# Hierarchical Bayesian Model Specification

**Response Variable:**

The dependent variable in the Bayesian regression model is the monthly dividend-adjusted log returns for individual S&P 500 stocks:

Where:

rit: Return for stock i at time t

Xit: Vector of predictors (factors) for stock i at time t

B: Vector of regression coefficients

Eit: Normally distributed error term with variance

**Prior Distributions (Hierarchical Structure & Spike-and-Slab):**

1. **Hierarchical Priors by Economic Themes:**

For each coefficient (factor within theme ):

Here, each factor within a theme shares a common distribution, allowing information pooling and thematic coherence.

* Hyperpriors for Theme Means:
* Hyperpriors for Theme Variances:

**Spike-and-Slab Prior on Individual Factors**:

To facilitate automatic feature selection at the individual factor level, a spike-and-slab prior is applied:

* S2spike: very small variance, tightly concentrated around zero
* S2slab: Larger variance, allowing significant deviation from zero.
* Inclusion indicator : determining factor relevance. A high gamma means the factor as been ‘selected’. Significant contributor to return forecasting.

**Novel Approach: Informing Prior Inclusion Probabilities (gamma) Using Jensen et al. (2023):**

The innovative part of this approach is setting prior inclusion probabilities informed by the economic significance identified by Jensen et al. (2023):

* If Jensen et al. found theme to be economically highly significant, factors within this theme are given a higher prior probability of inclusion:
* Conversely, for themes Jensen et al. deemed less relevant, a lower prior probability of inclusion is assigned:

**Variance Prior:**

Error variances can initially be assumed common for simplicity or allowed to vary by stock with the following prior distribution:

**Posterior Distribution Estimation:**

Using Bayes' theorem, the posterior distribution combines the data and the prior information to provide the updated distribution of the parameters:

**Python Implementation:**

The spike-and-slab hierarchical Bayesian model will be implemented using the PyMC Python library, a powerful tool for Bayesian statistical modeling and Markov Chain Monte Carlo (MCMC) sampling, offering straightforward methods to specify complex priors and hierarchical structures clearly.

**Purpose:**

This detailed mathematical formulation is aimed at obtaining precise feedback regarding the Bayesian hierarchical structure, the spike-and-slab priors tailored with economic insights from Jensen et al. (2023), and the overall suitability of these modeling choices for my forecasting objectives.