Central Thailand
Eastern Thailand
Regression posterior mean
Observations credible band
Regression credible band
Monthly income
Years of education
27500
25000
22500
20000
17500
15000
12500
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5

We can explain the data at a national and regional level.

There are no negative values for β_j nor γ.


Hierarchical model
Complete pooling
Regional ratio of income per year-of-educuation, β_j
Regional average of monthly income, α_j
0
500
1000
1500
2000
2500
3000
3500
4000
15000
16000
17000
18000
19000
20000
21000
22000
23000


Hierarchical model
No pooling
Regional ratio of income per year-of-educuation, β_j
Regional average of monthly income, α_j
0
500
1000
1500
2000
2500
3000
3500
4000
15000
16000
17000
18000
19000
20000
21000
22000
23000

If you feel at all confused, it is only because you are paying attention.