Statistics and Computing (2025) 35:110
https://doi.org/10.1007/s11222-025-10629-3

ORIGINAL PAPER

Uncertainty Quantification in Bayesian Reduced-Rank Sparse Regressions

Maria F. Pintado[1] · Matteo Iacopini[2] · Luca Rossini[3,5] · Alexander Y. Shestopaloff[1,4]

Received: 18 October 2024 / Accepted: 29 April 2025

Abstract

Reduced-rank regression recognises the possibility of a rank-deficient matrix of coefficients. We propose a novel Bayesian model for estimating the rank of the coefficient matrix, which obviates the need for post-processing steps and allows for uncertainty quantification. Our method employs a mixture prior on the regression coefficient matrix along with a global-local shrinkage prior on its low-rank decomposition. Then, we rely on the Signal Adaptive Variable Selector to perform sparsification and define two novel tools: the Posterior Inclusion Probability uncertainty index and the Relevance Index. The validity of the method is assessed in a simulation study, and then its advantages and usefulness are shown in real-data applications on the chemical composition of tobacco and on the photometry of galaxies.

Keywords Mixture prior · Reduced-rank regression · Sparse estimation · Uncertainty quantification · Variable selection

1 Introduction

A common problem found in several fields ranging from economics and finance to biology and medicine is the need to study the relationship between multiple response variables and their predictors. Multivariate linear regression (MLR) is an extensively used model that provides a straightforward interpretation of the relationship between a group of responses and a common set of predictors. This model is particularly useful when dealing with data with multiple dependent variables, as it allows for the analysis of the joint effect of the predictors on these variables. For each observational unit $i = 1, \dots, n$ from a sample of size n, let $\boldsymbol{y}_i \in \mathbb{R}^q$ be a response variable explained by p possible predictors $\boldsymbol{x}_i \in \mathbb{R}^p$. Let $Y \in \mathbb{R}^{n\times q}$ be the matrix of responses with the ith row as $\boldsymbol{y}_i'$, and $X \in \mathbb{R}^{n\times p}$ the matrix of explanatory variables with the ith row as $\boldsymbol{x}_i'$. The multivariate linear regression model is defined as

$$Y = XC + E, \qquad E = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)', \tag{1}$$

where the rows $\boldsymbol{e}_i$ of the error matrix are independent and normally distributed with mean zero and $q \times q$ covariance matrix Σ. If we assume a relatively small share of nonzero entries of the coefficient matrix C, then a sparse multivariate linear regression (sparse MLR) is obtained, which allows for sparse estimation (e.g., see Goh et al. 2017,).

On the other hand, standard reduced-rank regression (RR) postulates that the dependence structure between the responses and the covariates can be represented through a small number of linear combinations of the latter. For

All authors contributed equally to this work.

✉ Maria F. Pintado
m.f.pintadoserrano@qmul.ac.uk

Matteo Iacopini
miacopini@luiss.it

Luca Rossini
luca.rossini@unimi.it

Alexander Y. Shestopaloff
a.shestopaloff@qmul.ac.uk

1 School of Mathematical Sciences, Queen Mary University of London, 327 Mile End Road, London E1 4NS, United Kingdom

2 Department of Economics and Finance, LUISS University, Viale Romania 32, Rome 00197, Italy

3 Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio 7, Milan 20122, Italy

4 Department of Mathematics and Statistics, Memorial University of Newfoundland, 230 Elizabeth Avenue, St. John's A1C 5S7, Canada

5 Fondazione Eni Enrico Matteo, Corso Magenta 63, Milan 20123, Italy

Published online: 04 June 2025

instance Reinsel et al. (2022) used multivariate linear regression to study the influence of certain covariates on the chemical properties of a urine specimen in the context of reduced-rank regression. Jöreskog and Goldberger (1975) applied a similar model in sociology considering unobservable latent variables. Gudmundsson (1977) constructed linear combinations of variables in an econometric model of the United Kingdom to capture aspects of the economic situation.

Specifically, an RR model assumes the coefficient matrix C to be rank deficient, implying a reduction in the number of parameters through linear restrictions on the entries of C. Consequently, a reduced-rank decomposition of the coefficient matrix allows to obtain $C = BA'$, with $A \in \mathbb{R}^{q\times r}$ and $B \in \mathbb{R}^{p\times r}$ dependent on the rank r of C. The first approach to this method was made by Anderson (1951), who proposed a class of regression models that restrict C to be rank deficient. Izenman (1975) introduced the term reduced rank regression for these models and provided a further study of the estimators.

A closely related topic of research is low-rank matrix completion, where the aim is to complete the missing entries of a matrix from a partial observation (see Alquier 2013, for a revision of Bayesian methods). The similarity with reduced-rank regression lies in the estimation of a low-rank matrix, where the rank is user-defined (Lim and Teh 2007; Salakhutdinov and Mnih 2008) or implicitly estimated (Babacan et al. 2011). The previous authors assign Gaussian priors with mean zero on A and B, which is similar to our method. This is in contrast to the proposed entry-wise uniform prior of Mai and Alquier (2015), which can similarly be used within our proposed mixture approach.

The reduced rank regression model is also connected to the econometrics literature on cointegration, where the rank r of a coefficient matrix identifies the number of long-run common trends among the endogenous response variables. In this framework, several attempts have been made to infer r from the data (e.g., see Kleibergen and Paap 2002; Chua and Tsiaplias 2018; Strachan 2003,) using techniques such as Bayes' factors to compare nested models with varying ranks, or estimating posterior probabilities of different ranks with approaches such as Markov Chain Monte Carlo (MCMC). Similarly, the low-rank decomposition $C = BA'$ is closely related to the principal components analysis (PCA) and sparse factor analysis (FA). Standard PCA can be obtained by setting $Y = X$ and introducing an intercept term (Izenman 2008, ch.7), where r represents the number of principal components used to capture most of the variability of X. Instead, in sparse FA, A is a vector of latent factors, and B corresponds to a sparse matrix of factor loadings, with r corresponding to the number of latent factors. Within these frameworks, a recent stream of literature considers the problem of rank estimation (i.e., determining the number of components or factors) in a Bayesian perspective, such as Šmídl and Quinn (2007); Sobczyk et al. (2017); Ning and Ning (2024) for PCA and Lopes and West (2004); Viroli (2009); Ghosh and Dunson (2009); Frühwirth-Schnatter et al. (2025) in FA, among others. We believe that an appropriate modification of the statistical approach proposed in this article could be directly applied to each of these frameworks.

From a Bayesian perspective, Geweke (1996) pioneered the early work on reduced-rank regression by assigning independent Gaussian priors on the elements of the coefficient matrix conditioning on the rank, assumed to be known. In addition, Geweke proposed to use predictive odds ratios for regression models with different ranks when r is unknown to identify a true model. Although further methods treating the rank, r, as unknown have been developed, they typically treat it as a parameter to be fixed before performing the inference (Chen and Huang 2012; Goh et al. 2017). The literature that treats the rank as an unknown quantity relies on post-processing steps to estimate it, for example, by thresholding the singular values (Bunea et al. 2011; Chen et al. 2013; Chakraborty et al. 2019). The performance of post-processing methods typically depends on some user-specified tuning parameters, whose choice is hardly justifiable; moreover, it does not allow uncertainty quantification. The literature on nonparametric reduced-rank regression treats the rank as a tuning parameter chosen by cross-validation (Lian and Ma 2013), estimates it by thresholding singular values along with parameters to be selected by the user (Mukherjee 2013) or by imposing regularization penalties (Foygel et al. 2012). Additionally, the nonparametric techniques equally fail to fully incorporate the quantification of uncertainty. To address these issues, we propose a Bayesian Rank Estimation and Covariate Selection (BRECS) method, with the crucial difference that the rank is estimated jointly with all the other parameters, in a single-step fully Bayesian approach. As such, the proposed method allows for uncertainty quantification and removes the need for a post-processing scheme (i.e., two-step approach). The key to our method is to assume a finite mixture prior on the coefficient matrix, with mixture components corresponding to fixed possible values of the rank variable.

Sparsity-inducing estimators of the coefficient matrix have been used to overcome the over-parametrization and potential over-fitting that typically characterise high-dimensional models. Therefore, several authors have expanded Bayesian methods for incorporating sparsity into reduced rank regression (Zhu et al. 2014; Goh et al. 2017; Chakraborty et al. 2019; Yang et al. 2022). We impose a global-local shrinkage prior on the columns of B, which encourages sparsity in the matrix of coefficients (Bhattacharya et al. 2015).

Global-local shrinkage priors ensure that the coefficients are pulled towards zero, but exact sparsification (i.e., estimated coefficients exactly equal to zero) is prevented by

Table 1 Comparison of linear models in terms of reduced rank and sparse estimation

Model	Dimension reduction	Rank estimation	Sparse estimation	Unc.Quant. on rank	Unc.Quant. on sparsity
MLR	✗	✗	✗	✗	✗
Sparse MLR	✗	✗	✓	✗	✗
RR	✓	✓	✗	✗	✗
BRECS	✓	✓	✓	✓	✓

the continuity of the prior. This requires an additional step for coefficient selection. Consequently, the uncertainty about this mechanism also becomes relevant. Ray and Bhattacharya (2018) proposed the Signal Adaptive Variable Selector (SAVS) to post-process a point estimate, such as the posterior mean, and group coefficients into exact zeros and non-zeros. Huber et al. (2021) applied the SAVS to each MCMC draw of the parameter of interest, thus obtaining a posterior inclusion probability (PIP) for each coefficient. We adopt a similar strategy and apply SAVS to each element of the coefficient matrix C; then, we derive a new PIP uncertainty index to quantify uncertainty in coefficient selection (see Table 1). A related work by Yuchi et al. (2023) provides another way of defining a prior (conditional on the rank) and different uncertainty quantification measures for a low-rank matrix in the context of matrix completion.

In addition to the PIP uncertainty index, we also introduce the Relevance Index (RI). The PIP uncertainty index allows us to quantify uncertainty about variable inclusion. The RI, which is the most important one, is a distribution representing the relevance of a covariate in terms of the share of response variables on which it has a significant impact. This index provides full uncertainty quantification. Moreover, we propose a rule of thumb for variable selection that summarises and complements the information embedded in the RI by means of its survival function.

Uncertainty quantification in both rank estimation and variable selection is currently underdeveloped in the literature. Yang et al. (2022) address this issue by using the Laplace approximation of the posterior distributions within a collapsed Gibbs sampler and obtaining complete sparse rows of C for covariate selection. By coupling our mixture prior on C with the shrinkage prior on B, our proposed method enables sparse and low-rank estimation of the coefficient matrix by sampling from the exact full-conditional posteriors, obviating the demand for an approximation. Our approach removes the need for post-processing while jointly incorporating the quantification of uncertainty. So, unlike current techniques, we do not rely on visual inspection of the plots and on user-specified thresholds to choose the rank. Instead, we allow for a simple and transparent interpretation based on the entire posterior distribution of the rank. Practically being able to quantify the uncertainty in the rank allows a user to assess how well the data guides in choosing a specific rank value in a concrete setting. Besides, differently from Yang et al. (2022), our approach is able to obtain a sparse estimate of C where either entire rows or only single entries are null. The PIP reports uncertainty quantification on the estimates of the rank and the coefficients, and then the RI is used to measure the uncertainty about variable selection when working with a multivariate response.

Simulation studies were conducted to evaluate the performance of the proposed methodology in various scenarios, and compared to other relevant methods. The effectiveness of the algorithm is validated as well in real data application to datasets on the chemical composition of tobacco and on the photometry of galaxies. For the tobacco dataset, we obtain comparable results with Izenman (2008) while requiring the estimation of a single model and uncertainty quantification without relying on a subjective judgement. For the galaxy experiment, we provide a strong and interesting variable selection. Furthermore, the proposed relevance index and its survival function allow us to select covariates with different degrees of uncertainty, which is not possible with the commonly used post-processing methods.

The remainder of the article is organised as follows. Section 2 introduces the model, and presents the proposed priors for rank selection. Then, Section 3 demonstrates our sampling algorithm and provides different definitions of uncertainty quantification. Section 4 illustrates the performance of the proposed methods in simulated experiments, while Section 5 applies them to two real datasets. Finally, Section 6 provides a discussion.

2 Reduced-rank regression model

This Section presents the proposed Bayesian approach to reduced-rank regression in the context of the multivariate linear regression model described in eq. (1). Reduced-rank regression (RR) identifies a smaller set of variables that can explain a large proportion of the variation in the data. To consider this model, an assumption on the rank of the matrix of coefficients C is defined as $\text{rank}(C) = r \leq \min(p, q)$. Such an assumption translates into fewer parameters, leading to a more parsimonious model. In the literature, the constraint on the rank has been done by assuming that r is known (Geweke 1996), or declaring it as unknown but fixed as in Chen and Huang (2012) and Goh et al. (2017). Typically, there is no guidance in fixing a specific rank for C in real data applications, motivating the choice of treating it as an unknown quantity. However, previous results following this approach rely on post-processing schemes to estimate the rank with user-tuned parameters (Chakraborty et al. 2019; Mai 2021). In the proceeding of the paper, we consider the low-rank decomposition $C = BA'$ with $B \in \mathbb{R}^{p\times r}$ and $A \in \mathbb{R}^{q\times r}$, where the rank r needs to be estimated.

2.1 Mixture prior for rank selection

In this paper, we contribute to the literature on reduced rank regression by proposing a new Bayesian approach for rank estimation and related uncertainty quantification. Our proposal relies on finite mixture priors, which combine two or more prior probability distributions, namely the mixture components, each with its own set of parameters. The main aspect that favours a mixture prior is that rank selection, which corresponds to an automatic model choice, is made along with parameter estimation. In the proposed mixture prior, each component corresponds to a different rank. The Bayesian approach to inference coupled with data augmentation allows for obtaining a posterior distribution for the rank. Besides deriving a point estimate as the maximum a posteriori (MAP), our approach permits uncertainty quantification, a novel feature of this method in contrast to existing literature.

In detail, we define a finite mixture prior on C made by a number of components equal to $r_{\max} = \min(p, q)$, assumed to be q. Employing the notation C_s for the matrix C under the restriction $\text{rank}(C) = s$, the prior is expressed as

$$p(C) = \sum_{s=1}^{q} w_s p(C_s), \tag{2}$$

where w_s is the prior probability of $\text{rank}(C) = s$. Denoting $\mathbf{w} = (w_1, \ldots, w_q)$, we assume a Dirichlet prior distribution with parameter $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)$, that is $\mathbf{w} \sim \mathcal{D}ir(\boldsymbol{\gamma})$. We introduce a latent allocation variable u which assumes values in the set $\{1, \ldots, q\}$, and follows a prior Categorical distribution with parameter $\mathbf{w}$, represented as

$$p(u|\mathbf{w}) = \prod_{s=1}^{q} w_s^{\mathbb{I}(u=s)}, \tag{3}$$

where $\mathbb{I}(u = s)$ is the indicator function, taking value 1 if $u = s$ and 0 otherwise.

To define a prior on C conditional on its rank u, we rely on the low-rank representation $C = BA'$, and assume prior independence between A and B, which results in

$$p(C_u) = p(A_u)p(B_u), \tag{4}$$

where A_u and B_u are the factors of the rank-u decomposition of C, reminding that their dimensions are rank dependent, being $q \times u$ and $p \times u$, respectively. For the entries of the matrix A_u, we consider a standard normal prior $a_{jh} \sim \mathcal{N}(0, 1)$, with $j = 1, \ldots, q$ and $h = 1, \ldots, u$. Regarding the prior specification for B_u, we use a global-local shrinkage prior on each column $\boldsymbol{b}_h = (b_{1h}, b_{2h}, \ldots, b_{ph})' \in \mathbb{R}^p$ of B_u, for $h = 1, \ldots, u$. This family of distributions consists of a hierarchical scale mixture of (multivariate) Gaussian distributions of the type

$$\begin{aligned} b_{lh}|\tau_h, \phi_{lh} &\sim \mathcal{N}(0, \tau_h\phi_{lh}), \\ \tau_h \sim \pi_{\tau_h}(\tau_h), &\qquad \phi_{lh} \sim \pi_{\phi_{lh}}(\phi_{lh}), \end{aligned} \tag{5}$$

where τ_h and ϕ_{lh} are the global and local components of the variance, respectively, with distributions $\pi_{\tau_h}(\cdot)$ and $\pi_{\phi_{lh}}(\cdot)$.[1] In this article, we consider a Dirichlet-Laplace prior (Bhattacharya et al. 2015; Cross et al. 2020), which can be represented as

$$\begin{aligned} b_{lh}|\psi_{lh}, \tau_h, \phi_{lh} &\sim \mathcal{N}(0, \psi_{lh}\tau_h^2\phi_{lh}^2), \quad l = 1, \ldots, p \\ \tau_h|\alpha_h &\sim \mathcal{G}a(\alpha_h p, 1/2), \\ \boldsymbol{\phi}_h|\alpha_h &\sim \mathcal{D}ir(\alpha_h, \ldots, \alpha_h), \\ \psi_{lh} &\sim \mathcal{E}xp(1/2), \\ \alpha_h &\sim \mathcal{U}(L_\alpha, U_\alpha), \end{aligned} \tag{6}$$

where $\mathcal{G}a(\cdot)$ and $\mathcal{D}ir(\cdot)$ denote the Gamma (with the shape-rate parametrization) and Dirichlet distributions, respectively. As usual with hierarchical priors, the performance of the DL prior depends on the hyperparameter values, particularly on α_h. To address this issue, similarly to Cross et al. (2020), we assume a continuous uniform prior for α_h to let data inform about the degree of shrinkage. For each entry l in column h, the local shrinkage parameter is ψ_{lh}, and the vector of global shrinkage is $(\tau_h\phi_{1h}, \ldots, \tau_h\phi_{ph})$, where $\boldsymbol{\phi}_h = (\phi_{1h}, \ldots, \phi_{ph})'$ is constrained to lie in the $(p-1)$ simplex Δ^{p-1}. Finally, we impose no restrictions on the covariance matrix Σ, and place an inverse Wishart prior, $\Sigma \sim \mathcal{IW}(\nu, \Upsilon)$.

As a consequence of the mixture prior on the matrix of coefficients C, the observed likelihood function is a mixture distribution with the same weights. Denoting $\mathbf{A} =$

[1] By specifying the distributions of the two variance components and eventually introducing additional layers in the hierarchy, it is possible to generate a wide range of prior distributions that shrink the coefficients toward zero while allowing the data to inform about large deviations from the origin thanks to the heavy tails of the marginal prior (obtained integrating out the global component τ).

$\{A_1, \ldots, A_q\}$ and $\mathbf{B} = \{B_1, \ldots, B_q\}$, the likelihood is represented as

$$p(Y|\mathbf{A}, \mathbf{B}, \Sigma, \mathbf{w}) = \sum_{s=1}^{q} w_s (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{n}{2}} \times \exp\left\{ -\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(Y - XB_sA_s')' \times (Y - XB_sA_s')]\right\} \tag{7}$$

and the complete data likelihood is

$$p(Y, u|\mathbf{A}, \mathbf{B}, \Sigma, \mathbf{w}) = p(Y|\mathbf{A}, \mathbf{B}, \Sigma, u, \mathbf{w})p(u|\mathbf{w}) = \prod_{s=1}^{q} \left(\frac{1}{(2\pi)^{nq/2}|\Sigma|^{n/2}} \exp -\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(Y - XB_sA_s')' (Y - XB_sA_s')] w_s \right)^{\mathbb{I}(u=s)}. \tag{8}$$

2.2 Alternative prior parametrizations

The previous section has introduced a generic prior structure for the matrices A_u and B_u, conditional on the rank of C being u. However, no direct connection was assumed between A_u and A_v, for $u \neq v$ (similarly for B_u). In this section, we leverage on the particular structure imposed by the reduced-rank assumption for C to define two alternative prior parametrizations for the matrices A_u and B_u. In particular, we propose two main parametrizations for A_u and B_u: the naïve (RRn) and the column-sharing (RRcs). The first case provides the best unconditional approximation of the matrix of coefficients with the estimated rank, a parametrization that results in a computationally intensive algorithm with $O(q^3 + q^2 p)$ parameters. In contrast, RRcs is based on sharing information across columns, leading to the best conditional approximation, a reduced number of parameters of $O(q^2 + pq)$, and a computationally faster MCMC for posterior inference compared to the former approach.

The auxiliary variable u represents the (unknown) rank of C, thus implicitly determining the number of columns of A and B. Within the RRn parametrization, each value of u is associated with a specific collection of matrices $A_u \in \mathbb{R}^{q\times u}$, $B_u \in \mathbb{R}^{p\times u}$ and the corresponding parameters of the hierarchical prior for each column of B_u, that is $(\phi, \tau, \psi, \alpha)$. As u ranges from 1 to q, there are in total q collections of parameters, also differing in the number of elements included in each collection (Figure 1a). Instead, in the RRcs parametrization, moving from u to $u+1$ implies sharing the same parameters as in u, plus an additional column of the matrices A_{u+1}, B_{u+1} (and the corresponding parameters ϕ, τ, ψ, α). Therefore, the sharing mechanism conditions the reduced-rank approximation of the matrix C_{u+1} to the *even lower* rank approximation C_u (Figure 1b).

To summarise, the crucial difference is the parametrization of the collection of matrices $\{A_u, B_u\}_{u=1}^q$. Denoting with $\mathbf{a}_h^{(u)}$ and $\mathbf{b}_h^{(u)}$ the hth column of the matrices A_u and B_u, the RRn parametrization assumes:

$$\begin{aligned} A_1 &= [\mathbf{a}_1^{(1)}] & B_1 &= [\mathbf{b}_1^{(1)}] \\ A_2 &= [\mathbf{a}_1^{(2)}, \mathbf{a}_2^{(2)}] & B_2 &= [\mathbf{b}_1^{(2)}, \mathbf{b}_2^{(2)}] \\ &\vdots \\ A_q &= [\mathbf{a}_1^{(q)}, \mathbf{a}_2^{(q)}, \ldots, \mathbf{a}_q^{(q)}] & B_q &= [\mathbf{b}_1^{(q)}, \mathbf{b}_2^{(q)}, \ldots, \mathbf{b}_q^{(q)}]. \end{aligned} \tag{9}$$

Conversely, the RRcs parametrization assumes:

$$\begin{aligned} A_1 &= [\mathbf{a}_1^{(1)}] & B_1 &= [\mathbf{b}_1^{(1)}] \\ A_2 &= [\mathbf{a}_1^{(1)}, \mathbf{a}_2^{(2)}] & B_2 &= [\mathbf{b}_1^{(1)}, \mathbf{b}_2^{(2)}] \\ &\vdots \\ A_q &= [\mathbf{a}_1^{(1)}, \mathbf{a}_2^{(2)}, \ldots, \mathbf{a}_q^{(q)}] & B_q &= [\mathbf{b}_1^{(1)}, \mathbf{b}_2^{(2)}, \ldots, \mathbf{b}_q^{(q)}] \end{aligned} \tag{10}$$

The first parametrization produces a different estimate of the hth column of each low-rank matrix: $\mathbf{a}_h^{(u)} \neq \mathbf{a}_h^{(v)}$, for $u \neq v$, and $h \leq \min(u, v)$. In contrast, the second parametrization assumes that $\mathbf{a}_h^{(u)} = \mathbf{a}_h^{(v)}$. The same rationale applies to matrix B.

Concerning the prior construction for the columns of either matrix, we assume the same distributions for the RRn and RRcs, that is $\mathbf{a}_h^{(u)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and eq. (6) for each $\mathbf{b}_h^{(u)}$.

3 Posterior sampling

In this section, we yield the estimation details of the proposed mixture prior and derive the model uncertainty quantification indexes. Initially, we provide the representation of the joint posterior distribution of the model parameters. Let us define $\Psi = (\boldsymbol{\psi}_1 = (\psi_{11}, \ldots, \psi_{p1}), \ldots, \boldsymbol{\psi}_q = (\psi_{1q}, \ldots, \psi_{pq}))$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$, $\Phi = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_q)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)$. The parameters can be included in $\Theta = (\Sigma, \mathbf{w}, \mathbf{A}, \mathbf{B}, \Psi, \boldsymbol{\tau}, \Phi, \boldsymbol{\alpha})$, and the joint posterior distribution is given by

$$p(\Theta, u|Y) \propto p(Y|\Theta, u)\, p(\mathbf{A}|u)\, p(\mathbf{B}|\Psi, \boldsymbol{\tau}, \Phi, u)\, p(u|\mathbf{w}) \times p(\boldsymbol{\tau}|\boldsymbol{\alpha})\, p(\Phi|\boldsymbol{\alpha})\, p(\Sigma)\, p(\mathbf{w})\, p(\boldsymbol{\alpha})\, p(\Psi). \tag{11}$$

The choice of the prior distributions allows for a straightforward implementation of an efficient Markov Chain Monte Carlo (MCMC) algorithm to update the parameters by sampling from the full conditional posterior distributions (see the Supplement for a detailed derivation). The proposed Gibbs sampler for RRn and RRcs are outlined in Algorithm 1 and

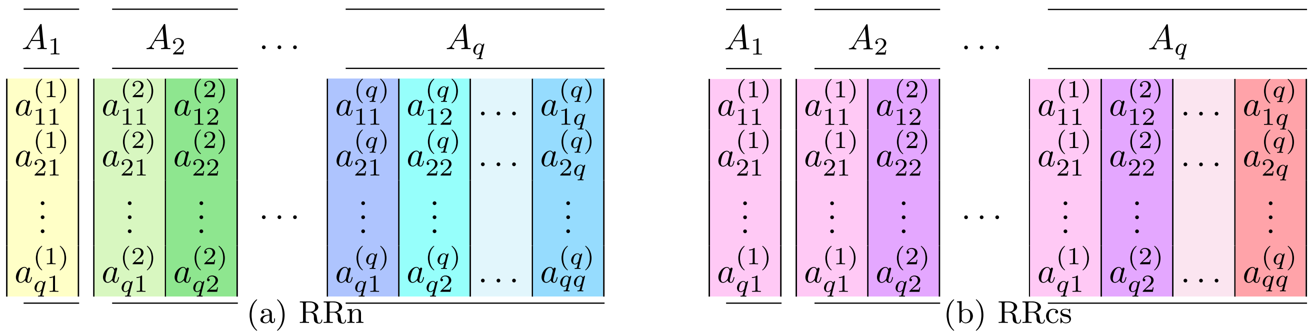


Fig. 1 Matrix A_u under both parametrizations: the naïve (panel a) and the column-sharing (panel b). Each colour represents a set of values for the corresponding columns of A_u. In RRn, the elements of the first (and only) column of A_1 are different from the first column of $A_2, \ldots, A_q$. In RRcs, the elements of the first (and only) column in A_1 are the same as those of the first column of $A_2, \ldots, A_q$

Algorithm 2, respectively, where GiG(·) and iG(·) denote the generalised inverse Gaussian and the inverse Gaussian distributions, and a $*$ superscript denotes the value of the hyper-parameters of the posterior distribution. A comprehensive description of hyperparameter specifications can be found in the Supplement.

Algorithm 1 Gibbs sampler for RRn specification

Sample u from the posterior distribution on a logarithmic scale through inverse transform sampling;
Sample $\mathbf{w}$ from $\mathcal{D}ir(\boldsymbol{\gamma}^*)$;
Sample Σ from $\mathcal{IW}(\nu^*, \Upsilon^*)$;
for $s = 1$ **to** $r_{\max}$ **do**
  if $s = u$ **then**
    Sample A_u by drawing vec(A_u) from $\mathcal{N}_{qu}(\boldsymbol{\mu}^*_{A_u}, \Sigma^*_{A_u})$;
    Sample B_u by drawing vec(B'_u) from $\mathcal{N}_{pu}(\boldsymbol{\mu}^*_{B_u}, \Sigma^*_{B_u})$;
  else
    Sample A_s and B_s from the prior;
  end if
  for $h = 1$ **to** s **do**
    Sample τ_h from GiG$(p^*_{\tau_h}, a^*_{\tau_h}, b^*_{\tau_h})$;
    Sample $\tilde{\psi}_{lh}$ from iG$(a^*_{\psi_{lh}}, b^*_{\psi_{lh}})$, then set $\psi_{lh} = \tilde{\psi}^{-1}_{lh}$, for each $l = 1, \ldots, p$;
    Sample T_{lh} from GiG$(p^*_{\phi_{lh}}, a^*_{\phi_{lh}}, b^*_{\phi_{lh}})$, then set $\phi_{lh} = T_{lh} / \sum_{i=1}^{p} T_{ih}$, for each $l = 1, \ldots, p$;
    Sample α_h from its full conditional using a griddy Gibbs sampler (Ritter and Tanner 1992).
  end for
end for

The algorithm of the column-sharing approach (RRcs) performs significantly faster than the naïve (RRn) parametrization since at each iteration of the MCMC, the number of columns updated (through the posterior distribution or the prior) for **A** and **B** is $r_{\max}$, the maximum possible rank, as opposed to $r_{\max}(r_{\max} + 1)/2$ in the naïve approach. In addition to faster computational performance, the column-sharing case presents a better mixing around the rank estimate (see Section 4).

Algorithm 2 Gibbs sampler for RRcs specification

Sample u from the posterior distribution on a logarithmic scale through inverse transform sampling;
Sample $\mathbf{w}$ from $\mathcal{D}ir(\boldsymbol{\gamma}^*)$;
Sample Σ from $\mathcal{IW}(\nu^*, \Upsilon^*)$;
for $s = 1$ **to** u **do**
  Sample A_u by drawing each column $\boldsymbol{a}_s$ from $\mathcal{N}_q(\mu^*_{\boldsymbol{a}_s}, \Sigma^*_{\boldsymbol{a}_s})$;
  Sample B_u by drawing each column $\boldsymbol{b}_s$ from $\mathcal{N}_p(\mu^*_{\boldsymbol{b}_s}, \Sigma^*_{\boldsymbol{b}_s})$;
end for
for $s = u + 1$ **to** $r_{\max}$ **do**
  Sample columns $\boldsymbol{a}_s$ and $\boldsymbol{b}_s$ from the prior;
end for
for $s = 1$ **to** $r_{\max}$ **do**
  Sample τ_s from GiG$(p^*_{\tau_s}, a^*_{\tau_s}, b^*_{\tau_s})$;
  Sample $\tilde{\psi}_{ls}$ from iG$(a^*_{\psi_{ls}}, b^*_{\psi_{ls}})$, then set $\psi_{ls} = \tilde{\psi}^{-1}_{ls}$, for each $l = 1, \ldots, p$;
  Sample T_{ls} from GiG$(p^*_{\phi_{ls}}, a^*_{\tau_{ls}}, b^*_{\tau_{ls}})$, then set $\phi_{ls} = T_{ls} / \sum_{i=1}^{p} T_{is}$, for each $l = 1, \ldots, p$;
  Sample α_s from its full conditional using a griddy Gibbs sampler (Ritter and Tanner 1992).
end for

3.1 Variable selection

Further interest is placed on obtaining exact zeros in the coefficient matrix C for covariate selection, and uncertainty quantification of covariates inclusion becomes relevant. Shrinkage priors allow for parameter estimates that are very close to zero but not exactly zero, and the nature of hierarchical shrinkage priors makes common MCMC methods for sparsification that rely on cross-validation to be computationally prohibitive. However, Ray and Bhattacharya (2018) introduced the signal adaptive variable selector (SAVS), a simple algorithm for the sparsification step on the posterior mean of the parameter of interest. Employing SAVS allows to have exact zeros, but uncertainty quantification remains uncovered. Huber et al. (2021) apply the SAVS method to sparsify every MCMC draw in the shrinkage step, thus allowing for parameter uncertainty. By following their approach to sparsify each draw, we allow for parameter uncertainty quantification. This yields the following estimate to obtain a sparse draw of C_{jk} at the mth iteration of the MCMC:

$$\bar{C}_{jk}^{(m)} = \text{sign}\left(C_{jk}^{(m)}\right) \|X_j\|^{-2} \left(|C_{jk}^{(m)}| \; \|X_j\|^2 - |C_{jk}^{(m)}|^{-2}\right)_+ \tag{12}$$

with $X_j = (X_{1j}, \ldots, X_{nj})'$ denoting the jth column of the matrix X, $(x)_+ = \max(x, 0)$ and $\text{sign}(x) = 1$ for $x \geq 0$ and -1 otherwise. Considering an MCMC of length M iterations, by applying the SAVS at each iteration m, we obtain a collection of M sparse estimates of every element C_{jk}. Therefore, the posterior probability that the coefficient C_{jk} is not zero is the proportion of MCMC iterations such that the estimate $\bar{C}_{jk} \neq 0$. We define this proportion as the posterior inclusion probability (PIP):

$$\text{PIP}_{jk} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{I}\left(\bar{C}_{jk}^{(m)} \neq 0\right). \tag{13}$$

The posterior estimate of the jkth entry of C is set to 0 if $\text{PIP}_{jk} \leq 0.5$ and to the posterior mean of $\bar{C}_{jk}$ if $\text{PIP}_{jk} > 0.5$.

By definition, the PIP is a value between 0 and 1, where a PIP close to 0 indicates that the corresponding entry is not likely to be important. In contrast, a PIP close to 1 suggests that the inclusion of the entry is well-supported by the data. Hence, for both high and low PIPs, the decision about including an element is less uncertain, opposite to PIPs that lie around 0.5. Even though PIPs are, by nature, a quantification of uncertainty, their direct interpretation may not appear evident to the reader as the degree of certainty is not monotonous. The need for a straightforward and easily interpreted manner to quantify uncertainty about variable inclusion leads us to the following definition.

Definition 1 (PIP uncertainty index) Let us assume the posterior inclusion probability (PIP) as in Eq. (13), the PIP uncertainty index, which takes values in [0, 1] is defined as:

$$\zeta_{jk} = 1 - 2\left|\text{PIP}_{jk} - 0.5\right|, \tag{14}$$

where ζ_{jk} close to 0 (to 1) means low (high) uncertainty about the decision of setting the jkth entry of C to an exact 0 or not.[2]

The following example illustrates the previous definition.

[2] The quantity ζ_{jk} is in strategy comparable to the Bernoulli variance. The linear transformation of PIP_{jk} allows equal weights for all probabilities, in contrast to different lengths in the range of values, pointing to different levels of variation in the Bernoulli variance. In this sense, ζ_{jk} facilitates the interpretability of the results. A sensitivity analysis about the choice of ζ_{jk} is included in the Supplement.

Example 1 Let us consider three different entries: jk, jk^* and jk^{**}. We assume $\text{PIP}_{jk} = 0.49$, $\text{PIP}_{jk^*} = 0.94$, and $\text{PIP}_{jk^{**}} = 0.01$. Then, the point estimates are $\hat{C}_{jk} = \hat{C}_{jk^{**}} = 0$, whereas $\hat{C}_{jk^*} = M^{-1} \sum_{m=1}^{M} \bar{C}_{jk*}^{(m)}$. The associated uncertainty indices are $\zeta_{jk} = 0.98$, $\zeta_{jk^*} = 0.12$, and $\zeta_{jk^{**}} = 0.02$, meaning that the decision of setting $\hat{C}_{jk^{**}}$ to 0 and $\hat{C}_{jk^*}$ to the posterior mean have low uncertainty (as $\zeta_{jk^{**}} = 0.02$ and $\zeta_{jk^*} = 0.12$). Conversely, the decision of setting to 0 the entry $\hat{C}_{jk}$ is very uncertain (since $\zeta_{jk} = 0.98$).

The element-wise PIP and the associated uncertainty index, ζ, provide information on what coefficients are nonzero and quantify the uncertainty about this statement. Instead, variable selection procedures are concerned with statistical techniques designed to identify and eliminate the subset of irrelevant covariates from the regression model. The PIP-based approach previously described can easily address this issue in univariate settings, but multivariate regressions call for the adoption of different methods as a prediction can have different impacts on each response variable. In fact, it may be possible that a covariate is irrelevant to predict a subset of the responses but exerts an influence on the remaining ones. In the Supplement, we report a variable selection method based on a single scalar quantity analogous to the group lasso of Chakraborty et al. (2019) and another one based on the PIP previously defined.

To address the limitation of the element-wise approach, we propose a novel index that assesses the relative importance of each covariate and is computationally inexpensive, as it relies on the output of the SAVS computed at every iteration of the MCMC.

Definition 2 (Relevance Index) The relevance index of the jth covariate, RI_j, with $j = 1, \ldots, p$, is defined as:

$$\text{RI}_j(k) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{I}\left(N_{\bullet j}^{(m)} = kq\right), \qquad k = 0, \frac{1}{q}, \frac{2}{q}, \ldots, 1, \tag{15}$$

where $N_{\bullet j}^{(m)} = \sum_{k=1}^{q} \mathbb{I}(\bar{C}_{jk}^{(m)} \neq 0)$ is the number of nonzero entries in the jth row of $\bar{C}$. Notice that RI_j is a discrete distribution supported on $\mathcal{D} = \{0, 1/q, 2/q, \ldots, 1\}$, with mass at each $k \in \mathcal{D}$ given by Eq. (15).

The RI can be interpreted as representing the distribution (across MCMC) of the "relevance" of a covariate, as measured by the share of response variables on which the covariate exerts a significant impact (i.e., nonzero). This motivates the support being the discrete grid between 0 and 1, with step size $1/q$. Therefore, a strongly right-skewed distribution suggests that the covariate is irrelevant or relevant just to a small share of response variables, whereas a left-skewed distribution is typical of common predictors that impact all the responses. An important feature of the RI is that it easily allows for uncertainty quantification by means of the variance of the distribution. For instance, take two right-skewed distributions, RI_j and RI_k, characterised by different variances, $\sigma_j^2 > \sigma_k^2$. In this case, we have evidence in favour of considering both x_j and x_k irrelevant, but with higher uncertainty of this statement about x_j. Eventually, for sufficiently large variance, a binary statement about the irrelevance of the covariate may appear hazardous.

The approach to variable selection based on the RI allows the assessment of the relevance of each variable and provides full uncertainty quantification. In practice, one is often concerned with identifying and excluding irrelevant predictors. The shape and dispersion of the RI_j probability mass function intrinsically inform the impact exerted by the covariate in the overall model and the degree of uncertainty about this belief. Besides, when the practitioner needs to make a binary decision about variable selection, it is possible to summarise the information content of the RI to answer this question. A possible heuristic for choosing whether to include or exclude a covariate relies on the tail distribution (or survival function) of RI_j, as described in the following rule of thumb.

Definition 3 (Rule of thumb for variable selection) Let $\overline{sr} \in [0, 1]$ be the share of response variables on which the jth covariate is required to have a significant impact, and $\overline{p} \in (0, 1)$ be the desired minimum probability. Then, a *rule of thumb* consists in excluding the jth covariate if the following statement is not satisfied:

$$S_{\mathrm{RI}_j}(\overline{sr}) = \mathbb{P}(\mathrm{RI}_j > \overline{sr}) \geq \overline{p}, \tag{16}$$

where the probability is computed with respect to the (discrete) distribution of RI_j.

However, we remark that any measure for a binary decision will inevitably lose part of the information embedded in the RI. Therefore, when variable selection is concerned, we suggest to *jointly* interpret the output of the binary rule and the RI distribution. The following example illustrates the application of the rule of thumb to make a binary decision about variable selection.

Example 2 We set $\overline{sr} = 0.70$ and $\overline{p} = 0.60$; thus, we require the RI of the jth covariate to assign at least 0.60 total probability mass on or above 0.70. In other words, this means that not to exclude the jth covariate, we require the probability of being relevant for more than the 70% of responses to be at least 0.60. This example is represented in Figure 2, which considers two covariates, characterised by a right-skewed and a left-skewed RI. The associated survival functions are plotted together with the values of $\overline{sr}$ and $\overline{p}$ (vertical and horizontal dashed lines, respectively). The *rule of thumb* in Eq. (16) states that the covariate should be considered irrelevant if the point $(\overline{sr}, \overline{p})$ lies above the curve of the survival function (i.e., outside the shaded area), and vice versa. Therefore, in this simple example, the first covariate, which typically impacts a few response variables (as shown by the right-skewed RI, first plot), is considered irrelevant (second plot). Conversely, the other covariate has a nonzero impact on most responses (see the left-skewed RI in the third plot), and the irrelevance hypothesis is rejected for it (fourth plot). However, notice that this decision does not account for the uncertainty quantified by the RI. In this case, the variance of the RI for the second covariate is high, thus implying that the selection decision for this variable should be taken with caution. Instead, the RI for the first covariate is quite low, which suggests high confidence in the decision to exclude it from the model.

4 Simulation study

We study the performance of the proposed reduced-rank model with the naïve (RRn) and the column-sharing (RRcs) priors across a range of simulation settings. Our primary objectives in conducting this simulation study are twofold: firstly, to assess the efficacy of the model in accurately estimating the rank in varied settings, including different data generating processes (DGP), and secondly, to evaluate the recovery of the coefficient matrix under distinct scenarios.

The data was generated from the multivariate linear model $Y = XC_0 + E$, where we considered correlated and uncorrelated structures on the errors and regressors of the model. The rows of X were independently drawn from $\mathcal{N}(0, \Sigma_X)$, with $\Sigma_X = I_p$ for independent regressors, and in the dependent case, the off-diagonal entries of Σ_X are set to 0.5. The rows of E were drawn from a zero mean multivariate normal distribution; under the assumption of uncorrelated errors, the covariance matrix is diagonal with elements sampled from $\mathcal{U}(0.5, 1.75)$; for correlated errors, we consider the compound symmetry as in X. We work with centred responses and exclude the intercept term for simplicity.

Recalling the decomposition of the matrix of coefficients C_0 into the product of two matrices A_0 and B_0, the entries of both matrices are generated from the standard Gaussian, and their dimensions depend on the true rank $r_0 < r_{\max}$. We consider two cases for the DGP of matrix B_0: non-sparse DGP, where the number of nonzero rows is equal to p, and sparse DGP, where the number of nonzero rows is $p^* < p$. Furthermore, our coefficient matrix estimation method is not limited to low-rank structures. We have also tested our approach on an additional "random zeros" DGP, where a share of entries z of the matrix C is randomly set to zero (Figures 4e, 4f). The performance evaluation of the estimator $\hat{C}$ of the coefficient matrix was conducted by considering the

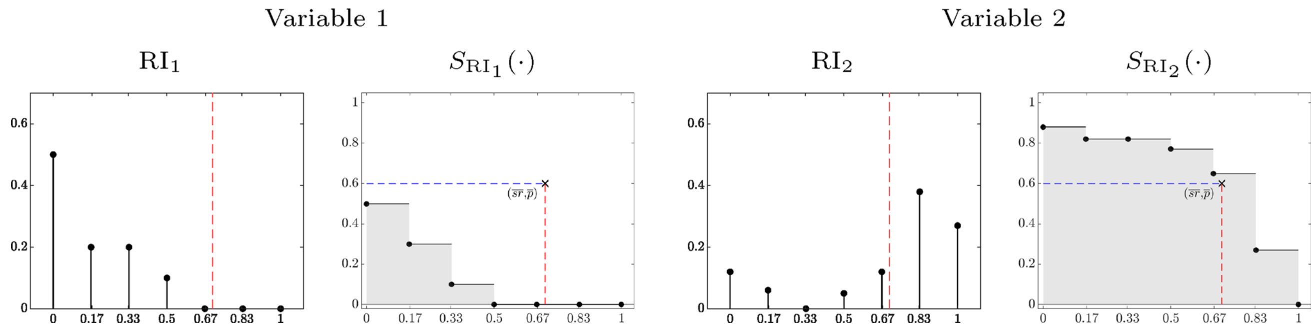


Fig. 2 Example of RI for two fictitious covariates: variable 1 (first and second plots) and variable 2 (third and fourth). The probability mass function (first and third plots) and the survival function (second and fourth) of the respective RI are reported for each covariate. Here we used $\overline{sr} = 0.70$ (red vertical line) and $\overline{p} = 0.60$ (blue horizontal line). The area below the survival function is shaded: if the point $(\overline{sr}, \overline{p})$ is located outside the shaded area, then the covariate is to be excluded

mean squared error, MSE $= \|\hat{C} - C_0\|_F^2/(pq)$, where $\|\cdot\|_F$ is the Frobenius norm.[3]

4.1 Simulation results

The RRn tends to underestimate the value of the rank across the majority of settings, particularly when the number of covariates and responses increases. As more data become available, the posterior variance of the rank parameter shrinks, resulting in the concentration of the posterior distribution. Therefore, the MCMC algorithm is further restricted in its ability to explore alternative values, resulting in a more pronounced underestimated rank. However, in all these cases, the estimated coefficient matrix $\hat{C}$ is quite close to the true C in all settings, succeeding in identifying the entries with high magnitude, and the estimates improve as the sample size grows.

After having shown the performance of the RRn, we provide evidence of the results for the RRcs approach. The column-sharing exhibits superior performance to the naïve parametrization in terms of the MSE, convergence to the true rank and computational resources. The posterior distribution of u tends to concentrates around the true value as n increases (see Figure 3). However, when B is sparse, the posterior distribution tends to put more mass on ranks smaller than r_0, resulting in a slight underestimation of the rank as the dimensionality of (q, p) increases. A possible motivation for these results is that having zero rows in B implies sparsity in C as well, which induces our method to prefer approximate C with a small r than to introduce additional parameters (higher r). This feature of the model can be interpreted as favouring more parsimonious parametrizations. Notice that the underestimation of r in these cases has little impact, as $\hat{C}$ is nonetheless close to C_0 (Figure 4). RRcs consistently achieves a better mixing of the MCMC chain compared to RRn. Notably, in both parameterizations, sparsity contributes to the improved mixing.[4]

The performance of RRcs deteriorates more rapidly for a fixed n as the number of responses q increases compared to the number of covariates p. Conversely, the performance decays slower as p increases and q remains unchanged. A change of p to p' means $(p' - p)r_{\max}$ more parameters to estimate. However, when q increases to q', the number of elements in A and Σ is directly affected, changing by $(q' - q)r_{\max}$ and $(q' - q)n$, respectively. The crucial point is that q represents the maximum rank in our context ($q \leq p$). Therefore, if q increases, it increases the number of the mixture prior components, their respective weights, and the dimension of B. Whether B is sparse or not, it appears to have no impact on this trend.

We emphasise that our analysis relies on a sparse estimate of C, obtained using the SAVS method described in Section 3.1. This estimation approach sets some entries to exact zeros, facilitating the variable selection. This can be considered a binary classification problem, wherein the positive class encompasses the nonzero entries (representing significant coefficients). In contrast, the null entries (indicating irrelevant coefficients) belong to the negative category. Consequently, our task involves identifying the position of zero and nonzero coefficients.

To evaluate the performance of this classification task, we employ the Matthews correlation coefficient (MCC), which is a more reliable statistical measure compared to commonly used metrics such as F_1 score and accuracy (Chicco et al. 2020). One notable advantage of MCC, particularly relevant to our study, is its robustness in scenarios where one class contains significantly more samples than the other, thus addressing the issue of imbalanced datasets. Our synthetic

3 See section 3.2 of the Supplement for a CODA analysis for the posterior distribution of the rank and the estimates of the matrix C.

4 Summary results comparing RRcs against RRn are included in Section 3.1 of the Supplement.

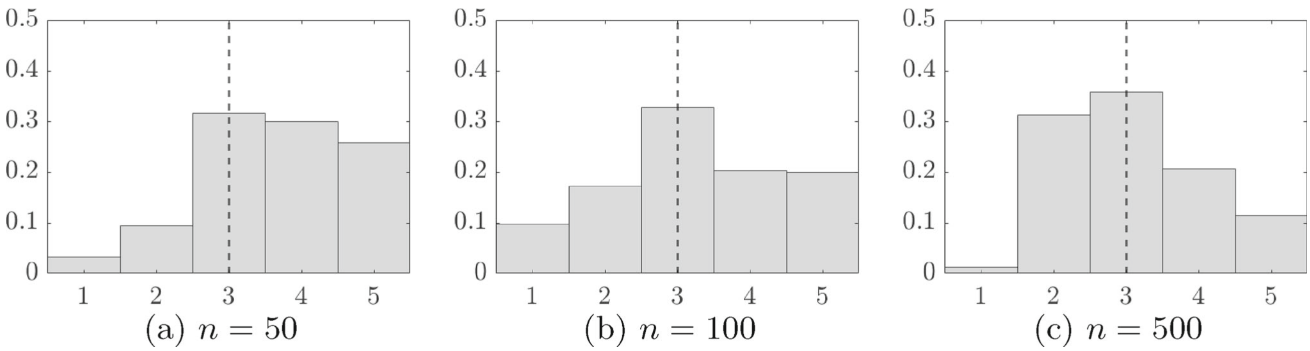


Fig. 3 Posterior distribution of the rank in the non-sparse simulation setting with $(q, p) = (5, 10)$ and true rank $r_0 = 3$ (dashed line) for different sample sizes, $n \in \{50, 100, 500\}$

data for C_0 repeatedly exhibits such imbalanced characteristics: in the non-sparse DGP when the majority of entries, if not all, deviate from zero, in the sparse DGP when only a few (or many) rows contain zero entries, and in the random zeros DGP when the percentage of zeros is low (or high). This varying distribution of zeros in different scenarios allows us to assess the classification performance of our method under varying levels of sparsity and imbalance.

The classification model predicts the class for each data instance, assigning a predicted label (positive or negative) to each sample. Depending on their actual class and their forecasted class, every sample is categorised into one of the following cases: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The MCC is given by:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \in [-1, 1]. \tag{17}$$

A value of -1 indicates poor performance, while a value of 1 represents the highest level of performance. MCC produces a high-quality score only if the prediction obtained good results in all categories (TP, TN, FP, FN), proportionally both to the size of positive and negative elements in the dataset. Instead, the measure is undefined when any of the factors in the denominator is 0, and specific mathematical reasoning should be considered. For instance, if C_0 comprises only nonzero (zero) entries, and they are all correctly identified in $\hat{C}$, then $\text{TN} = 0$ ($\text{TP} = 0$) and $\text{FN} = 0$ ($\text{FP} = 0$), resulting in an undefined MCC. Nonetheless, the classifier successfully identifies all samples in this case and achieves a perfect score of 1. If all samples belong to the same class and are all incorrectly predicted, then $\text{MCC} = -1$. We are left to study the cases when mixed samples are categorised into the same class or homogeneous samples are allocated to mixed classes. In either case, the correlation coefficient can be approximated by 0 (see Chicco et al. 2020, for a detailed description).

We report the MCC, the true positive rate (TPR = TP/(TP+FN)) of correctly identified nonzero entries, and the false negative rate (FNR = FN/(TP+FN)) of incorrectly identified zero entries across different simulation settings in Table 2. The non-negative MCCs obtained by our methods suggest a good performance in the estimation of the coefficient matrix, favouring once more the RRcs parametrization over RRn. Having attained MCCs close to 1, we show the robustness of the former method in accurately approximating the C matrix while incorporating sparsity in its estimation. The standard DGP illustrates the need for the corrections to the formula of the MCC when undefined since C_0 consisted of only nonzero entries; meanwhile, the estimated $\hat{C}$ had mixed values. The algorithm established 16% of the entries as 0, thus allowing for a sparse model with enhanced interpretability.

4.2 Comparison to other methods

The performance of the proposed mixture prior RRcs is comparable to other state-of-the-art methodologies in reduced-rank regression, as we showcase in this section. We evaluate our approach against the frequentist methods of Chen et al. (2013) and She and Chen (2017). The former utilises an adaptive nuclear norm penalisation approach (ANN), where the rank is estimated by the threshold of singular values. The latter proposes a robust reduced-rank regression approach (RRRR), which requires the user to choose the optimal rank, a task that is achieved through a suggested criterion.

The data was generated as outlined in Section 4. For each configuration, we conducted 20 replications of the experiment.

Consequently, the results presented reflect the average estimated rank, the share of replications where the rank was correctly selected, the MSE across these repetitions, and an additional error metric $\varrho(C) = \|\hat{C}(\hat{C}'\hat{C})^{+}\hat{C}' -$

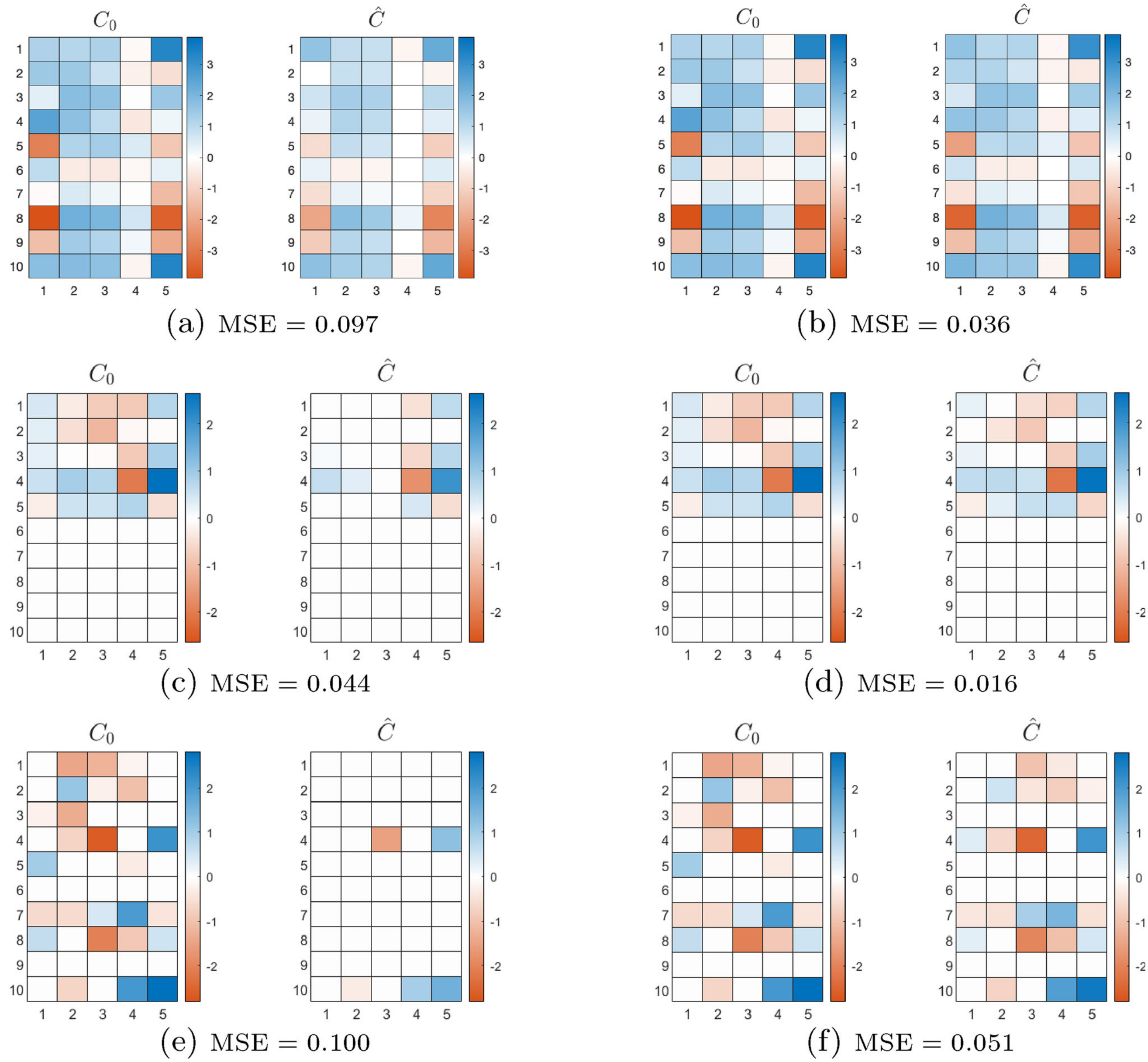


Fig. 4 True (C_0) and estimated ($\hat{C}$) coefficient matrix. Data generated as described in Section 4, with $(q, p) = (5, 10)$ and $n = 100$; non-sparse matrix B ($p^* = p$) and $r_0 = 3$ (top); sparse matrix B ($p^* = 5$) and $r_0 = 3$ (middle); $z = 50\%$ of entries of C set to 0 (bottom). Results for the RRn (left - (a),(c),(e)) and RRcs (right - (b),(d)(f)), together with the MSE

$C(C'C)^{+}C'\|_2$. Notably, our method achieves similar outcomes as both ANN and RRRR, particularly outperforming them when the sample size is $n = 50$ and the dimensions of q and p increase (see Table 3 and Table 4). While the hit rate for estimating the exact value of the rank tends to be low for all methods considered, our approach is able to provide an exact posterior estimate for r. Indeed, we find that with our method the true rank value tends to have substantial posterior mass. The metric $\varrho(C)$ of our method outperforms the one of its competitors or has comparable performance in almost all the simulated scenarios. We defer additional results to the Supplement.

5 Applications

This section showcases the efficacy of our proposed method through its practical implementation on two actual datasets. By providing empirical demonstrations of the method, we illustrate its value and effectiveness in addressing real-life situations.

Table 2 Measures of association between the true coefficient matrix C_0 and the sparse estimate $\hat{C}$ in the setting $(q, p) = (5, 10)$, $n = 100$, $r_0 = 3$ in the standard and sparse scenarios. p^* represents the number of nonzero rows in B (and consequently in C_0), and z is the proportion of randomly allocated zero entries in C_0. For RRn and RRcs, we report the MCC, TPR and FNR

DGP		**RRn** **MCC**	**TPR**	**FNR**	**RRcs** **MCC**	**TPR**	**FNR**
Standard	$p^* = 10$	—	0.84	0.16	—	0.94	0.06
Sparse	$p^* = 2$	0.00	0.00	1.00	0.81	0.70	0.30
	$p^* = 5$	0.53	0.44	0.56	0.78	0.76	0.24
	$p^* = 8$	0.13	0.08	0.92	0.54	0.68	0.32
	$p^* = 9$	0.38	0.62	0.38	0.83	0.96	0.04
Random zeros	$z = 0.20$	0.13	0.08	0.92	0.73	0.85	0.15
	$z = 0.50$	0.33	0.20	0.80	0.73	0.80	0.20
	$z = 0.80$	0.00	0.00	1.00	0.74	0.60	0.40
	$z = 0.90$	0.00	0.00	1.00	0.88	0.80	0.20

Table 3 Comparison of the estimated rank ($\hat{r}$), share of correctly estimated rank ($r_{\%}$), mean squared error of C (MSE(C)), and $\varrho(C) = \|\hat{C}(\hat{C}'\hat{C})^{+}\hat{C}' - C(C'C)^{+}C'\|_2$ obtained by RRcs against ANN (Chen et al. 2013) and RRRR (She and Chen 2017) for different values of (q, p) and true rank r_0. In all settings, $n = 50$, and the DGP is sparse with $p^* = 5$ if $p = 15$, while $p^* = 10$ if $p = 50$. We present the average estimates over 20 repetitions for independent errors (Σ_{ind}), correlated errors (Σ_{corr}), independent regressors (X_{ind}), and correlated regressors (X_{corr})

(q,p)	r_0	**X**	**Measure**	Σ_{ind} **ANN**	**RRRR**	**RRcs**	Σ_{corr} **ANN**	**RRRR**	**RRcs**
(5,15)	3	X_{ind}	$\hat{r}$	2.250	2.400	3.150	2.250	2.250	2.800
			$r_{\%}$	0.300	0.450	0.400	0.350	0.350	0.450
			MSE(C)	0.028	0.026	0.019	0.030	0.030	0.022
			$\varrho(C)$	0.789	0.714	1.000	0.759	0.762	0.992
		X_{corr}	$\hat{r}$	2.200	2.150	3.050	2.050	1.850	2.450
			$r_{\%}$	0.300	0.250	0.450	0.150	0.050	0.500
			MSE(C)	0.046	0.051	0.036	0.042	0.049	0.031
			$\varrho(C)$	0.797	0.831	1.000	0.923	0.977	0.992
	5	X_{ind}	$\hat{r}$	3.100	3.250	4.450	3.050	3.100	3.750
			$r_{\%}$	0.000	0.000	0.600	0.000	0.000	0.300
			MSE(C)	0.038	0.034	0.017	0.034	0.036	0.022
			$\varrho(C)$	1.000	1.000	0.000	1.000	1.000	0.046
		X_{corr}	$\hat{r}$	2.600	2.900	4.500	3.050	3.000	3.800
			$r_{\%}$	0.000	0.000	0.750	0.000	0.000	0.450
			MSE(C)	0.071	0.062	0.035	0.059	0.064	0.050
			$\varrho(C)$	1.000	1.000	0.049	1.000	1.000	0.155
(5,50)	3	X_{ind}	$\hat{r}$	0.050	5.000	2.350	0.050	5.000	1.650
			$r_{\%}$	0.000	0.000	0.250	0.000	0.000	0.150
			MSE(C)	1.754	16.800	0.108	0.451	4.279	0.083
			$\varrho(C)$	1.000	1.000	1.000	1.000	1.000	1.000
		X_{corr}	$\hat{r}$	0.250	5.000	2.800	0.200	5.000	1.900
			$r_{\%}$	0.000	0.000	0.350	0.000	0.000	0.150
			MSE(C)	0.984	34.732	0.068	0.954	37.709	0.100
			$\varrho(C)$	1.000	1.000	1.000	1.000	1.000	1.000
	5	X_{ind}	$\hat{r}$	0.100	5.000	4.400	0.000	5.000	3.600
			$r_{\%}$	0.000	1.000	0.700	0.000	1.000	0.300
			MSE(C)	3.178	1843.949	0.043	0.930	25.676	0.058
			$\varrho(C)$	1.000	0.993	0.807	1.000	0.990	0.800
		X_{corr}	$\hat{r}$	0.250	5.000	4.900	0.000	5.000	4.350
			$r_{\%}$	0.000	1.000	0.950	0.000	1.000	0.650
			MSE(C)	0.859	19.219	0.042	1.056	39.181	0.056
			$\varrho(C)$	1.000	0.991	0.796	1.000	0.987	0.830

Table 3 continued

(q,p)	r_0	X	Measure	Σ_{ind} ANN	RRRR	RRcs	Σ_{corr} ANN	RRRR	RRcs
(10,50)	3	X_{ind}	$\hat{r}$	1.650	10.000	1.000	1.650	10.000	1.000
			$r_{\%}$	0.300	0.000	0.000	0.450	0.000	0.000
			MSE(C)	5.800	165.912	0.242	23.358	415.744	0.265
			$\varrho(C)$	0.960	1.000	1.000	0.953	1.000	1.000
		X_{corr}	$\hat{r}$	0.950	10.000	1.100	1.650	10.000	1.000
			$r_{\%}$	0.100	0.000	0.000	0.250	0.000	0.000
			MSE(C)	2.824	15.210	0.257	2.787	49.317	0.301
			$\varrho(C)$	0.985	1.000	1.000	0.986	1.000	0.994
	5	X_{ind}	$\hat{r}$	0.150	10.000	1.100	0.050	10.000	1.000
			$r_{\%}$	0.000	0.000	0.000	0.000	0.000	0.000
			MSE(C)	1.034	100.673	0.531	0.891	18.591	0.503
			$\varrho(C)$	1.000	1.000	1.000	1.000	1.000	0.997
		X_{corr}	$\hat{r}$	0.400	10.000	1.350	0.450	10.000	1.050
			$r_{\%}$	0.000	0.000	0.000	0.000	0.000	0.000
			MSE(C)	0.990	78.371	0.538	3.082	52.587	0.543
			$\varrho(C)$	1.000	1.000	1.000	1.000	1.000	0.998

5.1 Chemical composition of tobacco

The dataset on the chemical composition of $n = 25$ tobacco leaf samples, taken from Anderson and Bancroft (1952), illustrates how our proposed methodology produces comparable results as those obtained in the literature with further advantages addressed subsequently.

Tobacco leaves are made up of organic and inorganic chemical constituents, and the typical interest is investigating the relationship between certain constituents. The $p = 6$ covariates are per cent nitrogen, per cent chlorine, per cent potassium, per cent phosphorus, per cent calcium, and per cent magnesium. We also include an intercept term. The $q = 3$ response variables are the rate of cigarette burn in inches per 1,000 seconds, the percentage of sugar in the leaf, and the percentage of nicotine in the leaf.

Under the column-sharing parametrization, the posterior distribution for the rank, as illustrated in Figure 5, shows no significant difference between any of the three values, identifying a potential uniform posterior distribution for r. Our result agrees with the findings of Izenman (2008), who uses the rank trace method. This procedure first estimates the coefficient matrix for each rank that minimises a weighted sum of squares criterion and the residual covariance matrix. Then, the rank is gradually increased, and the entries in both matrices will change significantly until the true rank is reached, where the matrices will stabilise. The change in the residual covariance matrix at each increment of r is plotted against the change in C in a scatterplot. Finally, the rank of C is assessed as the smallest rank for which the differences are close to 0. The rank-trace plot applied to the tobacco dataset shows that the rank-1 or rank-2 solutions have no discernible difference between them and the full-rank solution. The main drawback of this method is that conclusions on the effective dimensionality of the multivariate regression involve subjective judgement and visual interpretation of the rank trace plot (see Izenman 2008, for more details).

To statistically validate our hypothesis that the posterior distribution of the rank is uniform, we conduct a Pearson's chi-squared test for goodness of fit, obtaining a p-value of 0.1595, thus implying the non-rejection of the null hypothesis. Formal statistical testing for uniformity yields a more objective result as opposed to relying purely on visual analysis. Moreover, our method requires the estimation of one model, whereas the rank trace performs the estimation of three distinct models. Besides the estimates of the rank and the coefficient matrix, we provide in conjunction uncertainty quantification as an objective means of analysis, moving away from reliance on subjective judgement.

Regarding the estimated coefficient matrix for each rank, our algorithm produces consistent estimates compared to those reported in Izenman (2008), with the aforementioned advantages.

5.2 COMBO-17 galaxy photometric data

The second dataset consists of a subset of a public catalogue of astronomical objects, COMBO-17 (Classifying Objects by Medium-Band Observations in 17 filters), a project of international collaboration aimed at exploring the evolution of galaxies (Wolf et al. 2004). The present dataset is utilised herein to illustrate the proposed variable selection procedure

Table 4 Comparison of the estimated rank ($\hat{r}$), share of correctly estimated rank ($r_{\%}$), mean squared error of C (MSE(C)), and $\varrho(C) = \|\hat{C}(\hat{C}'\hat{C})^{+}\hat{C}' - C(C'C)^{+}C'\|_2$ obtained by RRcs against ANN (Chen et al. 2013) and RRRR (She and Chen 2017) for different values of (q, p) and true rank r_0. In all settings, $n = 50$, and the DGP is non-sparse with $p^* = p$. We present the average estimates over 20 repetitions for independent errors (Σ_{ind}), correlated errors (Σ_{corr}), independent regressors (X_{ind}), and correlated regressors (X_{corr})

(q,p)	r_0	X	Measure	Σ_{ind} ANN	RRRR	RRcs	Σ_{corr} ANN	RRRR	RRcs
(5,15)	3	X_{ind}	$\hat{r}$	2.900	2.800	3.650	2.950	3.000	3.350
			$r_{\%}$	0.900	0.800	0.300	0.950	1.000	0.400
			MSE(C)	0.022	0.026	0.129	0.023	0.023	0.089
			$\varrho(C)$	0.298	0.356	1.000	0.225	0.192	1.000
		X_{corr}	$\hat{r}$	2.900	2.900	3.900	3.000	3.000	3.050
			$r_{\%}$	0.900	0.900	0.300	1.000	1.000	0.150
			MSE(C)	0.040	0.042	0.102	0.037	0.041	0.195
			$\varrho(C)$	0.305	0.307	1.000	0.232	0.242	1.000
	5	X_{ind}	$\hat{r}$	4.000	4.400	4.850	3.950	4.350	4.900
			$r_{\%}$	0.000	0.400	0.900	0.000	0.400	0.900
			MSE(C)	0.059	0.034	0.151	0.044	0.031	0.102
			$\varrho(C)$	1.000	0.707	0.725	1.000	0.746	0.733
		X_{corr}	$\hat{r}$	3.900	4.400	4.950	4.000	4.350	5.000
			$r_{\%}$	0.000	0.400	0.950	0.000	0.350	1.000
			MSE(C)	0.092	0.066	0.174	0.074	0.059	0.128
			$\varrho(C)$	1.000	0.764	0.746	1.000	0.836	0.772
(5,50)	3	X_{ind}	$\hat{r}$	0.650	5.000	1.700	1.550	5.000	1.400
			$r_{\%}$	0.150	0.000	0.050	0.350	0.000	0.000
			MSE(C)	6.912	20.468	1.034	207.621	15.594	1.292
			$\varrho(C)$	0.985	1.000	1.000	0.957	1.000	1.000
		X_{corr}	$\hat{r}$	1.000	5.000	3.150	1.400	5.000	2.000
			$r_{\%}$	0.250	0.000	0.050	0.350	0.000	0.000
			MSE(C)	19.369	41.381	1.050	6.361	13.865	1.164
			$\varrho(C)$	0.953	1.000	1.000	0.977	1.000	1.000
	5	X_{ind}	$\hat{r}$	0.000	5.000	5.000	0.200	5.000	4.800
			$r_{\%}$	0.000	1.000	1.000	0.000	1.000	0.950
			MSE(C)	5.079	32.540	1.456	4.656	1399.716	1.653
			$\varrho(C)$	1.000	0.959	0.980	1.000	0.959	0.978
		X_{corr}	$\hat{r}$	0.150	5.000	5.000	0.200	5.000	5.000
			$r_{\%}$	0.000	1.000	1.000	0.000	1.000	1.000
			MSE(C)	22.558	42.882	1.869	5.952	176.258	1.874
			$\varrho(C)$	1.000	0.956	0.976	1.000	0.974	0.975
(10,50)	3	X_{ind}	$\hat{r}$	3.000	10.000	1.650	3.150	10.000	1.550
			$r_{\%}$	1.000	0.000	0.100	0.850	0.000	0.200
			MSE(C)	3473.521	1208.176	1.011	58.882	215.646	1.399
			$\varrho(C)$	0.697	1.000	1.000	0.793	1.000	1.000
		X_{corr}	$\hat{r}$	3.000	10.000	1.600	3.200	10.000	1.450
			$r_{\%}$	1.000	0.000	0.050	0.800	0.000	0.000
			MSE(C)	639.223	83.254	1.394	6.677	95.078	1.544
			$\varrho(C)$	0.758	1.000	1.000	0.836	1.000	1.000
	5	X_{ind}	$\hat{r}$	2.500	10.000	1.700	3.650	10.000	1.150
			$r_{\%}$	0.500	0.000	0.000	0.550	0.000	0.000
			MSE(C)	53.516	319.769	3.203	12.321	31.421	3.859
			$\varrho(C)$	0.923	1.000	1.000	0.933	1.000	1.000

Table 4 continued

(q,p)	r_0	X	Measure	Σ_{ind} ANN	RRRR	RRcs	Σ_{corr} ANN	RRRR	RRcs
		X_{corr}	$\hat{r}$	1.950	10.000	1.500	3.000	10.000	1.550
			$r_{\%}$	0.350	0.000	0.000	0.450	0.000	0.000
			MSE(C)	12.253	50.605	3.370	12.697	93.531	3.777
			$\varrho(C)$	0.946	1.000	1.000	0.917	1.000	1.000

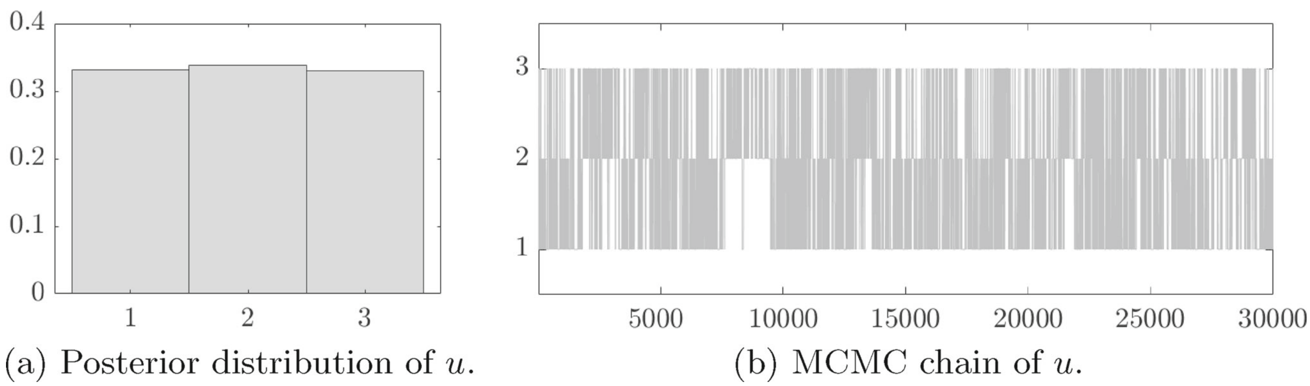


Fig. 5 Tobacco dataset: posterior distribution (panel a) and MCMC chain after burn-in (panel b) of the rank u for the coefficient matrix C

and the associated uncertainty.[5] The methodology serves as a reliable means of informing decision-making regarding selecting covariates, offering both suggestions and quantifying uncertainty in such selections.

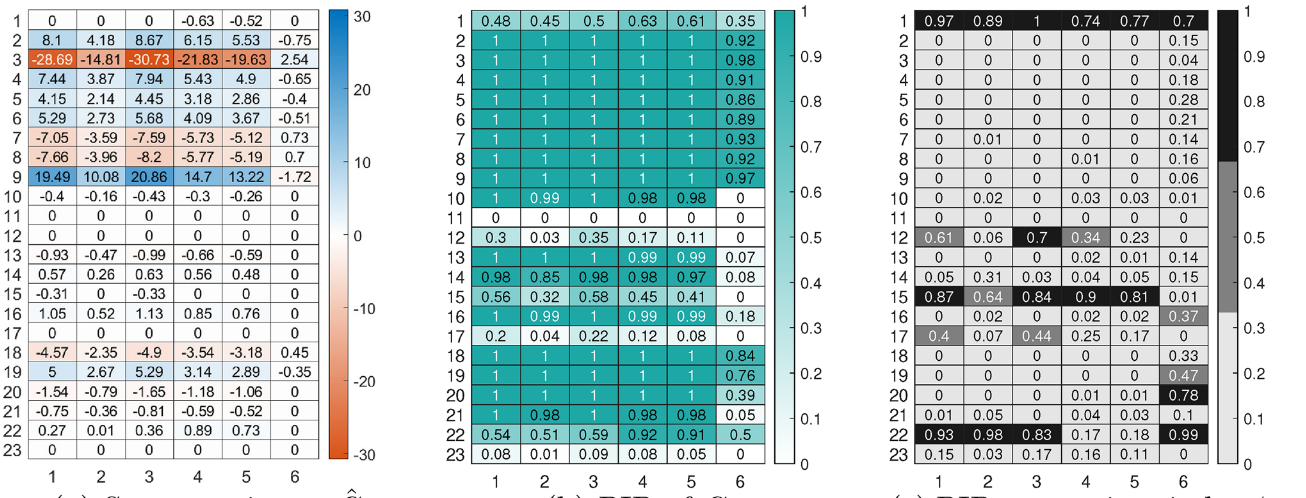
The original dataset consists of 63, 501 objects in the area of the sky named Chandra Deep Field South with brightness measurements in 17 passbands from 350 to 930 nm. We restrict the analysis to 3, 438 objects, all classified as "Galaxies" by Wolf et al. (2004), and with no missing values for any of the 65 variables. The measurement errors and five redundant variables were omitted, resulting in a total of 29 variables divided into $p = 23$ covariates and $q = 6$ responses, as done in Izenman (2008). Regarding the covariates, 10 variables correspond to the absolute magnitudes of the galaxy in 10 bands, while the remaining variables are the observed brightness in 13 bands across the range 420 − 915 nm. Meanwhile, the responses are the total R-band magnitude, the aperture difference of the R-band, the central surface brightness in the R-band, two redshift estimates, and the reduced chi-squared value of the best-fitting template galaxy spectrum.

Given that the whole subset of galaxies consists of a considerable number of $n = 3,438$ observations, the level of estimation uncertainty is likely to be quite small. Therefore, motivated by the intention of highlighting the ability to quantify the uncertainty of the proposed BRECS method, we apply it to a sub-sample of $n = 500$ observations randomly selected from the entire subset of data. The rank selection, sparse estimation, and variable selection procedures are also repeated for the full sample and for other randomly chosen sub-samples of size $n = 500$ and $n = 1,500$ (see the Supplement). In line with expectations, we find that the use of larger samples reduces the uncertainty, but the key insights about the advantages of using the BRECS method are unaltered.

The posterior distribution of the rank is right-skewed and achieves the maximum (MAP) at $\hat{u} = 2$. Considering all the 3, 438 observations, the posterior distribution is more concentrated around the same maximum point (see the Supplement for further details).

The effect of the covariates on the responses emphasises the importance of estimating the matrix C. For any pair (jk), a zero entry $\hat{C}_{jk} = 0$ means that there is no association between the jth covariate and the kth response. Therefore, zero entries facilitate interpretation as the nonzero rows of C identify the covariates that influence at least one response. We obtain a sparse estimate of C_{jk} by applying the SAVS algorithm at each iteration of the MCMC, computing its posterior inclusion probability PIP$_{jk}$, and set the element to 0 if PIP$_{jk} \leq 0.5$ or to its posterior mean otherwise, as described in Section 3. The matrix of PIPs takes values in [0, 1], and the elements of C with equal or less than 50% probability of inclusion are set to 0. We emphasise that even though some entries of the sparse estimate $\hat{C}$ are 0, the probability of including them is not exactly 0. At first inspection, the 12th covariate, the observed brightness of the galaxy in the corresponding band, is to be ruled out of the model since its coefficients are only zeros (see Figure 6a). However, not all of the PIPs of covariate 12 are close to 0, especially

[5] An additional forecasting exercise is presented in Section 4.3 of the Supplement.



(a) Sparse estimate $\hat{C}$. (b) PIP of C_{jk}. (c) PIP uncertainty index ζ_{jk}.

Fig. 6 Sparse estimate $\hat{C}$ of the coefficient matrix C of the linear regression model with the galaxy dataset of $n = 500$ observations (panel a), the uncertainty about this estimation through the posterior inclusion probabilities (panel b), and the PIP uncertainty index, in a grey-colour scale according to low ($\zeta_{jk} \leq 1/3$), medium ($1/3 < \zeta_{jk} \leq 2/3$), or high ($\zeta_{jk} > 2/3$) uncertainty (panel c)

$\text{PIP}_{12,1} = 0.30$ and $\text{PIP}_{12,3} = 0.35$ (see Figure 6b). The estimated coefficient matrix $\hat{C}$ of the galaxy dataset exposes 4 complete zero rows, thus reducing the number of covariates exerting a significant impact to 19.

Basing the decision of covariate selection solely on thresholding the PIP could lead to misleading interpretations of results. For this reason, we quantify the uncertainty about the decision of including the jkth element of C through the PIP uncertainty index ζ_{jk}. The PIP uncertainty index provides a straightforward interpretation of the probabilities observed in the matrix of PIPs as a way to quantify uncertainty about the effect of each covariate on every response, which is particularly advantageous when addressing each response separately. Notwithstanding, the decision about irrelevant covariates in the multivariate regression model is yet to be determined since some covariates may influence only a subset of the responses. For instance, covariates 1 and 15 have an apparent influence only on responses $\{4, 5\}$ and $\{1, 3\}$, respectively (Figure 6). Therefore, the relevance index RI is used to assess the relative importance of each covariate. Table 5 reports the summary statistics of the relevance index for each covariate.

As illustrated in the top panels of Figure 7, there is strong evidence for the exclusion of covariate 17 and for the inclusion of covariate 7. However, the exclusion or inclusion of covariate 22 is not apparent, for there is strong variation in the distribution, and additional study should be given due consideration.[6] The *rule of thumb* in (16) states not to exclude the jth covariate if $S_{\text{RI}_j}(\overline{sr}) = \mathbb{P}(\text{RI}_j > \overline{sr}) \geq \overline{p}$, for appropriate values of $\overline{sr}$ and $\overline{p}$. By setting $\overline{sr} = 0.70$ and $\overline{p} = 0.60$, we are excluding covariate 17 while maintaining covariate 7 in the model (bottom panels of Figure 7), as inferred from the probability mass function of RI_7 and RI_{22}, accordingly. Notice that for covariate 22, the point $(0.70, 0.60)$ lies in the rejection region for inclusion (above the curve of the tail distribution), a conclusion that would not have been reached had we required a minimum probability of 0.60 for more than 40% of the responses ($\overline{sr} = 0.60$, $\overline{p} = 0.40$). Under the previous specification of $\overline{sr}$ and $\overline{p}$, the covariates considered irrelevant in the majority of responses are 1, 11, 12, 15, 17, 22, 23.

The methodology applied to the entire subset of $n = 3,438$ objects reduces the percentage of zero elements in the sparse estimate of the coefficient matrix from 28% to 14% and the number of zero rows by 3. The quantity of covariates with medium and high PIP uncertainty indices decreases accordingly, in line with stronger information provided by the data to require more variables that explain the outcomes.

6 Discussion

We have proposed BRECS, a novel Bayesian approach for estimating the rank of the matrix of coefficients along with parameters in a reduced-rank regression model. Our method employs a mixture prior to rank estimation and shrinkage prior on the factor matrix resulting from the decomposition of the coefficient matrix for shrinking its irrelevant entries to 0. Furthermore, variable selection is achieved by adopting

[6] See Section 4.2 of the Supplement for the results based on alternative variable selection methods.

Table 5 Summary statistics of the distribution of RI in the sub-sample of $n = 500$ for each covariate: mode, mean, standard deviation (Std), and the 25th, 50th and 75th quartiles (Q25, Q50, Q75). The shaded rows identify covariates with medium uncertainty ($\text{Std}(\text{RI}_j) \geq 0.20$) and high uncertainty ($\text{Std}(\text{RI}_j) \geq 0.30$). A description of each covariate is found in the Supplement

x_j	Mode	Mean	Std	Q25	Q50	Q75
1	0	0.504	0.392	0	0.667	0.833
2	1	0.986	0.055	1	1	1
3	1	0.997	0.023	1	1	1
4	1	0.985	0.047	1	1	1
5	1	0.976	0.059	1	1	1
6	1	0.982	0.052	1	1	1
7	1	0.987	0.057	1	1	1
8	1	0.985	0.063	1	1	1
9	1	0.994	0.038	1	1	1
10	0.833	0.826	0.048	0.833	0.833	0.833
11	0	0	0.007	0	0	0
12	0	0.161	0.233	0	0	0.333

x_j	Mode	Mean	Std	Q25	Q50	Q75
13	0.833	0.843	0.049	0.833	0.833	0.833
14	0.833	0.806	0.112	0.833	0.833	0.833
15	0	0.387	0.348	0	0.333	0.667
16	0.833	0.859	0.073	0.833	0.833	0.833
17	0	0.111	0.211	0	0	0.167
18	1	0.972	0.062	1	1	1
19	1	0.96	0.071	1	1	1
20	0.833	0.897	0.083	0.833	0.833	1
21	0.833	0.831	0.068	0.833	0.833	0.833
22	0.833	0.659	0.234	0.5	0.667	0.833
23	0	0.052	0.147	0	0	0

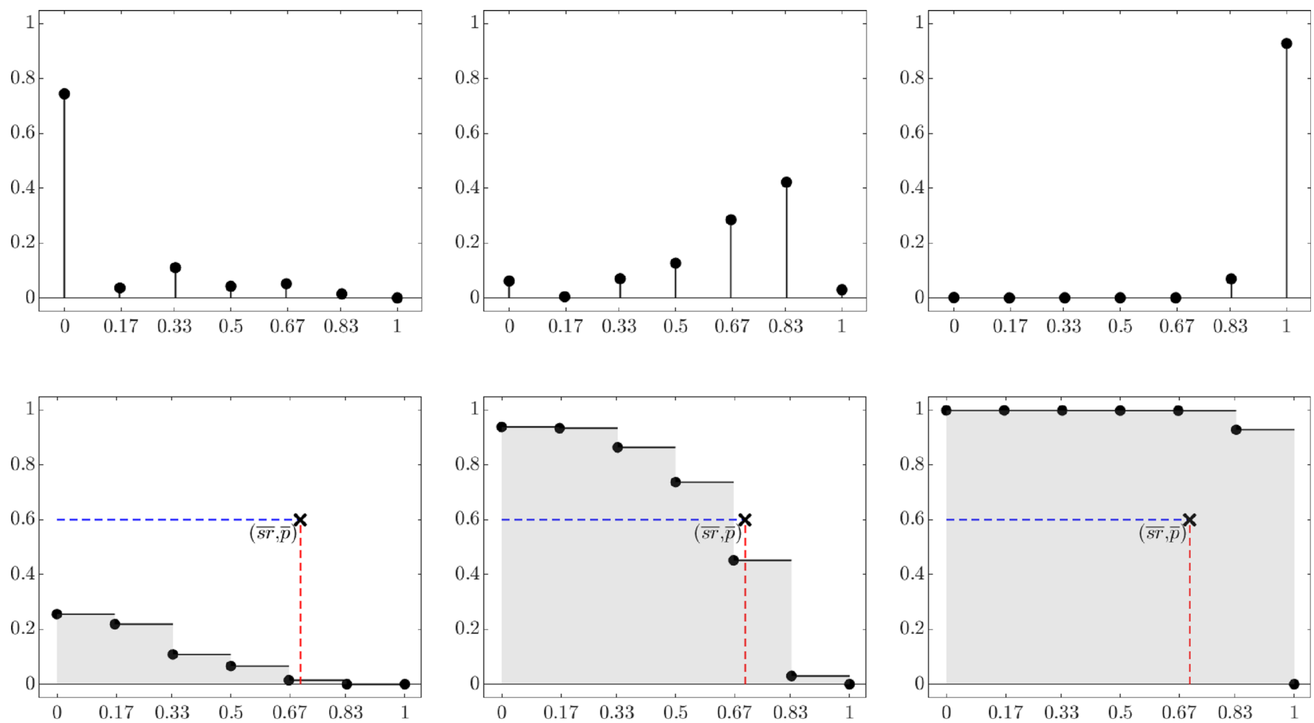


Fig. 7 Probability mass function (top) and survival function (bottom) of RI for covariates 17 (left), 22 (centre) and 7 (right). The x-axis represents the share of nonzero elements with the probabilities on the y-axis. If we set $\overline{sr} = 0.70$ (red vertical line) and $\overline{p} = 0.60$ (blue horizontal line), then covariates 17 and 22 are to be excluded, and we include covariate 7

SAVS to obtain a sparse estimate of C, in conjunction with the uncertainty about this estimation through the relevance index. By employing our method, researchers can avoid post-processing steps and obtain a quantification of uncertainty in estimating the rank, together with accurate statistical inference for the coefficient matrix.

The results of our simulation study suggest that RRcs exhibits superior performance over RRn in terms of converging more accurately to the true rank and being considerably faster in computational time. We observed that the algorithm's performance deteriorates more rapidly as the number of responses increases compared to the number of covariates. Thus, enhancing the model's scalability remains an area for future research. Overall, our method provides a reliable estimation of the coefficient matrix, even in cases where the rank is underestimated, effectively preventing over-fitting and resulting in a satisfactory approximation of the true matrix.

Our proposed approach was applied to real datasets about the chemical composition of tobacco leaves and photometric galaxy data. The obtained results are consistent with the findings presented in the literature, adding a quantification of the uncertainty about the obtained estimates. Additional domains for its applicability should be explored, including genomics (Hilafu et al. 2020) and macroeconomics (Reinsel et al. 2022). The latter explicitly suggests extending our research to the time series model and tensor regression (Billio et al. 2023; Luo 2024). Given the prevalence of models with incomplete response matrices in biostatistics, it is imperative to develop Bayesian approaches to address these scenarios. It is thus a promising avenue for further investigation (Mai and Alquier 2022).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-025-10629-3.

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

References

Anderson, R.L., Bancroft, T.A.: Statistical Theory in Research. McGraw-Hill, New York (1952)

Alquier, P.: Bayesian methods for low-rank matrix estimation: Short survey and theoretical study. In: Proc. 24th Int. Conf. Algorithmic Learning, pp. 309–23. Springer, New York (2013). https://doi.org/10.1007/978-3-642-40935-6_22

Anderson, T.W.: Estimating linear restrictions on regression coefficients for multivariate normal distributions. The Annals of Mathematical Statistics **22**, 327–351 (1951). https://doi.org/10.1214/aoms/1177729580

Billio, M., Casarin, R., Iacopini, M., Kaufmann, S.: Bayesian dynamic tensor regression. Journal of Business & Economic Statistics **41**(2), 429–439 (2023). https://doi.org/10.1080/07350015.2022.2032721

Babacan, S.D., Luessi, M., Molina, R., Katsaggelos, A.K.: Low-rank matrix completion by variational sparse Bayesian learning. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2188–2191 (2011)

Bhattacharya, A., Pati, D., Pillai, N.S., Dunson, D.B.: Dirichlet-Laplace priors for optimal shrinkage. Journal of the American Statistical Association **110**(512), 1479–1490 (2015). https://doi.org/10.1080/01621459.2014.960967

Bunea, F., She, Y., Wegkamp, M.H.: Optimal selection of reduced rank estimators of high-dimensional matrices. The Annals of Statistics **39**(2), 1282–1309 (2011). https://doi.org/10.1214/11-AOS876

Chakraborty, A., Bhattacharya, A., Mallick, B.K.: Bayesian sparse multiple regression for simultaneous rank reduction and variable selection. Biometrika **107**, 205–221 (2019). https://doi.org/10.1093/biomet/asz056

Chen, K., Dong, H., Chan, K.-S.: Reduced rank regression via adaptive nuclear norm penalization. Biometrika **100**, 901–920 (2013). https://doi.org/10.1093/biomet/ast036

Chen, L., Huang, J.Z.: Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. Journal of the American Statistical Association **107**(500), 1533–1545 (2012). https://doi.org/10.1080/01621459.2012.734178

Cross, J.L., Hou, C., Poon, A.: Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. International Journal of Forecasting **36**, 899–915 (2020). https://doi.org/10.1016/j.ijforecast.2019.10.002

Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics **21**(1) (2020) https://doi.org/10.1186/s12864-019-6413-7

Chua, C.L., Tsiaplias, S.: A Bayesian approach to modeling time-varying cointegration and cointegrating rank. Journal of Business & Economic Statistics **36**(2), 267–277 (2018). https://doi.org/10.1080/07350015.2016.1166117

Foygel, R., Horrell, M., Drton, M., Lafferty, J.: Nonparametric reduced rank regression. In: Bartlett, P., Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 25. Curran Associates, Inc., (2012)

Frühwirth-Schnatter, S., Hosszejni, D., Lopes, H.F.: Sparse Bayesian factor analysis when the number of factors is unknown. Bayesian Analysis **20**(1), 213–344 (2025). https://doi.org/10.1214/24-BA1423

Ghosh, J., Dunson, D.B.: Default prior distributions and efficient posterior computation in Bayesian factor analysis. Journal of Computational and Graphical Statistics **18**(2), 306–320 (2009). https://doi.org/10.1198/jcgs.2009.07145

Goh, G., Dey, D.K., Chen, K.: Bayesian sparse reduced rank multivariate regression. Journal of Multivariate Analysis **157**, 14–28 (2017). https://doi.org/10.1016/j.jmva.2017.02.007

Geweke, J.: Bayesian reduced rank regression in econometrics. Journal of Econometrics **75**(1), 121–146 (1996). https://doi.org/10.1016/0304-4076(95)01773-9

Gudmundsson, G.: Multivariate analysis of economic variables. Journal of the Royal Statistical Society. Series C (Applied Statistics) **26**(1), 48–59 (1977). https://doi.org/10.2307/2346867

Huber, F., Koop, G., Onorante, L.: Inducing sparsity and shrinkage in time-varying parameter models. Journal of Business & Economic Statistics **39**(3), 669–683 (2021). https://doi.org/10.1080/07350015.2020.1713796

Hilafu, H., Safo, S.E., Haine, L.: Sparse reduced-rank regression for integrating omics data. BMC Bioinformatics volume **21**(283) (2020) https://doi.org/10.1186/s12859-020-03606-2

Izenman, A.J.: Reduced-rank regression for the multivariate linear model. Journal of Multivariate Analysis **5**(2), 248–264 (1975)

Izenman, A.: Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer, New York (2008). https://doi.org/10.1007/978-0-387-78189-1
Jöreskog, K.G., Goldberger, A.S.: Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association **70**(351a), 631–639 (1975). https://doi.org/10.1080/01621459.1975.10482485
Kleibergen, F., Paap, R.: Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. Journal of Econometrics **111**(2), 223–249 (2002). https://doi.org/10.1016/S0304-4076(02)00105-7
Luo, Y., Griffin, J.E.: Bayesian inference of vector autoregressions with tensor decompositions. Journal of Business & Economic Statistics, 1–24 (2024) https://doi.org/10.1080/07350015.2024.2447302
Lian, H., Ma, S.: Reduced-rank regression in sparse multivariate varying-coefficient models with high-dimensional covariates. arXiv preprint arXiv:1309.6058 (2013)
Lim, Y., Teh, Y.: Variational Bayesian approach to movie rating prediction. In: Knowledge Discovery and Data Mining (2007)
Lopes, H.F., West, M.: Bayesian model assessment in factor analysis. Statistica Sinica, 41–67 (2004)
Mai, T.T., Alquier, P.: A Bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. Electronic Journal of Statistics **9**, 823–841 (2015). https://doi.org/10.1214/15-EJS1020
Mai, T.T., Alquier, P.: Optimal quasi-Bayesian reduced rank regression with incomplete response. arXiv preprint arXiv:2206.08619 (2022)
Mai, T.T.: Efficient Bayesian reduced rank regression using Langevin Monte Carlo approach. arXiv preprint arXiv:2102.07579 (2021)
Mukherjee, A.: Topics on reduced rank methods for multivariate regression. PhD thesis, The University of Michigan (2013)
Ning, Y.-C.B., Ning, N.: Spike and slab Bayesian sparse principal component analysis. Statistics and Computing **34**(3), 118 (2024). https://doi.org/10.1007/s11222-024-10430-8
Ray, P., Bhattacharya, A.: Signal adaptive variable selector for the horseshoe prior. arXiv preprint arXiv:1810.09004 (2018)
Ritter, C., Tanner, M.A.: Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. Journal of the American Statistical Association **87**(419), 861–868 (1992). https://doi.org/10.2307/2290225
Reinsel, G.C., Velu, R.P., Chen, K.: Multivariate Reduced-rank Regression: Theory, Methods and Applications. Lecture Notes in Statistics, vol. 225, 2nd edn. Springer, New York, NY (2022)
Sobczyk, P., Bogdan, M., Josse, J.: Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood. Journal of Computational and Graphical Statistics **26**(4), 826–839 (2017). https://doi.org/10.1080/10618600.2017.1340302
She, Y., Chen, K.: Robust reduced-rank regression. Biometrika **104**(3), 633–647 (2017). https://doi.org/10.1093/biomet/asx032
Salakhutdinov, R., Mnih, A.: Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In: Proceedings of the 25th International Conference on Machine Learning. ICML '08, pp. 880–887. Association for Computing Machinery, New York, NY, USA (2008). https://doi.org/10.1145/1390156.1390267
Šmídl, V., Quinn, A.: On Bayesian principal component analysis. Computational Statistics & Data Analysis **51**(9), 4101–4123 (2007). https://doi.org/10.1016/j.csda.2007.01.011
Strachan, R.W.: Valid Bayesian estimation of the cointegrating error correction model. Journal of Business & Economic Statistics **21**(1), 185–195 (2003). https://doi.org/10.1198/073500102288618883
Viroli, C.: Bayesian inference in non-Gaussian factor analysis. Statistics and Computing **19**, 451–463 (2009). https://doi.org/10.1007/s11222-008-9105-8
Wolf, C., Meisenheimer, K., Kleinheinrich, M., Borch, A., Dye, S., Gray, M., Wisotzki, L., Bell, E.F., Rix, H.-W., Cimatti, A., Hasinger, G., Szokoly, G.: A catalogue of the Chandra Deep Field South with multi-colour classification and photometric redshifts from COMBO-17. Astronomy & Astrophysics **421**(3), 913–936 (2004). https://doi.org/10.1051/0004-6361:20040525
Yang, D., Goh, G., Wang, H.: A fully Bayesian approach to sparse reduced-rank multivariate regression. Statistical Modelling **22**(3), 199–200 (2022). https://doi.org/10.1177/1471082X20948697
Yuchi, H.S., Mak, S., Xie, Y.: Bayesian uncertainty quantification for low-rank matrix completion. Bayesian Analysis **18**(2), 491–518 (2023). https://doi.org/10.1214/22-BA1317
Zhu, H., Khondker, Z., Lu, Z., Ibrahim, J.G.: Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. Journal of the American Statistical Association **109**(507), 977–990 (2014). https://doi.org/10.1080/01621459.2014.923775

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.