

Estimating the effect of GDP per capita on life expectancy between continents using Bayesian hierarchical model

Nabiel H. Efrem
Kevin Singh

Abstract

Hierarchical linear regression and separated models are standard models to analyze data from a Bayesian setting. Although the traditional separated model is commonly used, recently there have been many modifications to the traditional separated model. In this report, we will estimate the posterior statistics of parameters using both Hierarchical linear regression (HLRM) and one modification of the separated model (MSM). Further, we will compare the models using the Leave-one out cross validation (LOO) to test their predictive performance. It is shown that the posterior estimates are close to identical for both models but simultaneously MSM resulted in a higher predictive power. Real data application is included in the report.

1. Introduction

Recently, there has been an increase in the use of Hierarchical linear regression model (HLRM) and Separated model (SM) in Bayesian settings. One area that is driving the research on HLRM and SM is discriminant analysis and Classification. There is studies for HLRM and SM:s application in a wide range of areas and through these studies, several modification of the traditional HLRM and SM has been presented.

Among these modifications for the separated model, there is Bayesian hierarchical modeling outlined with un-pooled approach which we have chosen to call Modified Separated model (MSM). A model that, like HLRM, uses hyperparameters and hyperpriors. Both models can be categorized as multilevel models and can have enough parameters to fit the data structure without worrying about over-fitting which is often the case for non-multilevel models. Despite these similarities, there are several differences between the models not least when working with discriminant analysis. A difference worth emphasizing, HLRM assumes shared information, expressed in hyperpriors, for parameters between groups while the Separated model (and MSM) does not assume shared information between groups.

In this work we analyze the HLRM and MSM when having group-specific parameters where real data is being applied. Our HLRM is a generalization of the simple linear regression model with a Bayesian theory while MSM is a newer approach. Thus, we provide posterior estimates of the parameters of interest and compare their predictive power.

In section 2 we provide the data considered and for section 3 we present the formulas, notations, and distributions for HLRM and MSM with a Bayesian approach. In section 4 we will fit the mentioned models and compare them using Leave one out cross validation, followed by section 5 which summarizes and discusses the main results

2. Data Collection and Description

Gapminder data is originally collected from a Swedish foundation called Gapminder and the data set can be found in gapminder-package in R. However, we will use the subset of the gap-minder data. The total gap-minder data consists of 1704 observations for selected time points over a time period starting from 1952 and ending in 2007. This study will only use information from the year 2007 which contains information regarding 6 variables, which are either categorical or numerical, from 142 countries. The data description for gap-minder data year 2007 can be found in Table 2 below.

Table 1: Data description

| Variable | Data type |
|-----------|-------------|
| country | Categorical |
| continent | Categorical |
| year | Categorical |
| lifeExp | Numerical |
| pop | Numerical |
| gdpPercap | Numerical |

Throughout the report, we have decided to use the life expectancy variable as dependent and gdpPercap, short for GDP per capita, as covariate in both models. Due to one covariate explaining the variation of life expectancy we get two simple regression models which will be done in Bayesian settings.

The purpose is to evaluate if the effect of GDP per capita

on life expectancy differs between continents. Many studies have been done regarding this area, we have reason to believe that there is a structural difference between continents. Since, we want to estimate the life expectancy for different groups, the continent variable is being set as the group-index which results in 5 different groups for our models.

3. Methodology and Models

In this section, we will show some relevant parts of the simple linear regression model in Bayesian settings. We start by presenting the likelihood function that will be used in the HLRM and the modified MSM. Then we will do model specification by going into depth of the hyperprior, prior, and posterior distribution for both models including their main assumptions.

3.1. Likelihood - Normal Linear Regression model

We now consider having $m = 5$ groups in the model and each of the five groups consist of $n_j, j = 1, \dots, m$ independent normally distributed observations. Note, we get that $n_j = n \forall j$ when working with the balanced case for the data set. Let $y_{i,j}$ be characterized by a linear relationship of covariate $x_{i,j}$ which creates a group-and observation-specific normal simple linear regression model such as:

$$y_{i,j} = \alpha_j + x_{i,j}\beta + \epsilon_{i,j}$$

where $i = 1, 2, \dots, n_j, j = 1, \dots, m, \sum_{j=1}^m n_j = n$. Recall, in our case, we have that the sample size, number of groups, and number of covariates correspond to $n = 142, m = 5$ and $p = 1$ respectively. Further, $y_{i,j}, x_{i,j} : 1 \times 1$ and $\epsilon_{i,j}$ denotes the response variable, the covariate "gdpPercap" and the error term respectively for the i :th observation and the j :th group.

Given the independent and identically distributed (iid) assumption we have that $\epsilon_{i,j} | \sigma^2 \sim N(0, \sigma^2)$ and β_j is the regression coefficient in our model. Since the Normal linear regression model for $y_{i,j}$ is suitable $\forall (i, j)$ it can be rewritten as

$$y_{i,j} | x_{i,j}, \alpha_j, \beta_j, \sigma_j^2 \sim P(y_{i,j} | x_{i,j}, \theta_j) = N(\alpha_j + x_{i,j}\beta_j, \sigma_j^2)$$

where $\theta_j = (\alpha_j, \beta_j, \sigma_j^2)' : 3 \times 1$ is the parameter vector consisting of unknown model parameters. In order for us to compute HLRM and MSM with a Bayesian setting we need to assign the elements in θ_j to a prior distribution.

3.2. Hierarchical linear regression model

Now when the likelihood function has been stated, a natural next step when doing HLRM is to assign elements in $\theta_j : 3 \times 1$ a prior distribution. In addition regarding HLRM, similar is done to hyperparameters which is thus assigned

hyperpriors. Below, we propose the Hierarchical prior distributions that will be used in this report:

$$\beta_j | \mu_\beta, \sigma_\beta^2 \stackrel{\text{iid}}{\sim} N(\mu_\beta, \sigma_\beta^2) \text{ and } \alpha_j | \mu_\alpha, \sigma_\alpha^2 \stackrel{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$$

with

$$\mu_\beta \sim N(\mu_{0\beta}, \sigma_{0\beta}^2), \mu_\alpha \sim N(\mu_{0\alpha}, \sigma_{0\alpha}^2)$$

The model is structured in such a way that $\mu_{0\beta}, \sigma_{0\beta}^2$ determines the data generating process for β_j and similar for intercept α_j . We set $\mu_{0\beta} = 0.05, \sigma_{0\beta}^2 = 100, \mu_{0\alpha} = 1$ and $\sigma_{0\alpha}^2 = 50$ which is consistent with weakly informative hyperpriors

Recall that $\sigma_j^2 = \text{VAR}(y_{i,j} | x_{i,j}, \alpha_j, \beta_j, \sigma_j^2)$ given in section 3.1 and σ_j^2 is not being assigned a prior distribution above. Since we are using Stan for computation, σ_j^2 will on default be assigned a non-informative prior distribution. Thus, σ_j^2 includes minimum information about the groups resulting in σ_j^2 from now on, i.e $\sigma_j^2 = \sigma^2$. This can be denoted as $\sigma^2 \sim \pi(\cdot)$ where $\pi(\cdot)$ is a non-informative prior distribution.

3.3. Modified Separated Model

In recent times, many modifications of the traditional separated model has been presented and in this report we have decided to work with a Bayesian hierarchical modeling outlined with a un-pooled approach (see eg. (1)), referred to as MSM in this report. In traditional approach of separated model, we consider $y_{i,j}, j = 1, 2, \dots, m = 5$ being determined by parameter(s) and in Bayesian settings these parameter(s) is being assigned prior distribution(s). However for MSM, the hyperparameters is being assigned hyperpriors which can be summarized as:

$$\beta_j | \mu_{\beta_j}, \sigma_{\beta_j}^2 \stackrel{\text{iid}}{\sim} N(\mu_{\beta_j}, \sigma_{\beta_j}^2) \text{ and } \alpha_j | \mu_{\alpha_j}, \sigma_{\alpha_j}^2 \stackrel{\text{iid}}{\sim} N(\mu_{\alpha_j}, \sigma_{\alpha_j}^2)$$

with

$$\mu_{\beta_j} \sim N(\mu_{0\beta}, \sigma_{0\beta}^2), \mu_{\alpha_j} \sim N(\mu_{0\alpha}, \sigma_{0\alpha}^2)$$

where we set $\mu_{0\beta} = 1, \sigma_{0\beta} = 100, \mu_{0\alpha} = 1$ and $\sigma_{0\alpha} = 50$ in MSM. The difference between HLRM and MSM is that we are using the group-index μ_{β_j} and μ_{α_j} for MSM while there is no group-specification for HLRM.

3.4. Posterior Inference and Comparison method

The posterior inference for the models parameters $\theta_j = (\alpha_j, \beta_j, \sigma_j^2)' : 3 \times 1$ can be done using the NUTS algorithm. Since the main purpose of this study is to compare the posterior statistics between MSM and HLRM we need to state the posterior. Posterior distribution from Bayesian approach can be calculated as $P(\theta | y_j, x_j) \propto L(y_j | \theta_j, x_j) \cdot P(\theta), \forall j = 1, \dots, 5$.

In section 4 we will use Leave-one-out cross-validation (LOO) as the method for comparing the two models. LOO is a method that estimated the (out of sample) point-wise predictive power from a fitted Bayesian model using its log-likelihood function $\ell(\theta_j | y_j, x_j)$.

4. Results

To start this section we implement MSM and HLRM for which posterior distribution being sampled by the NUTS algorithm. For each model fitting, we set the numbers of iterations(iters), number of chains(chains) and warm-up period to iters = 8 000, chains = 4 and the first half of the iterations corresponding to the warm-up period.

The first subsection will be based on the Bayesian modeling fitting for posterior distribution. Since we want to compare the parameter estimated between the models we will compare the parameter estimates using LOO in the second section.

4.1. Model Fit

We fit the HLRM given in section 3.2 to the gapminder data for year 2007 and in Table 4.1 posterior summary statistics for the models parameters is being presented.

Table 2: Posterior statistics for HLRM

| Parameter | Mean | SD | Q2.5% | Q97.5% |
|------------|---------|---------|---------|----------|
| β_1 | 0.0004 | 0.0001 | 0.0002 | 0.0005 |
| β_2 | 0.0002 | 0.0001 | 0.0000 | 0.0004 |
| β_3 | 0.0010 | 0.0003 | 0.0005 | 0.0015 |
| β_4 | 0.0003 | 0.0001 | -0.0000 | 0.0005 |
| β_5 | 0.0002 | 0.0010 | -0.0017 | 0.0021 |
| α_1 | 65.9074 | 1.5485 | 62.9094 | 68.9118 |
| α_2 | 72.2118 | 2.8273 | 66.7507 | 77.9327 |
| α_3 | 51.6774 | 1.2039 | 49.3407 | 54.0886 |
| α_4 | 70.6135 | 2.0118 | 66.5703 | 74.6397 |
| α_5 | 74.0749 | 29.0309 | 17.4133 | 132.4849 |
| σ^2 | 6.5003 | 0.4140 | 5.7476 | 7.3597 |

Recall that the indexing for the parameter elements in Table 4.1 corresponds to the group-indexing, e.g. $\alpha_{j=1}$ is the intercept for the first group. When looking at the posterior summary statistics for HLRM we can see that the sample mean $E(\beta_j | \alpha_j, y_j, x_j, \sigma^2)$ of the slope coefficient $\beta_j \approx 0$ for all groups in the model. Also, the 95 % posterior interval for β_j is covering 0, $\forall j = 1, \dots, 5$.

On the other hand, this is not the case for the posterior estimate regarding the intercept α_j where we have that mean of α : 5×1 corresponds to $\hat{\alpha} = (\hat{\alpha}_1 = 65.9, \hat{\alpha}_2 = 72.3, \hat{\alpha}_3 = 51.64, \hat{\alpha}_4 = 70.6, \hat{\alpha}_5 = 74.4)$. Lastly, parameter σ^2 , that was given a non-informative prior, has a mean posterior estimate corresponding to 6.51.

Now we go further and fit the MSM given in section 3.3 thus Table 4.1 consist of the posterior statistics for element on gapminder data when computing MSM

Table 3: Posterior statistics for MSM

| Parameter | Mean | SD | Q2.5% | Q97.5% |
|------------|---------|---------|---------|----------|
| β_1 | 0.0004 | 0.0001 | 0.0002 | 0.0005 |
| β_2 | 0.0002 | 0.0001 | 0.0000 | 0.0004 |
| β_4 | 0.0003 | 0.0001 | -0.0000 | 0.0005 |
| β_3 | 0.0010 | 0.0003 | 0.0005 | 0.0015 |
| β_5 | 0.0003 | 0.0010 | -0.0016 | 0.0022 |
| α_1 | 65.8782 | 1.5292 | 62.8814 | 69.0013 |
| α_2 | 72.2143 | 2.7787 | 66.6917 | 77.8262 |
| α_3 | 51.6405 | 1.1840 | 49.3451 | 53.9570 |
| α_4 | 70.5648 | 2.0038 | 66.6260 | 74.4630 |
| α_5 | 72.2338 | 29.0169 | 16.2642 | 128.6245 |
| σ^2 | 6.5035 | 0.3935 | 5.7820 | 7.3294 |

The tables above indicates that the posterior statistics for the model parameters θ is approximately the same when computing HLRM and MSM separately on gapminder data. Hence, we have $\beta_j \approx 0$, $\forall j = 1, 2, \dots, 5$ when computing MSM. Furthermore, same analysis can be done for the intercept α_j and σ^2 for both Models.

4.2. Model Comparison

In order for us to be able to compare the models we will use LOO which estimates the out of sample point wise prediction performance for a given model. LOO will be computed on HLRM and MSM separately and the LOO estimates can be found in the table below:

| | Hierarchical | Separated |
|------------|--------------|-----------|
| elpd | -1322.8 | -472.7 |
| looi | 2645.6 | 945.3 |
| n k-values | 16 | 2 |

When evaluating the two models we are interested in comparing three different things. One is the elpd, another is looi and lastly the number of observations which have a pareto K value above 0.7.

The elpd for the hierarchical model is -1322.8, the elpd for the modified separated model is -472.7. When comparing the elpd value for the two models a higher value is an indication that the model is better. Thus the mixed separated model performs better in this remark.

The looi for the hierarchical model is 2645.6, for the mixed separated model the estimate is 945.3. When comparing the looi a lower value is desired. Thus it is once again concluded that the mixed separated model performs better than the hierarchical model.

Lastly it is compared whether the estimated elpd for the different models are reliable. This is analysed by inspecting

the pareto k-values for the different models. One k-value is calculated for every observation, if the k-value is above 0.7 it is a indication that the estimated elpd is not reliable. For the mixed separated model there are 2 out of 142 observations which had a k-value above 0.7. The number of observations which had a k-value above 0.7 for the hierarchical model is 16. Figure 1 and Figure 2, that can be found in appendix, gives a visualization of the Pareto k-values for HLRM and MSM.

5. Conclusions

For this report, we have estimated model parameters for two different multilevel model, HLRM and MSM, and compared their predictive performance on the gapminder data. One can argue that HLRM and MSM are both from the Bayesian hierarchical model framework but are structured in two different ways. Either the data structure can be modeled where the intercept α and slope coefficient β exhibit higher dependence between groups, sharing common information. Alternatively, the data can be constructed as a relatively independent model. The extent of this dependence hinges on several factors, with a critical one being the choice between using including the group-index for the hyperparameters or not. When assigning a group-index for the hyperparameters, a high variance results in a greater disparity between them, making them more dependent. However, the degree of dependence primarily relies on the variance specified when generating the different β estimates. If a very low Beta variance is chosen, the coefficients should exhibit a higher degree of interdependence

Based on the tables in section 4.1 we could notice close to identical posterior estimates for the models' parameters however the LOO estimates imply that MSM have higher predictive performance for the gapminder data which indicates that there is an independence between the continents, emphasizing the use for MSM

In conclusion, given the gapminder data for 2007 the report indicates that the MSM performed better compared to HLRM from a predictive perspective while having close to identical posterior estimates for the models' parameters.

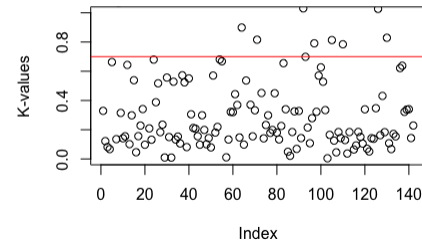
The result should not be seen as generalized, especially in terms of time generalization, since we are working with data from one time point, 2007. If panel data were to be used then we would be able to extend the model by including time-dependent covariates instead. We recommend this type of studies on the gapminder data for future studies.

References

- [1] K. Bhuwalka, E. Choi, E. A. Moore, R. Roth, R. E. Kirchain, and E. A. Olivetti. A hierarchical bayesian regression model that reduces uncertainty in material demand predictions. *Journal of Industrial Ecology*, 27(1):43–55, 2023.

Appendix

Pareto-k Plot for Hierarchical model



Pareto-k Plot for Modified separated model

