# Bayesian semiparametric analysis of multivariate continuous responses, with variable selection
# Supplement

Georgios Papageorgiou and Benjamin C. Marshall
Department of Economics, Mathematics and Statistics
Birkbeck, University of London, UK
g.papageorgiou@bbk.ac.uk

February 12, 2020

## 1 Additional simulation results

Tables 8 and 9 present results on relative bias and variance for the second choice of the mean model, $\mu_j = \beta_{0j} + \sum_{k=1}^{3} \beta_{jk} x_k$. We can observe that the patterns of relative bias and variance are the same as those seen in the main body of the paper for the first mean model choice, however, the gains for the second model are generally more pronounced.

Tables 10 and 11 present relative bias and variance results for the third choice of the mean model, $\mu_j = \beta_{0j} + \sum_{k=1}^{10} \beta_{jk} x_k$. The relative bias decreases as $\rho$ increases, for all $n$ and $d$. This decrease, for $n = 50$, is slower than that observed for the first and second mean models, while for $n = 150$, it is faster than that observed for the first and second mean models. The relative variance follows the same pattern as that observed for the other two mean models: it decreases as $\rho$ increases, for all $n$ and $d$. This decrease is faster than that observed for the first and second mean models.

Table 12 presents the comparison with the MRCE method for the third mean model. The proposed method has bias that is usually less than half of the bias of the MRCE approach.

Table 13 presents results on the variable selection performance of the proposed model for the third mean model. For this case, when fitting one-dimensional models, the irrelevant regressors where included 9.81% of the time when $n = 50$, and 6.87% of the time when $n = 150$. From this observation and from Table 13 we can see that probabilities of inclusion decrease as the sample size increases and that, for fixed dimension $d$, the probabilities decrease as the correlation coefficient increases.

Table 8: Simulation study results: the entries of the table are the relative biases $B(d)/B(1), d = 2, 4, 6, 10$, expressed as percentages. Rows refer to the sample size $n = 50, 150$, and columns to the correlation between the responses, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. Results are based on the second mean model, and 40 replicate datasets per sample size by correlation combination.

| $d = 2$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 4$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 91.85 | 83.54 | 74.45 | 62.65 | 48.83 | 50 | 92.17 | 78.90 | 65.24 | 54.41 | 45.88 |
| 150 | 98.26 | 95.02 | 91.95 | 88.25 | 78.35 | 150 | 98.79 | 99.44 | 95.10 | 88.32 | 78.06 |
| $d = 6$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 10$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 50 | 97.90 | 77.97 | 63.69 | 53.56 | 45.45 | 50 | 95.21 | 78.98 | 61.84 | 52.34 | 45.04 |
| 150 | 99.72 | 99.84 | 94.98 | 87.50 | 77.10 | 150 | 97.48 | 95.52 | 91.48 | 84.96 | 76.16 |

Table 9: Simulation study results: the entries of the table are the relative variances $V(d)/V(1), d = 2, 4, 6, 10$, expressed as percentages. Rows refer to the sample size $n = 50, 150$, and columns to the correlation between the responses, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. Results are based on the second mean model, and 40 replicate datasets per sample size by correlation combination.

| $d = 2$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 4$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 100.07 | 96.25 | 88.08 | 78.23 | 61.34 | 50 | 100.72 | 92.36 | 79.99 | 67.29 | 52.25 |
| 150 | 99.60 | 93.80 | 85.71 | 73.44 | 55.65 | 150 | 99.17 | 90.33 | 79.34 | 67.28 | 52.27 |
| $d = 6$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 10$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 50 | 99.19 | 89.42 | 76.23 | 64.52 | 53.94 | 50 | 94.92 | 84.59 | 74.53 | 65.06 | 52.55 |
| 150 | 98.65 | 88.43 | 77.53 | 65.44 | 51.61 | 150 | 97.26 | 86.97 | 76.49 | 63.93 | 51.49 |

Table 10: Simulation study results: the entries of the table are the relative biases $B(d)/B(1), d = 2, 4, 6, 10$, expressed as percentages. Rows refer to the sample size $n = 50, 150$, and columns to the correlation between the responses, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. Results are based on the third mean model, and 40 replicate datasets per sample size by correlation combination.

| $d = 2$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 4$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 97.49 | 97.11 | 95.34 | 84.16 | 63.14 | 50 | 99.82 | 91.03 | 84.68 | 74.34 | 58.50 |
| 150 | 100.49 | 91.53 | 80.04 | 58.41 | 34.69 | 150 | 102.94 | 93.67 | 81.13 | 57.69 | 33.55 |
| $d = 6$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 10$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 50 | 102.88 | 94.32 | 83.82 | 74.63 | 58.47 | 50 | 107.27 | 99.47 | 88.81 | 77.23 | 59.93 |
| 150 | 98.80 | 88.18 | 75.77 | 55.32 | 33.08 | 150 | 98.79 | 83.87 | 70.26 | 52.83 | 32.19 |

Table 11: Simulation study results: the entries of the table are the relative variances $V(d)/V(1), d = 2, 4, 6, 10$, expressed as percentages. Rows refer to the sample size $n = 50, 150$, and columns to the correlation between the responses, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. Results are based on the third mean model, and 40 replicate datasets per sample size by correlation combination.

| $d = 2$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 4$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 99.82 | 96.17 | 86.11 | 70.01 | 46.72 | 50 | 100.52 | 93.58 | 79.00 | 60.27 | 38.11 |
| 150 | 99.71 | 94.82 | 83.19 | 67.40 | 47.07 | 150 | 100.35 | 88.92 | 73.07 | 54.95 | 36.49 |
| $d = 6$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 10$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 50 | 99.90 | 89.40 | 74.20 | 56.48 | 37.92 | 50 | 97.65 | 85.21 | 71.05 | 54.65 | 36.13 |
| 150 | 98.06 | 84.78 | 68.21 | 51.49 | 35.61 | 150 | 97.81 | 84.20 | 67.20 | 51.46 | 35.15 |

Table 12: Simulation study results: the entries of the table are the relative biases $B(d)/B_M(d), d = 2, 4, 6, 10$, expressed as percentages. Rows refer to the sample size $n = 50, 150$, and columns to the correlation between the responses, $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$. Results are based on the third mean model, and 40 replicate datasets per sample size by correlation combination.

| $d = 2$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 4$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 42.57 | 39.61 | 45.52 | 50.37 | 49.86 | 50 | 29.88 | 33.41 | 42.21 | 51.35 | 61.39 |
| 150 | 55.00 | 53.50 | 48.59 | 34.37 | 27.48 | 150 | 56.49 | 56.42 | 36.55 | 38.20 | 35.64 |
| $d = 6$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | $d = 10$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 50 | 27.44 | 33.36 | 43.75 | 56.66 | 58.28 | 50 | 30.59 | 35.55 | 43.79 | 48.56 | 58.47 |
| 150 | 44.92 | 42.89 | 35.81 | 34.62 | 30.78 | 150 | 36.44 | 22.77 | 26.53 | 34.61 | 31.07 |

Table 13: Simulation study results: the entries of the table are the posterior probabilities, expressed as percentages, that at least one of $x_2, \ldots, x_{10}$ is included in the mean model of the first response. Rows refer to the dimension of the fitted model $d = 2, 4, 6, 10$, columns to the correlation coefficient $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, and the two parts of the table to the two sample sizes $n = 50, 150$. Results are based on 40 replicate datasets per sample size by correlation combination.

| | $n = 50$ | | | | | $n = 150$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 2 | 10.13 | 10.08 | 9.80 | 9.18 | 7.09 | 6.60 | 6.58 | 6.33 | 5.82 | 4.42 |
| 4 | 10.47 | 10.46 | 10.14 | 9.65 | 7.77 | 6.79 | 6.60 | 6.32 | 5.85 | 4.46 |
| 6 | 10.86 | 10.68 | 10.30 | 9.90 | 8.03 | 6.67 | 6.43 | 6.17 | 5.77 | 4.64 |
| 10 | 11.03 | 10.90 | 10.72 | 10.35 | 9.06 | 6.81 | 6.71 | 6.47 | 6.23 | 5.13 |

# 2 MCMC algorithm

Here we provide all details of the MCMC sampler of the three correlation models.

## 2.1 MCMC algorithm for the common correlations model

Starting from the common correlations model, the algorithm proceeds as follows:

1. As suggested by Chan et al. (2006), the elements of $\boldsymbol{\gamma}_{jk}, j = 1, \ldots, p, k = 1, \ldots, K$, are updated in random order and in blocks of random size. Let $\boldsymbol{\gamma}_{Bjk}$ be a block of elements of $\boldsymbol{\gamma}_{jk}$. The proposed value for $\boldsymbol{\gamma}_{Bjk}$ is obtained from its prior with the remaining elements of $\boldsymbol{\gamma}_{jk}$, denoted by $\boldsymbol{\gamma}_{Cjk}$, kept at their current value. The proposal pmf is obtained from the Bernoulli prior with $\pi_{\mu jk}$ integrated out

$$p(\boldsymbol{\gamma}_{Bjk}|\boldsymbol{\gamma}_{Cjk}) = \frac{p(\boldsymbol{\gamma}_{jk})}{p(\boldsymbol{\gamma}_{Cjk})} = \frac{\text{Beta}(c_{\mu jk} + N(\boldsymbol{\gamma}_{jk}), d_{\mu jk} + q_{\mu k} - N(\boldsymbol{\gamma}_{jk}))}{\text{Beta}(c_{\mu jk} + N(\boldsymbol{\gamma}_{Cjk}), d_{\mu jk} + q_{\mu k} - L(\boldsymbol{\gamma}_{Bjk}) - N(\boldsymbol{\gamma}_{Cjk}))},$$

where $L(\boldsymbol{\gamma}_{Bjk})$ denotes the length of $\boldsymbol{\gamma}_{Bjk}$ i.e. the size of the block. For this proposal pmf, the acceptance probability of the Metropolis-Hastings move reduces to the ratio of the likelihoods in (17) of the main body of the paper

$$\min\left\{1, (c_\beta + 1)^{\{N(\gamma^C) - N(\gamma^P)\}/2} \exp\{(S^C - S^P)/2\}\right\},$$

where superscripts $P$ and $C$ denote proposed and currents values respectively.

2. Pairs $(\boldsymbol{\delta}_{jk}, \boldsymbol{\alpha}_{jk}), j = 1, \ldots, p, k = 1, \ldots, Q$, are updated simultaneously. Similarly to the updating of $\boldsymbol{\gamma}_{jk}$, the elements of $\boldsymbol{\delta}_{jk}$ are updated in random order and in blocks of random size. Let $\boldsymbol{\delta}_{Bjk}$ denote a block. Blocks $\boldsymbol{\delta}_{Bjk}$ and the whole vector $\boldsymbol{\alpha}_{jk}$ are generated simultaneously. As was mentioned by Chan et al. (2006), generating the whole vector $\boldsymbol{\alpha}_{jk}$, instead of subvector $\boldsymbol{\alpha}_{Bjk}$, is necessary in order to make $\boldsymbol{\alpha}_{jk}$ consistent with the proposed value of $\boldsymbol{\delta}_{jk}$.

   Generating the proposed value for $\boldsymbol{\delta}_{Bjk}$ is done in a similar way as was done for $\boldsymbol{\gamma}_{Bjk}$. Let $\boldsymbol{\delta}_{jk}^P$ denote the proposed value of $\boldsymbol{\delta}_{jk}$. Next, we describe how the proposed vale for $\boldsymbol{\alpha}_{\delta_{jk}^P jk}$ is obtained. To avoid clutter, proposed values $\boldsymbol{\alpha}_{\delta_{jk}^P jk}^P$ will be denoted by the simpler $\boldsymbol{\alpha}_{jk}^P$. The development that follows is in the spirit of Chan et al. (2006) who built on the work of Gamerman (1997).

   Let $\hat{\boldsymbol{\beta}}_\gamma^C = \{c_\beta/(1+c_\beta)\}(\tilde{\boldsymbol{X}}_\gamma^\top \tilde{\boldsymbol{X}}_\gamma)^{-1}\tilde{\boldsymbol{X}}_\gamma^\top \tilde{\boldsymbol{Y}}$ denote the current value of the posterior mean of $\boldsymbol{\beta}_\gamma$. Define the current squared residuals

$$e_{ij}^C = (y_{ij} - \hat{\beta}_{0j}^C - \boldsymbol{x}_{\gamma_j i}^\top \hat{\boldsymbol{\beta}}_{\gamma_j j}^C)^2.$$

   These have an approximate $\sigma_{ij}^2 \chi_1^2$ distribution, where $\sigma_{ij}^2 = \sigma_j^2 \exp(\boldsymbol{z}_{\delta_j i}^\top \boldsymbol{\alpha}_{\delta_j j})$. The latter defines a Gamma generalized linear model (GLM) for the squared residuals with mean $\sigma_{ij}^2$, which, utilizing a log-link, can be thought of as Gamma GLM with an offset term: $\log(\sigma_{ij}^2) = \log(\sigma_j^2) + \boldsymbol{z}_{\delta_j i}^\top \boldsymbol{\alpha}_{\delta_j j}$. Given $\boldsymbol{\delta}_{jk}^P$, the proposal density for $\boldsymbol{\alpha}_{\delta_{jk}^P jk}$ is derived utilizing the one step iteratively reweighted least squares algorithm. This proceeds as follows. First define the transformed observations

$$d_{ij}^C(\boldsymbol{\alpha}_j^C) = \log(\sigma_j^2) + \boldsymbol{z}_i^\top \boldsymbol{\alpha}_j^C + \frac{e_{ij}^C - (\sigma_{ij}^2)^C}{(\sigma_{ij}^2)^C},$$

   where superscript $C$ denotes current values. Further, let $\boldsymbol{d}_j^C$ denote the vector of $d_{ij}^C$.

Next we define

$$\boldsymbol{\Delta}(\boldsymbol{\delta}_{jk}^P) = (c_{\alpha j}^{-1}\boldsymbol{I} + \boldsymbol{Z}_{\delta_{jk}^P}^\top \boldsymbol{Z}_{\delta_{jk}^P})^{-1} \text{ and } \hat{\boldsymbol{\alpha}}(\boldsymbol{\delta}_{jk}^P, \boldsymbol{\alpha}_j^C) = \boldsymbol{\Delta}_{\delta_{jk}^P} \boldsymbol{Z}_{\delta_{jk}^P}^\top \boldsymbol{d}_j^C,$$

where $\boldsymbol{Z}_{\delta_{jk}}$ is a submatrix of $\boldsymbol{Z}_{\delta_j}$ that was defined after (11), and it considers only the columns that pertain to the $k$th effect. The proposed value $\boldsymbol{\alpha}_{jk}^P$ is obtained from a multivariate normal distribution with mean $\hat{\boldsymbol{\alpha}}(\boldsymbol{\delta}_{jk}^P, \boldsymbol{\alpha}_j^C)$ and covariance $h\boldsymbol{\Delta}(\boldsymbol{\delta}_{jk}^P)$, denoted as $N(\boldsymbol{\alpha}_{jk}^P; \hat{\boldsymbol{\alpha}}(\boldsymbol{\delta}_{jk}^P, \boldsymbol{\alpha}_j^C), h_{jk}\boldsymbol{\Delta}(\boldsymbol{\delta}_{jk}^P))$, where $h_{jk}$ is a free parameter that we introduce and select adaptively (Roberts and Rosenthal, 2009) in order to achieve an acceptance probability of $20\% - 25\%$ (Roberts and Rosenthal, 2001).

Let $N(\boldsymbol{\alpha}_{jk}^C; \hat{\boldsymbol{\alpha}}(\boldsymbol{\delta}_{jk}^C, \boldsymbol{\alpha}_j^P), h_{jk}\boldsymbol{\Delta}(\boldsymbol{\delta}_{jk}^C))$ denote the proposal density for taking a step in the reverse direction, from model $\boldsymbol{\delta}_{jk}^P$ to $\boldsymbol{\delta}_{jk}^C$. Then the acceptance probability of the pair $(\boldsymbol{\delta}_{jk}^P, \boldsymbol{\alpha}_{\delta_{jk}^P jk}^P)$ is

$$\min\left\{1, \frac{|\boldsymbol{\Sigma}(\boldsymbol{R}, \boldsymbol{\alpha}^P, \boldsymbol{\delta}^P, \boldsymbol{\sigma}^2)|^{-\frac{1}{2}} \exp(-S^P/2)}{|\boldsymbol{\Sigma}(\boldsymbol{R}, \boldsymbol{\alpha}^C, \boldsymbol{\delta}^C, \boldsymbol{\sigma}^2)|^{-\frac{1}{2}} \exp(-S^C/2)} \frac{(2\pi c_{\alpha j})^{-\frac{N(\delta_{jk}^P)}{2}} \exp\{-\frac{1}{2c_{\alpha j}}(\boldsymbol{\alpha}_{jk}^P)^\top \boldsymbol{\alpha}_{jk}^P\}}{(2\pi c_{\alpha j})^{-\frac{N(\delta_{jk}^C)}{2}} \exp\{-\frac{1}{2c_{\alpha j}}(\boldsymbol{\alpha}_{jk}^C)^\top \boldsymbol{\alpha}_{jk}^C\}} \frac{N(\boldsymbol{\alpha}_{jk}^C; \hat{\boldsymbol{\alpha}}_{\delta_{jk}^C}, h_{jk}\boldsymbol{\Delta}_{\delta_{jk}^C})}{N(\boldsymbol{\alpha}_{jk}^P; \hat{\boldsymbol{\alpha}}_{\delta_{jk}^P}, h_{jk}\boldsymbol{\Delta}_{\delta_{jk}^P})}\right\},$$

where the determinants, for centred variables, are equal to one, otherwise, the ratio of the determinants may be computed as $\prod_{i=1}^n \{(\sigma_{ij}^2)^C/(\sigma_{ij}^2)^P\}^{1/2}$.

3. The full conditional of $\sigma_j^2, j = 1, \ldots, p$, is given by

$$f(\sigma_j^2|\ldots) \propto |\boldsymbol{\Sigma}(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2)|^{-\frac{1}{2}} \exp(-S/2)\xi(\sigma_j^2),$$

where $\xi(\sigma_j^2)$ denotes either the IG or half-normal prior. We follow a random walk algorithm obtaining proposed values $(\sigma_j^2)^{(P)} \sim N((\sigma_j^2)^{(C)}, f_{3j}^2)$, where $f_{3j}^2$ is a tuning parameter that we choose adaptively (Roberts and Rosenthal, 2009) in order to achieve an acceptance probability of $20\% - 25\%$ (Roberts and Rosenthal, 2001). Proposed values are accepted with probability $f((\sigma_j^2)^{(P)}|\ldots)/f((\sigma_j^2)^{(C)}|\ldots)$, which reduces to

$$\{(\sigma_j^2)^C/(\sigma_j^2)^P\}^{n/2} \exp\{(S^C - S^P)/2\}\xi((\sigma_j^2)^{(P)})/\xi((\sigma_j^2)^{(C)}).$$

4. Parameter $c_\beta$ is updated from the marginal (17) and the IG$(a_\beta, b_\beta)$ prior

$$f(c_\beta|\ldots) \propto (c_\beta + 1)^{-\frac{N(\gamma)+p}{2}} \exp(-S/2)(c_\beta)^{-a_\beta - 1} \exp(-b_\beta/c_\beta).$$

To sample from the above, we utilize a normal approximation. Let $\ell(c_\beta) = \log\{f(c_\beta|\ldots)\}$. We utilize a normal proposal density $N(\hat{c}_\beta, -g^2/\ell''(\hat{c}_\beta))$ where $\hat{c}_\beta$ is the mode of $\ell(c_\beta)$, found using a Newton-Raphson algorithm, $\ell''(\hat{c}_\beta)$ is the second derivative of $\ell(c_\beta)$ evaluated at the mode, and $g^2$ is a tuning variance parameter that we choose adaptively (Roberts and Rosenthal, 2009) to achieve an acceptance probability of $20\% - 25\%$ (Roberts and Rosenthal, 2001). With superscripts $P$ and $C$ denoting proposed and currents values, the acceptance probability is the minimum between one and

$$\frac{f(c_\beta^P|\ldots)}{f(c_\beta^C|\ldots)} \frac{N(c_\beta^C; \hat{c}_\beta, -g^2/\ell''(\hat{c}_\beta))}{N(c_\beta^P; \hat{c}_\beta, -g^2/\ell''(\hat{c}_\beta))}.$$

5. Concerning parameter $c_{\alpha j}, j = 1, \ldots, p$, the full conditional corresponding to the IG$(a_{\alpha j}, b_{\alpha j})$ prior is another inverse Gamma density IG$(a_{\alpha j} + N(\delta_j)/2, b_{\alpha j} + \boldsymbol{\alpha}_{\delta_j j}^\top \boldsymbol{\alpha}_{\delta_j j}/2)$.

The full conditional corresponding to the half-normal prior $\sqrt{c_{\alpha j}} \sim N(0, \phi^2_{c_{\alpha j}})I[\sqrt{c_{\alpha j}} > 0]$ is

$$f(c_{\alpha j}|\dots) \propto c_{\alpha j}^{-N(\delta_j)/2} \exp(-\boldsymbol{\alpha}_{\delta_j j}^\top \boldsymbol{\alpha}_{\delta_j j}/2c_{\alpha j}) \exp(-c_{\alpha j}/2\phi^2_{c_{\alpha j}})I[\sqrt{c_{\alpha j}} > 0].$$

We obtain proposed values $c_{\alpha j}^{(P)} \sim N(c_{\alpha j}^{(C)}, f_{2j}^2)$, where $c_{\alpha j}^{(C)}$ denotes the current value. Proposed values are accepted with probability $f(c_{\alpha j}^{(P)}|\dots)/f(c_{\alpha j}^{(C)}|\dots)$, where $f_{2j}^2$ is a tuning parameter. We select its value adaptively (Roberts and Rosenthal, 2009) so as to achieve an acceptance probability of $20\% - 25\%$ (Roberts and Rosenthal, 2001).

6. Using likelihood (16) and prior (15), we find the posterior $\boldsymbol{\beta}^*$ to be

$$\boldsymbol{\beta}^*_\gamma|\dots \sim N(\frac{c_\beta}{1+c_\beta}(\tilde{\boldsymbol{X}}_\gamma^\top \tilde{\boldsymbol{X}}_\gamma)^{-1}\tilde{\boldsymbol{X}}_\gamma^\top \tilde{\boldsymbol{Y}}, \frac{c_\beta}{1+c_\beta}(\tilde{\boldsymbol{X}}_\gamma^\top \tilde{\boldsymbol{X}}_\gamma)^{-1}).$$

7. Update $\boldsymbol{R}$ as described in the main body of the paper.

8. To sample from the full conditional of $\boldsymbol{\theta}$, write $f(\boldsymbol{r}|\boldsymbol{\theta}, \tau^2) = \nu(\boldsymbol{\theta}, \tau^2)N(g(\boldsymbol{r}); \boldsymbol{\theta}, \tau^2\boldsymbol{I})$ for the likelihood in (18). Further, the prior for $\boldsymbol{\theta}$ is given in (19), $\boldsymbol{\theta} \sim N(\mu_R \boldsymbol{1}, \sigma_R^2 \boldsymbol{I})$. Hence, it is easy to show that the posterior is

$$f(\boldsymbol{\theta}|\dots) = \nu(\boldsymbol{\theta}, \tau^2)N\left(\boldsymbol{\theta}; \boldsymbol{A}(\tau^{-2}g(\boldsymbol{r}) + \sigma_R^{-2}\mu_R\boldsymbol{1}), \boldsymbol{A} \equiv (\tau^{-2} + \sigma_R^{-2})^{-1}\boldsymbol{I}\right). \quad (24)$$

At iteration $u+1$ we sample $\boldsymbol{\theta}^{(u+1)}$ utilizing as proposal the normal distribution that appears on the right hand side of (24). The proposed $\boldsymbol{\theta}^{(u+1)}$ is accepted with probability

$$\min\left\{1, \frac{\nu(\boldsymbol{\theta}^{(u+1)}, \tau^2)}{\nu(\boldsymbol{\theta}^{(u)}, \tau^2)}\right\},$$

which, for a small value of $\tau^2$ can reasonably be assumed to be unity (Liechty et al., 2004; Yu et al., 2014; Liechty et al., 2009).

9. Update $\mu_R$ from $\mu_R \sim N((d/\sigma_R^2 + 1/\varphi_r^2)^{-1}(d/\sigma_R^2)\bar{\theta}, (d/\sigma_R^2 + 1/\varphi_r^2)^{-1})$, where $\bar{\theta}$ is the mean of the elements of vector $\boldsymbol{\theta}$.

10. We update $\sigma_R^2$ utilizing the following full conditional

$$f(\sigma_R^2|\dots) \propto (\sigma_R^2)^{-\frac{d}{2}} \exp\{-\sum_{i=1}^d (\theta_i - \mu_R)^2/(2\sigma_R^2)\} \exp\{-\sigma_R^2/(2\phi_R^2)\}I[\sigma_R > 0].$$

Proposed values are obtained from $(\sigma_R^2)^{(p)} \sim N((\sigma_R^2)^{(c)}, f_1^2)$ where $(\sigma_R^2)^{(c)}$ denotes the current value. Proposed values are accepted with probability $f((\sigma_R^2)^{(p)}|\dots)/f((\sigma_R^2)^{(c)}|\dots)$. We treat $f_1^2$ as a tuning parameter and we select its value adaptively (Roberts and Rosenthal, 2009) in order to achieve an acceptance probability of $20\% - 25\%$ (Roberts and Rosenthal, 2001).

## 2.2 MCMC algorithm for the grouped correlations model

With the introduction of the shadow prior, model (14) becomes the same as in (18). The difference is in the distribution of $\theta_{kl}$, which are now independently distributed with conditional distribution $\theta_{kl}|\lambda_{kl} = h \sim N(\mu_{R,h}, \sigma_R^2)$. Here we point out the additional MCMC steps needed for the 'grouped correlations' models:

1. Let $\boldsymbol{\theta}_h$ denote the vector of $\theta_{kl}$ that have been assigned to cluster $h$, $h = 1, \dots, H$. The posterior of $\boldsymbol{\theta}_h$ is

$$f(\boldsymbol{\theta}_h | \dots) \propto N\left(\boldsymbol{\theta}_h; \boldsymbol{A}(\tau^{-2} g(\boldsymbol{r}) + \sigma_R^{-2} \mu_{R,h} \mathbf{1}), \boldsymbol{A} \equiv (\tau^{-2} + \sigma_R^{-2})^{-1} \boldsymbol{I}\right).$$

2. Update $\mu_{R,h}$ from $\mu_{R,h} \sim N((d_h/\sigma_R^2 + 1/\varphi_r^2)^{-1}(d_h/\sigma_R^2)\bar{\theta}_h, (d_h/\sigma_R^2 + 1/\varphi_r^2)^{-1})$, where $d_h$ is the number of $\theta_{kl}$ assigned to the $h$th cluster and $\bar{\theta}_h$ is their mean.

3. Update $v_h \sim \text{Beta}(d_h + 1, d - \sum_{l=1}^{h} d_h + \alpha^*)$, $h = 1, \dots, H-1$, where $d_h$ is the number of correlations allocated in the $h$th cluster. Given $v_h$, update the stick-breaking weights $w_h, h = 1, \dots, H$.

4. Posterior cluster assignment probabilities are computed using

$$P(\lambda_{kl} = h | \dots) \propto w_h N(\theta_{kl}; \mu_{R,h}, \sigma_R^2).$$

5. We update concentration parameter $\alpha^*$ using the method described by Escobar and West (1995). With the $\alpha^* \sim \text{Gamma}(a_{\alpha*}, b_{\alpha*})$ prior, the posterior can be expressed as a mixture of two Gamma distributions:

$$\alpha^* | \eta, k \sim \pi_\eta \text{Gamma}(a_{\alpha*} + k, b_{\alpha*} - \log(\eta)) + (1 - \pi_\eta)\text{Gamma}(a_{\alpha*} + k - 1, b_{\alpha*} - \log(\eta)), \quad (25)$$

   where $k$ is the number of non-empty clusters, $\pi_\eta = (a_{\alpha*} + k - 1)/\{a_{\alpha*} + k - 1 + n(b_{\alpha*} - \log(\eta))\}$ and

$$\eta | \alpha^*, k \sim \text{Beta}(\alpha^* + 1, d). \quad (26)$$

   Hence the algorithm proceeds as follows: with $\alpha^*$ and $k$ fixed at their current values, we sample $\eta$ from (26). Then, based on the same $k$ and the newly sampled value of $\eta$, we sample a new $\alpha$ value from (25).

## 2.3 MCMC algorithm for the grouped variables model

Here we point out the only difference between the MCMC algorithms for grouped correlations and grouped variables models:

1. Let $w_h$ be the prior probability that a variable is assigned to cluster $h$. Then cluster assignment probabilities are computed as follows

$$P(\lambda_k = h | \dots) \propto w_h \prod_{l \neq k} N(\theta_{kl}; \mu_{R,h,\lambda_l}, \sigma_R^2).$$

# References


Chan, D., Kohn, R., Nott, D., and Kirby, C. (2006). Locally adaptive semiparametric estimation of the mean and variance functions in regression models. *Journal of Computational and Graphical Statistics*, 15(4):915–936.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68.

Liechty, J. C., Liechty, M. W., and Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, 91(1):1–14.

Liechty, M. W., Liechty, J. C., and Müller, P. (2009). The shadow prior. *Journal of Computational and Graphical Statistics*, 18(2):368–383.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Yu, Philip, L. H., Li, W. K., and Ng, F. C. (2014). Formulating hypothetical scenarios in correlation stress testing via a Bayesian framework. *The North American Journal of Economics and Finance*, 27:17–33.