A Bayesian Analysis of Earthquake Data

Executive Summary

This report investigates the modeling of earthquake magnitudes using autoregressive (AR) and mixture autoregressive (Mixture AR) models. The objective is to assess the performance of these models for time series forecasting and compare their goodness of fit. We fit an AR(3) model and a Mixture AR(3) model with two components to the data, using a Gibbs sampler for Bayesian estimation of parameters. Model comparison is conducted using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC). Based on the results, the AR(3) model demonstrates a significantly better fit in terms of AIC and BIC, while the Mixture AR model shows poorer performance, suggesting the AR(3) model is preferable for this dataset.

Introduction

The analysis focuses on earthquake magnitudes, with the goal of understanding the underlying time series structure and making predictions. Time series models such as autoregressive (AR) models are useful for capturing the temporal dependencies in data. This report fits both an AR(3) model and a Mixture AR(3) model to evaluate their effectiveness in modeling earthquake magnitudes. The Mixture AR model is designed to account for potential subpopulations or regimes within the data. The comparison between the two models is based on AIC, BIC, and DIC, standard model selection criteria in Bayesian analysis.

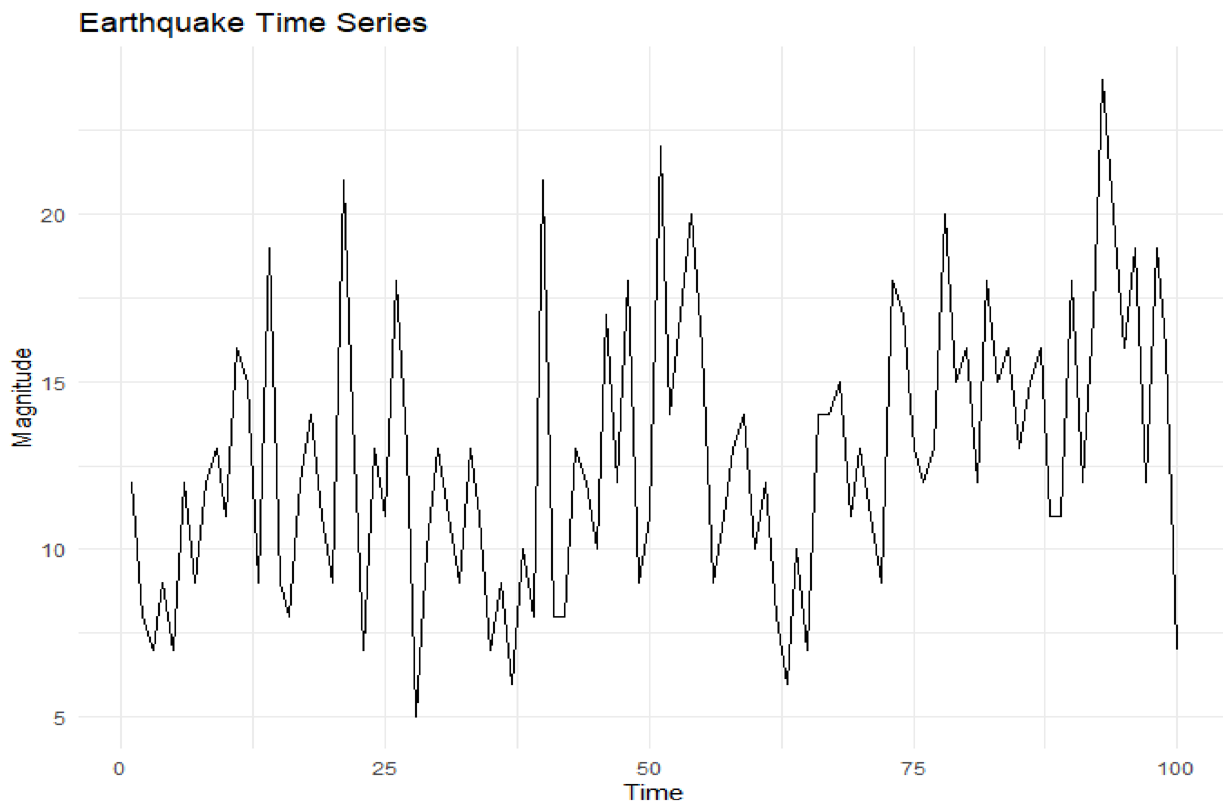
Data

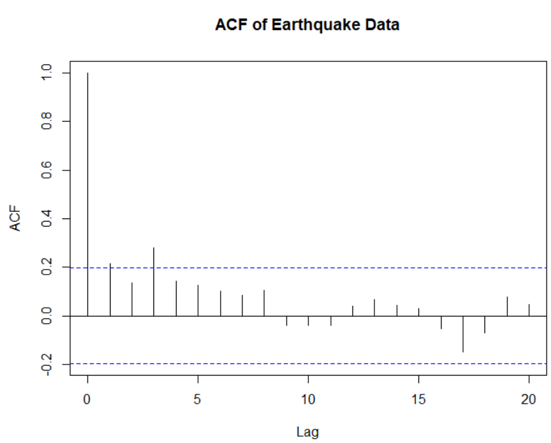
The dataset contains the magnitudes of earthquakes, with the first 100 data points used as the training set. The data was explored graphically to check for trends, seasonality, or stationarity. A time series plot of the earthquake magnitudes revealed no significant trends or periodic behavior, suggesting that the AR model could be an appropriate choice for modeling.

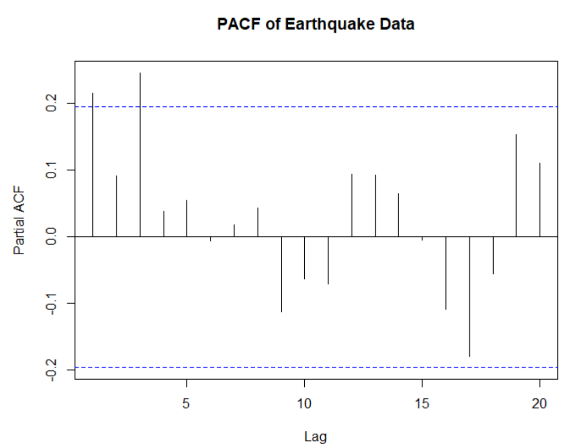
A plot of the data showed relatively stationary behavior with some large deviations, which could motivate considering a mixture of AR models. Even from the ADF test, we can reject the null hypothesis, indicating that there is evidence to support the claim that the time series is stationary.

<u>Augmented Dickey-Fuller Test</u>

Dickey-Fuller = -3.5881, Lag order = 4, p-value = 0.03778
alternative hypothesis: stationary


Earthquake Time Series
Magnitude
5
10
15
20
0
25
50
75
100
Time


ACF of Earthquake Data
ACF
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Lag


PACF of Earthquake Data
Partial ACF
-0.2
-0.1
0.0
0.1
0.2
5
10
15
20
Lag

AR Model

An AR(3) model was chosen for this analysis after examining the partial autocorrelation function (PACF) plot, which showed significant lags up to order 3. The model specification is as follows:

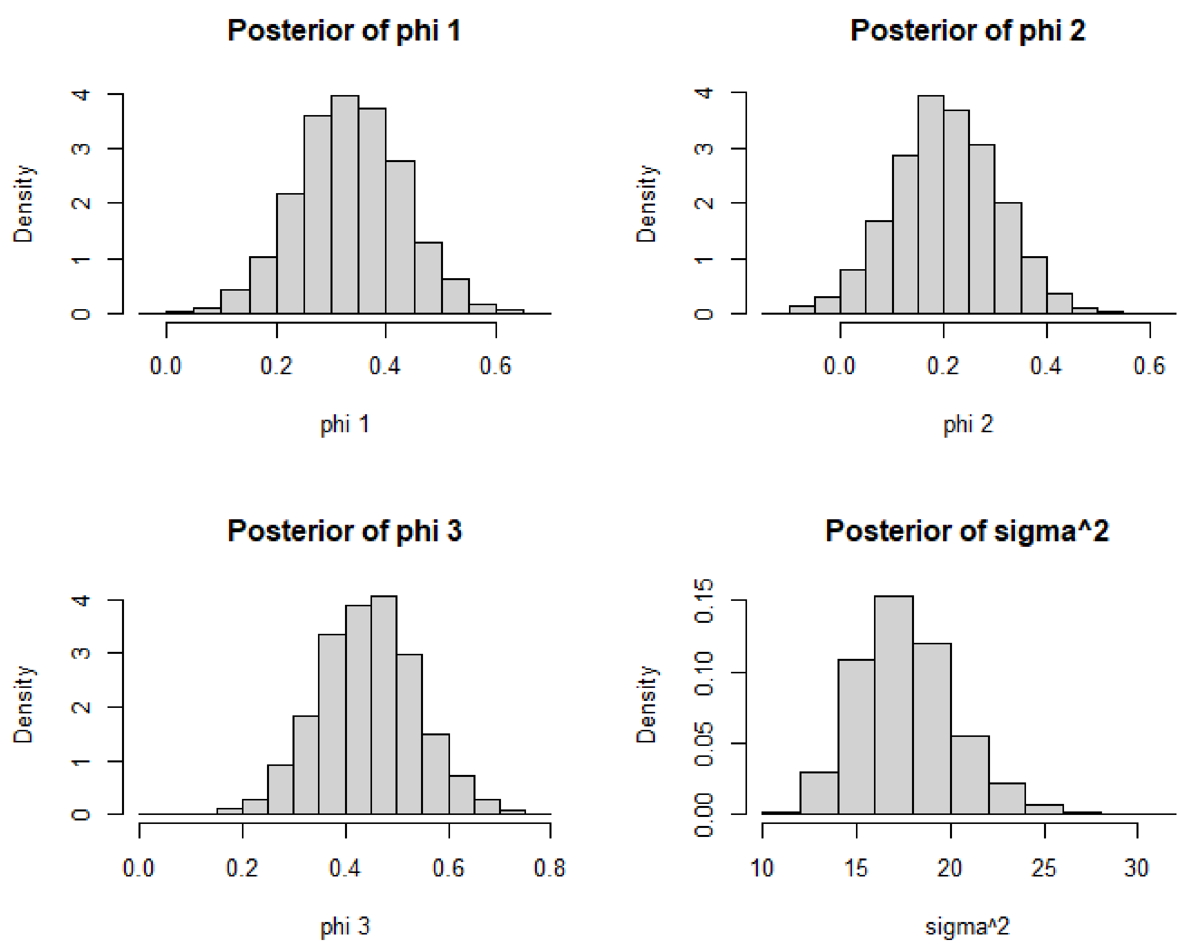
Model Specification:

- $Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t$

- $\epsilon_t \sim N(0,\sigma^2)$

Prior Distributions:

- $\Phi \sim N(0, \text{large variance})$
- $\sigma^2 \sim \text{Inverse-Gamma}(0.01, 0.01)$



Posterior Analysis: The prior for the AR coefficients ϕ\phiϕ was chosen to be a normal distribution centered at 0 with a diffuse variance (large). The prior for the error variance σ2\sigma^2σ2 was an inverse gamma distribution. A Gibbs sampler was implemented to estimate the posterior distributions of the AR coefficients and error variance. Trace plots showed good convergence after a burn-in period of 1000 iterations.

Posterior Mean of AR(3) Coefficients: 0.3338612 0.2078382 0.4399614

Posterior Mean of AR(3) Variance: 17.67252

Prediction: The model was used to make several steps ahead predictions, and it performed reasonably well, capturing the overall trend of the series with moderate uncertainty.

3-step ahead predictions from AR(3): 17.19669 14.57024 14.02171

Mixture of AR Model

The Mixture AR model introduces two components, each with its own set of AR(3) parameters and error variances. The mixture model accounts for different regimes or subgroups within the earthquake magnitudes, possibly reflecting different types of seismic activity.πππππ

Model Specification:

- $Y_t = \phi_{k,1} y_{t-1} + \phi_{k,2} y_{t-2} + \phi_{k,3} y_{t-3} + \epsilon_{t,k}$
- $\epsilon_t \sim N(0,\sigma^2)$

Prior Distributions:

- $\Phi \sim N(0, \text{large variance})$
- $\sigma^2 \sim$ Inverse-Gamma
- $\pi_k \sim$ Dirichlet(1)

The mixing proportions π_k are drawn from a Dirichlet prior, and the AR coefficients ϕ_k are modeled similarly to the single AR(3) model, with diffuse priors. A Gibbs sampler was employed to estimate the posterior distributions of the model parameters.

Posterior Analysis and Convergence: The posterior distributions for the mixture components converged slowly, with one component dominating the posterior estimates. This behavior suggests that the data may not strongly support multiple regimes, which is reflected in the poorer model fit.

Prediction: The Mixture AR model's predictions were highly uncertain, with wide credible intervals, indicating less reliable performance compared to the AR(3) model.

Model Comparison

The AIC, BIC, and DIC for both the AR(3) and Mixture AR models were calculated, and the results are summarized below:

Model	AIC	BIC	DIC
AR(3) Model	557.02	567.32	56475.33
Mixture of AR(3) Model	1757.364	1777.961	483397

The AR(3) model outperforms the Mixture AR model on all criteria (AIC, BIC, DIC). The extremely large DIC for the Mixture AR model suggests that it overfits the data, and its complexity is not warranted given the structure of the earthquake magnitudes.

Conclusions

The AR(3) model provides a better fit to the earthquake magnitude data compared to the Mixture AR model. While the Mixture AR model introduces more complexity by accounting for potential subpopulations, this added complexity does not result in improved performance, as reflected by the model comparison criteria. The AR(3) model, with fewer parameters and better predictive performance, is preferred for this dataset.

The analysis assumes that earthquake magnitudes can be captured by a simple autoregressive process. However, seismic events may exhibit more complex, nonlinear dynamics that are not captured by AR models. Future work could explore non-linear time series models or incorporate exogenous variables such as geographic or temporal factors.