# Hierarchical Baysian Modeling of Hurricane Trajectories

P8160 Group Project 3 - Markov chain Monte Carlo

Jingchen Chai, Yi Huang, Zining Qi, Ziyi Wang, Ruihan Zhang

May 2, 2023

## Contents

# Abstract

# Introduction

## Data

**ID**: ID of hurricanes

**Season**: In which year the hurricane occurred

**Month**: In which month the hurricane occurred

**Nature**: Nature of the hurricane

- ET: Extra Tropical
- DS: Disturbance
- NR: Not Rated
- SS: Sub Tropical
- TS: Tropical Storm

**Time**: dates and time of the record

**Latitude** and **Longitude**: The location of a hurricane check point

**Wind.kt**: Maximum wind speed (in Knot) at each check point

# Background

# Exploratory Data Analysis

## Data Pre-processing

- We have filtered observations that occurred on a 6-hour intervals. (e.g., hour 0, 6, 12, 18)
- Calculated the lag difference for latitude, longitude and wind speed.
- After data cleaning, we obtained 20293 observations and with 699 different hurricanes.

# Methodology

## Markov chain Monte Carlo

In our project, we employed a Markov chain Monte Carlo (MCMC) simulation to estimate the parameters of a model that predicts wind speed based on velocity trajectory data. The MCMC algorithm generates samples from the Markov Chain in a way that leads us closer to the desired posterior. In our study, we used two MCMC techniques: the Metropolis-Hastings algorithm and Gibbs sampling.

## Hierarchical Bayesian Model

Bayesian hierarchical modeling is a statistical approach that involves writing a model in multiple levels or a hierarchical form to estimate the parameters of the posterior distribution using Bayesian methodology. This technique assumes that the observed data are generated from a hierarchy of unknown parameters, and it estimates the posterior distribution of these parameters using a Bayesian approach. In other words, Bayesian hierarchical modeling is a way of modeling complex data structures by breaking them down into smaller, more manageable components and using Bayesian analysis to estimate the unknown parameters in each component.

From the Bayes' theorem:

$$posterior\ distribution \propto likelihood \times prior\ distribution$$

$$\pi(\theta|X) \propto \pi(X|\theta) \times \pi(\theta)$$

The Hierarchical Bayes

$$\pi(\theta, \alpha|X) \propto \pi(X|\theta) \times \pi(\theta|\alpha) \times \pi(\alpha)$$

The suggested Bayesian model is

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \boldsymbol{X_i}\gamma + \epsilon_i(t)$$

- where $Y_i(t)$ the wind speed at time $t$ (i.e. 6 hours earlier), $\Delta_{i,1}(t)$, $\Delta_{i,2}(t)$ and $\Delta_{i,3}(t)$ are the changes of latitude, longitude and wind speed between $t$ and $t-6$, and $\epsilon_{i,t}$ follows a normal distributions with mean zero and variance $\sigma^2$, independent across $t$.
- $\boldsymbol{X}_i = (x_{i,1}, x_{i,2}, x_{i,3})$ are covariates with fixed effect $\gamma$, where $x_{i,1}$ be the month of year when the $i$-th hurricane started, $x_{i,2}$ be the calendar year of the $i$ hurricane, and $x_{i,3}$ be the type of the $i$-th hurricane.
- $\boldsymbol{\beta}_i = (\beta_{0,i}, \beta_{1,i}, ..., \beta_{5,i})$, we assume that $\boldsymbol{\beta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

**Prior Distribution**

$$P(\boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi}|\boldsymbol{V}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{V}^{-1}\boldsymbol{\mu}\} \propto |\boldsymbol{V}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{V}^{-1}\boldsymbol{\mu}\}$$

where $\boldsymbol{V}$ is a variance-covariance matrix

$$P(\boldsymbol{\Sigma}) \propto |\Sigma|^{-\frac{(\nu+d+1)}{2}} \exp(-\frac{1}{2}tr(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}))$$

where $d$ is the dimension of $\boldsymbol{\beta}_i$, $\boldsymbol{S}$ is the scale matrix

$$P(\gamma) \propto exp(-\frac{\gamma^2}{2 \times (0.05)^2}) = e^{-200\gamma^2}$$

$$P(\sigma) = \frac{2\alpha}{\sigma^2 + \alpha^2} \propto \frac{1}{\sigma^2 + \alpha^2} = \frac{1}{\sigma^2 + 100}$$

**Joint Posterior Distribution**

Let $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$, derive the posterior distribution of the parameters $\Theta = (\mathbf{B}^\top, \boldsymbol{\mu}^\top, \sigma^2, \Sigma, \gamma)$.

Let$\boldsymbol{Z}_i(t)\boldsymbol{\beta}_i^\top = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \boldsymbol{X_i}\gamma + \epsilon_i(t)$

Where $\boldsymbol{Z_i}$ is the $n_i$*d covariate matrix for hurricane i.

We can find that

$$\mathbf{Y_i} \sim MVN(\boldsymbol{Z}_i\beta_i^T, \sigma^2 I)$$

The likelihood for Y is:

$$f(\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\mu}, \sigma, \boldsymbol{\Sigma}, \gamma) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{B}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma) =$$

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)^\top (\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)\}$$

$$\propto (\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)^\top (\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)\Big\}$$

For simple notation, let $N{=}(\sum_i^n n_i)$, representing the total number of unique hurricanes.

The likelihood for $\boldsymbol{B}$ is:

$$f(\boldsymbol{B} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(\boldsymbol{B}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) =$$
$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)\}$$
$$\propto (|\Sigma|)^{-\frac{N}{2}} \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}((\beta_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\beta_i - \boldsymbol{\mu}))\Big\}$$

**Joint Posterior**

By using the Bayesian rule, we can show the posterior distribution for $\boldsymbol{\Theta}$ is,

$$\pi(\boldsymbol{\Theta}|\boldsymbol{Y}) = P(\boldsymbol{B}, \boldsymbol{\mu}, \sigma, \boldsymbol{\Sigma}, \gamma|\boldsymbol{Y}) \propto \underbrace{L(\boldsymbol{Y}|\boldsymbol{B}, \sigma)}_{\text{likelihood of } \boldsymbol{Y}} \underbrace{L(\boldsymbol{B}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}_{\text{likelihood of } \mathbf{B}} \underbrace{p(\boldsymbol{\mu})p(\sigma)p(\boldsymbol{\Sigma})p(\gamma)}_{\text{priors}}$$
$$\propto \frac{1}{\sigma^N(\sigma^2 + 10^2)} \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma)\Big\}$$
$$\times \exp\{-\frac{1}{2}\sum_i^n(\beta_i - \boldsymbol{\mu})^\top\Sigma^{-1}(\beta_i - \boldsymbol{\mu})\}|\Sigma^{-1}|^{\frac{N+d+v+1}{2}} \exp\{-\frac{1}{2}tr(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\}|\boldsymbol{V}|^{-\frac{1}{2}}$$
$$\times \exp\{-\frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{V}^{-1}\boldsymbol{\mu}\}$$
$$\times \exp\{-200\gamma^2\}$$

**Conditional Posterior Distribution**

1. The posterior distribution of $\mathbf{B}$ Let $\boldsymbol{A}{=}\ \boldsymbol{\Sigma}^{-1}$

$$\pi(\mathbf{B}|\boldsymbol{Y}, \boldsymbol{\mu}^\top, \sigma, \boldsymbol{\Sigma}) \propto \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\gamma_i)\Big\}$$
$$\times \exp\{-\frac{1}{2}\sum_i^n(\beta_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\beta_i - \boldsymbol{\mu})\}$$
$$\propto \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)^\top(\sigma^2 I)^{-1}(\boldsymbol{Y}_i - \boldsymbol{Z}_i\beta_i - \boldsymbol{X_i}\boldsymbol{\gamma}_i)\Big\} + (\beta_i - \boldsymbol{\mu})^\top A(\beta_i - \boldsymbol{\mu})\}$$
$$\propto \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}(\beta_i(\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}Z_i + \boldsymbol{A})\beta_i^\top - 2(\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i - \boldsymbol{Z}_i^\top\gamma\boldsymbol{X}_i(\sigma^2 I)^{-1} + \boldsymbol{\mu}\boldsymbol{A})\beta_i$$
$$\propto \prod_{i=1}^{n} \exp\Big\{ -\frac{1}{2}[\beta_i - (\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Z}_i + \boldsymbol{A})^{-1}(\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i - \boldsymbol{Z}_i^\top\gamma\boldsymbol{X}_i(\sigma^2 I)^{-1} + \boldsymbol{\mu}_i\boldsymbol{A})^\top]^\top$$
$$\times (\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Z}_i + \boldsymbol{A})[\beta_i - (\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Z}_i + \boldsymbol{A})^{-1}(\boldsymbol{Z}_i^\top(\sigma^2 I)^{-1}\boldsymbol{Y}_i - \boldsymbol{Z}_i^\top\gamma\boldsymbol{X}_i(\sigma^2 I)^{-1} + \boldsymbol{\mu}_i\boldsymbol{A})^\top]\Big\}$$

$\beta_i \sim MVN(N^{-1}M, N^{-1})$, where $N = \frac{Z_i^\top Z_i}{\sigma^2} + \boldsymbol{A}$, $M = \frac{Z_i^\top Y_i - Z_i^\top X_i\gamma}{\sigma^2} + \mu\boldsymbol{A}$

2. The posterior distribution of $\boldsymbol{\mu}$

$$\begin{aligned}
\pi(\boldsymbol{\mu}|\boldsymbol{B},\sigma,A,\gamma) &\propto \exp\{-\frac{\boldsymbol{\mu}^\top \boldsymbol{V}^{-1}\boldsymbol{\mu}}{2}\}\prod_{i=1}^{N}\exp\{-\frac{(\beta_i-\boldsymbol{\mu})^\top A(\beta_i-\boldsymbol{\mu})}{2}\} \\
&= \exp\{\sum_{i}^{N}-\frac{1}{2}(\mu^\top(A-\frac{1}{N}\boldsymbol{V}^{-1})\mu-2\mu^\top A\beta_i+\beta_i^\top A\beta_i)\} \\
&= \exp\{-\frac{1}{2}(\boldsymbol{\mu}^\top(NA-\boldsymbol{V}^{-1})\boldsymbol{\mu}-2\boldsymbol{\mu}^\top\sum_{i}^{N}(\boldsymbol{A}\beta_i)+\beta_i^\top\boldsymbol{A}\beta_i)\}
\end{aligned}$$

$\boldsymbol{\mu} \sim MVN(M^{-1}N, M^{-1})$, where $M = NA - \boldsymbol{V}^{-1}$ and $N = \sum_i^n(\boldsymbol{A}\beta_i)$

3. The posterior distribution of $\boldsymbol{\Sigma}$

$$\begin{aligned}
\pi(\boldsymbol{\Sigma}|\boldsymbol{B},\boldsymbol{\mu},\gamma,\sigma,\boldsymbol{Y}) &\propto |\boldsymbol{\Sigma}|^{-\frac{(N+v+d+1)}{2}}\exp\{-\frac{1}{2}(\sum_{i}^{N}(\beta_i-\boldsymbol{\mu})^\top\boldsymbol{A}(\beta_i-\boldsymbol{\mu})+tr(\boldsymbol{SA}))\} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{(N+v+d+1)}{2}}\exp\{-\frac{1}{2}tr(S+\sum_{i}^{N}(\beta_i-\boldsymbol{\mu})(\beta_i-\boldsymbol{\mu})^\top)\boldsymbol{A}\}
\end{aligned}$$

$$\boldsymbol{\Sigma} \sim w^{-1}(\boldsymbol{S}+\sum_{i}^{N}(\beta_i-\boldsymbol{\mu})(\beta_i-\boldsymbol{\mu})^\top, N+v)$$

4. The posterior distribution of $\gamma$

$$\begin{aligned}
\pi(\gamma|\boldsymbol{B},\boldsymbol{\mu},\boldsymbol{\Sigma},\sigma,\boldsymbol{Y}) &\propto \prod_{i=1}^{N}\exp\{-\frac{1}{2}(Y_i-Z_i\beta_i-X_i\gamma)^\top(\sigma^2 I)^{-1}(Y_i-Z_i\beta_i-X_i\gamma)\}\times\exp\{-\frac{400\gamma^\top\gamma}{2}\} \\
&\propto \exp\{-\frac{1}{2}\sum_{i}^{N}\gamma^\top(X_i^\top\sigma^{-2}IX_i+400N^{-1}I)\gamma-2\gamma^\top(X_i^\top\sigma^{-2}IY_i-X_i^\top\sigma^{-2}IZ_i\beta_i) \\
&+Y_i^\top\sigma^{-2}IY_i-2Y_i^\top\sigma^{-2}IZ_i\beta_i+\beta_i^\top Z_i^\top\sigma^{-2}IZ_i\beta_i\}
\end{aligned}$$

$\gamma \sim MVN(M^{-1}N, M^{-1})$, where $M = \frac{\sum_i^N X_i^\top X_i}{\sigma^2} + 400I$ and $N = \frac{\sum_i^N(X_i^\top Y_i - X_i^\top Z_i\beta_i)}{\sigma^2}$

5. The posterior distribution of $\sigma$

$$\begin{aligned}
\pi(\sigma|\boldsymbol{Y},\mathbf{B},\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\gamma}) &\propto \frac{1}{\sigma^N(\sigma^2+10^2)} \\
&\times\prod_{i=1}^{n}\exp\Big\{-\frac{1}{2(\sigma^2 I)}(\boldsymbol{Y}_i-\boldsymbol{Z}_i\beta_i-\boldsymbol{X_i}\gamma)^\top(\boldsymbol{Y}_i-\boldsymbol{Z}_i\beta_i-\boldsymbol{X_i}\gamma)\Big\}
\end{aligned}$$

$\sigma$ does not have a close distribution.

## MCMC Algorithm

After deriving the conditional posterior of parameter that we want to estimate, the next step is to apply these conditional posterior to the MCMC Algorithm. Our MCMC algorithm is a hybrid of Metropolis-Hastings and Gibb Sampling.

**Algorithm 1** MCMC: Metropolis-Hastings

**Require:** Target distribution $\pi(\sigma)$
**for** $i$ = 1 to 1000 **do**
1. Proposed $\sigma_{proposed} = \sigma^{(i-1)} + (U - 0.5) * 2 * a$, where $U \sim$ Uniform(0,1), $a$ is step size
2. Calculate acceptance rate$\alpha_{XY} = \min(0, \frac{\pi(\sigma_{proposed})}{\pi(\sigma^{(i-1)})}$
3. If $U < \alpha_{XY}$: $\sigma^{(i)} = \sigma_{proposed}$, else $\sigma^{(i)} = \sigma^{(i-1)}$

**end for**
$\sigma_k = \sum_{i=801}^{1000} \frac{\sigma^{(i)}}{200}$, where $k$ is the iteration of Gibb Sampling

**Metropolis-Hastings**

From the conditional posterior of $\sigma$, it is hard to find a closed form distribution for it, unlike other parameters. Here, we apply Metropolis-Hastings to generate new $\sigma$. The detailed steps of Metropolis-Hasting is shown below:

The target distribution is the conditional posterior of $\sigma$. By setting the step size to 0.5, the acceptance rate reaches 43.5%, which is acceptable. The new sigma generated for Gibb Sampling will be the mean of next 200 values in the chain.

**Gibb Sampling**

After defining the Metropolis-Hastings algorithm to generate $\sigma$, we combine the Metropolis-Hastings with Gibb Sampling. We first initialized the parameters to start the algorithm. The parameters in Gibb Sampling will be updated component-wise. For each parameter to be updated, it always conditioned on the most recent values of other parameters. More precisely,

**Algorithm 2** MCMC: Gibb Sampling

**Require:** Initialize $\boldsymbol{B}, \boldsymbol{\mu}, \sigma, \boldsymbol{\Sigma}, \boldsymbol{\gamma}$
**for** $k$ = 1 to 5000 **do**
1. Generate $\boldsymbol{\beta}_i^k$ for $i^{th}$ hurricane from $\pi(\mathbf{B}|\boldsymbol{Y}, \boldsymbol{\mu}^{k-1}, \sigma^{k-1}, \boldsymbol{\Sigma}^{k-1}, \boldsymbol{\gamma}^{k-1})$
2. Generate $\boldsymbol{\mu}^k$ from $\pi(\boldsymbol{\mu}|\boldsymbol{Y}, \mathbf{B}^k, \sigma^{k-1}, \boldsymbol{\Sigma}^{k-1}, \boldsymbol{\gamma}^{k-1})$
3. Generate $\sigma_k$ from the Metropolis-Hastings steps
4. Generate $\boldsymbol{\Sigma}_k$ from $\pi(\boldsymbol{\Sigma}|\boldsymbol{Y}, \mathbf{B}_k, \boldsymbol{\mu}_k, \sigma_k, \boldsymbol{\gamma}_{k-1})$
5. Generate $\boldsymbol{\gamma}_k$ from $\pi(\boldsymbol{\gamma}|\boldsymbol{Y}, \mathbf{B}_k, \boldsymbol{\mu}_k, \sigma_k, \boldsymbol{\Sigma}_k)$

**end for**

We have tested different start values for MCMC algorithm, the result chain behave similarly. We finally decide to initialize the parameters by using the results from fitting generalized linear mixed model in R. $\boldsymbol{B}$ is a 5 * 699 matrix, $\boldsymbol{\mu}$ is a 5 * 1 matrix, $\sigma$ is a number, $\boldsymbol{\Sigma}$ is a 5 * 5 matrix, and $\boldsymbol{\gamma}$ is a 14 * 1 matrix.

# Results

In Markov Chain Monte Carlo, determining the appropriate number of iterations can depend on many factors such as the complexity of the model, the size of the dataset, etc. Therefore, it is difficult to make a general statement about a specific number of iterations that will be sufficient for all MCMC simulations. In our algorithm, we believe for most of the parameters, 5000 iterations reached the stationary of their posterior distribution. For convergence diagnostics, we generate trace plots, autocorrelation plots, and histograms.

## Random effect parameter

Figure 1 shows convergence plots of $\boldsymbol{B}$. $\beta_1$ takes only a few steps to reach stationary. The trace plot shows the history of our parameter $\beta_1$ across iterations of the chain. This chain appears most likely to converge with an average value of about 0.95. Similarly for $\beta_4$, the chain appears most likely to converge with an average value of about 0.48. $\beta_2$ and $\beta_3$ need to take more iterations to achieve convergence. In Figure 2, we extract Hurricane George 1951 from the data to check its $\beta$ convergence plots and distribution. We can see on trace plot of 5000 iterations for the selected parameter, each of the distributions are relatively normal with some heavy tails in $\beta_0$.

The Convergence plot of $\sigma^2$ suggests that the chain is mixing well and that the algorithm is converging to its posterior distribution of sigmaˆ2. After iteration 500, this chain appears to converge with an average value of 34.5 This also indicates that the estimated values of $\sigma^2$ are becoming more independent and less influenced by their past values as iterations increase.

Figure 1

Figure 2 shows the selected $\boldsymbol{B}$: Hurricane GEORGE.1951

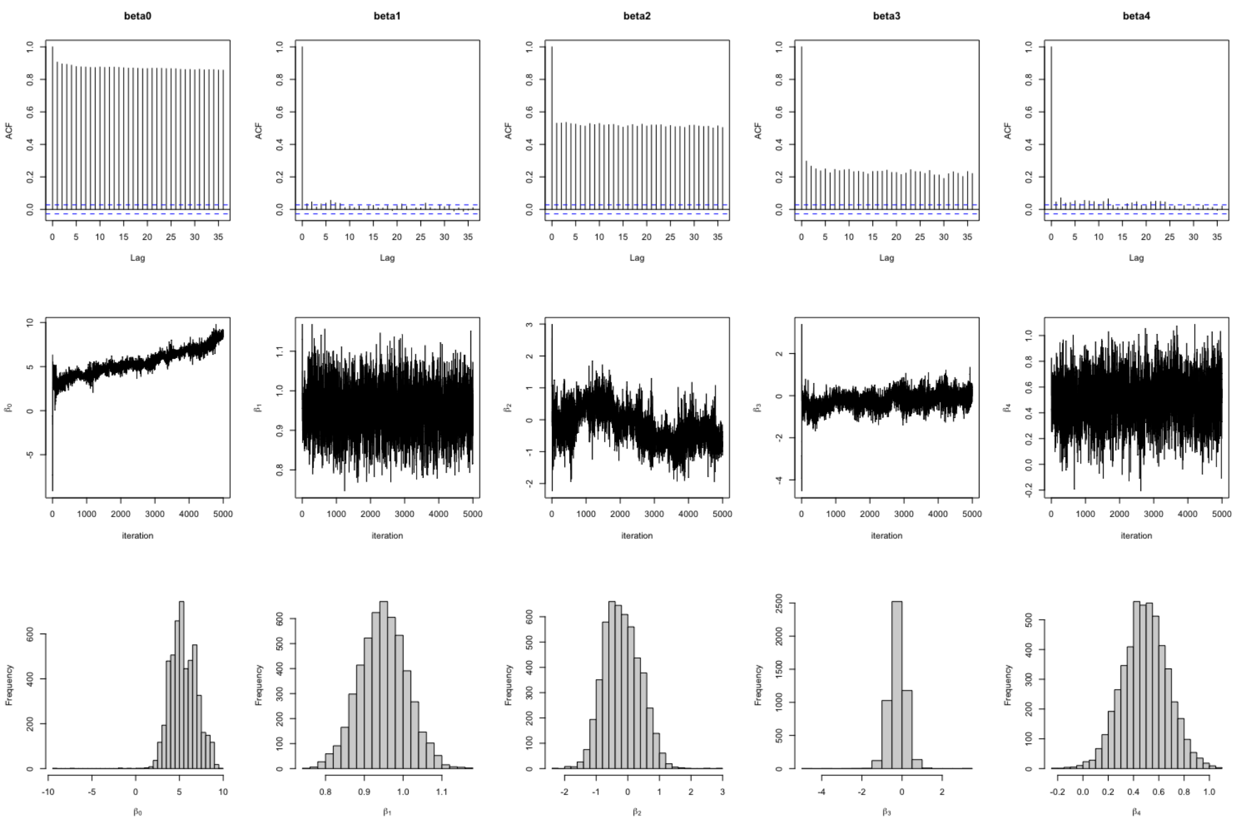


Figure 2

### Fix effect parameter $\gamma$

## Limitation and Discussion

## Conclusion

## References


Taboga, Marco (2021). "Markov Chain Monte Carlo (MCMC) diagnostics", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix.

Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter.

Zhang, Z. (2021). A Note on Wishart and Inverse Wishart Priors for Covariance Matrix. Journal of Behavioral Data Science, 1(2), 119–126.

# Contributions

# A First section of Appendix

# B Section section of Appendix