Bayesian Time Series Modeling of Tata Motors Stock Returns Using an AR(1) Process

1 Introduction

Financial markets are inherently volatile and influenced by a wide range of unpredictable factors. This project investigates whether Bayesian time series modeling, specifically an AR(1) process, can detect and quantify short-term autocorrelation in the daily log returns of Tata Motors Ltd stock. Since returns in liquid markets tend to exhibit low but persistent autocorrelation, an autoregressive framework is well-suited to explore such dynamics.

Our approach leverages Bayesian inference, which allows us to quantify uncertainty around model parameters while incorporating prior beliefs. The analysis was conducted in R and Stan, and all code is available at:
`https://github.com/bluenoclue/bayesian-time-series-tata-motors/blob/main/bayesian-time-series-tata-motors/scripts/STAT_447_Anita_Keshri_Final_Project.R`.

2 Data Description

The dataset used in this analysis contains daily stock prices for Tata Motors Ltd, covering the period from December 2006 to January 2024. The dataset was taken from Kaggle and can be found within the repository or at: `https://www.kaggle.com/datasets/raunakpandey030/tata-motors-limited-stock-prices-2006-2024`. We calculated log returns from the adjusted closing prices using the formula $r_t = \log(P_t/P_{t-1})$, which stabilizes variance and helps make the series more suitable for linear modeling.

Visual inspection of the time series and the autocorrelation function (ACF) showed mild but persistent autocorrelation at lag 1. This empirical finding provided a strong motivation for fitting an AR(1) model as a baseline.

3 Model Specification

Let y_t denote the log return on day t. The AR(1) model we use is given by:

$$
\begin{aligned}
y_t &\sim \mathcal{N}(\phi y_{t-1}, \sigma) \\
\phi &\sim \mathcal{N}(0, 0.5) \\
\sigma &\sim \mathcal{N}^+(0, 1)
\end{aligned}
$$

The prior on ϕ is centered at 0 with a relatively tight spread, expressing the belief that returns are only weakly autocorrelated. A half-normal prior was used for σ to ensure non-negativity while remaining weakly informative.

The model was implemented in Stan and fit using four MCMC chains, each running 2000 iterations with 500 warm-up steps. Diagnostics including trace plots and $\hat{R}$ statistics confirmed good convergence.

4 Literature Review

Bayesian methods have long been favored in time series analysis for their ability to incorporate uncertainty and prior knowledge. West and Harrison (1997) introduced a general framework for Bayesian dynamic models, highlighting their adaptability to evolving data structures. Koop (2003) extended Bayesian methods to macroeconomic models, illustrating how autoregressive structures can be flexibly estimated in uncertain environments.

In financial econometrics, Jacquier et al. (1994) applied Bayesian inference to stochastic volatility models, demonstrating its robustness in capturing latent market behaviors. Their work inspired many subsequent studies that modeled financial time series using hierarchical or state-space approaches.

Gelman et al. (2013) stressed the importance of careful prior selection and model checking, which informed our own approach to sensitivity analysis. Lastly, Carpenter et al. (2017) presented Stan, a modern platform for Bayesian computation. Stan's implementation of Hamiltonian Monte Carlo (HMC) allows efficient sampling from complex posteriors, making it well-suited for our model.

5 Prior Sensitivity Analysis

To evaluate the influence of our prior choice on ϕ, we fit a second model with a wider prior: $\phi \sim \mathcal{N}(0, 1)$. The resulting posterior was nearly identical to that from the original model, centered around 0.066, with only slightly increased variance.

This suggests that the data contains enough information to dominate the prior and confirms the robustness of our inference. The posterior distribution remained stable under both settings, providing additional confidence in the model's reliability.

6 Model Evaluation

The key posterior summaries for the AR(1) model are:

- $E[\phi] = 0.066$ with a 95% credible interval of [0.045, 0.087]
- $E[\sigma] = 0.028$, indicating relatively low noise
- $\hat{R} \approx 1.00$ for all parameters
- Effective sample sizes were high
- No divergent transitions were detected

Trace plots and posterior densities were smooth and unimodal. Visual checks confirmed that the MCMC chains were well-mixed. Although the signal is weak, the consistent positive value of ϕ aligns with theoretical expectations for daily returns.

We plotted the model's fitted values against observed data and observed good alignment for one-step-ahead forecasts. Residual analysis showed approximate normality but revealed mild heteroskedasticity during high-volatility periods. Posterior predictive checks further confirmed that the model captured the overall shape of the data, though predictive uncertainty grew over longer horizons.

7 Model Comparison: AR(1) vs AR(2)

To test whether a more complex model offered substantial gains, we fit an AR(2) process:

$$y_t \sim \mathcal{N}(\phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma)$$

Model comparison using Leave-One-Out Cross-Validation (LOO) showed no meaningful improvement in predictive accuracy. The posterior mean of ϕ_2 was very close to zero, suggesting that the second lag did not contribute much beyond what was already captured by ϕ_1.

In this context, the AR(1) model is preferable due to its parsimony and similar performance.

8 Methodological Discussion

The Bayesian AR(1) model offers a principled way to analyze time-dependent data with clear uncertainty quantification. Its implementation in Stan allowed for flexible prior specification and robust diagnostics. Our approach, though simple, lays the groundwork for more sophisticated models, such as those incorporating volatility or covariates.

Bayesian inference is especially suitable for financial data, where small signals must be separated from noise. The ability to incorporate domain knowledge through priors also opens opportunities for tailoring models to specific assets or market conditions.

9 Discussion and Limitations

Our results indicate weak but measurable autocorrelation in daily returns for Tata Motors Ltd. The inference proved robust to changes in prior specification, and model diagnostics supported the validity of the AR(1) framework.

However, the model does not capture volatility clustering, a hallmark of financial time series. Moreover, it does not account for market events or external variables that might explain some of the variance in returns. These limitations could be addressed by incorporating GARCH components or multivariate extensions in future work.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC press.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 371–389.

Koop, G. (2003). *Bayesian econometrics*. John Wiley & Sons.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). Springer.

Appendix

Stan Code for AR(1)

```
data {
  int<lower=1> N;
  vector[N] y;
}
parameters {
  real phi;
  real<lower=0> sigma;
}
model {
  phi ~ normal(0, 0.5);
  sigma ~ normal(0, 1);
  for (t in 2:N)
    y[t] ~ normal(phi * y[t-1], sigma);
}
```

Plots

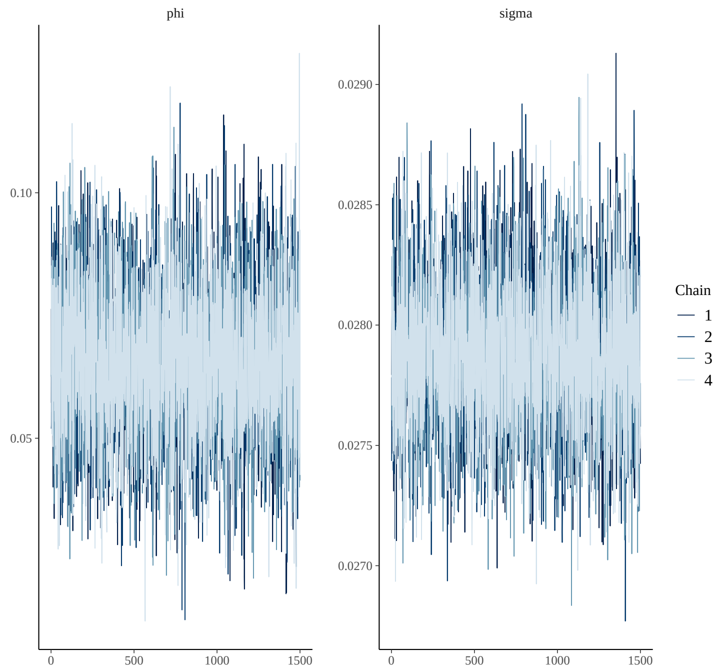

Figure 1: Trace Plot

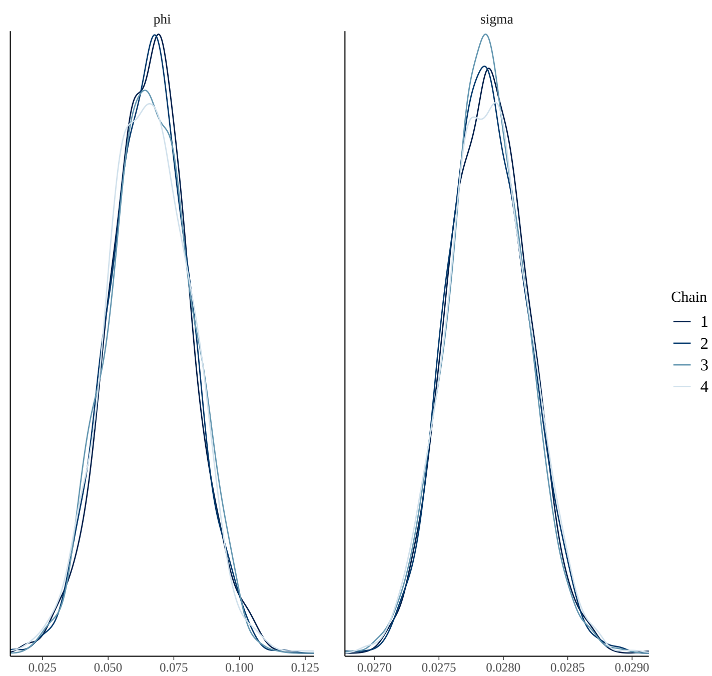

Figure 2: Posterior Density Overlay

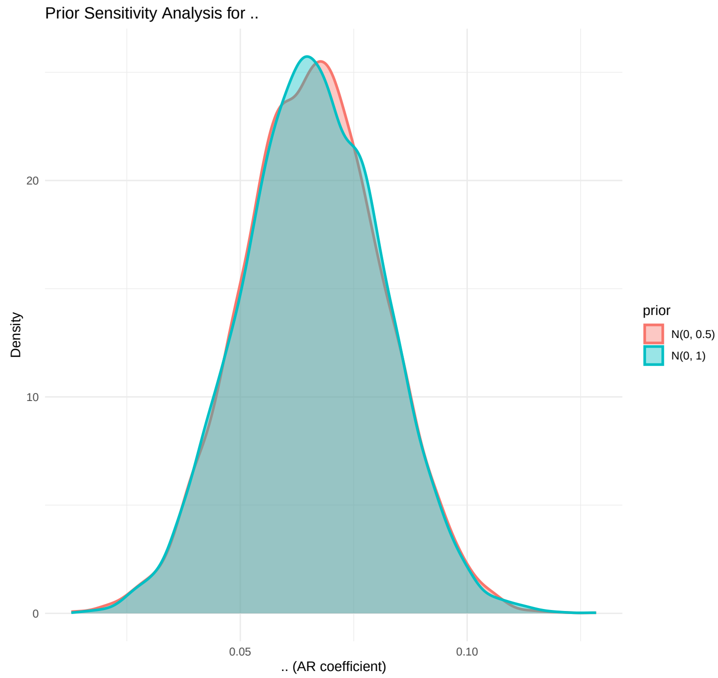

Figure 3: Prior Sensitivity Analysis

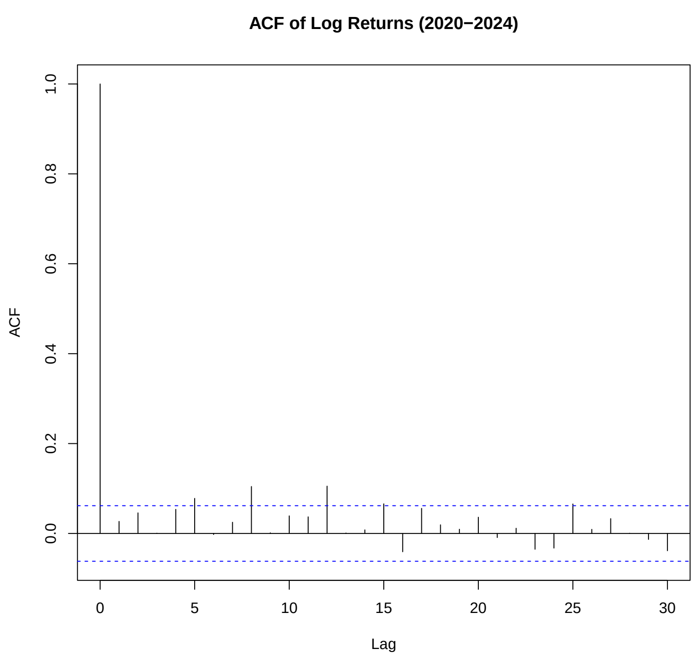

Figure 4: ACF

LOO Comparison Table

Model	LOOIC
AR(1)	2456.1
AR(2)	2455.3