

Bayesian Partial Reduced-Rank Regression

Maria F. Pintado, Matteo Iacopini, Luca Rossini & Alexander Y. Shestopaloff

To cite this article: Maria F. Pintado, Matteo Iacopini, Luca Rossini & Alexander Y. Shestopaloff (2025) Bayesian Partial Reduced-Rank Regression, Journal of Computational and Graphical Statistics, 34:4, 1344-1355, DOI: [10.1080/10618600.2024.2446357](https://doi.org/10.1080/10618600.2024.2446357)

To link to this article: <https://doi.org/10.1080/10618600.2024.2446357>



© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 04 Mar 2025.



[Submit your article to this journal](#)



Article views: 1593



[View related articles](#)



[View Crossmark data](#)

Bayesian Partial Reduced-Rank Regression

Maria F. Pintado^a, Matteo Iacopini^b, Luca Rossini^{c,d}, and Alexander Y. Shestopaloff^{a,e}

^aSchool of Mathematical Sciences, Queen Mary University of London, London, UK; ^bDepartment of Economics and Finance, LUISS University, Rome, Italy; ^cDepartment of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy; ^dFondazione Eni Enrico Mattei, Milan, Italy; ^eDepartment of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Canada

ABSTRACT

Reduced-rank (RR) regression may be interpreted as a dimensionality reduction technique able to reveal complex relationships among the data parsimoniously. However, RR regression models typically overlook any potential group structure among the responses by assuming a low-rank structure on the coefficient matrix. To address this limitation, a Bayesian Partial RR (BPRR) regression is exploited, where the response vector and the coefficient matrix are partitioned into low- and full-rank sub-groups. As opposed to the literature, which assumes known group structure and rank, a novel strategy is introduced that treats them as unknown parameters to be estimated.

The main contribution is 2-fold: an approach to infer the low- and full-rank group memberships from the data is proposed, and then, conditionally on this allocation, the corresponding (reduced) rank is estimated. Both steps are carried out in a Bayesian approach, allowing for full uncertainty quantification and based on a partially collapsed Gibbs sampler. It relies on a Laplace approximation of the marginal likelihood and the Metropolized Shotgun Stochastic Search to estimate the group allocation efficiently. Applications to synthetic and real-world data reveal the potential of the proposed method to reveal hidden structures in the data. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received July 2024
Accepted December 2024

KEYWORDS

Group learning; Laplace approximation; Rank estimation; Uncertainty quantification

1. Introduction



Intrinsic group structures are prevalent in certain data types, particularly in those involving complex information. These structures signify the presence of correlations among variables within these groups, and ignoring them can lead to an inefficient use of the available data. In the context of multivariate response-predictor analysis, a commonly adopted strategy is to perform covariate selection for each response variable, which addresses the predictor group structure (see Buch et al. 2023, for a review of such methods). However, the group structure among response variables is typically overlooked.


The reduced-rank (RR) regression model (Anderson 1951; Izenman 1975; Reinsel, Velu, and Chen 2022) offers a more natural means of handling block structures and dimension reduction, achieved by imposing a lower-rank constraint on the matrix of coefficients, C . This assumption translates into having a smaller number of relevant linear combinations of the predictor variables as latent factors that explain the variation in all the response variables. Different variants of the reduced-rank regression model have been explored in the literature. For instance, Anderson (1951) examines a partitioned coefficient matrix associated with a low-rank group and a full-rank group in the covariates using a predefined grouping. This result is further elaborated upon by Velu (1991), who extends it into two sets of regressors characterized by low-rank structures.

Recently, Li, Liu, and Chen (2018) investigated an integrative reduced-rank regression model for analysing multi-view data, where each view consists of several predictors and has its own low-rank coefficient matrix. On a different direction, Chen and Huang (2012) proposed a sparse reduced-rank regression that introduces row-wise sparsity in C , enabling predictors with no association to latent factors. Kim and Jung (2024) combined the latter two approaches in a reduced-rank regression setting with multi-source data, where each predictor set is associated with a low-rank coefficient matrix while simultaneously allowing for sparsity in both covariates and responses.

We explore an alternative generalization of the standard RR regression model, where the reduced-rank coefficient structure applies to only an unknown subset of the response variables. In contrast, the remaining subset maintains a full-rank coefficient submatrix. This more flexible approach has been called partially reduced-rank (PRR) regression by Reinsel and Velu (2006). In this scenario, the set of response variables is divided into two subsets Y_1 and Y_2 , and the reduced-rank structure is imposed on a submatrix C_1 of C , driving the relationship between Y_1 and the covariates, X . The reduced-rank assumption on C_1 implies that the regression of Y_1 on X is influenced by only a limited number of predictive variables constructed as linear combinations of X .

Considering group structures within the response variables enhances our comprehension of the data in diverse fields such as macroeconomics (Reinsel and Velu 2006) and genetics (Li,

CONTACT Maria F. Pintado  m.f.pintadoserrano@qmul.ac.uk  School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Nan, and Zhu 2015; Luo and Chen 2020). The PRR model, incorporating the proposed response groups, has a potential utility in multi-view data scenarios that consider two views in the responses, thereby enhancing model fit through the specialized structure in \mathbf{C} . Additionally, PRR adds flexibility to the model in accommodating complex relationships observed in real-world datasets, such as those found in economic contexts.

To the best of our knowledge, the literature on PRR regression summarized in Reinsel and Velu (2006) has always relied on an *a priori fixed* grouping structure, justified by application-specific theoretical considerations or the researcher’s intuition. We argue that this represents a possible reason hindering the large applicability of PRR regression in several real-world problems. To overcome these issues, we propose the first Bayesian approach to PRR regression, where the low-rank and full-rank response groupings are unknown and directly inferred from the data. This approach considers an agnostic position about the optimal allocation of response variables and opens the possibility of using PRR models even in the absence of strong and reliable information to dictate the grouping. Moreover, we design and implement a partially collapsed Gibbs sampler that, at each iteration, first samples both the grouping structure and the (reduced) rank relying on the Laplace method, then the remaining parameters in subsequent basic steps. We call the proposed method, Bayesian partial reduced-rank (BPRR) regression and we compare its performance against well-known specifications available in the literature.

A simulation study underscores the strong performance of our proposed model across various scenarios. It yields adequate estimates of the low-rank group allocation and rank and achieves a lower error compared to benchmark methods in linear regression such as the pre-specified PRR or the full-rank model. The effectiveness of the model is further demonstrated in an application to macroeconomic data from the United States, which provides evidence of a significant shift in the response grouping (and rank) following the COVID-19 pandemic, both in terms of point estimates and the associated uncertainty.

The remainder of the article is as follows. Section 2 briefly presents the PRR framework and describes the structure of the prior distributions. Then, Section 3 describes in detail the challenges encountered in the design of the algorithm to perform posterior sampling together with the proposed solutions. The performance of the algorithm is tested on synthetic data in Section 4, and then it is applied to a real-world dataset about macroeconomics in Section 5. Section 6 draws the conclusions.

2. Partial Reduced-Rank Regression Model

Let $\mathbf{Y} \in \mathbb{R}^{n \times q}$ be the matrix of responses with the i th row denoted as $\mathbf{y}'_{(i)}$, $\mathbf{X} \in \mathbb{R}^{n \times p}$ the matrix of explanatory variables with the i th row as $\mathbf{x}'_{(i)}$, and $\mathbf{E} \in \mathbb{R}^{n \times q}$ the matrix of innovation terms with $\mathbf{e}_{(i)}$ as the noise vector associated to the i th observation. The multivariate linear regression model is defined as

$$\mathbf{Y} = \mathbf{X}\mathbf{C} + \mathbf{E}, \quad \mathbf{E} = (\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(n)})',$$

$$\mathbf{e}_{(i)} \stackrel{iid}{\sim} \mathcal{N}_q(\mathbf{0}, \Sigma). \tag{1}$$

We assume that the response variables can be split into two different groups \mathbf{Y}_1 and \mathbf{Y}_2 of dimensions $n \times q_\gamma$ and $n \times (q - q_\gamma)$, respectively, where $q_\gamma \in \{2, \dots, q - 1\}$. Moreover, we assume that the relationship between \mathbf{Y}_1 and \mathbf{X} admits a low-rank structure, while the regression of \mathbf{Y}_2 on \mathbf{X} has full-rank. Under this assumption, the coefficient matrix $\mathbf{C} \in \mathbb{R}^{p \times q}$ can be partitioned as $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$, with $\mathbf{C}_1 \in \mathbb{R}^{p \times q_\gamma}$ having reduced rank $r = \text{rank}(\mathbf{C}_1) \leq \min(p, q_\gamma) - 1$ and $\mathbf{C}_2 \in \mathbb{R}^{p \times (q - q_\gamma)}$ with full rank $r_2 = \text{rank}(\mathbf{C}_2) = \min(p, q - q_\gamma)$.

Therefore, the model in (1) can be represented with partitioned matrices as

$$[\mathbf{Y}_1, \mathbf{Y}_2] = \mathbf{X}[\mathbf{C}_1, \mathbf{C}_2] + [\mathbf{E}_1, \mathbf{E}_2]. \tag{2}$$

Notice that each of the n response vectors $\mathbf{y}_{(i)}$, $i = 1, \dots, n$, is of the form $\mathbf{y}_{(i)} = (y_{i,1}, \dots, y_{i,q_\gamma}, y_{i,q_\gamma+1}, \dots, y_{i,q})' \in \mathbb{R}^q$. We write $\mathbf{e}_{(i)} = (\mathbf{e}'_{1i}, \mathbf{e}'_{2i})'$ with $\mathbf{e}_{1i} = (e_{i,1}, \dots, e_{i,q_\gamma})'$ and $\mathbf{e}_{2i} = (e_{i,q_\gamma+1}, \dots, e_{i,q})'$, and assume $\mathbf{e}_{(i)}$ is normally distributed with mean zero and a partitioned covariance matrix $\Sigma = \text{cov}(\mathbf{e}_{(i)})$ such that $\Sigma_{11} = \text{cov}(\mathbf{e}_{1i})$, $\Sigma_{22} = \text{cov}(\mathbf{e}_{2i})$, and $\Sigma_{12} = \text{cov}(\mathbf{e}_{1i}, \mathbf{e}_{2i})$.

2.1. Prior Specifications

The model in (2) classifies the response variable into two groups. Differently from Reinsel and Velu (2006), we assume the grouping structure to be unknown and aim at inferring it from the data. Therefore, 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 , $j = 1, \dots, q$, 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

$$p(\boldsymbol{\gamma}|\rho) = \left[\prod_{j=1}^q \text{Bern}(\gamma_j|\rho) \right] \mathbb{I}(1 < q_\gamma < q), \tag{3}$$

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. The constraint imposed by the indicator function in (3) allows for the existence of the low-rank group and, thus, of a PRR model. In fact, if $q_\gamma = 1$, \mathbf{Y}_1 comprises only one response, making \mathbf{C} full-rank. Conversely, when $q_\gamma = q$, all responses are part of the low-rank group, which collapses into the standard RR model. Additionally, we employ a hierarchical prior structure, where ρ is assigned a Beta prior distribution, $\rho \sim \text{Be}(\rho|\underline{a}_\rho, \underline{b}_\rho)$.

Remark 1. Reinsel and Velu (2006) fix the grouping structure in the estimation, then propose a stepwise backwards-elimination procedure to identify the optimal one. This approach is based on a sequence of likelihood ratio tests performed after estimating models with different grouping structures. Conversely, our model offers a unified framework where model selection and parameter estimation are performed jointly. Moreover, the posterior distribution for the allocation vector $\boldsymbol{\gamma}$ allows us to compute a point estimate of the grouping structure and to quantify the uncertainty around it, which is not obtainable from the test-based procedure in Reinsel and Velu (2006).

Moving to the coefficients matrix, \mathbf{C} , we can provide a specification for both the low and the full rank matrix. In particular, the submatrix of coefficients \mathbf{C}_1 is assumed to have reduced rank $r \leq r_{\max} = \min(p, q_Y) - 1$, which depends on the binary parameter $\boldsymbol{\gamma}$. Therefore, conditional on q_Y (hence, on $\boldsymbol{\gamma}$), we assume an uninformative uniform prior distribution for r over the discrete set $\{1, \dots, r_{\max}\}$, that is $r|\boldsymbol{\gamma} \sim \mathcal{U}(r|\{1, \dots, r_{\max}\})$.

Given that \mathbf{C}_1 is a low-rank matrix, we can express it as the product of two full-rank matrices $\mathbf{A} \in \mathbb{R}^{q_Y \times r}$ and $\mathbf{B} \in \mathbb{R}^{p \times r}$, such that $\mathbf{C}_1 = \mathbf{B}\mathbf{A}'$. This decomposition is not unique, since for any orthogonal $r \times r$ matrix \mathbf{P} , we have that $\mathbf{C}_1 = (\mathbf{B}\mathbf{P})(\mathbf{P}'\mathbf{A}')$. To achieve a unique decomposition of \mathbf{C}_1 , we follow Geweke (1996) and impose an identifying restriction by assuming the first r rows of \mathbf{A} are the identity matrix \mathbf{I}_r , that is

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{F} \end{bmatrix}, \tag{4}$$

where \mathbf{F} is a $(q_Y - r) \times r$. Denoting with $\text{vec}(\cdot)$ the vectorization operator, we assume a multivariate Gaussian prior distribution on $\boldsymbol{\alpha}_F = \text{vec}(\mathbf{F}')$, that is

$$\boldsymbol{\alpha}_F|\boldsymbol{\gamma}, r \sim \mathcal{N}_{(q_Y-r)r}(\boldsymbol{\alpha}_F|\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\alpha), \tag{5}$$

where $\underline{\boldsymbol{\Sigma}}_\alpha = a \mathbf{I}_{(q_Y-r)r}$, for fixed $a > 0$. Similarly, defining $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ and $\boldsymbol{\delta} = \text{vec}(\mathbf{C}_2)$, we assume the multivariate Gaussian prior distributions:

$$\boldsymbol{\beta}|\boldsymbol{\gamma}, r \sim \mathcal{N}_{pr}(\boldsymbol{\beta}|\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\beta), \tag{6}$$

$$\boldsymbol{\delta}|\boldsymbol{\gamma} \sim \mathcal{N}_{p(q-q_Y)}(\boldsymbol{\delta}|\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\delta), \tag{7}$$

where $\underline{\boldsymbol{\Sigma}}_\beta = b \mathbf{I}_{pr}$, and $\underline{\boldsymbol{\Sigma}}_\delta = d \mathbf{I}_{p(q-q_Y)}$, for fixed $b, d > 0$. Lastly, we adhere to the conventional practice and assign a conjugate inverse Wishart prior to Σ , that is $\Sigma \sim \mathcal{IW}_q(\Sigma|\underline{\nu}, \underline{\Psi})$, with $\underline{\nu}$ and $\underline{\Psi}$ being the fixed degrees of freedom and scale matrix.

3. Posterior Sampling

In this section, we design an MCMC algorithm to draw samples from the joint posterior distribution $p(\mathbf{A}, \mathbf{B}, \mathbf{C}_2, \Sigma, r, \boldsymbol{\gamma}, \rho|\mathbf{Y})$. The most critical issue to tackle is that the dimensions of the matrices \mathbf{A} , \mathbf{B} , and \mathbf{C}_2 depend on the states of $\boldsymbol{\gamma}$ and r , which implies that the dimension of the parameter space may change across the iterations of the MCMC algorithm. Consequently, the traditional Gibbs sampler is invalid in this setting, whereas a reversible jump MCMC, although theoretically feasible, poses significant challenges in terms of implementation and proper execution (Robert and Casella 1999).¹ To address this challenge,

¹Implementing a reversible jump MCMC in our setting would require the definition of a cross-model move involving the allocation vector $\boldsymbol{\gamma}$ and the rank r since both have an impact on the size of the parameter space. This move requires defining (a) the mapping function and (b) the proposal distribution for an auxiliary vector, w , which are highly arbitrary. The mapping function expresses a functional relationship between the parameters of two different models (the current and the proposed one), implying that a good choice for this function is likely to improve the sampler's performance in terms of between-model acceptance rates and chain mixing. However, the major difficulty is that good relationships can be hard to define even in the simplest model settings. In general cases like our framework, where two models are characterized by different binary vector $\boldsymbol{\gamma}$ and/or integer variable r , parameters between models may not be obviously comparable, thus, making the definition of a relationship between them even harder to define. Finally, there are no easy-to-use criteria for choosing the proposal distributions for the auxiliary vectors w .

Algorithm 1 PCG for Bayesian PRR model

- 1: Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{Y}, \Sigma, \rho)$.
 - 2: Sample r from $p(r|\boldsymbol{\gamma}, \mathbf{Y}, \Sigma)$.
 - 3: Sample $\boldsymbol{\delta} = \text{vec}(\mathbf{C}_2)$ from $p(\boldsymbol{\delta}|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}, \mathbf{A}, \mathbf{B}) = \mathcal{N}_{p(q-q_Y)}(\bar{\boldsymbol{\mu}}_\delta, \bar{\boldsymbol{\Sigma}}_\delta)$.
 - 4: Sample $\boldsymbol{\alpha}_F = \text{vec}(\mathbf{F}')$ from $p(\boldsymbol{\alpha}_F|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}, r, \mathbf{B}, \mathbf{C}_2) = \mathcal{N}_{(q_Y-r)r}(\bar{\boldsymbol{\mu}}_\alpha, \bar{\boldsymbol{\Sigma}}_\alpha)$, then set $\mathbf{A} = [\mathbf{I}_r, \mathbf{F}']'$.
 - 5: Sample $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ from $p(\boldsymbol{\beta}|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}, r, \mathbf{A}, \mathbf{C}_2) = \mathcal{N}_{pr}(\bar{\boldsymbol{\mu}}_\beta, \bar{\boldsymbol{\Sigma}}_\beta)$.
 - 6: Sample Σ from $p(\Sigma|\mathbf{Y}, \boldsymbol{\gamma}, \mathbf{A}, \mathbf{B}, \mathbf{C}_2) = \mathcal{IW}_q(\bar{\nu}, \bar{\Psi})$.
 - 7: Sample ρ from $p(\rho|\boldsymbol{\gamma}) = \mathcal{Be}(\bar{a}_\rho, \bar{b}_\rho)$.
-

we implement a partially collapsed Gibbs sampler (PCG, see van Dyk and Park 2008), which generalizes the composition of the conditional distributions in Gibbs samplers, relying on three basic tools: marginalization, permutation, and trimming. Specifically, we avoid the need for transdimensional samplers by drawing $(\boldsymbol{\gamma}, r)$ from a joint distribution marginalized over the parameters $(\mathbf{A}, \mathbf{B}, \mathbf{C}_2)$ whose size depends on $(\boldsymbol{\gamma}, r)$. Subsequently, we sample $(\mathbf{A}, \mathbf{B}, \mathbf{C}_2)$ conditionally on the updated values of $(\boldsymbol{\gamma}, r)$. The entire sampling process is summarized in Algorithm 1.

In the remainder of this section, we describe the procedures adopted to integrate out $(\mathbf{A}, \mathbf{B}, \mathbf{C}_2)$ from the likelihood, then we explain the main computational challenges and the proposed solutions. The first problem arises in Step 1, the most computational intensive step, where sampling $\boldsymbol{\gamma}$ entails the exploration of a 2^q dimensional parameter space and the computation of analytically intractable integrals, a limitation also encountered in Step 2. The second issue concerns the dimensions of matrices \mathbf{A} , \mathbf{B} , and \mathbf{C}_2 in practically implementing Steps 3 and 4, since their size depends on the states of $(\boldsymbol{\gamma}, r)$, producing a potential incompatibility between the dimension of the parameters at the previous MCMC iteration and the current one corresponding to the updated $(\boldsymbol{\gamma}, r)$.

Remark 2. The proposed BPRR model can be extended to the case of non-Gaussian noise while keeping its main features. In particular, the BPRR can be applied in the generalized linear model (GLM) framework by choosing an appropriate link function to responses from the natural exponential family. In several cases, it is possible to design a suitable data augmentation scheme to obtain a full conditional posterior for the coefficients of the same type as the Gaussian BPRR model (i.e., preserving the conjugacy of the prior). For example, regression models using a binomial likelihood can leverage the Pólya-gamma data augmentation of Polson, Scott, and Windle (2013) to obtain conditional conjugacy for the coefficients of the (latent) linear regression. Instead, for Poisson and Gamma likelihoods, the full conditional distribution for the coefficients is not conjugate, thus, requiring the use of Metropolis-Hastings algorithms.

3.1. Preliminaries

Let us define the likelihood function as

$$f(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \mathbf{C}_2, \Sigma, r, \boldsymbol{\gamma}) = \frac{1}{(2\pi)^{nq/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \|\mathbf{Y} - \mathbf{X}\mathbf{C}\Sigma^{-1/2}\|_F^2 \right\}, \quad (8)$$

where $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$, $\mathbf{C}_1 = \mathbf{B}\mathbf{A}'$, and $\|\cdot\|_F$ denotes the Frobenius norm. The first step involves explicitly expressing the likelihood in terms of \mathbf{C}_1 and \mathbf{C}_2 . Therefore, we exploit the partitioning of \mathbf{C} and rewrite the model in (2) equivalently as

$$\mathbf{Y} = \mathbf{X}\mathbf{C}_1\mathbf{V}_1 + \mathbf{X}\mathbf{C}_2\mathbf{V}_2 + \mathbf{E}, \quad (9)$$

where $\mathbf{V}_1 = [\mathbf{I}_{q_\gamma}, \mathbf{0}_{q_\gamma \times (q-q_\gamma)}]$, and $\mathbf{V}_2 = [\mathbf{0}_{(q-q_\gamma) \times q_\gamma}, \mathbf{I}_{q-q_\gamma}]$. By vectorizing (9), we obtain

$$\mathbf{y} = \mathbf{U}_1\mathbf{c}_1 + \mathbf{U}_2\boldsymbol{\delta} + \mathbf{e}, \quad (10)$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{c}_1 = \text{vec}(\mathbf{C}_1)$, $\boldsymbol{\delta} = \text{vec}(\mathbf{C}_2)$, $\mathbf{e} = \text{vec}(\mathbf{E})$, and $\mathbf{U}_i = \mathbf{V}_i' \otimes \mathbf{X}$, for each $i = 1, 2$. It follows that $\mathbf{y}|\mathbf{A}, \mathbf{B}, \mathbf{C}_2, \Sigma, r, \boldsymbol{\gamma} \sim \mathcal{N}_{nq}(\mathbf{y}|\mathbf{U}_1\mathbf{c}_1 + \mathbf{U}_2\boldsymbol{\delta}, \tilde{\Sigma})$, where $\tilde{\Sigma} = \Sigma \otimes \mathbf{I}_n$.

Under the vectorized model in (10), the likelihood can be marginalized over \mathbf{C}_2 analytically to obtain

$$\begin{aligned} f(\mathbf{y}|\mathbf{A}, \mathbf{B}, \Sigma, \boldsymbol{\gamma}, r) &= \int f(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \mathbf{C}_2, \Sigma, \boldsymbol{\gamma}, r) p(\mathbf{C}_2|\boldsymbol{\gamma}) d\mathbf{C}_2 \\ &= \int \mathcal{N}_{nq}(\mathbf{y}|\mathbf{U}_1\mathbf{c}_1 + \mathbf{U}_2\boldsymbol{\delta}, \tilde{\Sigma}) \\ &\quad \times \mathcal{N}_{p(q-q_\gamma)}(\boldsymbol{\delta}|\mathbf{0}, \underline{\Sigma}_\delta) d\boldsymbol{\delta} \\ &= \mathcal{N}_{nq}(\mathbf{y}|\mathbf{U}_1\mathbf{c}_1, \tilde{\Sigma} + \mathbf{U}_2\underline{\Sigma}_\delta\mathbf{U}_2'). \end{aligned} \quad (11)$$

This distribution represents the starting point in the design of the proposed PCG sampler.

3.2. Sampling the Response Allocation, $\boldsymbol{\gamma}$, and Rank, r

Starting from (11), we are left with the task of marginalizing \mathbf{A} and \mathbf{B} . As analytical integration is unfeasible, we obtain an approximation to the (marginal) posterior of $\boldsymbol{\gamma}$ via the Laplace method, which provides a trade-off between computational speed and accuracy (e.g., see Tierney and Kadane 1986; Tierney, Kass, and Kadane 1989; Kass and Raftery 1995, among others). Then, a sample from the approximate posterior is obtained through the Metropolized Shotgun Stochastic Search (MSSS) algorithm (Hans, Dobra, and West 2007). Our strategy is similar in spirit to Yang, Goh, and Wang (2022), which is concerned with rank estimation in a RR model.

The posterior of $\boldsymbol{\gamma}$ given by Bayes' theorem after integrating out $\mathbf{A}, \mathbf{B}, \mathbf{C}_2$, and r is

$$p(\boldsymbol{\gamma}|\mathbf{Y}, \Sigma, \rho) = \frac{f_\gamma(\mathbf{Y}|\Sigma, \boldsymbol{\gamma})p(\boldsymbol{\gamma}|\rho)}{\sum_{\boldsymbol{\gamma}^\dagger \in (0,1)^q} f_\gamma(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}^\dagger)p(\boldsymbol{\gamma}^\dagger|\rho)}, \quad (12)$$

where $f_\gamma(\mathbf{Y}|\Sigma, \boldsymbol{\gamma})$ is obtained from (13) by marginalizing over \mathbf{A}, \mathbf{B} , and r , that is

$$\begin{aligned} f_\gamma(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}) &= \sum_{r=1}^{r_{\max}} \frac{1}{r_{\max}} \iint f(\mathbf{Y}|\mathbf{A}, \mathbf{B}, \Sigma, \boldsymbol{\gamma}, r) p(\mathbf{A}, \mathbf{B}|r, \boldsymbol{\gamma}) d\mathbf{A} d\mathbf{B} \\ &= \sum_{r=1}^{r_{\max}} \frac{1}{r_{\max}} f_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r). \end{aligned} \quad (13)$$

Note that the integration with respect to r is performed analytically since the latter is a discrete parameter with finite support. Conversely, we use the Laplace method (Raftery 1995) to approximate the analytically intractable integration of \mathbf{A}, \mathbf{B} to obtain

$$\begin{aligned} \log f_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r) &\approx \log f(\mathbf{Y}|\hat{\mathbf{A}}, \hat{\mathbf{B}}, \Sigma, \boldsymbol{\gamma}, r) \\ &\quad - \frac{1}{2}(pr + (q_\gamma - r)r) \log n, \end{aligned} \quad (14)$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are the maximum likelihood estimators (MLEs) of \mathbf{A} and \mathbf{B} , given r and $\boldsymbol{\gamma}$. The main difficulty relies on the computation of these MLEs, although Reinsel, Velu, and Chen (2022) make available these estimators for iid Gaussian errors, our setting differs from this baseline in two main aspects. First, the Gaussian density in (11) has a covariance matrix $\Sigma_\mathbf{y} = \tilde{\Sigma} + \mathbf{U}_2\underline{\Sigma}_\delta\mathbf{U}_2'$ that incorporates heteroscedastic errors through the dependency of \mathbf{U}_2 on \mathbf{V}_2 . As a consequence, the block of the covariance matrix corresponding to \mathbf{C}_2 introduces a different variance for each observation. Second, we are imposing an identification restriction on the matrix \mathbf{A} and an additional restriction on the vectorized linear model's coefficient via the binary matrix \mathbf{V}_1 .

The first restriction requires the first r rows of \mathbf{A} to be the identity matrix. The second restriction pertains to the representation of \mathbf{Y} in (9) as the sum of a low-rank component and its full-rank counterpart, which clearly demands the introduction of zero factors through \mathbf{V}_1 and \mathbf{V}_2 to accommodate the desired structure. Consequently, an alternative procedure to compute the MLEs is required.

Hansen (2002) proposed an ML estimation technique for a general class of reduced rank regression models (called GRRR), including models with a generic structure of the covariance matrix and potential restrictions on the coefficient matrices. We exploit the GRRR setting to accommodate the heteroscedasticity and the aforementioned restrictions in the computation of the MLEs.

The GRRR problem considers the regression given by the vectorized model of (11)

$$\mathbf{y}_{(i)} = \mathbf{V}_1'\mathbf{A}\mathbf{B}'\mathbf{x}_{(i)} + \tilde{\mathbf{e}}_{(i)}, \quad (15)$$

where $\tilde{\mathbf{e}}_{(i)}$ is the i th column of $\tilde{\mathbf{E}} \in \mathbb{R}^{q \times n}$ and $\text{vec}(\tilde{\mathbf{E}}) \sim \mathcal{N}_{nq}(\mathbf{0}, \Sigma_\mathbf{y})$, subject to the restriction

$$\text{vec}(\mathbf{V}_1'\mathbf{A}) = \mathbf{G}\boldsymbol{\psi} + \mathbf{g}, \quad (16)$$

where $\boldsymbol{\psi}$ is the true vector of parameters to be estimated, \mathbf{G} is a binary $qr \times r(q_\gamma - r)$ matrix and \mathbf{g} is the qr -dimensional binary vector of restrictions (see Section 2 of the supplement for details). Then, the MLEs following the GRRR method are obtained as

$$\hat{\alpha}_{\mathbf{V}_1} = \text{vec}(\mathbf{V}'_1 \mathbf{A}) = \mathbf{G}(\mathbf{G}' \mathbf{M}_B \mathbf{G})^{-1} \mathbf{G}'(\mathbf{n}_B - \mathbf{M}_B \mathbf{g}) + \mathbf{g}, \quad (17)$$

$$\hat{\beta} = \text{vec}(\mathbf{B}) = \mathbf{M}_A^{-1} \mathbf{n}_A, \quad (18)$$

where $\mathbf{M}_B = (\mathbf{X}\mathbf{B} \otimes \mathbf{I}_q)' \tilde{\Sigma}_Y^{-1} (\mathbf{X}\mathbf{B} \otimes \mathbf{I}_q)$, $\mathbf{n}_B = (\mathbf{X}\mathbf{B} \otimes \mathbf{I}_q)' \tilde{\Sigma}_Y^{-1} \text{vec}(\mathbf{Y}')$, $\mathbf{M}_A = \mathbf{K}'_{p,r} (\mathbf{X} \otimes \mathbf{V}'_1 \mathbf{A})' \tilde{\Sigma}_Y^{-1} (\mathbf{X} \otimes \mathbf{V}'_1 \mathbf{A}) \mathbf{K}_{p,r}$, $\mathbf{n}_A = \mathbf{K}'_{p,r} (\mathbf{X} \otimes \mathbf{V}'_1 \mathbf{A})' \tilde{\Sigma}_Y^{-1} \text{vec}(\mathbf{Y}')$, $\tilde{\Sigma}_Y = \mathbf{K}_{n,q} \Sigma_Y \mathbf{K}'_{n,q}$, and $\mathbf{K}_{m,n}$ is the $mn \times mn$ commutation matrix, which transforms the vectorization of a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ into the vectorization of its transpose, such that $\mathbf{K}_{m,n} \text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M}')$.

Noticing that the expressions in (17) and (18) depend on each other, the practical implementation of the GRRR method is done in a recursive algorithm. After a random initialization of the parameter values, $\alpha_{\mathbf{V}_1}$ and β are iteratively updated until convergence. Once a solution $\hat{\alpha}_{\mathbf{V}_1}, \hat{\beta}$ is obtained, it suffices to transform the vectorized MLEs back to their matrix forms $\mathbf{V}_1 \hat{\mathbf{A}}'$ and $\hat{\mathbf{B}}$ to obtain the MLEs of the low-rank coefficient matrix as $\hat{\mathbf{C}}_1 = \hat{\mathbf{B}}(\mathbf{V}'_1)^+ \mathbf{V}'_1 \hat{\mathbf{A}} = \hat{\mathbf{B}} \hat{\mathbf{A}}'$, where \mathbf{M}^+ refers to the Moore-Penrose pseudoinverse of \mathbf{M} .²

Consequently, we obtain the approximation

$$f_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}) \approx \sum_{r=1}^{r_{\max}} \frac{1}{r_{\max}} \tilde{f}_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r) \equiv \tilde{f}_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}), \quad (19)$$

where $\tilde{f}_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r)$ is the Laplace approximation of $f_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r)$ obtained from (14) to the integral in (13). Therefore, the posterior distribution of $\boldsymbol{\gamma}$ is approximated by

$$\tilde{p}(\boldsymbol{\gamma}|\mathbf{Y}, \Sigma, \rho) = \frac{\tilde{f}_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}) p(\boldsymbol{\gamma}|\rho)}{\sum_{\boldsymbol{\gamma}^\dagger \in \{0,1\}^q} \tilde{f}_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}^\dagger) p(\boldsymbol{\gamma}^\dagger|\rho)}. \quad (20)$$

Given that, $\boldsymbol{\gamma}$ is a q -dimensional binary vector, the collection of all possible configurations for the response allocation encompasses 2^q distinct elements, which quickly becomes exceedingly large even for moderate q . This calls for the adoption of an approximate method to sample $\boldsymbol{\gamma}$. Given the discreteness of the support of $\boldsymbol{\gamma}$, we use a Metropolized Shotgun Stochastic Search (MSSS) algorithm proposed by Hans, Dobra, and West (2007). The MSSS approach explores regions of the high-dimensional parameter space by examining a selection of neighbors of the current iteration's $\boldsymbol{\gamma}$ and rapidly identifying those with high posterior probability. Defining the set of neighbors to contain only a subset of all the possible values of $\boldsymbol{\gamma}$ allows for a tradeoff between the space exploration speed and the computational time. Similar to Yang, Goh, and Wang (2022), we take the neighborhood to be every binary vector that is a one-variable change to the current allocation $\boldsymbol{\gamma}$ and at the same time complies with the existence of a low-rank group. For example, if $\boldsymbol{\gamma} = (1, 0, 1, 0)$, then $\text{nbd}(\boldsymbol{\gamma}) = \{(1, 1, 1, 0), (1, 0, 1, 1)\}$, but $(0, 1, 1, 0)$, $(0, 0, 1, 0)$ and $(1, 0, 0, 0)$ are not a neighbors.³ This restriction improves computational efficiency compared to an unrestricted neighborhood comprising all elements while allowing the (reduced) exploration of the space. We define a proposal distribution by

$$g(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(m)}) \propto \tilde{p}(\boldsymbol{\gamma}|\mathbf{Y}, \Sigma, \rho) \mathbb{I}(\boldsymbol{\gamma} \in \text{nbd}(\boldsymbol{\gamma}^{(m)})), \quad (21)$$

² \mathbf{V}_1 is not an invertible matrix given that its dimensionality is $q_Y \times q$, with $q_Y < q$.

³The element $(0, 1, 1, 0)$ is not a neighbor because two variables changed; $(0, 0, 1, 0)$ and $(1, 0, 0, 0)$ are not neighbors because they violate the constraint $q_Y \in \{1, \dots, q-1\}$.

where $\boldsymbol{\gamma}^{(m)}$ is the value of $\boldsymbol{\gamma}$ at the m th iteration of the MCMC. Summarizing, the first step of the proposed PCG sampler generates a draw from the marginal posterior $p(\boldsymbol{\gamma}|\mathbf{Y}, \Sigma, \rho)$ with the following procedure:

1. Generate $\boldsymbol{\gamma}^*$ from $g(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(m)})$.
2. Accept $\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^*$ with probability

$$\rho_{\boldsymbol{\gamma}} = \min \left\{ 1, \frac{\sum_{\boldsymbol{\gamma} \in \text{nbd}(\boldsymbol{\gamma}^{(m)})} \tilde{f}_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}) p(\boldsymbol{\gamma}|\rho)}{\sum_{\boldsymbol{\gamma}^\dagger \in \text{nbd}(\boldsymbol{\gamma}^*)} \tilde{f}_{\mathcal{Y}}(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}^\dagger) p(\boldsymbol{\gamma}^\dagger|\rho)} \right\}, \quad (22)$$

and otherwise, set $\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)}$.

Similarly to $\boldsymbol{\gamma}$, the conditional posterior of r is approximated through the Laplace method. Specifically, we compute the approximated posterior

$$\tilde{p}(r|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}) = \frac{\tilde{f}_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r) p(r|\boldsymbol{\gamma})}{\sum_{r^\dagger=1}^{r_{\max}} \tilde{f}_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r^\dagger) p(r^\dagger|\boldsymbol{\gamma})}, \quad (23)$$

where $\tilde{f}_r(\mathbf{Y}|\Sigma, \boldsymbol{\gamma}, r) p(r|\boldsymbol{\gamma})$ is the same as in (14). Then, a new value of r is sampled from the discrete distribution on $\{1, \dots, r_{\max}\}$ with the probabilities given in (23).

3.3. Sampling the Matrices \mathbf{A} and \mathbf{C}_2

The proposed PCG sampler introduces a particular challenge related to the dimensions of matrices \mathbf{A} , \mathbf{B} , and \mathbf{C}_2 in Steps 3 and 4 of Algorithm 1. Suppose that at the end of iteration m , we have generated $(\boldsymbol{\gamma}^{(m)}, r^{(m)}, \mathbf{A}^{(m)}, \mathbf{B}^{(m)}, \mathbf{C}_2^{(m)})$. Then, at iteration $m+1$, we obtain new values $(\boldsymbol{\gamma}^{(m+1)}, r^{(m+1)})$ in Steps 1 and 2. Afterwards, we shall update \mathbf{C}_2 by sampling $\delta^{(m+1)}$ conditioned on the value $\boldsymbol{\gamma}^{(m+1)}$ just generated. To this aim, notice that $\mathbf{C}_2^{(m+1)}$ should be of "new" dimension $p \times (q - q_{\boldsymbol{\gamma}^{(m+1)}})$, and its posterior distribution depends on the matrix \mathbf{C}_1 , which should have "new" dimension $p \times q_{\boldsymbol{\gamma}^{(m+1)}}$. However, the matrix $\mathbf{C}_1^{(m)}$ available at this step is of dimension $p \times q_{\boldsymbol{\gamma}^{(m)}}$. An analogous issue is encountered in Step 4 when sampling $\mathbf{A}^{(m+1)}$, as the posterior of the latter parameter would require a matrix \mathbf{B} with "new" dimension $p \times r^{(m+1)}$, whereas the available matrix $\mathbf{B}^{(m)}$ has $r^{(m)}$ columns.

It is worth emphasizing that these dimensionality issues stem from considering $(\boldsymbol{\gamma}, r)$ as parameters to be estimated, thus, varying quantities across the MCMC iterations. Moreover, changing the order of the Gibbs steps (while keeping the PCG sampler) would not circumvent the problem.

A possible way out of the issue in Step 3 can be found by recalling the decomposition of the coefficient matrix in (2), that is $\mathbf{C}^{(m)} = [\mathbf{C}_1^{(m)}, \mathbf{C}_2^{(m)}]$. Importantly, for any iteration $m = 1, \dots, M$, this matrix of coefficients has a fixed dimension $p \times q$, whereas the number of columns of $\mathbf{C}_1^{(m)}$ and $\mathbf{C}_2^{(m)}$ are possibly changing across iterations in consequence of varying $\boldsymbol{\gamma}^{(m)}$. Hence, to sample $\delta^{(m+1)}$ conditioned on $\boldsymbol{\gamma}^{(m+1)}, \mathbf{A}^{(m)}, \mathbf{B}^{(m)}$, we construct an auxiliary matrix $\mathbf{C}_{1*}^{(m)}$ that is consistent with the newly sampled value of $\boldsymbol{\gamma}^{(m+1)}$, formed by selecting the first $q_{\boldsymbol{\gamma}^{(m+1)}}$ columns of the available $\mathbf{C}^{(m)}$:

$$\mathbf{C}_{1*}^{(m)} = [\mathbf{C}_{\bullet 1}^{(m)}, \dots, \mathbf{C}_{\bullet q_{\boldsymbol{\gamma}^{(m+1)}}}^{(m)}], \quad (24)$$

where $\mathbf{C}_{\bullet j}^{(m)}$ is the j th column of matrix $\mathbf{C}^{(m)}$. At iteration $m + 1$, the auxiliary matrix $\mathbf{C}_{1*}^{(m)}$ is used to compute the updated parameters of the posterior distribution of $\delta^{(m+1)}$.

To address the dimensionality inconsistency in Step 4, let us recall the restriction in (4); then, substituting $\mathbf{A}^{(m)} = [\mathbf{I}_{r^{(m)}}, \mathbf{F}^{(m)'}]'$ and $\mathbf{B}^{(m)}$ in the low-rank matrix representation yields

$$\mathbf{C}_1^{(m)} = [\mathbf{B}^{(m)}, \mathbf{B}^{(m)}\mathbf{F}^{(m)'}]. \tag{25}$$

Based on (25), it is evident that the first $r^{(m)}$ columns of $\mathbf{C}_1^{(m)}$ coincide with the matrix $\mathbf{B}^{(m)}$. Consequently, to update $\mathbf{A}^{(m+1)}$, we define the auxiliary matrix $\mathbf{B}_*^{(m)}$ as

$$\mathbf{B}_*^{(m)} = [\mathbf{C}_{\bullet 1}^{(m)}, \dots, \mathbf{C}_{\bullet r^{(m+1)}}^{(m)}]. \tag{26}$$

It is important to remark that the auxiliary matrices $\mathbf{C}_1^{(m)}$ and $\mathbf{B}_*^{(m)}$ have appropriate dimensions, corresponding to the updated values $\boldsymbol{\gamma}^{(m+1)}$ and $r^{(m+1)}$, and contain elements already available at iteration $m + 1$, that is $\mathbf{C}^{(m)}, \mathbf{B}^{(m)}, \mathbf{F}^{(m)}$.

In more detail, in Step 3, the full-rank coefficient matrix is sampled in vectorized form. Denoting $\mathbf{c}_{1*} = \text{vec}(\mathbf{C}_{1*})$, the posterior distribution of δ is proportional to the multivariate Gaussian distribution $p(\delta|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}, \mathbf{c}_{1*}) \propto p(\delta)p(\boldsymbol{\gamma}|\Sigma, \delta, \mathbf{c}_{1*}) \sim \mathcal{N}_{p(q-q_\gamma)}(\delta|\bar{\boldsymbol{\mu}}_\delta, \bar{\Sigma}_\delta)$ with mean $\bar{\boldsymbol{\mu}}_\delta = \bar{\Sigma}_\delta \mathbf{U}'_2 \bar{\Sigma}^{-1}(\mathbf{y} - \mathbf{U}_1 \mathbf{c}_{1*})$ and covariance matrix $\bar{\Sigma}_\delta = (\bar{\Sigma}_\delta^{-1} + \mathbf{U}'_2 \bar{\Sigma}^{-1} \mathbf{U}_2)^{-1}$.

The update of $\mathbf{A} = [\mathbf{I}_r, \mathbf{F}']'$ given $(\mathbf{Y}, \boldsymbol{\gamma}, r, \Sigma, \mathbf{B}, \mathbf{C}_2)$ is performed by sampling $\boldsymbol{\alpha}_F = \text{vec}(\mathbf{F}')$. The posterior distribution of $\boldsymbol{\alpha}_F$ is proportional to the multivariate Gaussian distribution $p(\boldsymbol{\alpha}_F|\mathbf{Y}, \Sigma, \boldsymbol{\gamma}, r, \mathbf{C}_2, \mathbf{B}_*) \propto p(\boldsymbol{\alpha}_F|\boldsymbol{\gamma}, r)p(\mathbf{Y}|\Sigma, \mathbf{A}, \mathbf{B}_*, \mathbf{C}_2) \sim \mathcal{N}_{(q_\gamma-r)r}(\boldsymbol{\alpha}_F|\bar{\boldsymbol{\mu}}_\alpha, \bar{\Sigma}_\alpha)$ with mean $\bar{\boldsymbol{\mu}}_\alpha = \bar{\Sigma}_\alpha(\mathbf{m}_J - \mathbf{H}_{[J,J]}\mathbf{v})$ and covariance matrix $\bar{\Sigma}_\alpha = (\bar{\Sigma}_\alpha^{-1} + \mathbf{H}_{[J,J]})^{-1}$, where $\mathbf{v} = \text{vec}(\mathbf{I}_r)$, $\mathbf{m} = \mathbf{M}'_\alpha \bar{\Sigma}^{-1} \tilde{\mathbf{y}}_2$, $\mathbf{H} = \mathbf{M}'_\alpha \bar{\Sigma}^{-1} \mathbf{M}_\alpha$, $\tilde{\mathbf{y}}_2 = \mathbf{y} - \mathbf{U}_2 \delta$, and $\mathbf{M}_\alpha = \mathbf{U}_1(\mathbf{I}_{q_\gamma} \otimes \mathbf{B})$. Moreover, $\mathbf{H}_{[J,J]}$ indicates the J th row and the J th column in \mathbf{H} for the sequence $J = \{r^2 + 1, r^2 + 2, \dots, q_\gamma r\}$.

3.4. Sampling the Other Parameters

Regarding Steps (5)–(7) of Algorithm 1, the conditional posterior distribution of $\boldsymbol{\beta}$, given $(\mathbf{Y}, \boldsymbol{\gamma}, r, \Sigma, \mathbf{A}, \mathbf{C}_2)$, is proportional to the multivariate Gaussian distribution $\mathcal{N}_{pr}(\boldsymbol{\beta}|\bar{\boldsymbol{\mu}}_\beta, \bar{\Sigma}_\beta)$, where $\bar{\Sigma}_\beta = (\bar{\Sigma}_\beta^{-1} + \mathbf{M}'_\beta \bar{\Sigma}^{-1} \mathbf{M}_\beta)^{-1}$ and $\bar{\boldsymbol{\mu}}_\beta = \bar{\Sigma}_\beta \mathbf{M}'_\beta \bar{\Sigma}^{-1} \tilde{\mathbf{y}}_2$, with $\mathbf{M}_\beta = \mathbf{U}_1(\mathbf{A} \otimes \mathbf{I}_p)$.

The conditional posterior of the innovation covariance matrix Σ , given $(\mathbf{A}, \mathbf{B}, \mathbf{C}_2, \mathbf{Y})$, is the inverse Wishart $\mathcal{IW}_q(\Sigma|\bar{\mathbf{v}}, \bar{\Psi})$, where $\bar{\mathbf{v}} = \underline{\mathbf{v}} + n$ and $\bar{\Psi} = \underline{\Psi} + (\mathbf{Y} - \mathbf{X}\mathbf{C})'(\mathbf{Y} - \mathbf{X}\mathbf{C})$. Finally, the posterior distribution of ρ , the probability of a response variable belonging to the low-rank group, is the Beta distribution $\mathcal{Be}(\rho|\bar{a}_\rho, \bar{b}_\rho)$, with $\bar{a}_\rho = \underline{a}_\rho + q_\gamma$, and $\bar{b}_\rho = \underline{b}_\rho + q - q_\gamma$. A detailed derivation of the posterior distributions is available in the supplement.

4. Simulation Study

This section is devoted to examining the proposed model's performance in terms of the ability to recover the true group

allocation, that is, the classification of variables into the low- and full-rank groups in different simulation studies. Next, the performance of the sampler is tested about rank estimation and the recovery of the overall coefficient matrix, \mathbf{C} .

The data was generated from the multivariate linear model $\mathbf{Y}_0 = \mathbf{X}\mathbf{C}_0 + \mathbf{E}_0$. The rows of \mathbf{X} were independently drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and the rows of \mathbf{E}_0 were drawn from $\mathcal{N}(\mathbf{0}, \Sigma_0)$, where the covariance matrix Σ_0 is diagonal with elements sampled from $\mathcal{U}(0.5, 1.75)$. We work with centred responses and exclude the intercept term for simplicity. To generate the coefficient matrix \mathbf{C}_0 , we first recall its partition into the low-rank and full-rank matrices $\mathbf{C}_1 = \mathbf{B}\mathbf{A}'$ and \mathbf{C}_2 , then draw each free entry of \mathbf{A} , \mathbf{B} , and \mathbf{C}_2 from a standard Gaussian. Notice that the dimensions of these matrices depend on the fixed number of responses in the low-rank group, $q_\gamma < q$, and the true rank, $r \leq r_{\max}$.

The allocation of the response variables to the reduced-rank group is randomly selected given q_γ , and represented by the binary vector $\boldsymbol{\gamma}$. The columns of the matrix \mathbf{Y}_0 are then permuted following the allocation imposed by $\boldsymbol{\gamma}$. The response matrix so generated, \mathbf{Y} , need not necessarily be partitioned as $\mathbf{Y}_0 = [\mathbf{Y}_1, \mathbf{Y}_2]$, which is the representation postulated by our BPRR model. For example, if $q = 5$ and $q_\gamma = 3$, the data generating process (DGP) initially comprises the partitioned response matrix $\mathbf{Y}_0 = [\mathbf{Y}_1, \mathbf{Y}_2]$, with $\mathbf{Y}_1 = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3] \in \mathbb{R}^{n \times 3}$ and $\mathbf{Y}_2 = [\mathbf{y}_4, \mathbf{y}_5] \in \mathbb{R}^{n \times 2}$. However, after a random $\boldsymbol{\gamma}$ has been generated, say $\boldsymbol{\gamma}_0 = (0, 1, 0, 1, 1)$, then the final generated response matrix that is fed into our model is $\mathbf{Y} = [\mathbf{y}_4, \mathbf{y}_1, \mathbf{y}_5, \mathbf{y}_2, \mathbf{y}_3]$, aiming to reorder the responses to their true form, \mathbf{Y}_0 , if $\hat{\boldsymbol{\gamma}}$ is estimated correctly.

The hyperparameters are set to consider noninformative priors, specifically $\underline{a}_\rho = \underline{b}_\rho = 1$, $\underline{a} = \underline{b} = \underline{d} = 0.5$, $\underline{\mathbf{v}} = \mathbf{q} + 1$, and $\underline{\Psi} = \mathbf{I}_q$. The starting value of Σ is the identity matrix, while the coefficient matrix and the response allocation vector are initialized at random.

We consider different simulation settings with varying dimensionality, number of low-rank responses and true rank. Our method (BPRR) is compared with the following competitors: full-rank (FR), full low-rank (RR), and pre-specified allocation partial low-rank (PRR*). The first one is a standard linear regression model, where no low-rank structure is assumed. The full low-rank concerns a usual reduced-rank regression without an imposed partition. The last competitor is a partially reduced-rank regression model in which the low-rank group is fixed at random, a feature that serves two purposes. First, it accommodates scenarios where the researcher might have prior knowledge about the grouping structure, enabling the estimation procedure of the partial reduced-rank model to be conducted with a constraint by the imposed $\boldsymbol{\gamma}$. Second, it allows us to examine whether the automatic model selection in BPRR offers advantages over a random grouping choice.

The performance of the estimator $\hat{\mathbf{C}}$ of the coefficient matrix is evaluated using the mean squared error, $\text{MSE} = \|\hat{\mathbf{C}} - \mathbf{C}_0\|_{\text{F}}^2/(pq)$, where $\hat{\mathbf{C}}$ is the posterior mean of the predicted coefficients in their original ordering, warranting a fair comparison if the estimated allocation vector, $\hat{\boldsymbol{\gamma}}$, is not the correct grouping. The point estimates of the binary vector and the rank

Table 1. Average MSE ($\times 10^2$) over 100 replicates of the listed simulation settings in columns 1–5, for each model: Bayesian partial reduced-rank regression (BPRR), full low-rank regression (FR), reduced-rank regression (RR) and pre-specified allocation partial low-rank regression (PRR*).

p	q	q_γ	r	BPRR metrics				MSE $\times 10^2$					
				n	\hat{q}_γ	\hat{r}	Accuracy	F_1 score	BPRR	FR	RR	PRR*	Oracle
5	5	3	1	20	3.240	1.740	0.584	0.659	11.831	10.369	13.350	15.275	8.566
5	6	3	1	20	3.650	2.030	0.612	0.653	10.218	8.900	12.420	13.712	8.227
5	6	3	1	50	3.950	2.120	0.698	0.743	3.434	2.646	4.136	6.354	2.510
5	6	5	2	50	4.790	2.090	0.862	0.913	2.646	3.252	3.480	9.673	2.353
5	6	5	2	100	4.910	2.090	0.925	0.953	1.202	1.277	1.580	5.366	1.108
5	7	3	1	20	3.770	2.090	0.576	0.560	11.356	10.782	16.966	15.689	9.302
5	7	3	1	50	4.650	2.610	0.610	0.657	3.182	2.673	5.172	5.900	2.972
5	7	5	2	50	5.320	2.500	0.786	0.849	3.503	3.186	4.632	7.062	2.936
5	8	6	2	20	5.100	2.120	0.713	0.777	14.684	16.345	17.112	21.816	14.467
10	5	3	1	20	3.400	1.910	0.636	0.714	14.925	16.998	18.838	23.548	10.309
10	5	3	1	100	3.750	1.920	0.766	0.831	1.948	1.297	1.953	8.659	1.712
10	6	3	1	20	3.490	2.100	0.622	0.656	16.145	16.182	19.166	23.283	11.830
10	6	5	2	20	4.600	2.310	0.797	0.866	30.161	38.745	29.880	43.679	26.247
10	6	5	2	50	4.980	2.250	0.893	0.936	4.892	5.561	5.573	10.051	4.563
10	7	3	1	20	3.620	2.280	0.591	0.580	15.644	14.609	17.508	20.657	10.661
10	7	5	2	20	4.870	2.650	0.696	0.769	23.839	26.104	25.028	38.296	21.142
10	8	3	1	20	3.030	1.810	0.664	0.561	18.609	14.082	16.886	23.783	12.490
10	8	6	2	20	5.290	2.590	0.696	0.765	22.094	26.685	23.590	35.390	19.869
20	5	3	1	20	2.610	1.480	0.570	0.610	32.304	30.856	36.725	39.718	24.326
20	5	3	1	50	3.920	2.390	0.608	0.719	6.866	4.820	8.097	14.118	5.980
20	6	3	1	20	2.410	1.290	0.595	0.545	32.099	28.815	35.023	38.934	23.014
20	6	5	2	20	2.530	1.310	0.528	0.609	56.350	54.898	54.315	58.261	44.784
20	7	3	1	20	2.200	1.070	0.609	0.463	31.839	27.640	34.408	37.256	22.802
20	7	5	2	20	2.350	1.220	0.521	0.535	57.864	56.544	59.930	67.723	51.176
50	5	3	1	50	3.490	2.140	0.570	0.661	23.463	25.552	32.526	30.296	16.145

NOTE: For each combination of parameters, columns 6–9 provide the average estimates of the number of low-rank responses, the rank of C_1 , the accuracy and the F_1 score.

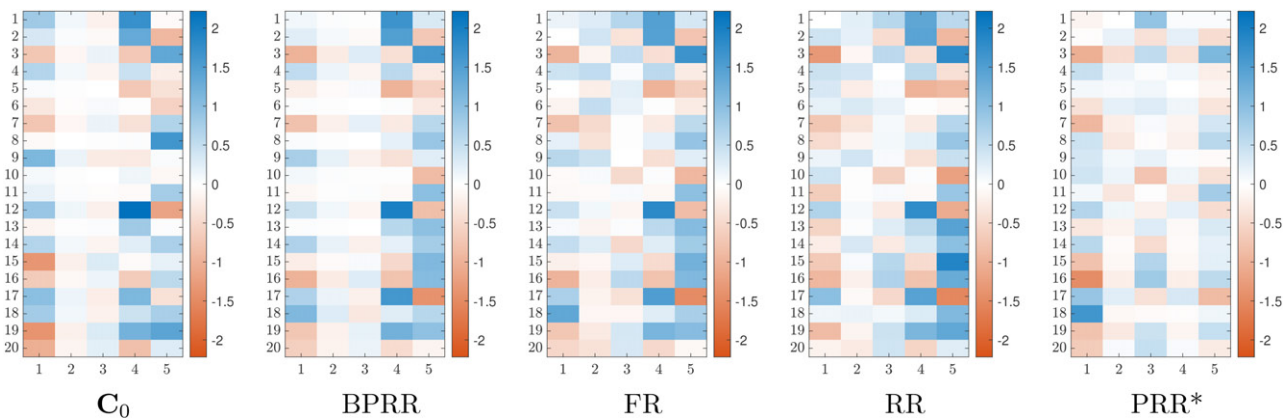


Figure 1. True coefficient matrix (first left) and estimated C matrix by each model in the simulation scenario where $p = 20, q = 5, q_\gamma = 3, r = 1,$ and $n = 20$.

($\hat{\gamma}$ and \hat{r} , respectively) are their corresponding maximum a posteriori.

Table 1 summarizes the simulation results by providing the average MSE over 100 independent experiments of each scenario. Our method predominantly achieves the minimum mean squared error, and increasing the number of observations results in a smaller error. The former result is visually explored in Figure 1, which examines the similarity of the coefficient matrix estimated by each of the four models with the true ordering. BPRR approximates C_0 more accurately than its competitors, and the automatic grouping choice eliminates the need for a pre-processing step to select a potentially incorrect γ , which could result in an inferior estimation as in this case. Additionally, we present the average estimated number of low-rank response variables with their corresponding rank estimate. As a measure of classification for the allocation parameter, we employ the accuracy and the F_1 score, both of which range from 0 to 1.

A higher value indicates a more accurate classification of the responses.

We observe that misspecifying a PRR regression model, either by choosing an incorrect regression model or by inaccurately determining the reduced-rank grouping, can lead to substantive performance loss, as demonstrated earlier.

4.1. Convergence Diagnostic Analysis

We assess the computational performance of the proposed MCMC algorithm by means of a convergence diagnostic analysis (CODA). Specifically, for each parameter under investigation, we consider Geweke’s convergence diagnostic, Heidelberger and Welch’s stationarity and half-width tests (see Heidelberger and Welch 1983; Geweke 1992, for further details). Since the coefficient matrix C has pq entries, Table 2 reports the share of entries that pass each of the aforementioned tests (i.e., p -value

Table 2. Convergence diagnostics of the rank and the coefficient matrix C: Geweke and the Heidelberger and Welch’s (HW) stationarity tests p -values of the rank (> 0.05), and the rank’s HW half-width test ratio (< 0.10).

p	q	q_γ	r	n	Geweke test		HW Stationarity test		HW Half-width test	
					r (p -value)	C (share)	r (p -value)	C (share)	r (ratio)	C (share)
5	5	3	1	20	0.814	0.920	0.376	1.000	0.011	0.800
5	5	3	1	40	0.450	0.800	0.072	0.680	0.018	0.882
5	8	3	1	20	0.193	0.750	0.216	1.000	0.037	0.900
5	8	6	2	20	0.019	0.975	0.223	1.000	0.008	0.950
5	8	6	4	20	0.203	0.975	0.290	0.975	0.003	0.949
10	5	3	1	20	0.275	0.920	0.360	0.980	0.016	0.918
10	5	3	1	40	0.795	1.000	0.306	0.980	0.015	1.000
10	8	3	1	20	0.486	0.825	0.218	0.963	0.132	0.870
10	8	6	2	20	0.819	0.975	0.550	1.000	0.033	0.813
10	8	6	4	20	0.268	0.913	0.148	0.988	0.104	0.848
20	5	3	1	20	0.011	0.830	0.433	0.990	0.004	0.909
20	5	3	1	40	0.072	0.620	0.234	1.000	0.017	0.930

NOTE: Share of entries of the coefficient matrix C that pass the Geweke test, HW stationarity and half-width tests (in column).

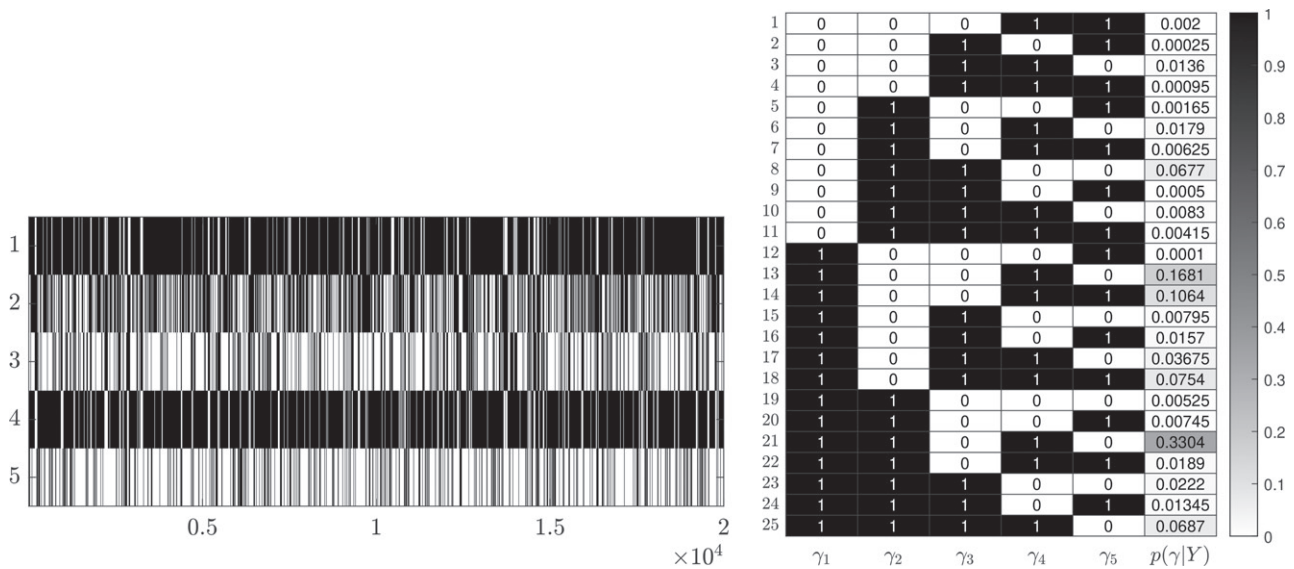


Figure 2. Trace plot (left) and posterior distribution (right) of γ in the simulation scenario where $p = 20, q = 5, q_\gamma = 3, r = 1$, and $n = 20$.

> 0.05 ; ratio < 0.10). The convergence diagnostics of the rank of the low-rank matrix C_1 are equally included in Table 2. The results are satisfactory and suggest convergence of the chains for both parameters.⁴

Instead, as pertains to the binary allocation vector γ , similar tests are not available. Therefore, we rely on the visual inspection of the trace plot and posterior distribution in Figure 2 to assess the convergence and mixing of the chain. It is worth emphasizing that we also performed independent runs of the algorithm with random initialization and obtained similar plots. Therefore, the results in Figure 2 suggest a good exploration of the space of configurations and a good acceptance rate of the MSSS step. In particular, the posterior distribution assigns positive mass on several configurations while displaying a unique maximum at $\hat{\gamma} = (1, 1, 0, 1, 0)$, which corresponds to the true allocation (see Section 3 of the supplement for further results).

4.2. Forecasting Exercise

As a further evaluation step, we conduct a forecasting exercise to evaluate the predictive performance of our proposed method. The dimensions of the artificially generated data for this purpose were $p = 10, q = 5, q_\gamma = 3, r = 1$. Two different sample sizes were considered, $n = 60$ and $n = 100$. We make a one-step-ahead prediction for n_{test} observations using our model and the three competitors: full-rank (FR), full low-rank (RR), and pre-specified allocation partial low-rank (PRR*). A fitted matrix of responses, \hat{Y} , was constructed with all n_{test} predictions, and compared against the true values using the mean squared error (MSE) and the mean absolute error (MAE), defined as

$$\text{MSE} = \sum_{i=1}^{n_{\text{test}}} \sum_{j=1}^q (\hat{y}_{ij} - y_{ij})^2 / (n_{\text{test}}q), \text{ and}$$

$$\text{MAE} = \sum_{i=1}^{n_{\text{test}}} \sum_{j=1}^q |\hat{y}_{ij} - y_{ij}| / (n_{\text{test}}q). \tag{27}$$

The forecast error metrics of the models are presented in Table 3, demonstrating that our model achieves superior predictive performance compared to the alternatives. The model exhibiting the smallest error is highlighted in gray.

⁴We performed the CODA analysis also on other independent runs of the algorithm, finding analogous results.

5. An Application to Macroeconomic Data

This section aims to demonstrate the usefulness of our method when applied to real-world data. We consider quarterly macroeconomic data for the United States from 2014Q1 to 2023Q4, which were retrieved from FRED, Federal Reserve Bank of St. Louis, and the OECD, Organisation for Economic Co-operation and Development (see Section 4 of the supplement for details).

The $q = 5$ responses are the 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), and volume index of exports of goods and services (y_5). The $p = 5$ covariates are 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), and price index of final consumption expenditure (x_5). All variables were standardized before conducting the analysis.

To account for possible temporal dependence, we introduce time variation in the innovation covariance, by assuming $e_{(i)} \sim \mathcal{N}_q(\mathbf{0}, \Sigma_i)$, where we assume the decomposition $\Sigma_i = \mathbf{W}\mathbf{D}_i\mathbf{W}'$ with \mathbf{W} being a lower triangular matrix with ones on the diagonal and $\mathbf{D}_i = \text{diag}(\exp(h_{1i}), \dots, \exp(h_{qi}))$ is a diagonal matrix of variances (e.g., see Carriero, Clark, and Marcellino 2019). Each log-variance is assumed to follow a random walk process

$$h_{ji} = h_{j,i-1} + \varepsilon_{ji}, \quad \varepsilon_{ji} \sim \mathcal{N}(0, \sigma_j^2), \quad j = 1, \dots, q. \quad (28)$$

Table 3. Mean squared error (MSE) and mean absolute error (MAE) of the true responses versus the fitted values through a rolling forecast with the models BPRR, FR, RR, and PRR*. The model achieving the smallest error is highlighted in gray.

(n, n_{test})		BPRR	FR	RR	PRR*
(60, 20)	MSE	1.405	1.501	1.535	2.059
	MAE	0.903	0.940	0.963	1.043
(100, 40)	MSE	1.467	1.529	1.456	2.797
	MAE	0.941	0.959	0.945	1.100

The model specification is completed by assuming a Gaussian prior for the free entries of \mathbf{W} , denoted $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Omega)$, a conjugate prior for the variance, $\sigma_j^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$, and the initial point, $h_{j0} \sim \mathcal{N}(0, \underline{\sigma}_j^2)$.

We investigate whether the pre-COVID period and the years following the outbreak of the COVID-19 pandemic have similar drivers. Therefore, we consider two sub-periods by splitting the sample into the pre-COVID (2014Q1 to 2018Q4) and the (post)-COVID (2019Q1 to 2023Q4) periods, where the latter includes the outbreak of the pandemic and the subsequent recovery. Each period consists of $n = 20$ quarterly observations. Our interest lies, in particular, in investigating whether and how the estimation of the response allocation vector changes over time and quantifying the associated uncertainty.

In the first period, the estimated allocation is $\hat{\gamma} = (0, 1, 1, 1, 1)$. However, upon inspecting the posterior distribution of γ in Figure 3, it is evident that this result is highly uncertain. The allocation has posterior probability of 0.16, followed closely by $(0, 1, 0, 1, 0)$ at 0.12, and $(0, 1, 1, 1, 0)$ at 0.10. In contrast, the posterior distribution of γ for the period including the COVID-19 pandemic exhibits a mode at $\hat{\gamma} = (1, 0, 0, 0, 1)$. These results suggest a significant shift in the low-rank structure of the coefficient matrix, which moved from an almost reduced rank to an almost full rank structure. It is worth emphasizing that the two periods differ not only in the point estimate of the allocation vector, $\hat{\gamma}$, but also in the uncertainty about the estimate, which is higher in the pre-COVID period. The proposed method allows us to uncover both findings directly from the data, as opposed to the traditional PRR model with *a-priori fixed* allocation.

The uncertainty regarding parameter estimates for the 2014–2018 period is less pronounced in the rank's posterior distribution, which shows high probabilities for 1 and 2, with a clear single mode at 1 (see Figure 4). Conversely, in the second period, we obtain a solid conclusion for a rank 1 model.

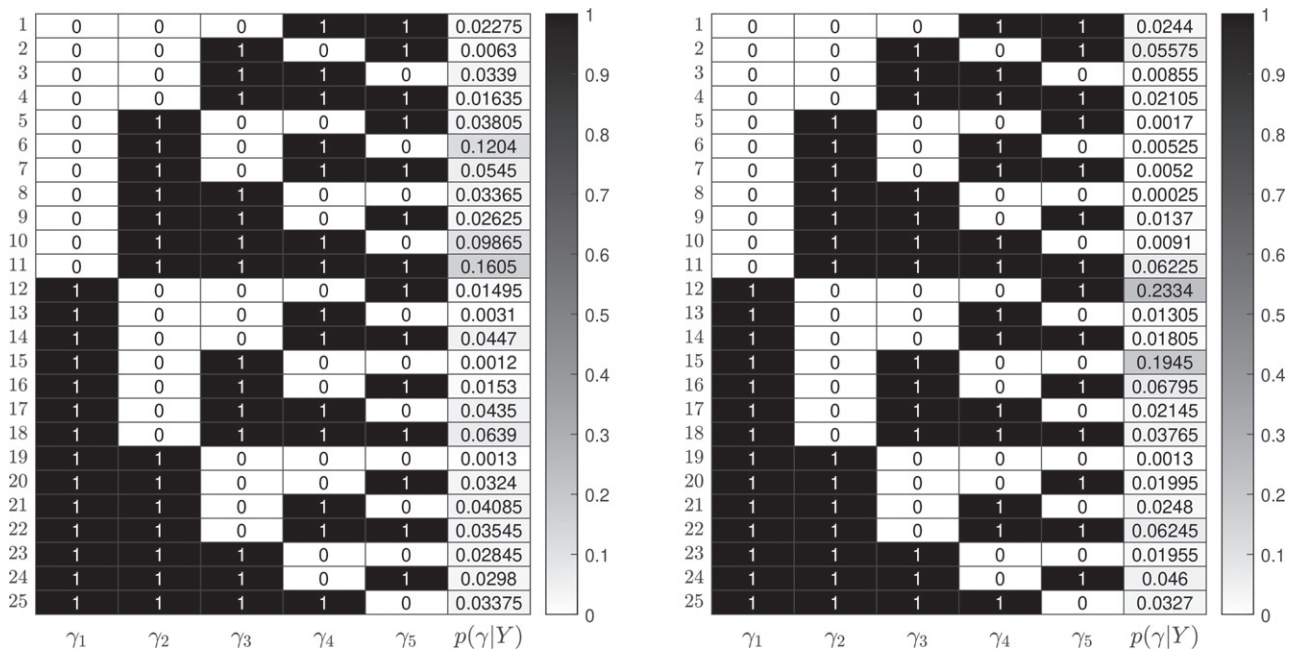


Figure 3. Posterior distribution of the allocation vector, γ , for the period 2014Q1-2018Q4 (left) and 2019Q1-2023Q4 (right).

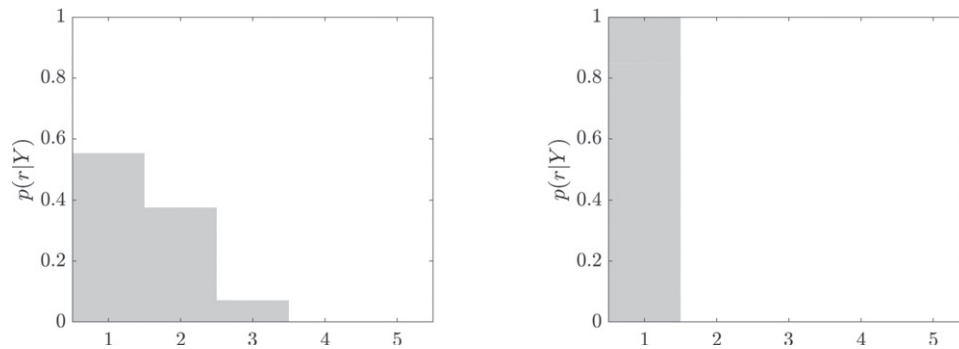


Figure 4. Posterior distribution of the rank, r , for the period 2014Q1-2018Q4 (left) and 2019Q1-2023Q4 (right).

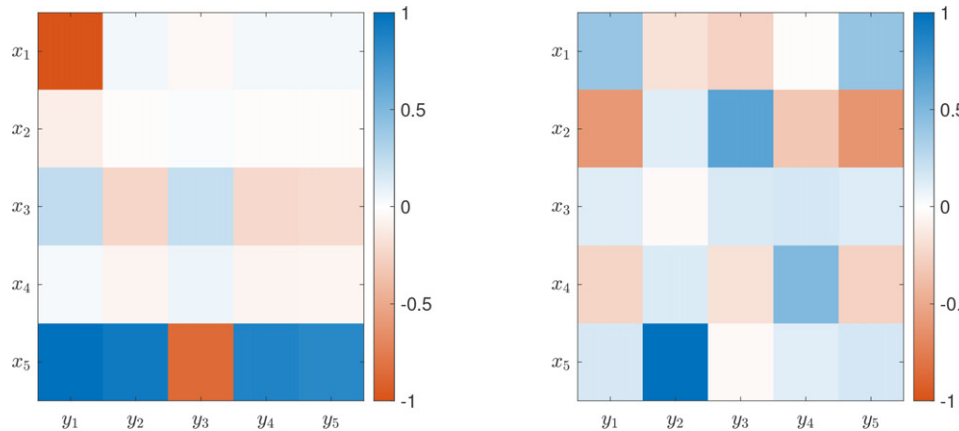


Figure 5. Posterior mean of the coefficient matrix, C , for the period 2014Q1-2018Q4 (left) and 2019Q1-2023Q4 (right). Responses are labeled in the horizontal axis, and covariates in the vertical axis.

Overall, these findings suggest that in the 2014–2018 period the relationship between the responses and covariates is rather simple, as most responses are allocated to the low-rank group with an estimated rank of 1. However, the high uncertainty with posterior mass concentrated on models with large low-rank groups indicates the lack of a clear grouping and suggests that the most relevant feature to account for in this period is the overall reduced-rank structure of the coefficient matrix, which yields a total rank of 2 under the model with the maximum a posteriori γ . A standard reduced-rank regression estimated a rank 3 coefficient matrix. Conversely, the 2019–2023 period features a small low-rank group with $\hat{r} = 1$ and a three-dimensional full-rank group, which is indicative of a significantly more complex structure of the relationship. The rank obtained under the traditional reduced-rank setting is again 3, failing to capture any increase in the complexity of the data from the previous period to the current one.

A possible reason for this change in the uncertainty of the posterior distribution for the grouping structure is the shift brought by COVID-19. Prior to the outbreak of the pandemic, the data were characterized by a large group with a simple structure, as pointed by $q_\gamma = 4$ and a low rank with small r across all models with high posterior probability. The uncertainty highlights the prominence of the reduced-rank feature over the qualitative information about which variables should be included in the group. Second, the outbreak of the pandemic

could also explain the change towards more complex relationships, since in 2019–2023 we find $q_\gamma = 2$ with high probability (more than 45%).

Pertaining to the regression coefficients, in Figure 5 we find evidence of a change in the relationship structure between the two periods. In the first period, median weekly earnings (x_2), and the price index of exports of goods and services (x_4) appear to have a negligible impact on explaining the responses of the low-rank group (y_2 to y_5). This pattern changes in the subsequent quarters, where the index of industrial production (y_1) and the volume index of exports of goods and services (y_5) exhibit a relationship with the covariates that are effectively captured by a rank-1 coefficient matrix. Furthermore, the weak signal of the covariates in the first period has strengthened in the second, and this estimation is more accurate in terms of the MSE compared to the other reduced-rank models presented in Section 4.⁵

The variation between the two periods and the associated uncertainty involved suggest a research direction concerning the incorporation of time-varying parameters within a time-series framework, which could potentially facilitate the identification of structural breaks.

⁵MSE_{BPRR} = 0.052, MSE_{FR} = 0.051, MSE_{RR} = 0.065, MSE_{PRR*} = 0.062.

6. Conclusions

We have proposed a novel Bayesian approach to inference for a partial reduced rank regression model (BPRR). To circumvent the need for transdimensional samplers, we rely on a partially collapsed Gibbs sampler, where the allocation vector and the rank parameters are drawn from their joint distribution marginalized over the coefficient matrix. Then, a Metropolis-Hasting step with local exploration is used to draw the allocation vector, reducing the unfeasible exploring of the entire space of configurations to a computationally manageable local search.

The simulation study has highlighted the good performance of the model and the proposed partially collapsed Gibbs sampler algorithm. BPRR outperforms its competitors regarding the MSE and effectively estimates the allocation vector, the rank of the reduced-rank matrix, and the regression coefficients. The MCMC convergence diagnostics support the efficacy of the algorithm. Our approach's usefulness has also been demonstrated in real macroeconomic data, showing a significant shift in both the point estimates and posterior uncertainty about the allocation vector and the rank parameters since the outbreak of the COVID-19 pandemic.

The proposed approach can be extended in several future directions. For instance, when dealing with time series data, it would be interesting to allow the allocation vector to vary over time, that is, γ_t . Another point worth exploring is the design of computational tools to speed up the inferential algorithm when sampling the allocation vector (e.g., Geels, Pratola, and Herbei 2023).

Supplementary Materials

The supplementary material reports the derivation of the posterior distributions for each parameter of the BPRR model in Section 3, the procedure to obtain the MLE estimates of (A, B) in Section 3.2, additional results from the simulation study in Section 4 and details about the macroeconomic data in Section 5.

Acknowledgments

The authors thank Co-Editor Laura Sangalli, the Associate Editor, and two anonymous referees for their valuable comments and suggestions. The authors gratefully acknowledge the participants at the 52nd SIS Conference, the Bayesian Young Statisticians Meeting 2024 (BAYSM), and the ISBA 2024 World Meeting for their helpful feedback. This research used the Computational resources provided by the Core Facility INDACO, a project of High-Performance Computing at the University of Milan.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Maria F. Pintado gratefully acknowledges CONAHCyT (Mexico), grant no. 2021-000007-01EXTF-00090. Alexander Y. Shestopaloff gratefully acknowledges support from NSERC Discovery Grant RGPIN-2024-05996.

References

- Anderson, T. W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22, 327–351. [1344]
- Buch, G., Schulz, A., Schmidtman, I., Strauch, K., and Wild, P. S. (2023), "A Systematic Review and Evaluation of Statistical Methods for Group Variable Selection," *Statistics in Medicine*, 42, 331–352. [1344]
- Carriero, A., Clark, T. E., and Marcellino, M. (2019), "Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *Journal of Econometrics*, 212, 137–154. [1352]
- Chen, L., and Huang, J. Z. (2012), "Sparse Reduced-Rank Regression for Simultaneous Dimension Reduction and Variable Selection," *Journal of the American Statistical Association*, 107, 1533–1545. [1344]
- Geels, V., Pratola, M. T., and Herbei, R. (2023), "The Taxicab Sampler: MCMC for Discrete Spaces with Application to Tree Models," *Journal of Statistical Computation and Simulation*, 93, 753–774. [1354]
- Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. DeGroot, 1931–1989*. [1350]
- (1996), "Bayesian Reduced Rank Regression in Econometrics," *Journal of Econometrics*, 75, 121–146. [1346]
- Hans, C., Dobra, A., and West, M. (2007), "Shotgun Stochastic Search for "large p" Regression," *Journal of the American Statistical Association*, 102, 507–516. [1347,1348]
- Hansen, P. R. (2002), "Generalized Reduced Rank Regression," Social Science Research Network. Working Paper No. 2002-02. [1347]
- Heidelberger, P., and Welch, P. D. (1983), "Simulation Run Length Control in the Presence of an Initial Transient," *Operations Research*, 31, 1109–1144. [1350]
- Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264. [1344]
- Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [1347]
- Kim, K., and Jung, S. (2024), "Integrative Sparse Reduced-Rank Regression via Orthogonal Rotation for Analysis of High-Dimensional Multi-Source Data," *Statistics and Computing*, 34, 2. [1344]
- Li, G., Liu, X., and Chen, K. (2018), "Integrative Multi-View Regression: Bridging Group-Sparse and Low-Rank Models," *Biometrics*, 75, 593–602. [1344]
- Li, Y., Nan, B., and Zhu, J. (2015), "Multivariate Sparse Group Lasso for the Multivariate Multiple Linear Regression with an Arbitrary Group Structure," *Biometrics*, 71, 354–363. [1345]
- Luo, S., and Chen, Z. (2020), "Feature Selection by Canonical Correlation Search in High-Dimensional Multiresponse Models with Complex Group Structures," *Journal of the American Statistical Association*, 115, 1227–1235. [1345]
- Polson, N. G., Scott, J. G., and Windle, J. (2013), "Bayesian Inference for Logistic Models Using Pólya-Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349. [1346]
- Raftery, A. E. (1995), "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111–163. [1347]
- Reinsel, G. C., and Velu, R. P. (2006), "Partially Reduced-Rank Multivariate Regression Models," *Statistica Sinica*, 16, 899–917. [1344,1345]
- Reinsel, G. C., Velu, R. P., and Chen, K. (2022), *Multivariate Reduced-Rank Regression: Theory, Methods and Applications* (2nd ed.), volume 225 of Lecture Notes in Statistics. New York, NY: Springer. [1344,1347]
- Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods* (Vol. 2), New York: Springer. [1346]
- Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86. [1347]

- Tierney, L., Kass, R. E., and Kadane, J. B. (1989), “Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions,” *Journal of the American Statistical Association*, 84, 710–716. [1347]
- van Dyk, D. A., and Park, T. (2008), “Partially Collapsed Gibbs Samplers,” *Journal of the American Statistical Association*, 103, 790–796. [1346]
- Velu, R. P. (1991), “Reduced Rank Models with Two Sets of Regressors,” *Journal of the Royal Statistical Society, Series C*, 40, 159–170. [1344]
- Yang, D., Goh, G., and Wang, H. (2022), “A Fully Bayesian Approach to Sparse Reduced-Rank Multivariate Regression,” *Statistical Modelling*, 22, 199–200. [1347,1348]