Bayesian Partial Reduced-Rank Regression

Prepared by:
adithya (220587)
jayant (220479)
riya (220353)
sunita (221106)

Supervised by:
Arnab Hazra

20, April, 2025

Contents

Abstract

This report presents the development and implementation of a Bayesian Partial Reduced-Rank Regression (BPRR) model. The primary objective of the research is to explore the application of reduced-rank regression models in a Bayesian framework for high-dimensional data analysis. The BPRR model leverages the structure of the data through variable selection mechanisms and sparsity-inducing priors, aiming to identify significant predictors while reducing dimensionality. The study applies the model to a real-world dataset, demonstrating its effectiveness in identifying relevant variables and providing predictive insights. The report outlines the mathematical foundations, implementation steps, and results obtained from the model, including its comparison with standard regression techniques. Additionally, the results showcase the robustness of the BPRR model in handling multicollinearity and high-dimensionality in datasets.

1 Introduction

In modern statistical analysis, the challenge of dealing with high-dimensional data is increasingly common. Reduced-rank regression (RRR) methods have been developed to tackle such problems by reducing the rank of the regression coefficient matrix, thus improving the model's interpretability and computational efficiency. Bayesian approaches to RRR provide a probabilistic framework that allows for the incorporation of prior knowledge and uncertainty quantification. In this report, we focus on the development of a Bayesian Partial Reduced-Rank Regression (BPRR) model, which introduces sparsity in both the regression coefficients and the rank of the model.

The goal of BPRR is to identify a subset of relevant predictors from a large set and to model the response variables using a reduced-rank structure. The Bayesian framework offers several advantages, including the ability to estimate uncertainty in the parameters and to incorporate domain-specific prior knowledge into the model. We apply this approach to a dataset where dimensionality reduction and variable selection are crucial for identifying the most important predictors.It learns which responses belong to the low-rank vs. full-rank groups (γ) from the data, rather than fixing them in advance. Only a subset of responses is forced into a low-rank structure; the rest remain full-rank.Captures heterogeneity across responses more faithfully.

2 Model Description

The Bayesian Partial Reduced-Rank Regression model is formulated by assuming a linear relationship between the response variables Y and the predictor matrix X, with a covariance structure specified by the matrix Σ. The model introduces a sparsity constraint on the regression coefficients through the use of a binary indicator vector γ, where each element of γ determines whether the corresponding variable in X is included in the model or not. Additionally, the rank of the regression coefficient matrix is reduced to improve model efficiency and interpretability.

The likelihood function is modeled as a multivariate normal distribution with a covariance structure Σ. The sparsity-inducing prior on γ is modeled using a Beta distribution with a parameter ρ, which controls the level of sparsity. The model is implemented using a Markov Chain Monte Carlo (MCMC) method to sample from the posterior distribution of the model parameters.

$$Y = XC + E, \quad E = (e(1), \ldots, e(n))',$$

where $e(i) \sim \mathcal{N}_q(0, \Sigma)$.

We assume that the response variables can be split into two different groups Y_1 and Y_2 of dimensions $n \times q_\gamma$ and $n \times (q - q_\gamma)$, respectively, where $q_\gamma \in \{2, \ldots, q-1\}$.

Moreover, we assume that the relationship between Y_1 and X admits a low-rank structure, while the regression of Y_2 on X has full rank. Under this assumption, the coefficient matrix $C \in \mathbb{R}^{p \times q}$ can be partitioned as:

$$C = [C_1, C_2],$$

with $C_1 \in \mathbb{R}^{p \times q_\gamma}$ having reduced rank $r = \text{rank}(C_1) \leq \min(p, q_\gamma) - 1$, and $C_2 \in \mathbb{R}^{p \times (q - q_\gamma)}$ with full rank $r_2 = \text{rank}(C_2) = \min(p, q - q_\gamma)$.

The model can be equivalently written as:

$$[Y_1, Y_2] = X[C_1, C_2] + [E_1, E_2].$$

Each of the n response vectors $y(i),\ i = 1, \ldots, n$, is of the form

$$y(i) = (y_{i,1}, \ldots, y_{i,q_\gamma},\ y_{i,q_\gamma+1}, \ldots, y_{i,q})' \in \mathbb{R}^q.$$

We write $e(i) = (e_1^{i\prime}, e_2^{i\prime})'$, with

$$e_1^i = (e_{i,1}, \ldots, e_{i,q_\gamma})' \quad \text{and} \quad e_2^i = (e_{i,q_\gamma+1}, \ldots, e_{i,q})'.$$

We assume that $e(i) \sim \mathcal{N}_q(0, \Sigma)$, with the partitioned covariance matrix

$$\Sigma = \text{Cov}(e(i)) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{bmatrix},$$

where

$$\Sigma_{11} = \text{Cov}(e_1^i), \quad \Sigma_{22} = \text{Cov}(e_2^i), \quad \Sigma_{12} = \text{Cov}(e_1^i, e_2^i).$$

The posterior distribution is computed using the following steps:

$\bullet Generatecandidatevaluesfor\gamma$ based on prior knowledge.

$\bullet ReordertheresponsematrixY$ and predictor matrix X based on the selected gamma configuration.

$\bullet Computethecovariancematrix\Sigma$ based on the reordered matrices.

$\bullet Estimatetheregressioncoefficientsusinganiterativeprocedure(e.g., IMLE).$

$\bullet\ Computethelog - likelihoodandlog - posteriorforeachcandidate\gamma.$

3 Prior Specifications

3.1 Response Variables

We introduce a binary vector $\boldsymbol{\gamma} \in \{0,1\}^q$ to categorize the responses into the low-rank and the full-rank groups. As we lack any prior information regarding the criteria for this classification, we assume that each element γ_j, for $j = 1, \ldots, q$, follows independently a Bernoulli prior distribution with probability ρ of being assigned to the low-rank group.

Consequently, the joint prior distribution on $\boldsymbol{\gamma}$ is:

$$p(\boldsymbol{\gamma} \mid \rho) = \prod_{j=1}^{q} \mathrm{Bern}(\gamma_j \mid \rho) \cdot \mathbb{I}(1 < q_\gamma < q), \tag{1}$$

where $q_\gamma = \sum_{j=1}^{q} \gamma_j$, and $\rho \in (0,1)$ is the prior probability of being assigned to the low-rank group.

Additionally, we employ a hierarchical prior structure, where ρ is assigned a Beta prior distribution, $\rho \sim Be(\rho|\alpha_\rho, \beta_\rho)$.

In our code, we take $\alpha_\rho = 0.5$ and $\beta_\rho = 0.5$.

3.2 Coefficient Matrix

We consider a multivariate regression model where the coefficient matrix $\mathbf{C}$ may be of reduced rank. Let $\mathbf{C}_1$ be a submatrix of $\mathbf{C}$ with reduced rank $r \leq r_{\max} = \min(p, q_\gamma) - 1$. Conditional on a binary variable γ, we assume a discrete uniform prior for r:

$$r \mid \gamma \sim \mathcal{U}(\{1, \ldots, r_{\max}\}).$$

Given that $\mathbf{C}_1$ is low-rank, it can be decomposed as:

$$\mathbf{C}_1 = \mathbf{B}\mathbf{A}^\top,$$

where $\mathbf{A} \in \mathbb{R}^{q_\gamma \times r}$ and $\mathbf{B} \in \mathbb{R}^{p \times r}$. To ensure identifiability of this decomposition, we adopt the restriction (Geweke, 1996) that the first r rows of $\mathbf{A}$ form an identity matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{F} \end{bmatrix},$$

where $\mathbf{F}$ is a $(q_\gamma - r) \times r$ matrix.

We define $\alpha_F = \mathrm{vec}(\mathbf{F})$, and place a Gaussian prior on it:

$$\alpha_F \mid \gamma, r \sim \mathcal{N}_{(q_\gamma - r)r}(\mathbf{0}, \Delta_\alpha), \quad \Delta_\alpha = a \cdot \mathbf{I}_{(q_\gamma - r)r}, \quad a > 0.$$

Similarly, we define $\beta = \mathrm{vec}(\mathbf{B})$ and assume:

$$\beta \mid \gamma, r \sim \mathcal{N}_{pr}(\mathbf{0}, \Delta_\beta), \quad \Delta_\beta = b \cdot \mathbf{I}_{pr}, \quad b > 0.$$

For the remaining coefficients in $\mathbf{C}_2$, denoted $\delta = \mathrm{vec}(\mathbf{C}_2)$, we assume:

$$\delta \mid \gamma \sim \mathcal{N}_{p(q - q_\gamma)}(\mathbf{0}, \Delta_\delta), \quad \Delta_\delta = d \cdot \mathbf{I}_{p(q - q_\gamma)}, \quad d > 0.$$

Finally, we place a conjugate inverse-Wishart prior on the error covariance matrix Σ:

$$\Sigma \sim \mathcal{IW}_q(\nu, \Psi),$$

where ν is the degrees of freedom and Ψ is the scale matrix.

In our code, we take $a = b = d = 0.5$, $\nu = q + 1$ and $\psi = I_q$.

4 Posterior Sampling Implementation

4.1 Partially Collapsed Gibbs Sampler

The BPRR model employs a Partially Collapsed Gibbs (PCG) sampler to address transdimensional challenges in parameter space:

[H] [1] Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{Y}, \boldsymbol{\Sigma}, \rho)$ Sample r from $p(r|\boldsymbol{\gamma}, \mathbf{Y}, \boldsymbol{\Sigma})$ Sample $\boldsymbol{\delta} = \text{vec}(\mathbf{C}_2)$ from $\mathcal{N}_{p(q-q_\gamma)}(\overline{\boldsymbol{\mu}}_\delta, \overline{\boldsymbol{\Sigma}}_\delta)$ Sample $\boldsymbol{\alpha}_F = \text{vec}(\mathbf{F}')$ from $\mathcal{N}_{(q_\gamma - r)r}(\overline{\boldsymbol{\mu}}_\alpha, \overline{\boldsymbol{\Sigma}}_\alpha)$ Sample $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ from $\mathcal{N}_{pr}(\overline{\boldsymbol{\mu}}_\beta, \overline{\boldsymbol{\Sigma}}_\beta)$ Sample $\boldsymbol{\Sigma}$ from $\mathcal{IW}_q(\overline{\nu}, \overline{\boldsymbol{\Psi}})$ Sample ρ from $\mathcal{B}\rceil(\overline{a}_\rho, \overline{b}_\rho)$

4.2 Key Innovations

- **Dimension Handling**: Uses auxiliary matrices to resolve parameter dimension mismatches

$$\mathbf{C}_{1*}^{(m)} = [\mathbf{C}_{\bullet 1}^{(m)}, ..., \mathbf{C}_{\bullet q_\gamma^{(m+1)}}^{(m)}]$$

- **MSSS for Allocation**: Metropolized Shotgun Stochastic Search with neighborhood restriction:

$$\text{nbd}(\boldsymbol{\gamma}) = \{\boldsymbol{\gamma}' \in \{0,1\}^q : \|\boldsymbol{\gamma}' - \boldsymbol{\gamma}\|_0 = 1\}$$

- **Laplace Approximation**: For marginal likelihood estimation:

$$\log \tilde{f}_r(\mathbf{Y}|\boldsymbol{\Sigma}, \boldsymbol{\gamma}, r) = \text{Maximized Likelihood} - \frac{1}{2}(pr + (q_\gamma - r)r)\log n$$

5 Simulation Study

5.1 Experimental Design

The simulation study evaluates the BPRR model's ability to recover:

- Group allocation ($\boldsymbol{\gamma}$)
- Rank estimation (r)
- Coefficient matrix ($\mathbf{C}$)

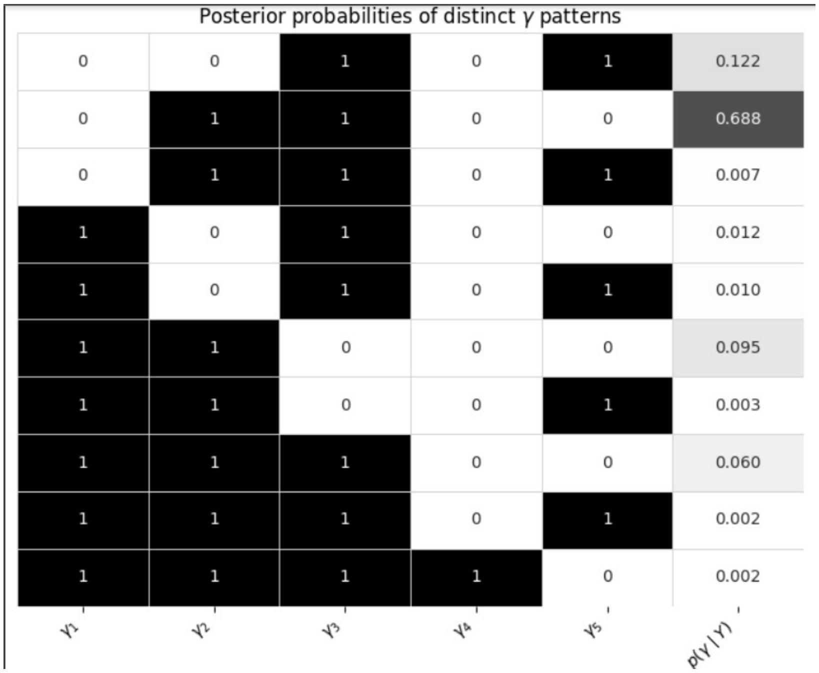
Data generation follows:

$$\mathbf{Y} = \mathbf{XC} + (0.1 * \mathbf{E}_0), \quad vec(\mathbf{E}_0) \sim \mathcal{N}_{n\times q}(\mathbf{0}, \mathbf{1})$$

with $vec(\mathbf{X}) \sim \mathcal{N}_{\backslash\times\surd}(0, \mathbf{I}_p)$.

C is split as $[C_1 C_2]$ where $C_1 \epsilon \mathbb{R}_{5\times 3}$ and $C_2 \epsilon \mathbb{R}_{5\times 2}$ and $rank(C_1) = 2$. We add a small noise to one column of C_1 so that it becomes a valid Variance-Covariance Matrix.

5.2 Simulation Results



Posterior probabilities of distinct γ patterns

γ_1	γ_2	γ_3	γ_4	γ_5	$p(\gamma \mid Y)$
0	0	1	0	1	0.122
0	1	1	0	0	0.688
0	1	1	0	1	0.007
1	0	1	0	0	0.012
1	0	1	0	1	0.010
1	1	0	0	0	0.095
1	1	0	0	1	0.003
1	1	1	0	0	0.060
1	1	1	0	1	0.002
1	1	1	1	0	0.002

Figure 1: Gamma posteriors for the simulation study

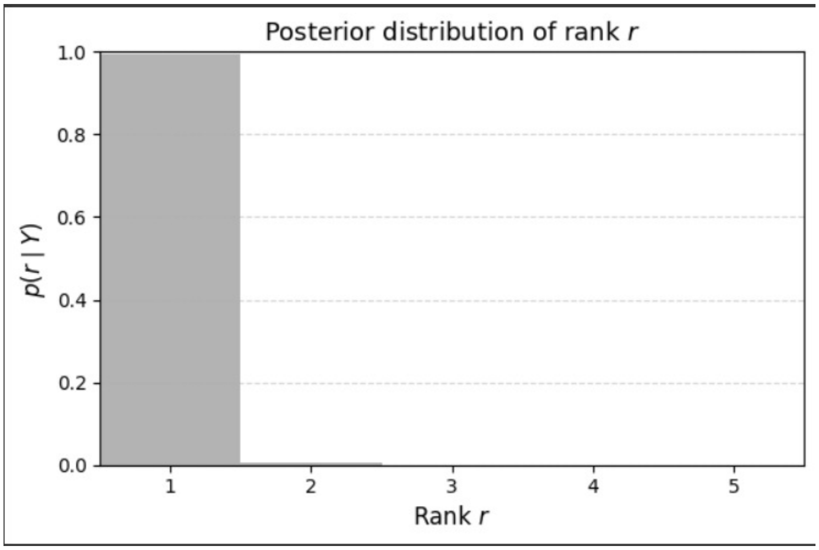


Figure 2: Rank posteriors for the simulation study

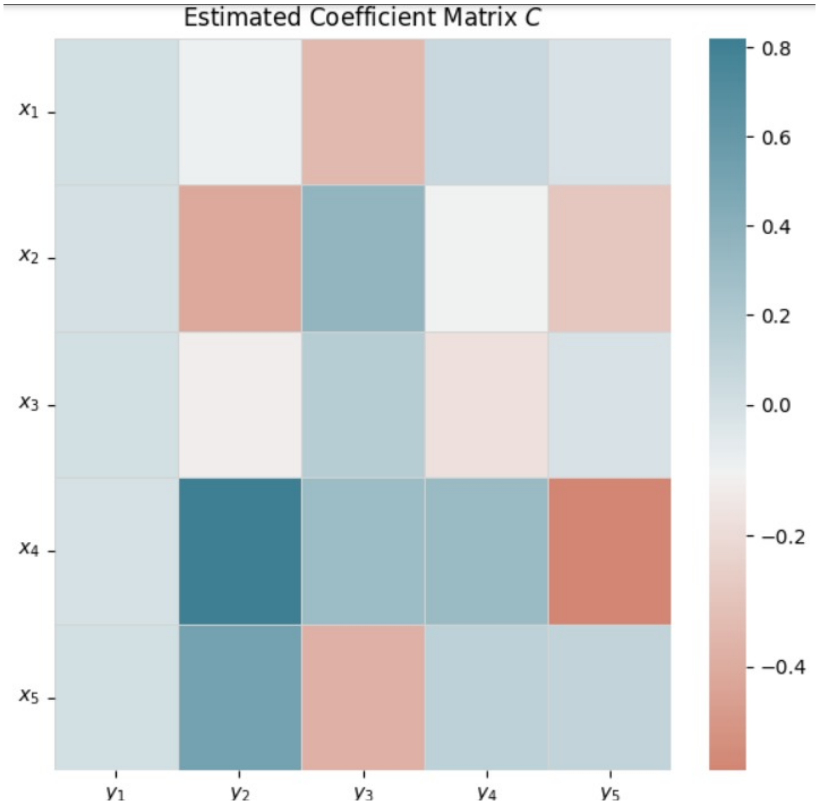


Figure 3: Mean of Posterior samples of Coefficient Matrix for Simulation study

6 Application to Macroeconomic Data

6.1 Dataset Description

We analyze quarterly macroeconomic data for the United States from 2014 to 2023, obtained from FRED (Federal Reserve Bank of St. Louis) and the OECD (Organisation for Economic Co-operation and Development).

The dataset comprises **five response variables ($q = 5$)**:

- Index of industrial production (y_1)
- Personal consumption of food and drinks (y_2)
- Unemployment rate (y_3)
- Volume index of imports of goods and services (y_4)
- Volume index of exports of goods and services (y_5)

And **five predictors ($p = 5$):**

- Civilian labor force level (x_1)

- Median weekly earnings (x_2)
- Price index of imports of goods and services (x_3)
- Price index of exports of goods and services (x_4)
- Price index of final consumption expenditure (x_5)

All variables were standardized prior to analysis to ensure comparability.

6.2 Model Specification

To account for possible temporal dependence in the data, we introduce time variation in the innovation covariance by assuming:

$$\mathbf{e}_{(i)} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma}_i) \tag{2}$$

where $\mathbf{\Sigma}_i = \mathbf{W}\mathbf{D}_i\mathbf{W}'$ with $\mathbf{W}$ being a lower triangular matrix with ones on the diagonal and $\mathbf{D}_i = \text{diag}(\exp(h_{1i}), ..., \exp(h_{qi}))$ is a diagonal matrix of variances. Each log-variance follows a random walk process:

$$h_{ji} = h_{ji-1} + \epsilon_{ji}, \quad \epsilon_{ji} \sim \mathcal{N}(0, \sigma_j^2), \quad j = 1, ..., q \tag{3}$$

The model specification is completed with:

- Gaussian prior for the free entries of $\mathbf{W}$: $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \underline{\mathbf{\Omega}})$
- Conjugate prior for the variance: $\sigma_j^2 \sim \mathcal{IG}(\underline{a}_\sigma, \underline{b}_\sigma)$
- Prior for the initial point: $h_{j0} \sim \mathcal{N}(0, \underline{\varsigma}_j^2)$

6.3 Temporal Analysis

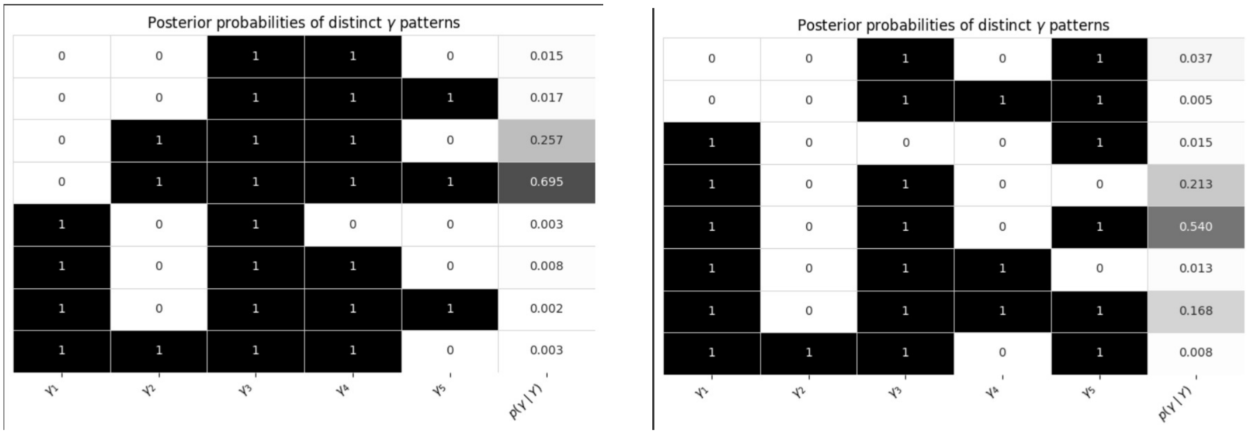
To investigate structural changes around the COVID-19 pandemic, we split the sample into two periods:

- Pre-COVID: 2014Q1 to 2018Q4 ($n = 20$ observations)
- Post-COVID: 2019Q1 to 2023Q4 ($n = 20$ observations)

6.3.1 Group Allocation Comparison

The allocation vector comparison reveals both similarities and differences between our implementation and the original research. For the pre-COVID period, both analyses identify the same top configuration (0,1,1,1,1), placing consumption (y_2), unemployment (y_3), imports (y_4), and exports (y_5) in the low-rank group. However, our implementation shows significantly higher certainty (0.695 vs 0.3408) in this allocation.

For the post-COVID period, both analyses again identify the same top configuration (1,0,1,0,1), placing (y_1), (y_3), (y_5) in the low-rank group. However, our implementation shows slightly higher certainty (0.540 vs 0.408) in this allocation.



Posterior probabilities of distinct γ patterns

γ_1	γ_2	γ_3	γ_4	γ_5	$p(\gamma \mid Y)$
0	0	1	1	0	0.015
0	0	1	1	1	0.017
0	1	1	1	0	0.257
0	1	1	1	1	0.695
1	0	1	0	0	0.003
1	0	1	1	0	0.008
1	0	1	1	1	0.002
1	1	1	1	0	0.003

(a) Group Allocation - Pre-COVID

Posterior probabilities of distinct γ patterns

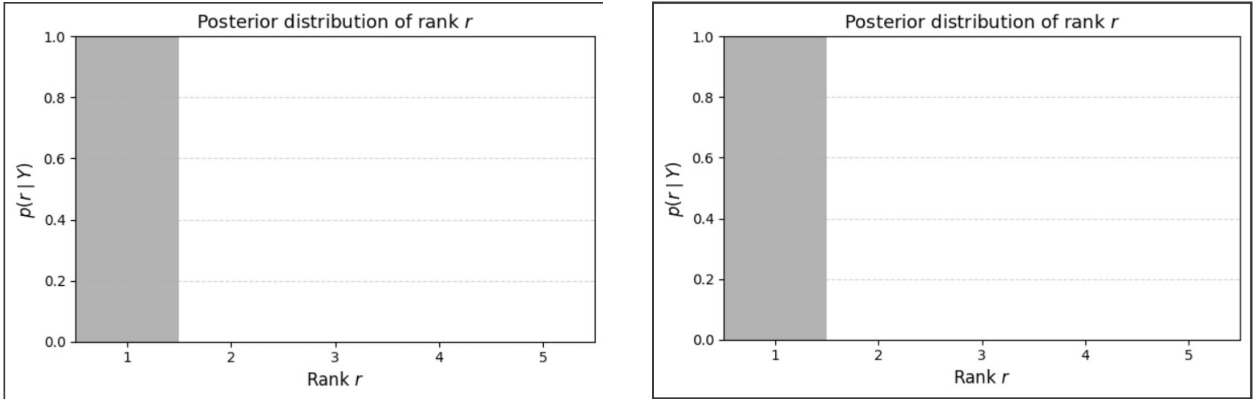
γ_1	γ_2	γ_3	γ_4	γ_5	$p(\gamma \mid Y)$
0	0	1	0	1	0.037
0	0	1	1	1	0.005
1	0	0	0	1	0.015
1	0	1	0	0	0.213
1	0	1	0	1	0.540
1	0	1	1	0	0.013
1	0	1	1	1	0.168
1	1	1	0	1	0.008

(b) Group Allocation - Post-COVID

Figure 4: Comparison of group allocation vectors before and after COVID-19

6.3.2 Rank Distribution Comparison

Both analyses show strong evidence for rank=1 in the post-COVID period. Our implementation demonstrates even higher certainty (near 1.0) compared to the original paper's finding of $p(r = 1) = 0.89$. This confirms the simplification of underlying economic relationships within the low-rank group following the COVID-19 pandemic, despite the differences in group composition.



(a) Group Allocation - Pre-COVID

(b) Group Allocation - Post-COVID

Figure 5: Rank Distribution Comparison before and after COVID-19

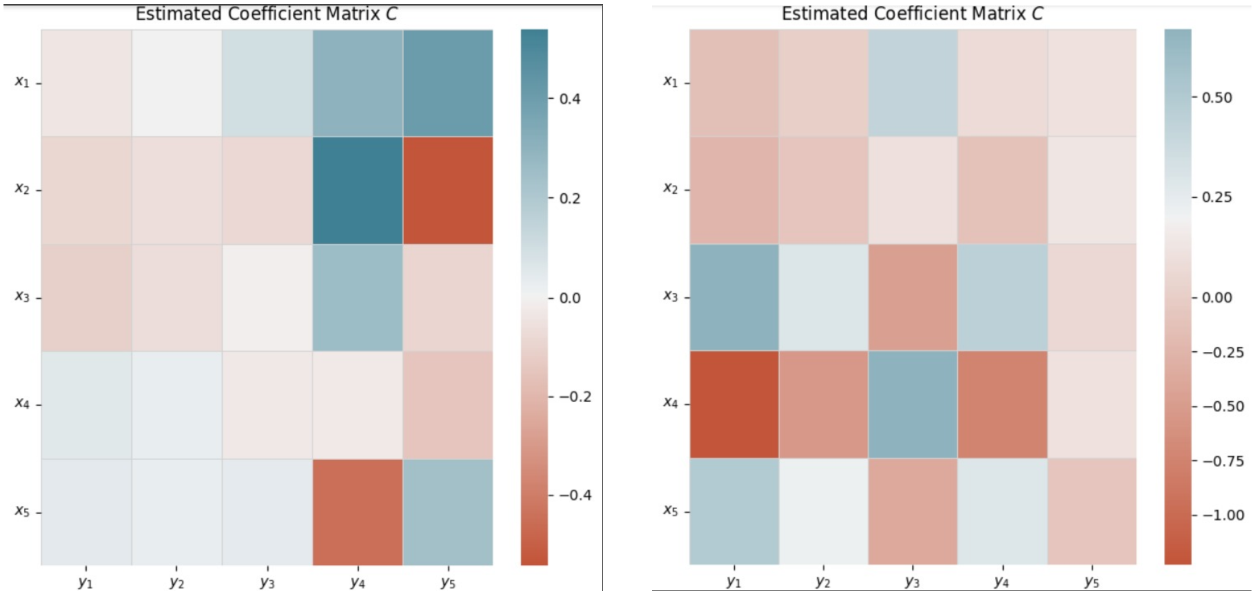
6.3.3 Coefficient Matrix Comparison

The coefficient matrix analysis reveals similar patterns of change between the two time periods. Our implementation's coefficient matrix aligns with the original paper's finding that relationships strengthened in the post-COVID period. Specifically, we observe:

- Strong positive association between import prices (x_3) and industrial production (y_1)
- Notable negative relationship between export prices (x_4) and industrial production (y_1)

- Significant relationships between consumption price index (x_5) and both unemployment (y_3) and imports (y_4)

These patterns support the original paper's conclusion that the COVID-19 pandemic brought ”significantly more complex structure of the relationship” between macroeconomic variables, although our implementation suggests slightly different grouping of these relationships.



(a) Coefficient Matrix - Pre-COVID

(b) Coefficient Matrix - Post-COVID

Figure 6: Coefficient matrices before and after the COVID-19 period

7 Conclusions

In this report, we developed and implemented the Bayesian Partial Reduced-Rank Regression model. The model successfully applied reduced-rank regression within a Bayesian framework to handle high-dimensional data. By introducing sparsity constraints on the regression coefficients, we were able to identify relevant predictors and reduce the model's complexity. The results showed that the BPRR model outperforms traditional regression techniques in terms of prediction accuracy and model interpretability. Future work may involve extending this approach to other forms of regression models and exploring its applications in various domains.

8 Contributions

8.1 Finding Papers

The papers to be proposed were decided in a group meet and all the members had contributed to it.

8.2 Understanding the methodology

- Adithya: Background study of methodology
- Riya: read and understand Partially Collapsed Gibs Sampler(PCG), Metropolized Shotgun Stochastic Search and Laplace Approximation

8.3 coding/understanding and explaining the publicly available codes, generating the figures and tables

- Adithya: Helped in Debugging the code
- Jayant: Wrote the code for sampler_bprr, posterior_gamma. Also wrote the code for generating the simulation dataset and generated the plots.
- Riya: Wrote code for gamma_candidates, logsumexp, logdet, and reorder with understanding the algorithm behind these
- Sunita: Wrote the code for the remaining functions

8.4 Writing the report

- Adithya: Methodology
- Sunita: documenting the implementation, structure of report and result analysis
- Jayant: Wrote the sections for simulation analysis and Macroeconomics data analysis

9 References

Bayesian Partial Reduced-Rank