Supplementary Material for "Uncertainty Quantification in Bayesian Reduced-Rank Sparse Regressions"

May 7, 2025

Abstract

The present document provides additional information to complement the paper "Uncertainty Quantification in Bayesian Reduced-Rank Sparse Regressions". Section 1 includes detailed derivations of the full conditional distributions in the Gibbs sampler. Section 2 provides a sensitivity analysis for the PIP uncertainty index. Further simulation study results are presented in Section 3. Finally, Section 4 contains additional results on the application to the COMBO-17 galaxy dataset.

1 Details on posterior full conditionals

This section provides a comprehensive derivation of the posterior full conditional distributions of the parameters in the specified model outlined in the Main paper. The naïve parametrization, (RRn), and the column-sharing (RRcs) approach differ solely in the update process of the rank-u matrices $\mathbf{A}_u \in \mathbb{R}^{q\times u}$ and $\mathbf{B}_u \in \mathbb{R}^{q\times u}$ when the coefficient matrix has a rank of u, i.e. $\mathbf{C}_u = \mathbf{B}_u\mathbf{A}_u' \in \mathbb{R}^{q\times p}$. The specific details and distinctions between these approaches are elucidated in the respective sections below.

First, we clarify the notation used in the Supplement in accordance to the Main paper. Small bold letters represent vectors, and capital letters are reserved for matrices. Let $\mathbf{A} = \{\mathbf{A}_s : s = 1,\dots,q\}$ and $\mathbf{B} = \{\mathbf{B}_s : s = 1,\dots,q\}$ be the collections of matrices of each possible rank, $\mathbf{w} = (w_1,\dots,w_q)$, $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_q)$, $\boldsymbol{\Psi} = \{\psi_{lh} : l = 1,\dots,p \text{ and } h = 1,\dots,p\}$, $\boldsymbol{\Phi} = \{\phi_{lh} : l = 1,\dots,p \text{ and } h = 1,\dots,p\}$, $\boldsymbol{\tau} = (\tau_1,\dots,\tau_q)$, and $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_q)$. The full conditional distributions of the parameters are:

$$p(u|\mathbf{A},\mathbf{B},\mathbf{w}) \propto p(u|\mathbf{w})\, p(\mathbf{A}|u)\, p(\mathbf{B}|u),$$
$$p(\mathbf{w}|u) \propto p(\mathbf{w})\, p(u|\mathbf{w}),$$

$$
\begin{aligned}
p(\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{A},\mathbf{B},u) &\propto p(\boldsymbol{\Sigma})\, p(\mathbf{Y}|\mathbf{A},\mathbf{B},\boldsymbol{\Sigma},u),\\
p(\mathbf{A}|\mathbf{Y},\mathbf{B},\boldsymbol{\Sigma},u) &\propto p(\mathbf{A}|u)\, p(\mathbf{Y}|\mathbf{A},\mathbf{B},\boldsymbol{\Sigma},u),\\
p(\mathbf{B}|\mathbf{Y},\boldsymbol{\Sigma},\mathbf{A},\Psi,\boldsymbol{\tau},\Phi,u) &\propto p(\mathbf{B}|\Psi,\boldsymbol{\tau},\Phi,u)\, p(\mathbf{Y}|\mathbf{A},\mathbf{B},\boldsymbol{\Sigma},u),\\
p(\boldsymbol{\tau}|\mathbf{B},\Phi,\boldsymbol{\alpha},u) &\propto p(\boldsymbol{\tau}|\boldsymbol{\alpha})\, p(\mathbf{B}|\boldsymbol{\tau},\Phi,u),\\
p(\Psi|\mathbf{B},\Phi,\boldsymbol{\tau},u) &\propto p(\Psi)\, p(\mathbf{B}|\Psi,\boldsymbol{\tau},\Phi,u),\\
p(\Phi|\mathbf{B},\boldsymbol{\alpha},u) &\propto p(\Phi|\boldsymbol{\alpha})\, p(\mathbf{B}|\Phi,u),\\
p(\boldsymbol{\alpha}|\boldsymbol{\tau},\Phi) &\propto p(\boldsymbol{\alpha})\, p(\boldsymbol{\tau}|\boldsymbol{\alpha})\, p(\Phi|\boldsymbol{\alpha}).
\end{aligned}
$$

The derivation in detail of the full conditional distributions for the Gibbs sample is found hereafter.

Update u

The posterior probability that $\mathbf{C}$ has been generated from the sth component of the mixture distribution ($\text{rank}(\mathbf{C}) = s$) is

$$
\mathbb{P}(u = s|C, \mathbf{w}) = \frac{w_s\, p(\mathbf{C}_s)}{\sum_{j=1}^{q} w_j\, p(\mathbf{C}_j)}. \tag{1.1}
$$

However, the straightforward implementation of Eq. (1.1) may lead to numerical problems when the denominator is close to zero. Hence it is better to work on the logarithmic scale (Frühwirth-Schnatter, 2006).

Define $L_{\max} = \max_{s\in\{1,\dots,q\}} \log p(\mathbf{C}_s)$, and compute for each $s = 1, \dots, q$:

$$
p^*(\mathbf{C}_s) = \exp\{\log p(\mathbf{C}_s) - L_{\max}\}, \tag{1.2}
$$

as well as the sum $p^*(\mathbf{C}) = \sum_{j=1}^{q} w_j\, p^*(\mathbf{C}_j)$. Then the probability for each value of the rank is given by

$$
\mathbb{P}(u = s|C, \mathbf{w}) = \frac{w_s\, p^*(\mathbf{C}_s)}{p^*(\mathbf{C})}. \tag{1.3}
$$

To sample from Eq. 1.3, we first compute $p(\mathbf{C}_s)$ needed in Eq. 1.2 as

$$
\begin{aligned}
p(\mathbf{C}_s) &\propto p(\mathbf{A}_s)\, p(\mathbf{B}_s)\, p(\mathbf{Y}|\mathbf{A},\mathbf{B},\boldsymbol{\Sigma},u,\mathbf{w})\\
&\propto \prod_{j=1}^{q}\prod_{h=1}^{s} p(a_{jh}) \cdot \prod_{h=1}^{s}\prod_{l=1}^{p} p(b_{lh}) \cdot p(\mathbf{Y}|\mathbf{A},\mathbf{B},\boldsymbol{\Sigma},u,\mathbf{w}),
\end{aligned} \tag{1.4}
$$

where a_{jh} is the jhth entry of matrix $\mathbf{A}_s \in \mathbb{R}^{q\times s}$, and b_{lh} corresponds to the lhth element

of matrix $\mathbf{B}_s \in \mathbb{R}^{p\times s}$.

We define $\tilde{\mathbf{B}}_s$ as the s-rank matrix whose element in row l and column h is given by $b_{lh}(\psi_{lh}\tau_h^2\phi_{lh}^2)^{-1/2}$, and $\mathbf{\Lambda}_s = \text{diag}(\psi_{11}\tau_1^2\phi_{11}^2, \ldots, \psi_{1s}\tau_s^2\phi_{1s}^2, \ldots, \psi_{p1}\tau_1^2\phi_{p1}^2, \ldots, \psi_{ps}\tau_s^2\phi_{ps}^2)$. Then Eq. 1.4 transforms into

$$\begin{aligned} p(\mathbf{C}_s) \propto (2\pi)^{-s(p+q)/2}|\mathbf{\Lambda}_s|^{-1/2}\exp\left\{-\frac{1}{2}\text{tr}[\mathbf{A}_s'\mathbf{A}_s + \tilde{\mathbf{B}}_s{}'\tilde{\mathbf{B}}_s]\right\} \\ \times \exp\left\{-\frac{1}{2}\text{tr}[\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{B}_s\mathbf{A}_s')'(\mathbf{Y}-\mathbf{X}\mathbf{B}_s\mathbf{A}_s')]\right\}. \end{aligned} \tag{1.5}$$

Update w and $\mathbf{\Sigma}$

It is straightforward to derive the posterior distribution of $\mathbf{w}$ and $\mathbf{\Sigma}$ due to the conjugacy of the priors. The weights of the mixture distribution of $\mathbf{C}$ are updated from $\prod_{s=1}^{q} w_s^{\gamma_s+\mathbb{I}(u=s)-1}$, a Dirichlet distribution with parameter $\boldsymbol{\gamma}^* = (\gamma_1+\mathbb{I}(u=1), \ldots, \gamma_q+\mathbb{I}(u=q))$.
The error covariance matrix is updated by sampling from $\mathcal{IW}(\nu^*, \mathbf{\Upsilon}^*)$, where $\nu^* = \nu+n$ and $\mathbf{\Upsilon}^* = \mathbf{\Upsilon} + (\mathbf{Y}-\mathbf{X}\mathbf{C})'(\mathbf{Y}-\mathbf{X}\mathbf{C})$ are the degrees of freedom and the scale matrix, respectively.

Update $\mathbf{A}_u$

Naïve parametrization

To obtain the full conditional distribution of $\mathbf{A}_u$, we use the vectorization of the multivariate linear regression model $\mathbf{Y} = \mathbf{X}\mathbf{B}_u\mathbf{A}_u' + \mathbf{E}$, denoting by $\text{vec}(\mathbf{M})$ the vectorization of a matrix $\mathbf{M}$:

$$\text{vec}(\mathbf{Y}') = \text{vec}(\mathbf{A}_u\mathbf{B}_u'\mathbf{X}') + \text{vec}(\mathbf{E}'). \tag{1.6}$$

For three matrices $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$ with appropriate dimensions, the following identity holds:

$$\text{vec}(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3) = (\mathbf{M}_3' \otimes \mathbf{M}_1)\,\text{vec}(\mathbf{M}_2), \tag{1.7}$$

where $\otimes$ is the Kronecker product. Setting $\mathbf{M}_1 = \mathbf{I}_q$, $\mathbf{M}_2 = \mathbf{A}_u$, and $\mathbf{M}_3 = \mathbf{B}_u'\mathbf{X}'$, from Eq. 1.6 follows

$$\text{vec}(\mathbf{Y}') = (\mathbf{X}\mathbf{B}_u \otimes \mathbf{I}_q)\,\text{vec}(\mathbf{A}_u) + \text{vec}(\mathbf{E}'). \tag{1.8}$$

Let $\mathbf{y} = \text{vec}(\mathbf{Y}') \in \mathbb{R}^{nq\times 1}$, $\tilde{\mathbf{a}}_u = \text{vec}(\mathbf{A}_u) \in \mathbb{R}^{qu\times 1}$, and $\boldsymbol{e} = \text{vec}(\mathbf{E}') \in \mathbb{R}^{nq\times 1}$, where $\boldsymbol{e} \sim N_{nq}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$ with $\tilde{\boldsymbol{\Sigma}} = \text{diag}(\boldsymbol{\Sigma}, \ldots, \boldsymbol{\Sigma})$. Eq. 1.8 is equivalent to

$$\mathbf{y} = (\mathbf{X}\mathbf{B}_u \otimes \mathbf{I}_q)\, \tilde{\mathbf{a}}_u + \boldsymbol{e}. \tag{1.9}$$

Since $\text{vec}(\mathbf{Y}) \sim \mathcal{N}_{nq}(\text{vec}(\mathbf{X}\mathbf{B}_u\mathbf{A}_u'), \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, then $\mathbf{y} = \text{vec}(\mathbf{Y}')$ follows a multivariate normal distribution with mean $\text{vec}(\mathbf{A}_u\mathbf{B}_u'\mathbf{X}') = (\mathbf{X}\mathbf{B}_u \otimes \mathbf{I}_q)\, \tilde{\mathbf{a}}_u$ and covariance matrix $\mathbf{I}_n \otimes \boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}$. We compute the full conditional of $\mathbf{A}_u$ in terms of $\tilde{\mathbf{a}}_u$. If $\tilde{\mathbf{y}} = \tilde{\boldsymbol{\Sigma}}^{-1/2}\mathbf{y}$, $\tilde{\mathbf{X}} = \tilde{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}\mathbf{B}_u \otimes \mathbf{I}_q)$, and $\boldsymbol{\Omega}_{\mathbf{A}_u} = (\mathbf{I}_{qu} + \tilde{\mathbf{X}}'\tilde{\mathbf{X}})$, subsequently the vector $\tilde{\mathbf{a}}_u$ is sampled from $\mathcal{N}_{qu}(\boldsymbol{\mu}^*_{\mathbf{A}_u}, \boldsymbol{\Sigma}^*_{\mathbf{A}_u})$, where $\boldsymbol{\mu}^*_{\mathbf{A}_u} = \boldsymbol{\Omega}^{-1}_{\mathbf{A}_u}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$ and $\boldsymbol{\Sigma}^*_{\mathbf{A}_u} = \boldsymbol{\Omega}^{-1}_{\mathbf{A}_u}$. Samples are made resorting to the algorithm developed in Rue (2001).

Column-sharing parametrization

The rank-u matrix $\mathbf{A}_u$ is defined as having exactly u columns, i.e. $\mathbf{A}_u = (\mathbf{a}_1, \ldots, \mathbf{a}_u)$, where $\mathbf{a}_h = (a_{1h}, \ldots, a_{qh})$ for each $h = 1, \ldots, u$. To obtain the full conditional distribution of $\mathbf{A}_u$, we sample each column independently from its respective posterior distribution. Therefore, it is necessary to express the likelihood function in terms of the columns of $\mathbf{A}_u$.

Define $\mathbf{y}$, $\tilde{\mathbf{y}}$ and $\tilde{\boldsymbol{\Sigma}}$ as in the previous section, and let $\mathbf{M}_{\mathbf{B}_u}$ denote the resulting matrix from $\mathbf{X}\mathbf{B}_u \otimes \mathbf{I}_q$. We allow the superscript (i, q) to denote the iqth columns of a matrix $\mathbf{M} = (m_1, \ldots, m_k)$ with $k \geq q$ columns. For example, if $\mathbf{M}$ comprises 12 columns, then $\mathbf{M}^{(2,3)} = (m_4, m_5, m_6)$, while $\mathbf{M}^{(4,3)} = (m_{10}, m_{11}, m_{12})$. Therefore, the model in Eq. 1.9 expressed in relation to the columns of $\mathbf{A}_u$ is

$$\mathbf{y} = \sum_{i=1}^{u} \mathbf{M}^{(i,q)}_{\mathbf{B}_u}\mathbf{a}_i + \boldsymbol{e} \tag{1.10}$$

and the distribution of $\mathbf{y}$ is $\mathcal{N}_{nq}(\mathbf{M}_u, \tilde{\boldsymbol{\Sigma}})$, where $\mathbf{M}_u = \sum_{i=1}^{u} \mathbf{M}^{(i,q)}_{\mathbf{B}_u}\mathbf{a}_i$.

Let $\tilde{\mathbf{X}}_i = \tilde{\boldsymbol{\Sigma}}^{-1/2}\mathbf{M}^{(i,q)}_{\mathbf{B}_u}$, $\boldsymbol{\Omega}_{\mathbf{a}_h} = \mathbf{I}_q + \tilde{\mathbf{X}}_h'\tilde{\mathbf{X}}_h$ and $\Delta_{\mathbf{a}_h} = \left(\tilde{\mathbf{y}}' - \sum_{i\neq h}^{u} \mathbf{a}_i'\tilde{\mathbf{X}}_i'\right)\tilde{\mathbf{X}}_h$. Finally, each column $\mathbf{a}_h$ is drawn from a multivariate normal distribution with mean $\boldsymbol{\mu}^*_{\mathbf{a}_h} = \boldsymbol{\Omega}^{-1}_{\mathbf{a}_h}\Delta'_{\mathbf{a}_h}$, and covariance matrix $\boldsymbol{\Sigma}^*_{\mathbf{a}_h} = \boldsymbol{\Omega}^{-1}_{\mathbf{a}_h}$.

Update $\mathbf{B}_u$

Naïve parametrization

To obtain the full conditional distribution of $\mathbf{B}_u$, we employ the vectorization of the multivariate linear regression model $\mathbf{Y} = \mathbf{X}\mathbf{B}_u\mathbf{A}_u' + \mathbf{E}$, and the identity in Eq. 1.7 with

$\mathbf{M}_1 = \mathbf{A}_u$, $\mathbf{M}_2 = \mathbf{B}_u'$, and $\mathbf{M}_3 = \mathbf{X}'$. Let $\tilde{\mathbf{b}}_u = \text{vec}(\mathbf{B}_u') \in \mathbb{R}^{pu \times 1}$, thus the model is

$$\mathbf{y} = (\mathbf{X} \otimes \mathbf{A}_u)\, \tilde{\mathbf{b}}_u + \boldsymbol{e}. \tag{1.11}$$

Note that $\tilde{\mathbf{b}}_u \sim \mathcal{N}_{pu}(\mathbf{0}, \boldsymbol{\Lambda}_u)$, where $\boldsymbol{\Lambda}_u$ is defined as in Eq. 1.4. Through similar reasoning as in the previous step, we obtain that $\tilde{\mathbf{b}}_u$ is multivariate normally distributed with mean $\boldsymbol{\mu}^*_{\mathbf{B}_u} = \boldsymbol{\Omega}^{-1}_{\mathbf{B}_u} \tilde{\mathbf{X}}' \tilde{\mathbf{y}}$ and covariance matrix $\boldsymbol{\Sigma}^*_{\mathbf{B}_u} = \boldsymbol{\Omega}^{-1}_{\mathbf{B}_u}$, with $\boldsymbol{\Omega}_{\mathbf{B}_u} = (\boldsymbol{\Lambda}_u^{-1} + \tilde{\mathbf{X}}'\tilde{\mathbf{X}})$, $\tilde{\mathbf{X}} = \tilde{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X} \otimes \mathbf{A}_u)$, and $\tilde{\mathbf{y}}$ is defined as in the previous step. The algorithm from Bhattacharya et al. (2016) is adapted to obtain the desired draws.

Column-sharing parametrization

For every $h = 1. \ldots, u$, the prior distribution of column $\mathbf{b}_h$ of matrix $\mathbf{B}_u = (\mathbf{b}_1, \ldots, \mathbf{b}_u)$ is a multivariate normal with mean the zero vector and the covariance matrix denoted by $\boldsymbol{\Lambda}_{\mathbf{b}_h} = \text{diag}(\psi_{1h}\tau_{1h}^2\phi_{1h}^2, \ldots, \psi_{ph}\tau_{ph}^2\phi_{ph}^2)$. We follow the same reasoning as in the update of $\mathbf{A}_u$ in RRcs to derive the posterior distribution of each column $\mathbf{b}_h$. Consequently, the likelihood function requires to be formulated in relation to the columns of $\mathbf{B}_u$.

Let $\mathbf{y}^* = \text{vec}(\mathbf{Y})$, $\boldsymbol{e}^* = \text{vec}(\mathbf{E})$, $\mathbf{M}_{\mathbf{A}_u} = \mathbf{A}_u \otimes \mathbf{X}$, and $\tilde{\boldsymbol{\Sigma}}^* = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$. Notice that $\mathbf{y}^*$ follows the multivariate normal distribution $\mathcal{N}_{nq}(\mathbf{M}_{\mathbf{A}_u}\, \text{vec}(\mathbf{B}_u), \tilde{\boldsymbol{\Sigma}}^*)$. Using the previous superscript notation, we express $\mathbf{y}^*$ in terms of the columns of $\mathbf{B}_u$ as:

$$\mathbf{y}^* = \sum_{i=1}^{u} \mathbf{M}^{(i,p)}_{\mathbf{A}u} \mathbf{b}_i + \boldsymbol{e}^*. \tag{1.12}$$

Lastly, by denoting $\tilde{\mathbf{y}} = \tilde{\boldsymbol{\Sigma}^*}^{-1/2} \mathbf{y}^*$, $\tilde{\mathbf{X}}^*_i = \tilde{\boldsymbol{\Sigma}^*}^{-1/2} \mathbf{M}^{(i,p)}_{\mathbf{A}_u}$, $\boldsymbol{\Omega}_{\mathbf{b}_h} = \boldsymbol{\Lambda}^{-1}_{\mathbf{b}_h} + \tilde{\mathbf{X}}^{*\prime}_h \tilde{\mathbf{X}}^*_h$ and $\Delta_{\mathbf{b}_h} = \left(\tilde{\mathbf{y}^{*\prime}} - \sum_{i \neq h}^{u} \mathbf{b}_i' \tilde{\mathbf{X}}_i^{*\prime}\right) \tilde{\mathbf{X}}^*_h$, we sample $\mathbf{b}_h$ from $\mathcal{N}_p(\boldsymbol{\mu}^*_{\mathbf{b}_h}, \boldsymbol{\Sigma}^*_{\mathbf{b}_h})$, where $\boldsymbol{\mu}^*_{\mathbf{b}_h} = \boldsymbol{\Omega}^{-1}_{\mathbf{b}_h} \Delta'_{\mathbf{b}_h}$ and $\boldsymbol{\Sigma}^*_{\mathbf{b}_h} = \boldsymbol{\Omega}^{-1}_{\mathbf{b}_h}$.

Update the hyperparameters $\boldsymbol{\tau}$, Ψ, Φ and $\boldsymbol{\alpha}$

The posterior distributions of the global shrinkage parameters are

$$\tau_h | b_{lh}, \phi_{lh}, \alpha_h \sim \text{GiG}\left(p(\alpha_h - 1),\, 1,\, 2\sum_{l=1}^{p} \frac{|b_{lh}|}{\phi_{lh}}\right),$$
$$\psi_{lh} | b_{lh}, \phi_{lh}, \tau_h \sim \text{iG}\left(\frac{\phi_{lh}\tau_h}{|b_{lh}|},\, 1\right)$$

for $l = 1, \ldots, p$ and $h = 1, \ldots, q$, where $\text{GiG}(\cdot)$ denotes the generalised inverse Gaussian distribution, and $\text{iG}(\cdot)$ the inverse Gaussian distribution. The update of ϕ_{lh} is carried

out by first obtaining samples for $T_{l1}, \ldots, T_{lq}$ from

$$T_{lh}|b_{lh}, \alpha_h \sim \text{GiG}\left((\alpha_h - 1),\, 1,\, 2|b_{lh}|\right)$$

afterwards set $\phi_{lh} = \frac{T_{lh}}{\sum_{i=1}^{p} T_{ih}}$. Finally, the log of the full conditional distribution for α_h is given by

$$\log p(\alpha_h|-) = -p \log \Gamma(\alpha_h) + \alpha_h \left(p \log \tau_h - p \log 2 + \sum_{l=1}^{p} \phi_{lh} \right) + c,$$

where c is a constant independent of α_h.

2 PIP uncertainty index sensitivity analysis

This section examines four different definitions of the PIP uncertainty index, including the one presented in the Main paper and three alternative formulations.

The sparse posterior estimate of the coefficient matrix allows exact zero elements. For the jkth coefficient ($j = 1, \ldots, p,\ \ k = 1, \ldots, q$), we obtain a sparse draw at each iteration of the MCMC, and subsequently set the posterior estimate of the jkth entry to 0 if the proportion of nonzero draws is less than 0.5 ($\text{PIP}_{jk} \leq 0.5$), or to the posterior mean of these samples otherwise. For ease of notation, we omit the jk subscript in the remaining lines of this section, allowing PIP to denote any PIP_{jk}. Values close to 0 (or 1) indicate low uncertainty about the exclusion (or inclusion) of the coefficient, while a probability around 0.5 signals insufficient evidence for either conclusion. To provide a straightforward and easily interpreted manner to quantify uncertainty about this decision, we defined the PIP uncertainty index as:

$$\zeta_{(1)} = 1 - 2\big|\text{PIP} - 0.5\big|. \tag{2.1}$$

This choice of ζ is based primarily on two features. First, the image of ζ is the interval $[0, 1]$, where 0 represents low uncertainty and 1 means high uncertainty. Second, the PIP values of 0 and 1 should be mapped to 0 (lowest uncertainty), as opposed to a PIP of 0.5 corresponding to a PIP uncertainty index of 1 (highest uncertainty). Evidently, our restrictions on the definition of ζ permits alternative representations. Therefore, we provide the additional formulations of ζ:

$$\zeta_{(2)} = 1 - 4(\text{PIP} - 0.5)^2, \tag{2.2}$$

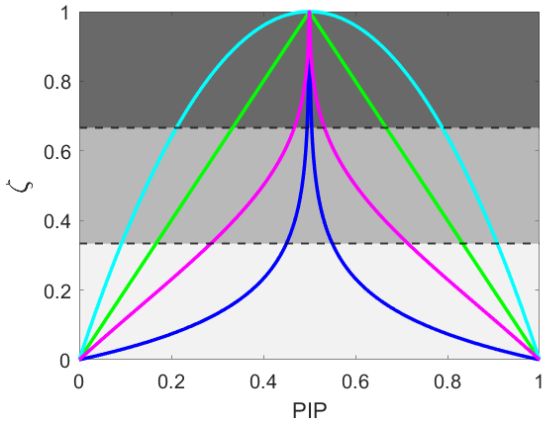


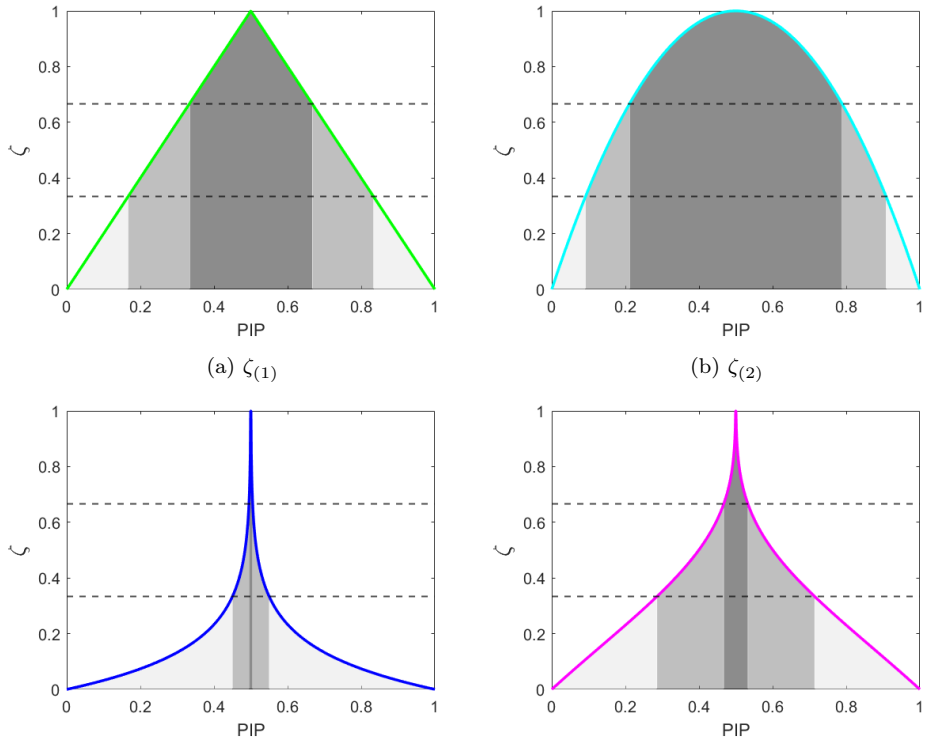
Figure 1: Plot of the four definitions of PIP uncertainty index, ζ, as a function of the PIP: $\zeta_{(1)}$ (green), $\zeta_{(2)}$ (cyan), $\zeta_{(3)}$ (blue), $\zeta_{(4)}$ (magenta). The shaded regions in a grey-colour scale indicate low ($\zeta \leq 1/3$), medium ($1/3 < \zeta \leq 2/3$), or high ($\zeta > 2/3$) uncertainty.

$$\zeta_{(3)} = \frac{\log\left(\left|\text{PIP} - 0.5\right| + \epsilon\right) + \log(0.5)}{\min(\log\left(\left|\text{PIP} - 0.5\right| + \epsilon\right) + \log(0.5))}, \tag{2.3}$$

$$\zeta_{(4)} = \frac{\log\left(\log\left(\left|\text{PIP} - 0.5\right| + \epsilon\right) + \epsilon\right) + \log\left(\log\left(0.5 + \epsilon\right)\right)}{\max(\log\left(\log\left(\left|\text{PIP} - 0.5\right| + \epsilon\right) + \epsilon\right) + \log\left(\log\left(0.5 + \epsilon\right)\right))}, \tag{2.4}$$

where ϵ is an arbitrarily small positive number for numerical stability. The four specifications of ζ are plotted in Figure 1 and in Figure 2.

The first definition we presented in Eq. (2.1), a piecewise linear function, poses an advantage over the remaining three through the linearity property. The set of PIP values that are evaluated by $\zeta_{(1)}$ to a low uncertainty ζ ($\zeta \leq 1/3$) is the same size as those PIP values that result in medium ($1/3 < \zeta \leq 2/3$), and high ($\zeta > 2/3$) uncertainty. The remaining three specifications of ζ are nonlinear functions, and therefore the portions in the domain corresponding to each level of uncertainty are of different lengths. Consequently, more than half of the points are allocated to a high uncertainty region in $\zeta_{(2)}$, whereas 24% of PIP values are assigned medium uncertainty, and 18.4% evaluate to a low level. The functions $\zeta_{(3)}$ and $\zeta_{(4)}$ exhibit the opposite behaviour, since the set of PIP values corresponding to low uncertainty is bigger than the set of points mapped to a medium range, which in turn comprises more points than the PIP interval of high uncertainty (see Figure 2). Hence, the choice of $\zeta_{(1)}$ as the PIP uncertainty index is not biased towards greater portions of the domain corresponding to a particular level of uncertainty, and the relationship between PIP and ζ becomes transparent. Where the particular needs of the researcher demand for a nonlinear behaviour of ζ that could result from an optimistic or pessimistic approach to the quantification of uncertainty, the three aforementioned specifications could be considered as suitable options.



(a) $\zeta_{(1)}$ (b) $\zeta_{(2)}$

(c) $\zeta_{(3)}$ (d) $\zeta_{(4)}$

Figure 2: Function ζ in each of the cases considered: the piece-wise linear function $\zeta_{(1)}$ (panel a), and the nonlinear definitions $\zeta_{(2)}$, $\zeta_{(3)}$ and $\zeta_{(4)}$ (panels b, c and d, respectively). The width of the shaded areas is in accordance with the portion in the domain that evaluates to low (light grey), medium (medium grey) or high (dark grey) uncertainty. The values of ζ that delimit these regions are plotted in horizontal dashed lines.

3 Simulation experiments

3.1 Additional simulation results

In Section 4 of the Main paper, we conducted a comparative analysis of our methodology under the RRcs parametrization against other state-of-the-art methodologies. Here, we present the findings pertaining to the RRn parametrization across various simulation scenarios (see Table 1). As elucidated in the Main manuscript, RRcs yields superior estimations of the rank and of the coefficient matrix compared to RRn, as evidenced by the Mean Squared Error (MSE) metric. In instances of smaller dimensions, RRn tends to underestimate the rank, whereas RRcs exhibits improved estimations. With increasing dimensions of p and q, both parametrizations yield more conservative rank estimates in sparse scenarios. However, RRcs achieves overall significantly lower errors while promoting a more parsimonious model.

			Sparse DGP				Non-sparse DGP			
			$\boldsymbol{\Sigma}_{ind}$		$\boldsymbol{\Sigma}_{corr}$		$\boldsymbol{\Sigma}_{ind}$		$\boldsymbol{\Sigma}_{corr}$	
(q,p)	**X**	**Measure**	**RRcs**	**RRn**	**RRcs**	**RRn**	**RRcs**	**RRn**	**RRcs**	**RRn**
(5,10)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.280	1.020	1.320	1.020	1.280	1.040	1.380
		MSE	0.453	0.297	0.527	3.841	1.163	0.878	1.080	0.776
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.120	1.000	1.140	1.000	1.340	1.020	1.420
		MSE	0.490	0.456	0.484	0.375	1.284	1.006	1.328	0.939
(5,15)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.040	1.000	1.360	1.020	1.420	1.060	1.540
		MSE	0.261	0.241	0.288	2.592	1.204	0.763	1.052	0.634
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.240	1.000	1.280	1.040	1.280	1.020	1.620
		MSE	0.346	0.262	0.367	0.270	1.051	0.905	1.376	0.878
(5,20)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.100	1.000	1.280	1.020	1.460	1.000	1.640
		MSE	0.182	0.168	0.237	0.192	1.057	0.760	1.272	0.736
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.200	1.000	1.080	1.000	1.180	1.020	1.420
		MSE	0.235	0.188	0.297	0.256	1.163	0.983	1.383	0.985
(5,50)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.120	1.000	1.140	1.000	1.220	1.000	1.560
		MSE	0.218	0.187	0.217	0.200	1.112	0.787	1.203	0.666
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.060	1.000	1.060	1.000	1.300	1.000	1.300
		MSE	0.231	0.197	0.252	0.243	1.353	0.998	1.529	1.097
(10,10)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.080	1.000	1.220	1.000	1.260	1.000	1.380
		MSE	0.579	0.523	0.441	0.374	1.243	1.017	1.046	0.785
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.120	1.000	1.140	1.000	1.280	1.000	1.420
		MSE	0.544	0.471	0.588	0.493	1.252	0.926	1.308	0.955
(10,15)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.140	1.000	1.160	1.000	1.380	1.000	1.520
		MSE	0.311	0.269	0.371	0.318	1.259	0.953	1.151	0.835
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.140	1.000	1.100	1.000	1.160	1.000	1.480
		MSE	0.436	0.372	0.349	0.324	1.272	1.158	1.583	1.082
(10,20)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.040	1.000	1.040	1.000	1.380	1.000	1.940
		MSE	0.213	0.203	0.234	0.227	1.210	0.888	1.383	0.747
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.060	1.000	1.060	1.000	1.400	1.000	1.600
		MSE	0.303	0.289	0.309	0.279	1.480	1.071	1.395	0.827
(10,50)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.040	1.000	1.040	1.000	1.660	1.000	1.760
		MSE	0.217	0.208	0.217	0.207	1.419	0.815	1.730	0.909
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.040	1.000	1.040	1.000	1.460	1.000	1.520
		MSE	0.301	0.286	0.256	0.245	1.637	1.143	1.701	0.978
(10,75)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.000	1.000	1.120	1.000	1.700	1.000	1.600
		MSE	0.261	0.260	0.248	0.230	1.597	0.894	1.471	1.002
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.040	1.000	1.100	1.000	1.520	1.000	1.660
		MSE	0.275	0.266	0.298	0.280	2.090	1.320	2.065	1.210
(10,100)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.040	1.000	1.080	1.000	1.420	1.000	1.420
		MSE	0.261	0.253	0.248	0.229	1.949	1.452	1.997	1.473
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.060	1.000	1.060	1.000	1.380	1.000	1.600
		MSE	0.267	0.257	0.327	0.311	2.088	1.597	2.223	1.382
(10,500)	$\mathbf{X}_{ind}$	$\hat{r}$	1.000	1.480	1.000	1.000	1.000	3.200	1.000	3.220
		MSE	0.239	0.190	0.250	0.280	4.379	11.559	4.546	8.214
	$\mathbf{X}_{corr}$	$\hat{r}$	1.000	1.400	1.000	1.140	1.000	2.540	1.000	2.880
		MSE	0.247	0.237	0.274	0.259	4.335	17.480	4.037	22.265

Table 1: Comparison of the estimated rank ($\hat{r}$) and mean squared error (MSE) obtained by RRcs against RRn for different values of (q, p). In all settings, $n = 100$ and $r_0 = 3$. When $p \in \{10, 15, 20\}$, $p^* = 5$, and in the case where $p \in \{50, 75, 100, 500\}$, $p^* = 0.2p$. We present the average estimates over 50 repetitions for independent errors ($\boldsymbol{\Sigma}_{ind}$), correlated errors ($\boldsymbol{\Sigma}_{corr}$), independent regressors ($\mathbf{X}_{ind}$), and correlated regressors ($\mathbf{X}_{corr}$).

3.2 CODA analysis

We perform a CODA analysis for the posterior distribution of the rank (Table 2), and the estimates of the matrix $\mathbf{C}$ (Table 3). In both studies, we consider an MCMC chain of $50,000$ iterations to perform the Geweke convergence diagnostic test and the Heidelberger and Welch's convergence diagnostic test.

Table 2 shows that the p-values for the Geweke convergence test and Heidelberg's stationarity test applied to the rank estimation analysis are greater than 0.05, suggesting adequate convergence of the estimated rank chain in all the scenarios considered. These findings hold across all scenarios and thinning levels. By inspecting the ratio between the half-width of the 95% confidence interval for the mean and the mean (last column), a value greater than 0.1 suggests the need for a longer chain. This happens only for the scenario in the third row of Table 2, which however does pass all the other CODA tests.

Regarding the convergence diagnostics for the coefficient matrix estimates, given the high number of entries, pq, instead of reporting all the individual results, we show some summary measures as follows. Specifically, for each of the above-mentioned tests, we compute in Table 3 the proportion of the elements of $\mathbf{C}$ passing the specified test. Therefore, the closest the values in each column are to 1, the larger the share of coefficients that pass the corresponding convergence test. The results in Table 3 report a good performance according to Heidelberg's stationarity test and the half-width test, as shown by the high percentages (above 70%). Moreover, Geweke's diagnostic suggests an average good convergence for the entries of $\mathbf{C}$ in the first and third scenarios but suggests increasing the length of the MCMC.

Simulation	MCMC iterations	Thinning	Geweke p-value	Heidelberger	
				Stationarity test p-value	Ratio
$(q,p) = (10,20)$, $r_0 = 3$,$n = 100$, sparse	50,000	1	0.946	0.287	0.048
		10	0.976	0.317	0.054
$(q,p) = (10,20)$, $r_0 = 3$, $n = 100$, non-sparse	50,000	1	0.132	0.105	0.086
		10	0.092	0.171	0.093
$(q,p) = (10,15)$, $r_0 = 5$, $n = 100$, non-sparse	50,000	1	0.305	0.282	0.116
		10	0.271	0.252	0.126
$(q,p) = (10,30)$, $r_0 = 5$, $n = 100$, non-sparse	50,000	1	0.185	0.942	0.077
		10	0.217	0.968	0.080

Table 2: CODA analysis of u. We studied the various scenarios under the first column. For thinning, we consider the values 1 (no thinning) and 10. Geweke's diagnostic tests the assumption that the first 10% and the last 50% parts of the Markov chain are independent, reporting the results in the fourth column. We consider Heidelberger's stationarity test (5th column) and the half-width test. The latter indicates whether the length of the sample is long enough to estimate the mean with sufficient accuracy (< 0.1).

Simulation	MCMC iterations	Geweke Prop	Heidelberger	
			Stationarity test Prop	Half-width test Prop
$(q,p) = (10,20)$, $r_0 = 3$, $n = 100$, sparse	50,000	0.980	1.000	0.570
$(q,p) = (10,20)$, $r_0 = 3$, $n = 100$, non-sparse	50,000	0.315	0.915	0.874
$(q,p) = (10,15)$, $r_0 = 5$, $n = 100$, non-sparse	50,000	0.667	0.767	0.713
$(q,p) = (10,30)$, $r_0 = 5$, $n = 100$, non-sparse	50,000	0.280	0.990	0.781

Table 3: CODA analysis of $\hat{\mathbf{C}}$. We studied the settings presented in the first column, and report the proportion of times that the Markov chain passed Geweke's test (p-value ≥ 0.05) in the third column. The proportion of samples passing Heidelberg's stationarity test and half-width test are stated in the last two columns.

3.3 Bura and Cook test

In this subsection, we have implemented the test by Bura and Cook (2003). However, since the method elaborates on the OLS estimate of a multiple linear regression model, it could not be run for the cases $p \geq n$. The test results are reported in Table 4.

				sparse		non-sparse	
(q,p)	r_0	**X**	**Measure**	$\boldsymbol{\Sigma}_{ind}$	$\boldsymbol{\Sigma}_{corr}$	$\boldsymbol{\Sigma}_{ind}$	$\boldsymbol{\Sigma}_{corr}$
(5,15)	3	$\mathbf{X}_{ind}$	$\hat{r}$	3.000	2.980	3.180	3.080
			$r_{\%}$	0.640	0.780	0.820	0.920
		$\mathbf{X}_{corr}$	$\hat{r}$	3.300	3.300	3.220	3.220
			$r_{\%}$	0.560	0.600	0.780	0.780
	5	$\mathbf{X}_{ind}$	$\hat{r}$	3.960	4.240	4.700	4.700
			$r_{\%}$	0.140	0.280	0.700	0.700
		$\mathbf{X}_{corr}$	$\hat{r}$	4.040	3.920	4.700	4.660
			$r_{\%}$	0.260	0.200	0.700	0.660

Table 4: Comparison of the estimated rank ($\hat{r}$), share of correctly estimated rank ($r_{\%}$), obtained by the test of Bura and Cook (2003) for $(q, p) = (5, 15)$, true rank r_0 and $n = 50$. The sparse DGP considers $p^* = 5$, while the non-sparse DGP has $p^* = p$. We present the average estimates over 50 repetitions for independent errors ($\boldsymbol{\Sigma}_{ind}$), correlated errors ($\boldsymbol{\Sigma}_{corr}$), independent regressors ($\mathbf{X}_{ind}$), and correlated regressors ($\mathbf{X}_{corr}$).

3.4 HS prior

The results presented in Tables 3 and 4 of the main paper for our proposed approach under the column-sharing parameterization use the Dirichlet-Laplace (DL) prior on the columns of matrix $\mathbf{B}$. In this section, we also present the results using the horseshoe (HS) prior in Tables 5 and 6.

The MSE shows a good performance in estimating the coefficient matrix $\mathbf{C}$, compared to the other methods; however, the rank is typically overestimated across all the settings. This behaviour signals a potential over-shrinkage of the columns of $\mathbf{B}$, which motivates the need for a larger number of components (i.e., rank) to estimate $\mathbf{C} = \mathbf{B}\mathbf{A}'$. In the non-sparse DGP setting with $(q, p) = (10, 50)$ the performance of the DL and the HS priors is similar, whereas in the other cases, the accuracy of our method using the DL prior is improved.

(q,p)	r_0	X	Measure	$\boldsymbol{\Sigma}_{ind}$ ANN	RRRR	RRcs-DL	RRcs-HS	$\boldsymbol{\Sigma}_{corr}$ ANN	RRRR	RRcs-DL	RRcs-HS
(5,15)	3	$\mathbf{X}_{ind}$	$\hat{r}$	2.320	2.420	1.860	2.660	2.420	2.460	1.720	2.680
			$r_{\%}$	0.400	0.480	0.160	0.000	0.480	0.500	0.180	0.000
			MSE($\mathbf{C}$)	0.030	0.030	0.161	0.153	0.027	0.028	0.189	0.180
			$\varrho(\mathbf{C})$	0.745	0.709	0.984	0.985	0.674	0.670	0.988	0.989
		$\mathbf{X}_{corr}$	$\hat{r}$	2.280	2.240	1.820	2.700	2.020	1.940	1.580	2.740
			$r_{\%}$	0.340	0.300	0.180	0.000	0.180	0.140	0.200	0.000
			MSE($\mathbf{C}$)	0.043	0.047	0.251	0.192	0.046	0.051	0.173	0.139
			$\varrho(\mathbf{C})$	0.779	0.804	0.995	0.994	0.898	0.917	0.986	0.991
	5	$\mathbf{X}_{ind}$	$\hat{r}$	3.160	3.280	2.380	2.700	3.240	3.240	2.100	2.540
			$r_{\%}$	0.000	0.020	0.240	0.400	0.000	0.000	0.120	0.360
			MSE($\mathbf{C}$)	0.036	0.035	0.376	0.342	0.031	0.034	0.394	0.388
			$\varrho(\mathbf{C})$	1.000	0.984	0.380	0.400	1.000	1.000	0.419	0.539
		$\mathbf{X}_{corr}$	$\hat{r}$	2.700	2.940	2.400	2.720	2.900	2.920	2.120	2.660
			$r_{\%}$	0.000	0.000	0.300	0.400	0.000	0.020	0.180	0.400
			MSE($\mathbf{C}$)	0.072	0.063	0.470	0.400	0.063	0.065	0.463	0.434
			$\varrho(\mathbf{C})$	1.000	1.000	0.440	0.581	1.000	0.987	0.442	0.471
(5,50)	3	$\mathbf{X}_{ind}$	$\hat{r}$	0.160	5.000	1.540	2.600	0.240	5.000	1.260	2.440
			$r_{\%}$	0.000	0.000	0.100	0.000	0.020	0.000	0.060	0.040
			MSE($\mathbf{C}$)	1.010	144.533	0.148	0.107	0.667	17.328	0.183	0.155
			$\varrho(\mathbf{C})$	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	0.380	5.000	1.720	2.600	0.300	5.000	1.360	2.560
			$r_{\%}$	0.000	0.000	0.140	0.000	0.000	0.000	0.060	0.020
			MSE($\mathbf{C}$)	2.197	32.051	0.165	0.145	1.181	32.620	0.209	0.178
			$\varrho(\mathbf{C})$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	$\mathbf{X}_{ind}$	$\hat{r}$	0.060	5.000	2.360	2.540	0.020	5.000	2.040	2.580
			$r_{\%}$	0.000	1.000	0.280	0.380	0.000	1.000	0.120	0.380
			MSE($\mathbf{C}$)	1.849	813.710	0.309	0.302	0.917	102.402	0.298	0.277
			$\varrho(\mathbf{C})$	1.000	0.987	0.901	0.896	1.000	0.991	0.912	0.894
		$\mathbf{X}_{corr}$	$\hat{r}$	0.160	5.000	2.560	2.560	0.040	5.000	2.340	2.600
			$r_{\%}$	0.000	1.000	0.380	0.380	0.000	1.000	0.260	0.400
			MSE($\mathbf{C}$)	6.032	65.801	0.371	0.363	0.930	52.111	0.336	0.323
			$\varrho(\mathbf{C})$	1.000	0.992	0.904	0.929	1.000	0.988	0.925	0.906
(10,50)	3	$\mathbf{X}_{ind}$	$\hat{r}$	1.360	10.000	1.000	2.320	1.820	10.000	1.000	1.520
			$r_{\%}$	0.220	0.000	0.000	0.020	0.460	0.000	0.000	0.120
			MSE($\mathbf{C}$)	8.372	236.725	0.222	0.170	16.348	197.125	0.251	0.222
			$\varrho(\mathbf{C})$	0.962	1.000	1.000	1.000	0.943	1.000	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	1.260	10.000	1.040	2.640	1.600	10.000	1.000	1.980
			$r_{\%}$	0.220	0.000	0.000	0.020	0.220	0.000	0.000	0.120
			MSE($\mathbf{C}$)	36.481	13.903	0.300	0.232	3.239	43.166	0.281	0.228
			$\varrho(\mathbf{C})$	0.977	1.000	1.000	1.000	0.971	1.000	0.998	1.000
	5	$\mathbf{X}_{ind}$	$\hat{r}$	0.080	10.000	1.040	2.120	0.020	10.000	1.000	1.380
			$r_{\%}$	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.040
			MSE($\mathbf{C}$)	1.001	88.157	0.522	0.404	0.917	51.975	0.511	0.465
			$\varrho(\mathbf{C})$	1.000	1.000	0.998	0.999	1.000	1.000	0.999	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	0.480	10.000	1.140	2.960	0.500	10.000	1.020	2.060
			$r_{\%}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			MSE($\mathbf{C}$)	1.363	64.436	0.559	0.402	2.441	30.059	0.618	0.525
			$\varrho(\mathbf{C})$	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000

Table 5: Comparison of the estimated rank ($\hat{r}$), share of correctly estimated rank ($r_{\%}$), mean squared error of $\mathbf{C}$ (MSE($\mathbf{C}$)), and $\varrho(\mathbf{C}) = \|\hat{\mathbf{C}}(\hat{\mathbf{C}}'\hat{\mathbf{C}})^{+}\hat{\mathbf{C}}' - \mathbf{C}(\mathbf{C}'\mathbf{C})^{+}\mathbf{C}'\|_2$ obtained by RRcs with Dirichlet-Laplace and horseshoe prior (RRcs-DL and RRcs-HS) 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 50 repetitions for independent errors ($\boldsymbol{\Sigma}_{ind}$), correlated errors ($\boldsymbol{\Sigma}_{corr}$), independent regressors ($\mathbf{X}_{ind}$), and correlated regressors ($\mathbf{X}_{corr}$).

(q,p)	r_0	X	Measure	$\boldsymbol{\Sigma}_{ind}$ ANN	RRRR	RRcs-DL	RRcs-HS	$\boldsymbol{\Sigma}_{corr}$ ANN	RRRR	RRcs-DL	RRcs-HS
(5,15)	3	$\mathbf{X}_{ind}$	$\hat{r}$	2.940	2.880	2.060	2.540	2.940	2.960	1.960	2.360
			$r_{\%}$	0.940	0.880	0.120	0.000	0.940	0.960	0.160	0.000
			MSE($\mathbf{C}$)	0.023	0.026	0.799	0.764	0.021	0.023	0.675	0.797
			$\varrho(\mathbf{C})$	0.260	0.292	1.000	1.000	0.228	0.223	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	2.860	2.880	2.160	2.620	2.940	2.960	1.860	2.600
			$r_{\%}$	0.860	0.840	0.120	0.000	0.940	0.920	0.060	0.000
			MSE($\mathbf{C}$)	0.042	0.045	0.697	0.674	0.040	0.043	0.748	0.777
			$\varrho(\mathbf{C})$	0.340	0.357	1.000	1.000	0.296	0.319	1.000	1.000
	5	$\mathbf{X}_{ind}$	$\hat{r}$	4.000	4.360	2.540	2.600	3.940	4.380	2.600	2.660
			$r_{\%}$	0.000	0.380	0.360	0.380	0.000	0.420	0.360	0.400
			MSE($\mathbf{C}$)	0.058	0.037	1.594	1.545	0.051	0.033	1.624	1.722
			$\varrho(\mathbf{C})$	1.000	0.743	0.880	0.830	1.000	0.730	0.888	0.832
		$\mathbf{X}_{corr}$	$\hat{r}$	3.940	4.240	2.580	2.680	3.880	4.160	2.600	2.640
			$r_{\%}$	0.000	0.320	0.380	0.400	0.000	0.280	0.400	0.400
			MSE($\mathbf{C}$)	0.088	0.070	1.633	1.555	0.080	0.064	1.724	1.686
			$\varrho(\mathbf{C})$	1.000	0.806	0.891	0.855	1.000	0.855	0.898	0.860
(5,50)	3	$\mathbf{X}_{ind}$	$\hat{r}$	0.940	5.000	1.280	2.560	1.620	5.000	1.160	2.400
			$r_{\%}$	0.260	0.000	0.020	0.020	0.360	0.000	0.000	0.000
			MSE($\mathbf{C}$)	4.957	17.005	1.285	0.974	98.964	52.488	1.535	1.202
			$\varrho(\mathbf{C})$	0.953	1.000	1.000	1.000	0.954	1.000	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	0.800	5.000	1.860	2.380	1.620	5.000	1.400	2.600
			$r_{\%}$	0.200	0.000	0.020	0.020	0.420	0.000	0.000	0.000
			MSE($\mathbf{C}$)	15.744	155.578	1.395	1.217	303602.928	34.301	1.653	1.337
			$\varrho(\mathbf{C})$	0.969	1.000	1.000	1.000	0.973	1.000	1.000	1.000
	5	$\mathbf{X}_{ind}$	$\hat{r}$	0.000	5.000	2.600	2.560	0.160	5.000	2.520	2.560
			$r_{\%}$	0.000	1.000	0.400	0.380	0.000	1.000	0.380	0.360
			MSE($\mathbf{C}$)	5.179	228.422	2.586	2.233	5.526	580.312	2.741	2.301
			$\varrho(\mathbf{C})$	1.000	0.972	0.990	0.933	1.000	0.952	0.990	0.926
		$\mathbf{X}_{corr}$	$\hat{r}$	0.180	5.000	2.600	2.600	0.420	5.000	2.600	2.520
			$r_{\%}$	0.000	1.000	0.400	0.400	0.000	1.000	0.400	0.380
			MSE($\mathbf{C}$)	12.083	36.098	3.064	2.639	463.520	128.832	3.293	2.842
			$\varrho(\mathbf{C})$	1.000	0.962	0.989	0.928	1.000	0.978	0.989	0.931
(10,50)	3	$\mathbf{X}_{ind}$	$\hat{r}$	3.000	10.000	1.260	1.260	3.200	10.000	1.220	1.020
			$r_{\%}$	1.000	0.000	0.040	0.000	0.800	0.000	0.080	0.000
			MSE($\mathbf{C}$)	1391.891	494.411	1.570	1.693	25.856	99.000	1.711	1.756
			$\varrho(\mathbf{C})$	0.730	1.000	1.000	1.000	0.782	1.000	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	3.000	10.000	1.240	1.560	3.260	10.000	1.180	1.180
			$r_{\%}$	1.000	0.000	0.020	0.020	0.740	0.000	0.000	0.000
			MSE($\mathbf{C}$)	298.131	76.197	2.062	2.167	25.326	84.624	1.884	1.956
			$\varrho(\mathbf{C})$	0.806	1.000	1.000	1.000	0.839	1.000	1.000	1.000
	5	$\mathbf{X}_{ind}$	$\hat{r}$	2.420	10.000	1.280	1.000	4.060	10.000	1.060	1.000
			$r_{\%}$	0.480	0.000	0.000	0.000	0.500	0.000	0.000	0.000
			MSE($\mathbf{C}$)	25.933	171.279	3.548	3.697	18.649	43.215	4.064	3.898
			$\varrho(\mathbf{C})$	0.922	1.000	1.000	1.000	0.919	1.000	1.000	1.000
		$\mathbf{X}_{corr}$	$\hat{r}$	2.200	10.000	1.200	1.060	2.680	10.000	1.220	1.020
			$r_{\%}$	0.400	0.000	0.000	0.000	0.380	0.000	0.000	0.000
			MSE($\mathbf{C}$)	10.417	47.036	3.913	3.981	545.300	67.874	4.231	4.262
			$\varrho(\mathbf{C})$	0.953	1.000	1.000	1.000	0.948	1.000	1.000	1.000

Table 6: Comparison of the estimated rank ($\hat{r}$), share of correctly estimated rank ($r_{\%}$), mean squared error of $\mathbf{C}$ (MSE($\mathbf{C}$)), and $\varrho(\mathbf{C}) = \|\hat{\mathbf{C}}(\hat{\mathbf{C}}'\hat{\mathbf{C}})^{+}\hat{\mathbf{C}}' - \mathbf{C}(\mathbf{C}'\mathbf{C})^{+}\mathbf{C}'\|_2$ obtained by RRcs with Dirichlet-Laplace and horseshoe prior (RRcs-DL and RRcs-HS) 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 50 repetitions for independent errors ($\boldsymbol{\Sigma}_{ind}$), correlated errors ($\boldsymbol{\Sigma}_{corr}$), independent regressors ($\mathbf{X}_{ind}$), and correlated regressors ($\mathbf{X}_{corr}$).

j	Name	Description
1	UjMAG	$M_{abs,gal}$ in Johnson U
2	BjMAG	$M_{abs,gal}$ in Johnson B
3	VjMAG	$M_{abs,gal}$ in Johnson V
4	usMAG	$M_{abs,gal}$ in SDSS u
5	gsMAG	$M_{abs,gal}$ in SDSS g
6	rsMAG	$M_{abs,gal}$ in SDSS r
7	UbMAG	$M_{abs,gal}$ in Bessell U
8	BbMAG	$M_{abs,gal}$ in Bessell B
9	VbMAG	$M_{abs,gal}$ in Bessell V
10	S280MAG	$M_{abs,gal}$ in 280/40
11	W420FE	photon flux in filter 420 in run E
12	W462FE	photon flux in filter 462 in run E
13	W485FD	photon flux in filter 485 in run D
14	W518FE	photon flux in filter 518 in run E
15	W571FS	photon flux in filter 571 combined
16	W604FE	photon flux in filter 604 in run E
17	W646FD	photon flux in filter 646 in run D
18	W696FE	photon flux in filter 696 in run E
19	W753FE	photon flux in filter 753 in run E
20	W815FS	photon flux in filter 815 combined
21	W856FD	photon flux in filter 856 in run D
22	W914FD	photon flux in filter 914 in run D
23	W914FE	photon flux in filter 914 in run E

Table 7: Name and description of covariates in the COMBO-17 galaxy dataset. Covariates $1-10$ are the absolute magnitude $M_{abs,gal}$ of the galaxy in the indicated passband, and covariates $11-23$ are the photon flux in different filters and various runs of the experiment.

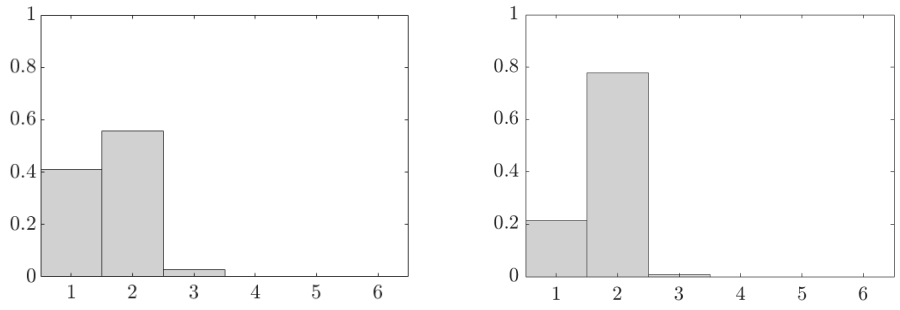
4 Further results on COMBO-17 dataset

The application of the methodology to the COMBO-17 galaxy dataset illustrates extensively the performance of the proposed approach. In this section, we provide further comprehension of the covariates studied in the model (Table 7), and show the posterior concentration of the rank when the number of data points increases (Figure 3), results obtained with the full dataset (Table 8 and Figure 4), and plots of the probability mass function and tail distribution of RI for all covariates (Figures 5, 6).

The COMBO-17 object catalogue of the Chandra Deep Field South lists photometry, astrometry and morphological features of more than $60,000$ celestial objects. We restrict the analysis to $3,438$ objects, all classified as "Galaxies", with no missing values. Measurement errors and redundant variables were omitted, resulting in a total of 29 variables divided into 23 covariates and 6 responses, as done in Izenman (2008). Table 7 provides the description of the 23 explanatory variables, whereas deeper comprehension of how the measurements were acquired, the mentioned passbands, filters and experimental runs is found in Wolf et al. (2004).

We demonstrated in the simulation results in the Main paper how the posterior distribution of the rank concentrates around the estimated value as the number of data points is increased. The application of the methodology to the actual dataset on galaxy photometry illustrates high concentration when the observations change from the sub-sample with $n = 500$ observations to the full set (Figure 3).

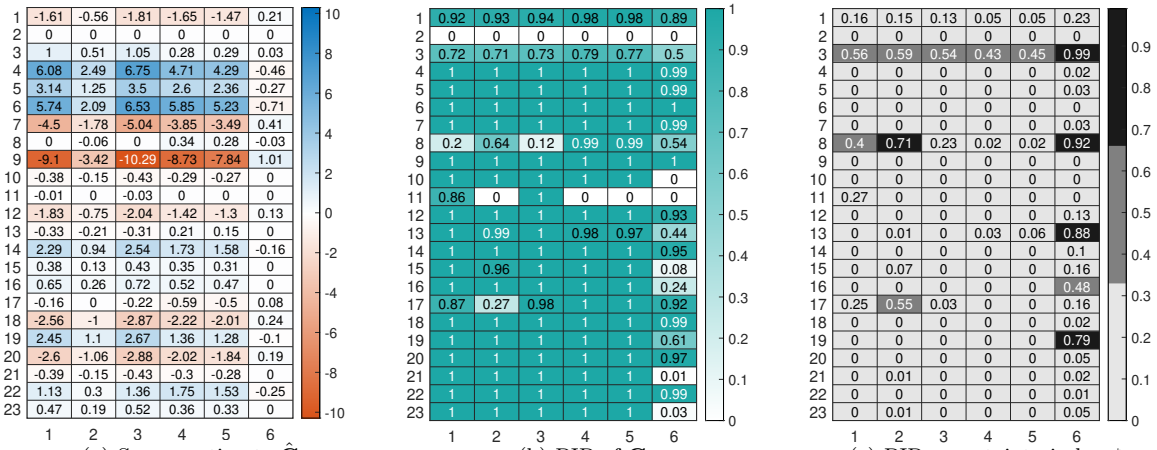
The estimated sparse coefficient matrix $\hat{\mathbf{C}}$ in the sub-sample with 500 data points sets 28% of the coefficients to zero and excludes four covariates from the model. Meanwhile, the full model encompasses 14% of its entries as zeros, and only one complete row of



(a) $n = 500$ (b) $n = 3,438$

Figure 3: Posterior distribution of the estimated rank in the sub-sample of 500 observations of the galaxy dataset (panel a), and the full set (panel b). The plots exhibit posterior concentration of the rank around 2 as the number of data points is increased.

zeros (Figure 4a). The uncertainty about this decision is low, as reported by the PIP and the PIP uncertainty index in Figures 4b and 4c.



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

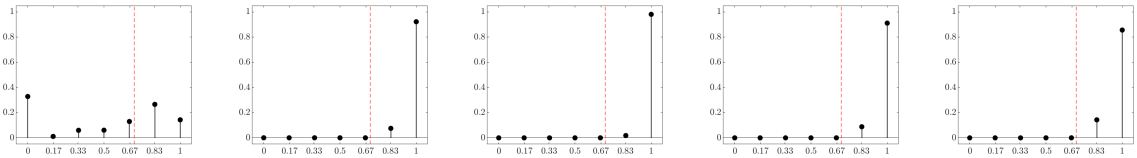
Figure 4: Sparse estimate $\hat{\mathbf{C}}$ of the coefficient matrix C of the linear regression model with the full galaxy dataset (panel a), the uncertainty about this estimation through the posterior inclusion probabilities (panel b), and the PIP uncertainty index, in a grey-colour scale according to low ($\zeta_{jk} \leq 1/3$), medium ($1/3 < \zeta_{jk} \leq 2/3$), or high ($\zeta_{jk} > 2/3$) uncertainty (panel c).

A measure that synthesises the relevance of each covariate over all responses is the relevance index. Table 8 presents the summary statistics of the distribution of RI in the full dataset, as a method to identify the relevant covariates in the model when considering their effect across all responses. The uncertainty about the inclusion of a covariate is given by the standard deviation of the relevance index: the higher the dispersion of RI, the higher the uncertainty about inclusion or exclusion. For instance, covariate x_3 is estimated to have an effect on the 6 responses (Figure 4a). Nonetheless, $\mathbf{C}_{3k} > 1/3$ for all k (Figure 4c), interpreted as a medium to high uncertainty for all of its entries, and the corresponding relevance index in Table 8 exhibits the highest variance (std $= 0.372$).

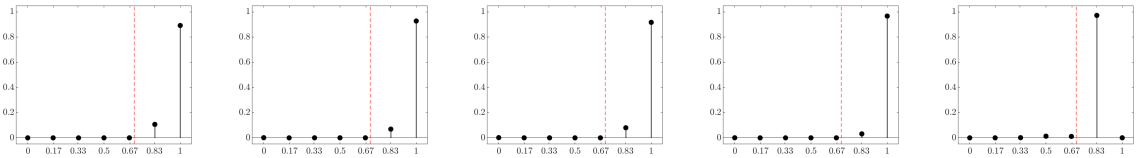
x_j	Mode	Mean	Std	Q25	Q50	Q75
1	1	0.937	0.165	1	1	1
2	0	0	0.011	0	0	0
3	1	0.704	0.372	0.5	0.833	1
4	1	0.998	0.016	1	1	1
5	1	0.998	0.019	1	1	1
6	1	1	0.004	1	1	1
7	1	0.998	0.019	1	1	1
8	0.667	0.58	0.177	0.5	0.667	0.667
9	1	1	0.005	1	1	1
10	0.833	0.833	0.004	0.833	0.833	0.833
11	0.333	0.311	0.058	0.333	0.333	0.333
12	1	0.989	0.042	1	1	1
13	0.833	0.897	0.106	0.833	0.833	1
14	1	0.992	0.036	1	1	1
15	0.833	0.841	0.042	0.833	0.833	0.833
16	0.833	0.873	0.071	0.833	0.833	0.833
17	0.833	0.842	0.068	0.833	0.833	0.833
18	1	0.999	0.015	1	1	1
19	1	0.934	0.081	0.833	1	1
20	1	0.995	0.027	1	1	1
21	0.833	0.834	0.022	0.833	0.833	0.833
22	1	0.999	0.013	1	1	1
23	0.833	0.837	0.026	0.833	0.833	0.833

Table 8: Summary statistics of the distribution of RI in the full dataset ($n = 3,438$) for each covariate: mode, mean, standard deviation (Std), and the 25th, 50th and 75th quartiles (Q25, Q50, Q75). The shaded rows identify covariates with high uncertainty ($\text{Std}(\text{RI}_j) \geq 0.30$). A description of each covariate is found in the Supplement.

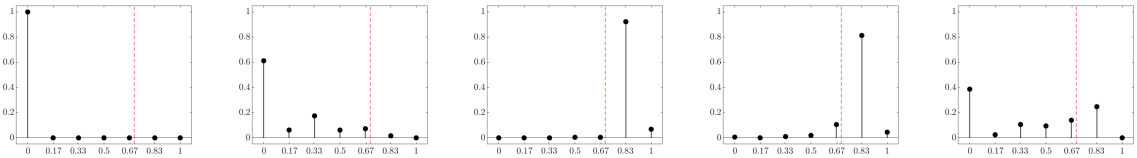
A visual representation of uncertainty in variable selection is illustrated by RI probability mass function, as depicted in Figure 5 for the analysed sub-sample. When the distribution is left-skewed or right-skewed, there is low uncertainty about the covariate inclusion or exclusion. For example, x_9 and x_{17} present this behaviour. However, a distribution without concentrated mass indicates high uncertainty, depicted in the plots of x_1, x_{15} and x_{22}. A binary decision about inclusion is given by the tail distribution or survival function of RI (Figure 6).



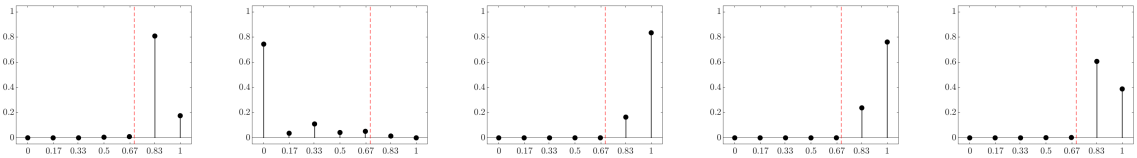
(a) RI_1. (b) RI_2. (c) RI_3. (d) RI_4. (e) RI_5.



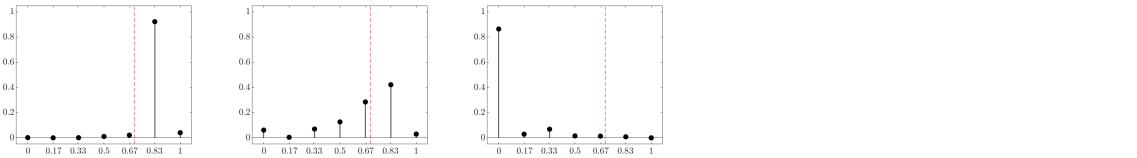
(f) RI_6. (g) RI_7. (h) RI_8. (i) RI_9. (j) RI_{10}.



(k) RI_{11}. (l) RI_{12}. (m) RI_{13}. (n) RI_{14}. (o) RI_{15}.

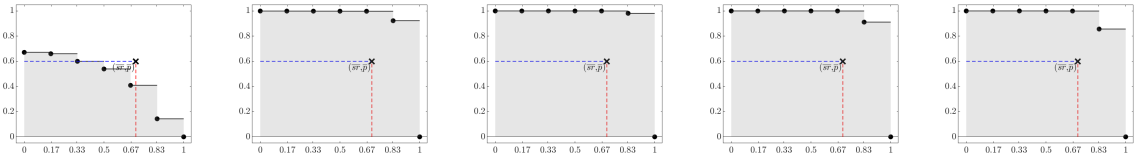


(p) RI_{16}. (q) RI_{17}. (r) RI_{18}. (s) RI_{19}. (t) RI_{20}.

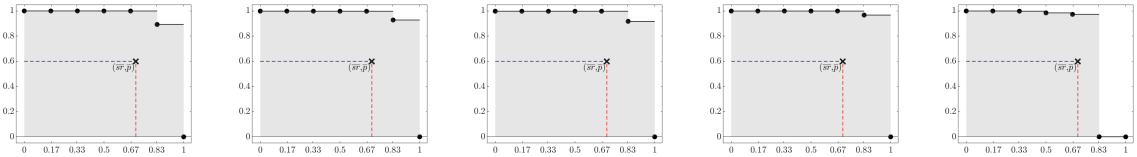


(u) RI_{21}. (v) RI_{22}. (w) RI_{23}.

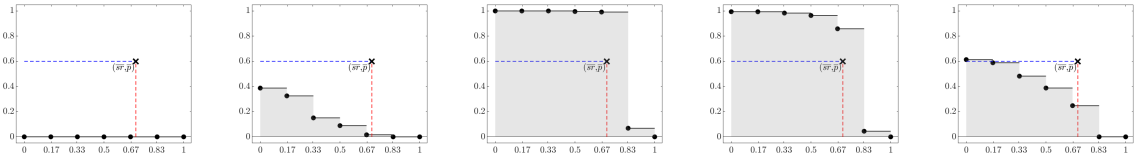
Figure 5: For each covariate in the sub-sample with $n = 500$ data points of the galaxy dataset, the probability mass function of the respective Relevance Index (RI) is reported. Here we used $\overline{sr} = 0.70$ (red vertical line): the covariates with *more* mass (to be specified through $\overline{p}$) located to the right of $\overline{sr}$ are preferred to be included over the counterpart.



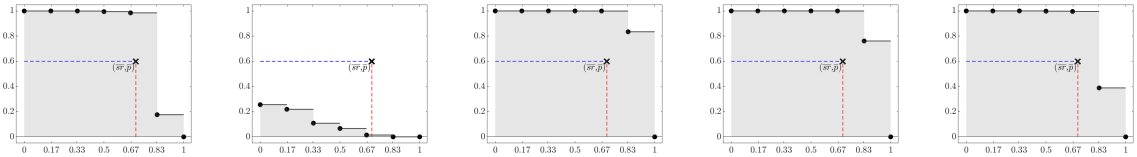
(a) RI_1. (b) RI_2. (c) RI_3. (d) RI_4. (e) RI_5.



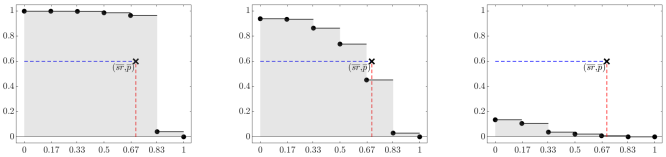
(f) RI_6. (g) RI_7. (h) RI_8. (i) RI_9. (j) RI_{10}.



(k) RI_{11}. (l) RI_{12}. (m) RI_{13}. (n) RI_{14}. (o) RI_{15}.



(p) RI_{16}. (q) RI_{17}. (r) RI_{18}. (s) RI_{19}. (t) RI_{20}.



(u) RI_{21}. (v) RI_{22}. (w) RI_{23}.

Figure 6: For each covariate in the sub-sample with $n = 500$ data points of the galaxy dataset, the survival function S_{RI} of the respective Relevance Index RI is reported. 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.

4.1 Results for $n = 500$ and $n = 1,500$

Section 4 of the Main article analysed a sub-sample of the COMBO-17 galaxy dataset consisting of $n = 500$ randomly chosen data points. This has been motivated by the intention of highlighting the ability to quantify the uncertainty of the proposed BRECS method.

As a single random sub-sample might not be representative of the full dataset, in this Section we report the results from the application of BRECS to 5 other randomly drawn sub-samples of size $n = 500$ and 6 sub-samples of size $n = 1,500$. The purpose is twofold: first, we aim to show that the difference between the results in the Main article and those with the full dataset (see the previous Section) is due to a small sub-sample size. Second, we argue that the "cross-sample" differences wipe out as the size of the sub-sample increases.

By comparing the plots in Figure 7, we find an overall similarity in the sparsity pattern of the estimated matrix $\hat{\mathbf{C}}$. However, the entry-wise PIPs and associated uncertainty index ζ show some differences in terms of the precise location of certain zero coefficients (an associated uncertainty). For example, in the 3rd run (subfigure 3.), the coefficients of the second covariate have PIPs close to 0 and very low uncertainty, whereas in the 6th run (subfigure 6.) all their PIPs except one are greater than 0.50 with medium uncertainty. This is due to the randomness in the selection of the sub-samples, which might result in selecting a handful of data that are affecting the overall results. However, notice that the overall degree of uncertainty shown by the (c) subplots seems similar across all the runs.

Comparing the results for $n = 1,500$ in Figure 8 to those for $n = 500$, we find evidence of two main changes. First, with $n = 1,500$ we notice a reduction in the number of zero rows compared to the sub-samples with fewer data points, although it remains higher than that of the full dataset. Second, the average level of uncertainty is lower with $n = 1,500$, yet higher than that of the full dataset.

Overall, these results are in line with expectations: the larger the size of the sub-samples, the closer the results get to the full sample. Moreover, as the sub-sample size increases, the uncertainty around parameter classification (zero versus non-zero) decreases.

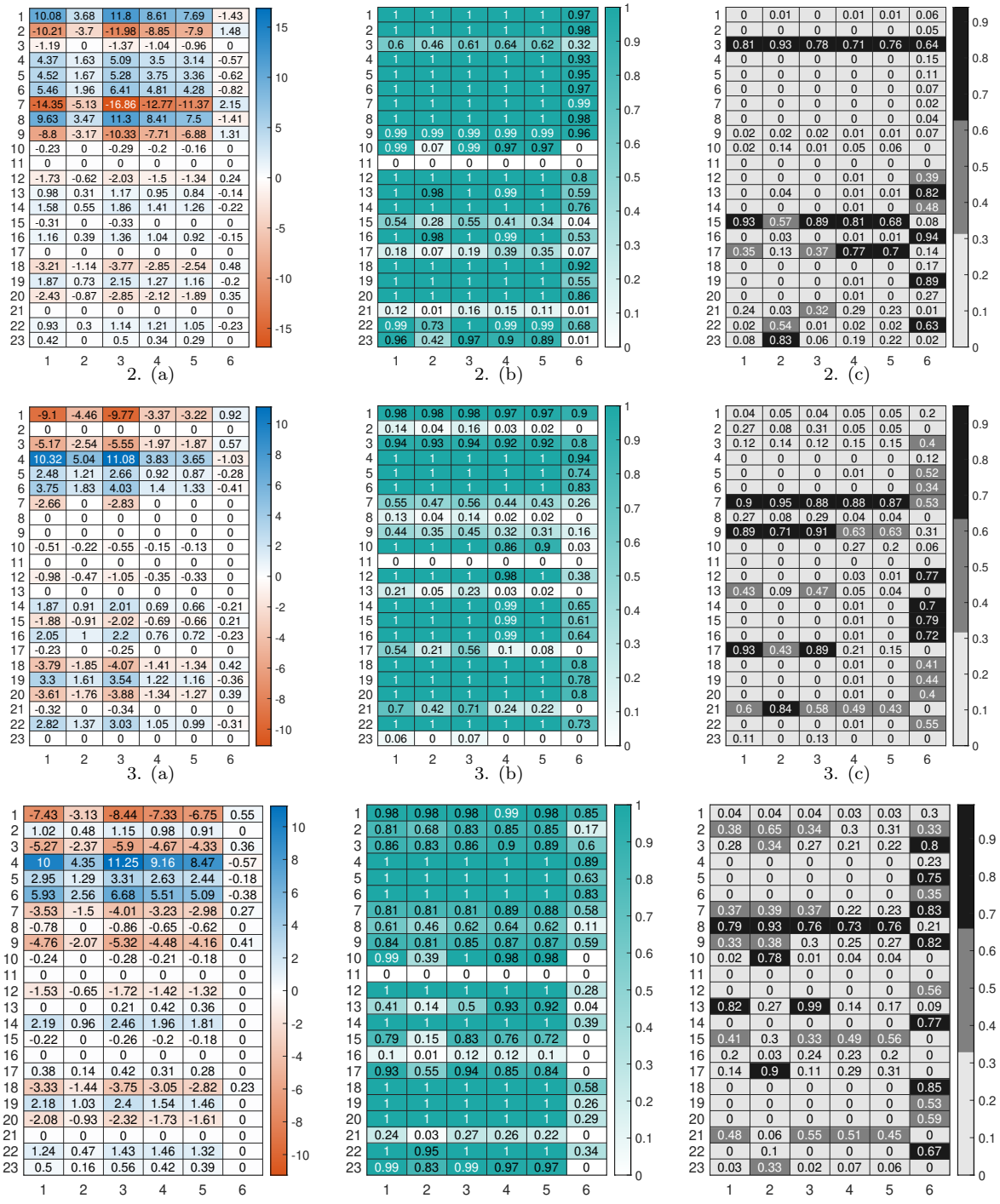
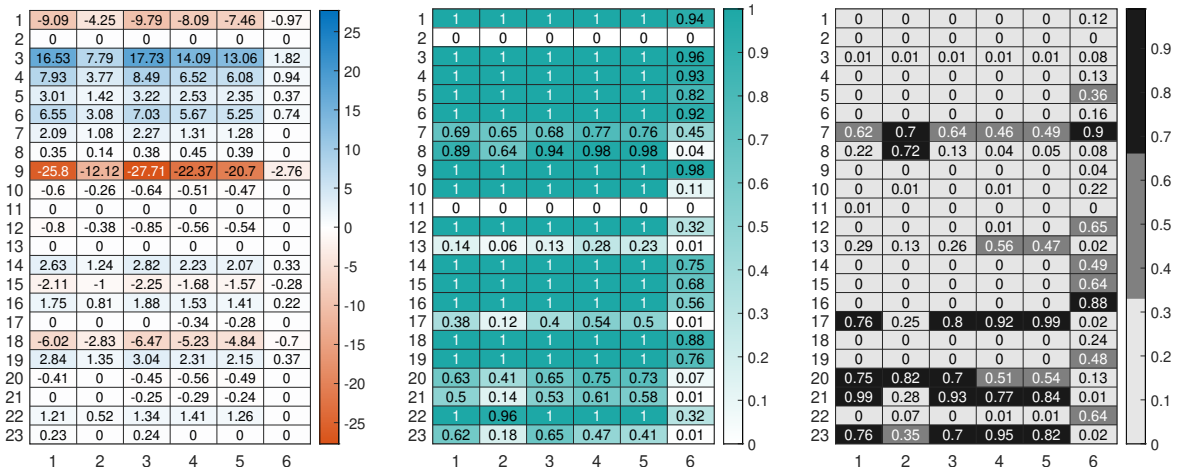
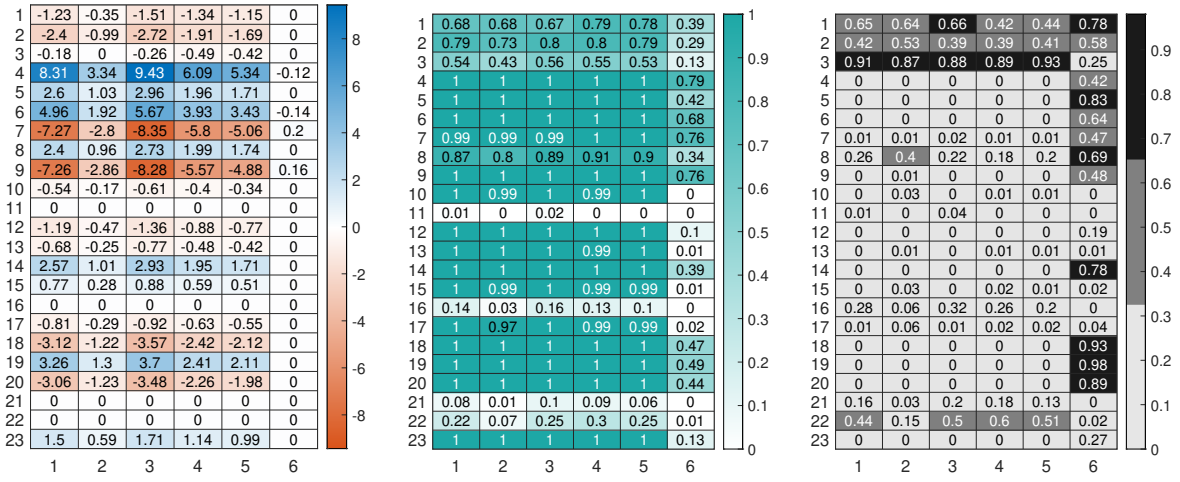


Figure 7: The methodology was applied six times to the dataset with $n = 500$. The rows correspond to each run (refer to the Main paper for the results of the first run), where we plot the sparse estimates $\hat{\mathbf{C}}$ of the coefficient matrix $\mathbf{C}$ (panel a), the uncertainty about this 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).



5. (a) 5. (b) 5. (c)



6. (a) 6. (b) 6. (c)

Figure 7: The methodology was applied six times to the dataset with $n = 500$. The rows correspond to each run (refer to the Main paper for the results of the first run), where we plot the sparse estimates $\hat{\mathbf{C}}$ of the coefficient matrix $\mathbf{C}$ (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) (cont.).

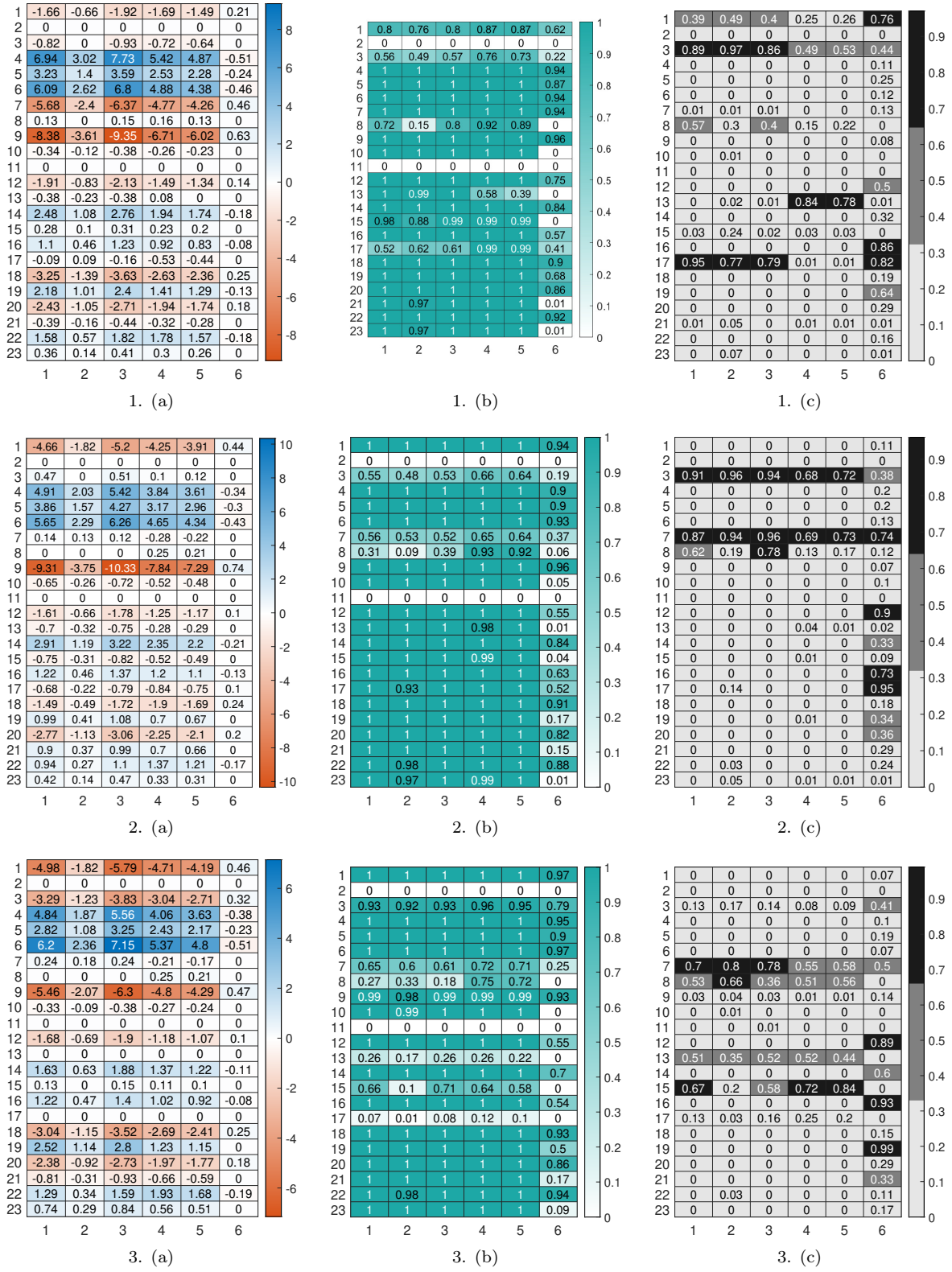
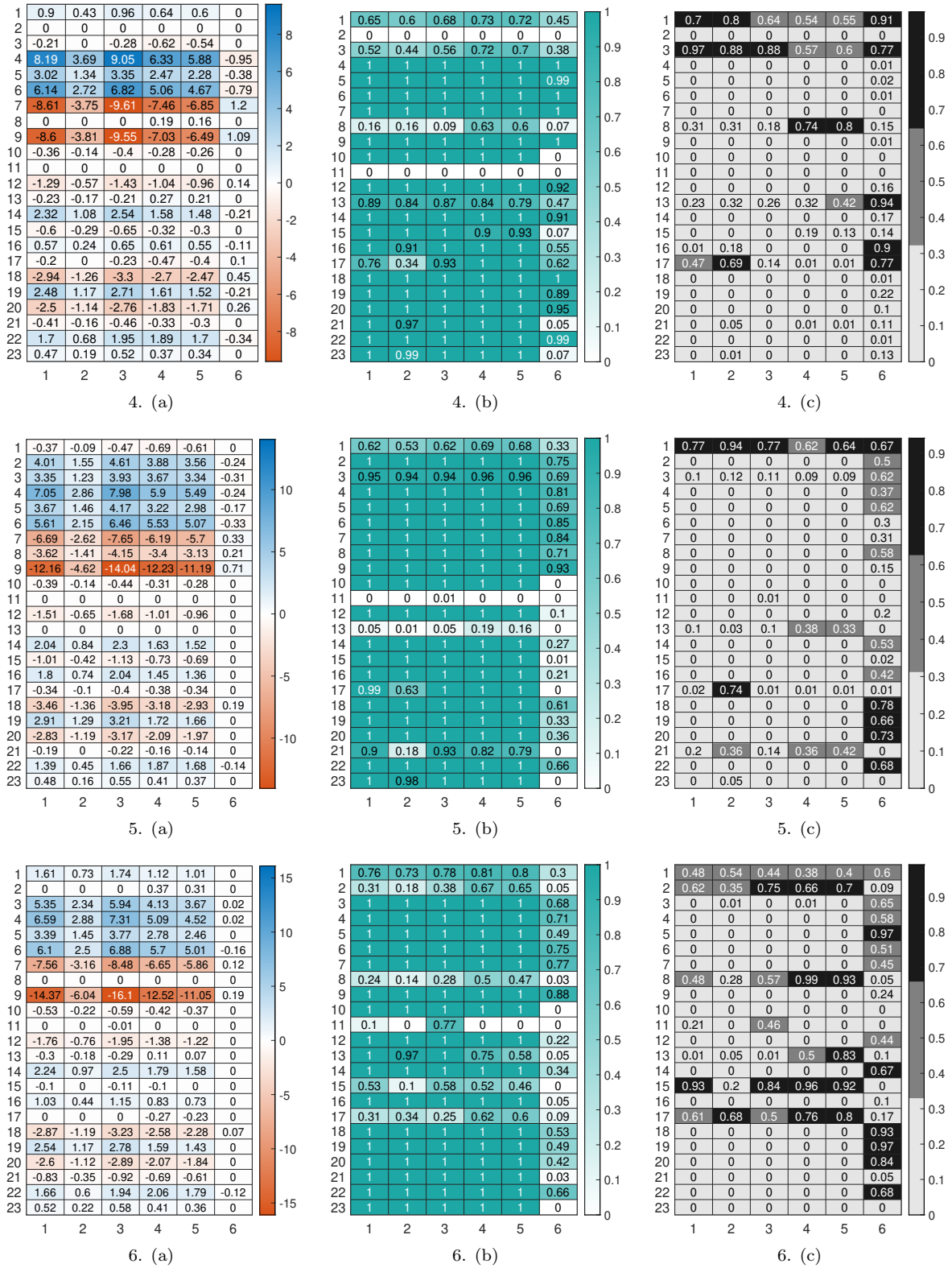


Figure 8: The methodology was applied six times to the dataset with $n = 1500$. The rows correspond to each run (refer to the Main paper for the results of the first run), where we plot the sparse estimates $\hat{\mathbf{C}}$ of the coefficient matrix $\mathbf{C}$ (panel a), the uncertainty about this 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).



4. (a) 4. (b) 4. (c)

5. (a) 5. (b) 5. (c)

6. (a) 6. (b) 6. (c)

Figure 8: The methodology was applied six times to the dataset with $n = 1500$. The rows correspond to each run (refer to the Main paper for the results of the first run), where we plot the sparse estimates $\hat{\mathbf{C}}$ of the coefficient matrix $\mathbf{C}$ (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) (cont.).

4.2 Results for alternative variable selection methods

Concerning variable selection, we propose the RI and suggest a rule-of-thumb based on two tuning parameters for which we provide a concrete interpretation. As an alternative, we have implemented the group lasso method used by Chakraborty et al. (2019) to obtain an indicator for each variable and MCMC iteration. This information was then used to define the quantity $\text{PIP}_j^* \in [0,1]$, which reports the posterior probability of including covariate j. We have also introduced another measure, PIP_j^{**}, which is based on the entry-wise probabilities PIP_{ij}.

Chakraborty et al. (2019) use a version of the SAVS to solve an optimisation problem inducing row sparsity via a group lasso penalty. Specifically, the jth (sparsified) row of the coefficient matrix is obtained as

$$\hat{\mathbf{C}}_R^{(j)} = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\Big(1 - \frac{\mu_j}{2\|\mathbf{X}_j'R_j\|}\Big)_+ \mathbf{X}_j'R_j, \tag{4.1}$$

with $R_j = \mathbf{XC} - \sum_{i\neq j}\mathbf{X}_i\hat{\mathbf{C}}_R^{(i)}$ and the weight $\mu_j = \|\mathbf{C}_j\|^{-2}$. By computing the row-sparsified matrix $\hat{\mathbf{C}}_R$ at each iteration, we can define the posterior inclusion probability of each row $j = 1,\ldots,p$ as

$$\text{PIP}_j^* = 1 - \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}(\hat{\mathbf{C}}_R^{(j)} = \mathbf{0}), \tag{4.2}$$

then define the associated uncertainty index, ζ_j, in a way analogous to the scalar ζ_{jk}.

An alternative way to define the PIP row-wise in the context of our proposed method is as follows

$$\text{PIP}_j^{**} = 1 - \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}\big(\bar{\mathbf{C}}_{j1}^{(m)} = 0 \wedge \ldots \wedge \bar{\mathbf{C}}_{jq}^{(m)} = 0\big) \tag{4.3}$$

$$= \frac{1}{M}\sum_{m=1}^{M}\mathbb{I}\big(\bar{\mathbf{C}}_{j1}^{(m)} \neq 0 \vee \ldots \vee \bar{\mathbf{C}}_{jq}^{(m)} \neq 0\big). \tag{4.4}$$

This would produce potentially a conservative estimate of the number of zero rows as it suffices to have one non-zero coefficient in order to include a covariate. The problem might be exacerbated when q is big and just one coefficient is non-zero, as that would imply the inclusion of the covariate even though $q-1$ response variables are unaffected by it.

Table 9 presents the application of the row-wise methods for uncertainty quantification about variable inclusion discussed above. The galaxy dataset of Section 5.2 in

the Main manuscript comprises 23 covariates, and the relevance of each one is assessed by PIP* and PIP**. We report as well the associated uncertainty indices ζ_j^* and ζ_j^{**}, computed in the same fashion as ζ_{ij} for PIP$_{ij}$.

j	$\mathbf{PIP_j^*}$	ζ_j^*	$\mathbf{PIP_j^{**}}$	ζ_j^{**}	j	$\mathbf{PIP_j^*}$	ζ_j^*	$\mathbf{PIP_j^{**}}$	ζ_j^{**}
1	0.7361	0.5278	0.6714	0.6571	13	1.0000	0.0000	1.0000	0.0000
2	0.9998	0.0004	0.9995	0.0011	14	0.9986	0.0029	0.9941	0.0118
3	1.0000	0.0000	1.0000	0.0000	15	0.7330	0.5339	0.6134	0.7731
4	1.0000	0.0000	1.0000	0.0000	16	0.9998	0.0004	0.9996	0.0007
5	1.0000	0.0000	1.0000	0.0000	17	0.4790	0.9579	0.2555	0.5110
6	1.0000	0.0000	1.0000	0.0000	18	1.0000	0.0000	1.0000	0.0000
7	0.9987	0.0025	0.9987	0.0025	19	1.0000	0.0000	1.0000	0.0000
8	0.9982	0.0036	0.9982	0.0036	20	1.0000	0.0000	1.0000	0.0000
9	0.9995	0.0011	0.9995	0.0011	21	0.9996	0.0007	0.9982	0.0036
10	0.9998	0.0004	0.9998	0.0004	22	0.9563	0.0874	0.9382	0.1235
11	0.9415	0.1171	0.0004	0.0007	23	0.3271	0.6543	0.1355	0.2711
12	0.6655	0.6689	0.3875	0.7749					

Table 9: Quantification of uncertainty about the inclusion of covariate $j = 1, \ldots, 23$ from the galaxy dataset (Section 5.2). PIP* and PIP** values close to 1 (0) indicate strong evidence that the covariate is relevant (irrelevant). ζ^* and ζ^{**} values close to 1 (0) indicate high (low) uncertainty about including or not a covariate.

In the Main manuscript, we apply our *rule of thumb* based on the RI to the galaxy dataset. Under the choice $\overline{sr} = 0.60$ (share of response variables on which the jth covariate is required to have a significant impact) and $\overline{p} = 0.40$ (minimum probability with which this occurs), we exclude covariates 1, 11, 12, 15, 17, 22 and 23 from the model. A decision about variable selection based on PIP$_j^*$ and PIP$_j^{**}$ can be achieved by choosing a threshold for these values. For instance, a minimum posterior probability of inclusion of 0.5 is a natural choice. In this case PIP$_j^*$ agrees on the exclusion of covariates 17 and 23, while PIP$_j^{**}$ points towards excluding covariates 11, 12, 17, and 23. Covariates 1 and 15 exhibit a high uncertainty under our RI approach (see Table 5 of the Main manuscript), which is equally observed in Table 9 above. The standard deviation of RI$_{22}$ is at a medium level, and the conclusion of excluding it from the model would not have been reached for slightly different values of $\overline{sr}$ and/or $\overline{p}$.

There is an apparent contrast between PIP$_{11}^*$ and PIP$_{11}^{**}$. The group lasso penalty of Chakraborty et al. (2019) is a conservative approach compared to our implementation of the SAVS. $\hat{\mathbf{C}}_R^{11}$ did produce zero rows in a few iterations of the MCMC, but for the most part, the estimated values of this covariate were close to zero[1]. The penalisation

[1]The norm of the average $\hat{\mathbf{C}}_R^{11}$ across iterations is 0.049. For comparison, the next smallest norm value is $\|\hat{\mathbf{C}}_R^{23}\| = 0.134$, and the maximum is $\|\hat{\mathbf{C}}_R^{3}\| = 53.428$.

imposed in $\bar{\mathbf{C}}_{11}$ continuously produced zero estimates, as the otherwise unpenalised values were not considered significant due to their small magnitude.

Finally, we remark that the 0.5 threshold (as any other) may lead to different conclusions according to the criterion chosen. For instance, covariate 12 should be included by PIP*, but excluded according to PIP**. This evidence is motivated by the high uncertainty in the posterior distributions of the coefficients associated to this covariate. However, we believe this uncertainty is hardly captured by the scalar-valued PIP* and PIP**, whereas the proposed RI is a better suited tool to explain and investigate this situation, also emphasising that in such non-extreme situations, the final decision to include/exclude a covariate is potentially highly subjective.

4.3 Comparison to other methods

We study a forecasting exercise where we predict a subset of n^* observations of the sub-sample with $n = 500$ examined in Section 5.2. Then, we compute the mean squared error (MSE) of the fitted values, $\hat{\mathbf{y}}_i$, as $\sum_{i=1}^{n^*}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2/q$, and the mean absolute error (MAE) as $\sum_{i=1}^{n^*}|\mathbf{y}_i - \hat{\mathbf{y}}_i|/q$. We perform this procedure by applying our method and the frequentist approaches of Chen et al. (2013) (ANN) and She and Chen (2017) (RRRR). We outperform the latter in most cases in terms of the MSE, with the remaining metrics exhibiting comparable values, albeit with slight differences. Table 10 provides a summary of the forecast results using an expanding window and a rolling window.

Forecasting window	**n***	**MSE**			**MAE**		
		ANN	**RRRR**	**RRcs**	**ANN**	**RRRR**	**RRcs**
Expanding	400	0.921	1.311	1.085	0.616	0.634	0.682
Expanding	250	0.660	1.925	0.835	0.553	0.585	0.647
Expanding	100	0.482	0.416	0.534	0.534	0.467	0.560
Rolling	250	0.576	1.506	0.726	0.550	0.606	0.632

Table 10: Mean squared error (MSE) and mean absolute error (MAE) of the true responses versus the fitted values through a forecast with expanding or rolling window, where n^* represents the number of out-of-sample points. The table reports the average errors across the predicted observations for each of the three methods, ANN, RRRR, and RRcs.

References

Bhattacharya, A., Chakraborty, A., and Mallick, B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika*, 103(4):985–991.

Bura, E. and Cook, R. (2003). Rank estimation in reduced-rank regression. *Journal of Multivariate Analysis*, 87(1):159–176.

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

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, 100:901–920.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer New York.

Izenman, A. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer, New York.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society Series B*, 63(2):325–338.

She, Y. and Chen, K. (2017). Robust reduced-rank regression. *Biometrika*, 104(3):633–647.

Wolf, C., Meisenheimer, K., Kleinheinrich, M., Borch, A., Dye, S., Gray, M., Wisotzki, L., Bell, E. F., Rix, H.-W., Cimatti, A., Hasinger, G., and Szokoly, G. (2004). A catalogue of the Chandra Deep Field South with multi-colour classification and photometric redshifts from COMBO-17. *Astronomy & Astrophysics*, 421(3):913–936.