Bayesian Structured Additive Distributional Regression for Multivariate Responses Supplement

Nadja Klein, Thomas Kneib
Chair of Statistics
Georg-August-University Göttingen

Stephan Klasen
Chair of Development Economics
Georg-August-University Göttingen

Stefan Lang
Department of Statistics
University of Innsbruck

B Score Vectors and Working Weights

In this section, we provide the score vectors v_i and the working weights w_i for the bivariate normal and Dirichlet distribution of the main paper which are needed for the MCMC algorithm of Section 2.3. Note that score vectors and weights for the bivariate t distribution and bivariate probit model are obtained in a similar way as for the bivariate normal distribution such that we omit detailed calculations here.

We usually compute first and second derivatives of all logarithmic densities $p(y_i|\vartheta_{i1},\dots,\vartheta_{iK})$ with respect to the predictors $\eta_i^{\vartheta_1},\dots,\eta_i^{\vartheta_K}$, show how to get the expectations of the negative second derivatives, and that the resulting working weights are positive. In the following we simply write $l \equiv l(y_i|\eta_i^{\vartheta_1},\dots,\eta_i^{\vartheta_K})$ for the log-likelihood $\log(p_i)$ of y_i as a function of the predictors $\eta_i^{\vartheta_1},\dots,\eta_i^{\vartheta_K}$ and drop the index i.

For $x > 0$, $\Gamma(x)$ denotes the gamma function, $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log(\Gamma(x))$ the digamma function and $\psi_1(x) = \frac{\mathrm{d}^2}{(\mathrm{d}x)^2}\log(\Gamma(x))$ the trigamma function.

B.1 Bivariate Normal Distribution

Parameters are $\mu_1 = \eta^{\mu_1}$, $\mu_2 = \eta^{\mu_2}$, $\sigma_1 = \exp(\eta^{\sigma_1})$, $\sigma_2 = \exp(\eta^{\sigma_2})$ and $\rho = \frac{\eta^{\rho}}{\sqrt{1+\eta^{2\rho}}}$

$$
\begin{aligned}
l = & -\log(2\pi) - \log(\sigma_1) - \log(\sigma_2) - \frac{1}{2}\log(1-\rho^2) \\
& - \frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\rho}{\partial\eta^\rho} &= \frac{1}{\sqrt{1+(\eta^\rho)^2}} - \frac{\eta^\rho\frac{1}{2}\left(1+(\eta^\rho)^2\right)^{-\frac{1}{2}}2\eta^\rho}{1+(\eta^\rho)^2} \\
&= \frac{1+(\eta^\rho)^2-(\eta^\rho)^2}{\left(1+(\eta^\rho)^2\right)^{\frac{3}{2}}} = \frac{1}{\left(1+(\eta^\rho)^2\right)^{\frac{3}{2}}} \\
1-\rho^2 &= 1-\left(\frac{\eta^\rho}{\sqrt{1+(\eta^\rho)^2}}\right)^2 = \frac{1}{1+(\eta^\rho)^2} \\
\frac{\partial(1-\rho^2)}{\partial\eta^\rho} &= -\frac{2\eta^\rho}{\left(1+(\eta^\rho)^2\right)^2} = -\frac{2\rho}{\left(1+(\eta^\rho)^2\right)^{\frac{3}{2}}} \\
\frac{1}{1-\rho^2} &= 1+(\eta^\rho)^2 \\
\frac{\partial\left(\frac{1}{1-\rho^2}\right)}{\partial\eta^\rho} &= 2\eta^\rho \\
\frac{\rho}{1-\rho^2} &= \rho\left(1+(\eta^\rho)^2\right) = \frac{\eta^\rho}{\sqrt{1+(\eta^\rho)^2}}\left(1+(\eta^\rho)^2\right) \\
&= \eta^\rho\sqrt{1+(\eta^\rho)^2} \\
\frac{\partial\left(\frac{\rho}{1-\rho^2}\right)}{\partial\eta^\rho} &= \sqrt{1+(\eta^\rho)^2} + \eta^\rho\frac{1}{2}\left(1+(\eta^\rho)^2\right)^{-\frac{1}{2}}2\eta^\rho \\
&= \frac{1+(\eta^\rho)^2+(\eta^\rho)^2}{\sqrt{1+(\eta^\rho)^2}} \\
&= \frac{1+2(\eta^\rho)^2}{\sqrt{1+(\eta^\rho)^2}} \\
&= \frac{1+(\eta^\rho)^2}{\sqrt{1+(\eta^\rho)^2}} + \rho\eta^\rho \\
&= \left(1+(\eta^\rho)^2\right)^{\frac{1}{2}} + \rho\eta^\rho
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}(y_1 y_2) = \mathrm{Cov}(y_1, y_2) + \mathbb{E}(y_1) + \mathbb{E}(y_2) \\
&\mathbb{E}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) = \mathrm{Cov}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right), \left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) \\
&= \frac{1}{\sigma_1\sigma_2}\mathrm{Cov}(y_1, y_2) = \frac{1}{\sigma_1\sigma_2}\rho\sigma_1\sigma_2 = \rho \\
&\mathbb{E}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right) = 1
\end{aligned}
$$

Score Vectors

$$
\begin{aligned}
\frac{\partial l}{\partial \eta^{\mu_1}} &= -\frac{1}{2(1-\rho^2)} 2\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{-1}{\sigma_1}\right) + 2\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{-1}{\sigma_1}\right)\left(\frac{1}{2(1-\rho^2)}\right) \\
&= \frac{1}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1^2}\right) - \frac{\rho}{\sigma_1}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{1}{1-\rho^2}\right) \\
&= \frac{1}{(1-\rho^2)\sigma_1}\left(\frac{y_1-\mu_1}{\sigma_1} - \rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) \\
\frac{\partial l}{d\eta^{\mu_2}} &= \frac{1}{(1-\rho^2)\sigma_2}\left(\frac{y_2-\mu_2}{\sigma_2} - \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\right) \\
\frac{\partial l}{\partial \sigma_1} &= -1 - \frac{1}{2(1-\rho^2)}(y_1-\mu_1)^2\left(\frac{-2}{\sigma_1^3}\right)\sigma_1 + \frac{2\rho}{2(1-\rho^2)}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1^2}\right)(-1)\sigma_1 \\
&= -1 + \frac{1}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - \frac{\rho}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) \\
&= -1 + \frac{1}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_1-\mu_1}{\sigma_1} - \rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) \\
\frac{\partial l}{\partial \sigma_2} &= -1 + \frac{1}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - \frac{\rho}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right) \\
&= -1 + \frac{1}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_2-\mu_2}{\sigma_2} - \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\right) \\
\frac{\partial l}{\partial \eta^{\rho}} &= \frac{1}{1-\rho^2}\left(\frac{\rho}{(1+(\eta^\rho)^2)^{\frac{3}{2}}}\right) - \eta^\rho\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right] \\
&\quad + \frac{1+2\left(\eta^\rho\right)^2}{\sqrt{1+\left(\eta^\rho\right)^2}}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)
\end{aligned}
$$

Second Derivatives

$$
\begin{aligned}
\frac{\partial^2 l}{(\partial\eta^{\mu_1})^2} &= \frac{1}{(1-\rho^2)\sigma_1}\left(\frac{-1}{\sigma_1}\right) = -\frac{1}{(1-\rho^2)\sigma_1^2}\\
\frac{\partial^2 l}{(\partial\eta^{\mu_2})^2} &= \frac{1}{(1-\rho^2)\sigma_2}\left(\frac{-1}{\sigma_2}\right) = -\frac{1}{(1-\rho^2)\sigma_2^2}\\
\frac{\partial^2 l}{(\partial\eta^{\sigma_1})^2} &= \frac{1}{1-\rho^2}\left(\frac{(y_1-\mu_1)^2}{\sigma_1^3}\right)(-2)\sigma_1 - \frac{\rho}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{(-1)(y_1-\mu_1)}{\sigma_1^2}\right)\sigma_1\\
&= \frac{-2}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \frac{\rho}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\\
\frac{\partial^2 l}{(\partial\eta^{\sigma_2})^2} &= \frac{1}{1-\rho^2}\left(\frac{(y_2-\mu_2)^2}{\sigma_2^3}\right)(-2)\sigma_2 - \frac{\rho}{1-\rho^2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{(-1)(y_2-\mu_2)}{\sigma_2^2}\right)\sigma_2\\
&= \frac{-2}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 + \frac{\rho}{1-\rho^2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\\
\frac{\partial^2 l}{(\partial\eta^{\rho})^2} &= \frac{1}{1+(\eta^\rho)^2} - \frac{\eta^\rho 2\eta^\rho}{\left(1+(\eta^\rho)^2\right)^2} - \left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\\
&\quad + \left(\frac{1}{2}\frac{2\eta^\rho}{\sqrt{1+(\eta^\rho)^2}} + \frac{\eta^\rho}{\left(1+(\eta^\rho)^2\right)^{\frac{3}{2}}} + \rho\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)\\
&= \frac{1-(\eta^\rho)^2}{\left(1+(\eta^\rho)^2\right)^2} - \left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\\
&\quad + \left(2\rho + \rho(1-\rho^2)\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)
\end{aligned}
$$

Working Weights

$$\mathbb{E}\left(-\frac{\partial^2 l}{(\partial\eta^{\mu_1})^2}\right) = \frac{1}{(1-\rho^2)\sigma_1^2} \tag{B.1}$$

$$\mathbb{E}\left(-\frac{\partial^2 l}{(\partial\eta^{\mu_2})^2}\right) = \frac{1}{(1-\rho^2)\sigma_2^2} \tag{B.2}$$

$$\mathbb{E}\left(-\frac{\partial^2 l}{(\partial\eta^{\sigma_1})^2}\right) = \frac{2}{1-\rho^2} - \frac{\rho^2}{1-\rho^2} = 1 + \frac{1}{1-\rho^2} \tag{B.3}$$

$$\mathbb{E}\left(-\frac{\partial^2 l}{(\partial\eta^{\sigma_2})^2}\right) = \frac{2}{1-\rho^2} - \frac{\rho^2}{1-\rho^2} = 1 + \frac{1}{1-\rho^2} \tag{B.4}$$

$$
\begin{aligned}
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial\eta^{\rho})^2}\right) &= -(1-\rho^2)^2 + \frac{\rho^2}{1+(\eta^\rho)^2} + 2 - \rho(3\rho-\rho^3)\\
&= -(1-\rho^2)^2 + \rho^2(1-\rho^2) + 2 - \rho^2(3-\rho^2)\\
&= 2 + (1-\rho^2)(\rho^2 - (1-\rho^2)) - \rho^2(3-\rho^2)\\
&= 2 - (1-\rho^2) - \rho^2(3-\rho^2) + 2\rho^2(1-\rho^2)\\
&= 2 - 1 + \rho^2 - 3\rho^2 + \rho^4 + 2\rho^2 - 2\rho^4 = 1 - \rho^4 \qquad \text{(B.5)}
\end{aligned}
$$

Lemma B.1. *The working weight matrices for the bivariate normal distribution with diagonal elements given in* (B.1) *to* (B.5), *are positive definite.*

Proof. The claim follows directly since we have $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho^2 \in (0, 1)$. $\square$

B.2 Bivariate t Distribution

Parameters are $\mu_1 = \eta^{\mu_1}$, $\mu_2 = \eta^{\mu_2}$, $\sigma_1 = \exp(\eta^{\sigma_1})$, $\sigma_2 = \exp(\eta^{\sigma_2})$, $\rho = \frac{\eta^\rho}{\sqrt{1+(\eta^\rho)^2}}$, $n = \exp(\eta^n)$ with $n_{df} > 2$.

$$
\begin{aligned}
l =& \log\left(\Gamma\left(\frac{n_{df}+2}{2}\right)\right) - \log\left(\Gamma\left(\frac{n_{df}}{2}\right)\right) - \log(\pi) - \log(n_{df}) - \log(\sigma_1) - \log(\sigma_2) - \frac{1}{2}\log(1-\rho^2) \\
& - \frac{n_{df}+2}{2}\log\underbrace{\left[1 + \frac{1}{n_{df}(1-\rho^2)}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right)\right]}_{c_1}
\end{aligned}
$$

Score Vectors

$$
\begin{aligned}
\frac{\partial l}{\partial \eta^{\mu_1}} &= \frac{\frac{y_1-\mu_1}{\sigma_1} - \rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)}{c_1}\left(\frac{n_{df}+2}{\sigma_1}\right)\left(\frac{1}{n_{df}(1-\rho^2)}\right)\\
\frac{\partial l}{\partial \eta^{\mu_2}} &= \frac{\frac{y_2-\mu_2}{\sigma_2} - \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)}{c_1}\left(\frac{n_{df}+2}{\sigma_2}\right)\left(\frac{1}{n_{df}(1-\rho^2)}\right)\\
\frac{\partial l}{\partial \eta^{\sigma_1}} &= -1 - \frac{n_{df}+2}{2}\left(\frac{(-2)\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)}{c_1}\right)\left(\frac{1}{n_{df}(1-\rho^2)}\right)\\
&= -1 - \frac{n_{df}+2}{n_{df}(1-\rho^2)c_1}\left(\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right) - \left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right)\\
\frac{\partial l}{\partial \eta^{\sigma_2}} &= -1 - \frac{n_{df}+2}{2}\left(\frac{(-2)\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 + 2\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)}{c_1}\right)\left(\frac{1}{n_{df}(1-\rho^2)}\right)\\
&= -1 - \frac{n_{df}+2}{n_{df}(1-\rho^2)c_1}\left(\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) - \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right)\\
\frac{\partial l}{\partial \eta^{\rho}} &= \frac{\rho}{\sqrt{1+(\eta^\rho)^2}}\\
&\quad - \frac{n_{df}+2}{n_{df}c}\eta^\rho\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right)\\
&\quad - \frac{n_{df}+2}{n_{df}c}\left(\rho\eta^\rho + \sqrt{1+(\eta^\rho)^2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\\
\frac{\partial l}{\partial \eta^{n_{df}}} &= \frac{n_{df}}{2}\left(\psi\left(\frac{n_{df}+2}{2}\right) - \psi\left(\frac{n_{df}}{2}\right) - \log(c_1)\right) - 1 + \frac{n_{df}+2}{2}\frac{c_1-1}{c_1}
\end{aligned}
$$

Second Derivatives

$$\frac{\partial^2 l}{(\partial\eta^{\mu_1})^2} = \frac{-(n_{df}+2)}{\sigma_1^2 n_{df}(1-\rho^2)c_1} + \frac{2(n_{df}+2)}{\sigma_1^2}\left(\frac{\left(\frac{y_1-\mu_1}{\sigma_1} - \rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right)^2}{n_{df}^2(1-\rho^2)^2 c_1^2}\right)$$

$$\frac{\partial^2 l}{(\partial\eta^{\mu_2})^2} = \frac{-(n_{df}+2)}{\sigma_2^2 n_{df}(1-\rho^2)c_1} + \frac{2(n_{df}+2)}{\sigma_2^2}\left(\frac{\left(\frac{y_2-\mu_2}{\sigma_2} - \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\right)^2}{n_{df}^2(1-\rho^2)^2 c_1^2}\right)$$

$$\begin{aligned}
\frac{\partial^2 l}{(\partial\eta^{\sigma_1})^2} &= \underbrace{-\frac{n_{df}+2}{n_{df}(1-\rho^2)c_1}\left(-\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right) + 2\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right)}_{c_2} \\
&\quad + \frac{n_{df}+2}{n_{df}^2(1-\rho^2)^2 c_1^2}\left(\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right) - \left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right) \\
&\quad\ * \left(-2\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) \\
&= c_2 + \frac{n_{df}+2}{n_{df}^2(1-\rho^2)^2 c_1^2}\left(-4\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)^3\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right) \\
&\quad + \left(2\left(\frac{y_1-\mu_1}{\sigma_1}\right)^4 + 2\rho^2\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right) \\
&= c_2 + \frac{2(n_{df}+2)}{c_1^2}\left(\frac{1}{n_{df}(1-\rho^2)}\left(-\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right)\right)^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{(\partial\eta^{\sigma_2})^2} &= \underbrace{-\frac{n_{df}+2}{n_{df}(1-\rho^2)c_1}\left(-\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + 2\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right)}_{c_3} \\
&\quad + \frac{n_{df}+2}{n_{df}^2(1-\rho^2)^2 c_1^2}\left(\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) - \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right) \\
&\quad\ * \left(-2\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 + 2\rho\left(\frac{y_2-\mu_2}{\sigma_2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\right) \\
&= c_3 + \frac{2(n_{df}+2)}{c_1^2}\left(\frac{1}{n_{df}(1-\rho^2)}\left(-\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 + \rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right)\right)^2
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l}{(\partial \eta^{\rho})^2} &= \left(1-\rho^2\right)^2 - \rho^2(1-\rho^2) - \frac{n_{df}+2}{n_{df}c}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right) \\
&\quad - \frac{n_{df}+2}{n_{df}c}\left(\rho\left(2\eta^{\rho}+1+\frac{1}{1+(\eta^{\rho})^2}\right)\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) \\
&\quad - \left(\eta^{\rho}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right) - \left(\rho\eta^{\rho}+\sqrt{1+(\eta^{\rho})^2}\right)\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right)\right)^2 \\
&\quad * \frac{2(n_{df}+2)}{n_{df}^2 c^2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l}{(\partial \eta^{n_{df}})^2} &= \underbrace{\frac{n_{df}}{2}\left(\psi\left(\frac{n_{df}+2}{2}\right) - \psi\left(\frac{n_{df}}{2}\right) - \log(c_1)\right)}_{c_4} + \frac{n_{df}}{2}\frac{c_1-1}{c_1} \\
&\quad + \frac{n_{df}^2}{4}\left(\psi_1\left(\frac{n_{df}+2}{2}\right) - \psi_1\left(\frac{n_{df}}{2}\right)\right) + \frac{n_{df}}{2}\frac{c_1-1}{c_1} + \frac{n_{df}+2}{2}\underbrace{\left(-\frac{(c_1-1)}{c_1} + \left(\frac{c_1-1}{c_1}\right)^2\right)}_{\frac{1-c_1}{c_1^2}}
\end{aligned}
$$

Working Weights For parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ we use approximate working weights since in practice they work well, compare also the simulation in Section C. Concrete, we simply use the weights obtained for the bivariate normal distribution for third reasons. First, the weights are only used for the proposal densities and using the approximate weights yield high acceptance rates and good mixing behaviour. Second, for moderately large n_{df} the bivariate normal distribution is an approximation for the t distribution. And last, note that for computing the expectations of higher order product moments and the second derivatives, again approximations would be needed.

For computation of the working weight for n_{df} we note that

$$
\mathbb{E}\left(\frac{1}{c_1}\right) = \frac{n_{df}}{n_{df}+2}
$$

$$
\mathbb{E}\left(\frac{1}{c_1^2}\right) = \frac{n_{df}}{n_{df}+4}
$$

Furthermore with $\mathbb{E}\left(\frac{\partial l}{\partial \eta^{n_{df}}}\right) = 0$ it follows

$$
\mathbb{E}(c_4) = 0
$$

Working Weights

$$
\begin{aligned}
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\mu_1})^2}\right) &= \frac{1}{(1-\rho^2)\sigma_1^2} \\
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\mu_2})^2}\right) &= \frac{1}{(1-\rho^2)\sigma_2^2} \\
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\sigma_1})^2}\right) &= 1 + \frac{1}{1-\rho^2} \\
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\sigma_2})^2}\right) &= 1 + \frac{1}{1-\rho^2} \\
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\rho})^2}\right) &= 1 - \rho^4 \\
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{n_{df}})^2}\right) &= -\frac{n_{df}^2}{4}\left(\psi_1\left(\frac{n_{df}+2}{2}\right) - \psi_1\left(\frac{n_{df}}{2}\right)\right) - \frac{2n_{df}}{n_{df}+2} - \frac{n_{df}(n_{df}+2)}{2(n_{df}+4)} + \frac{n_{df}}{2}
\end{aligned} \tag{B.6}
$$

Lemma B.2. *The working weight matrix for* n_{df} *of the bivariate t distribution with diagonal elements given in* (B.6) *is positive definite.*

Proof.

$$
\begin{aligned}
\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{n_{df}})^2}\right) - \frac{n_{df}^2}{4}\left(\psi_1\left(\frac{n_{df}+2}{2}\right) - \psi_1\left(\frac{n_{df}}{2}\right)\right) - \frac{2n_{df}}{n_{df}+2} - \frac{n_{df}^2+2n_{df}}{2(n_{df}+4)} + \frac{n_{df}}{2} \\
= -\frac{n^2}{4}\left(\psi_1\left(\frac{n}{2}\right) - \frac{2}{n} - \psi_1\left(\frac{n}{2}\right)\right) - \frac{2n_{df}}{n_{df}+2} - \frac{n_{df}^2+2n_{df}}{2(n_{df}+4)} + \frac{n_{df}}{2} \\
= n_{df} - \frac{2n_{df}}{n_{df}+2} - \frac{n_{df}^2+2n_{df}}{2(n_{df}+4)} \\
= n_{df}\left(\frac{(n_{df}+2)^2-8}{2(n_{df}+2)(n_{df}+4)}\right) \\
> 0
\end{aligned}
$$

□

B.3 Dirichlet Distribution

For a response vector $\boldsymbol{y} = (y_1, \ldots, y_D)'$, parameters are $\alpha_d = \exp(\eta^{\alpha_d})$, $d = 1, \ldots, D$ with $\alpha_1, \ldots, \alpha_D > 0$ and $D \geq 2$.

$$
l = -\sum_{d=1}^{D} \log(\Gamma(\alpha_d)) + \log\left(\Gamma\left(\sum_{d=1}^{D}\alpha_d\right)\right) + \sum_{d=1}^{D}(\alpha_d - 1)\log(y_d).
$$

Score Vector

$$\frac{\partial l}{\partial \eta^{\alpha_d}} = -\alpha_d \psi(\alpha_d) + \alpha_d \psi\left(\sum_{k=1}^{K} \alpha_k\right) + \alpha_d \log(y_d).$$

Second Derivative

$$\frac{\partial^2 l}{(\partial \eta^{\alpha_d})^2} = -\alpha_d \psi(\alpha_i) + \alpha_d \psi\left(\sum_{k=1}^{K} \alpha_k\right) + \alpha_d \log(y_d) - \alpha_d^2 \psi_1(\alpha_d) + \alpha_d^2 \psi_1\left(\sum_{k=1}^{K} \alpha_k\right).$$

Working Weights

$$\begin{aligned}\mathbb{E}\left(-\frac{\partial^2 l}{(\partial \eta^{\alpha_d})^2}\right) &= -\alpha_d \mathbb{E}\left(-\psi(\alpha_d) + \psi\left(\sum_{k=1}^{K} a_k\right) + \log(y_d)\right) + a_d^2\left(-\psi_1\left(\sum_{k=1}^{K} a_k\right) + \psi_1(\alpha_d)\right) \\ &= \alpha_d^2\left(-\psi_1\left(\sum_{k=1}^{K} \alpha_k\right) + \psi_1(\alpha_d)\right) \end{aligned} \tag{B.7}$$

Lemma B.3. *The working weight matrices for the Dirichlet distribution with diagonal elements given in* (B.7)*, are positive definite.*

Proof. The claim follows directly since $\psi_1(x)$ is monotone decreasing for $x \in \mathbb{R}^+$. □

C Simulation Study for the Bivariate Normal and Bivariate t Distribution

In order to validate the performance of Bayesian structured additive distributional regression for the bivariate normal and the bivariate t distribution we consider the following model setup:

(1) The predictors are specified based on the centred versions of the functions

$$\begin{aligned} \tilde{f}^{\mu_1}(x_1) &= \sin(x_1), & \tilde{f}^{\mu_2}(x_2) &= \cos(x_2) \\ \tilde{f}^{\sigma_1}(x_1) &= \sqrt{x_1}\cos(x_1), & \tilde{f}^{\sigma_2}(x_2) &= 0.3 x_2 \cos(x_2) \\ \tilde{f}^{\rho}(x_1) &= \log(x_1), & \tilde{f}^{n_{df}}(x_2) &= \frac{1}{3} x_2^2 \end{aligned}$$

where $\tilde{f}^{n_{df}}$ is only used in the bivariate t model.

(2) The covariates x_1 and x_2 are uniform random numbers from the intervals $[1, 6]$ and $[-3, 3]$, respectively.

(3) For each of the two models, we simulated 250 data sets with sample sizes $n = 100, 250, 500$ and $1,000$.

(4) The smooth functions are approximated by cubic Bayesian P-splines with 20 inner knots on an equidistant grid within the range of the covariates and second order random walk prior. The priors of the hyperparameters are weakly informative inverse gamma priors with parameters $a = b = 0.001$, compare Section 2 for further details and references.

(5) We use 12,000 iterations with a burnin phase of 2,000 iterations and a thinning parameter of size 10.

Figure C1 shows logarithmic mean squared errors (MSE) of all effects in the bivariate normal and the bivariate t model. Figures C2 and C6 show coverage rates of pointwise 95% credible intervals for each simulated effect in the bivariate normal model and the bivariate t model, respectively. The posterior mean estimates of all effects compared to the true simulated functions for three specific replications (chosen corresponding to the 5%, 50% and 95% quantile of the MSE distribution of the corresponding function) as well as 95% posterior probabilities can be compared in Figures C3 to C5 for the bivariate normal model and in Figures C7 to C9 for the bivariate t model. The sample size is $n = 1,000$ in these graphs. All presented results are predictions resulting from a grid of step size 0.01 within the covariate ranges $[1, 6]$ and $[-3, 3]$, respectively. Results can briefly be summarised as follows:

(1) Coverage rates are generally close to the nominal level such that they can be considered as a helpful tool in applied analyses to assess uncertainties of effects. One exception are the rates of f^{σ_1} for $n = 100$ where the nominal level is not kept in both the normal and the t model. An explanation here might be that given the complexity of the model (5 or 6 splines with 22 knots) $n = 100$ is too small to estimate all unknown parameters. Apart from the effect f^{μ_1}, our approach tends to slightly overestimate uncertainty in the simulated example for sample sizes $n > 250$ yielding somewhat conservative credible interval.

(2) From the MSEs depicted in Figure C1 the presumption of slightly better results for the bivariate normal model can be confirmed. As expected, increasing the sample size reduces the MSEs.

(3) For $n = 1,000$ the approach yields point estimates which are quite close to the true functions for all distribution parameters, even for the replication with the 95% worst MSE. Qualitatively similar results are also obtained for the smaller sample sizes.

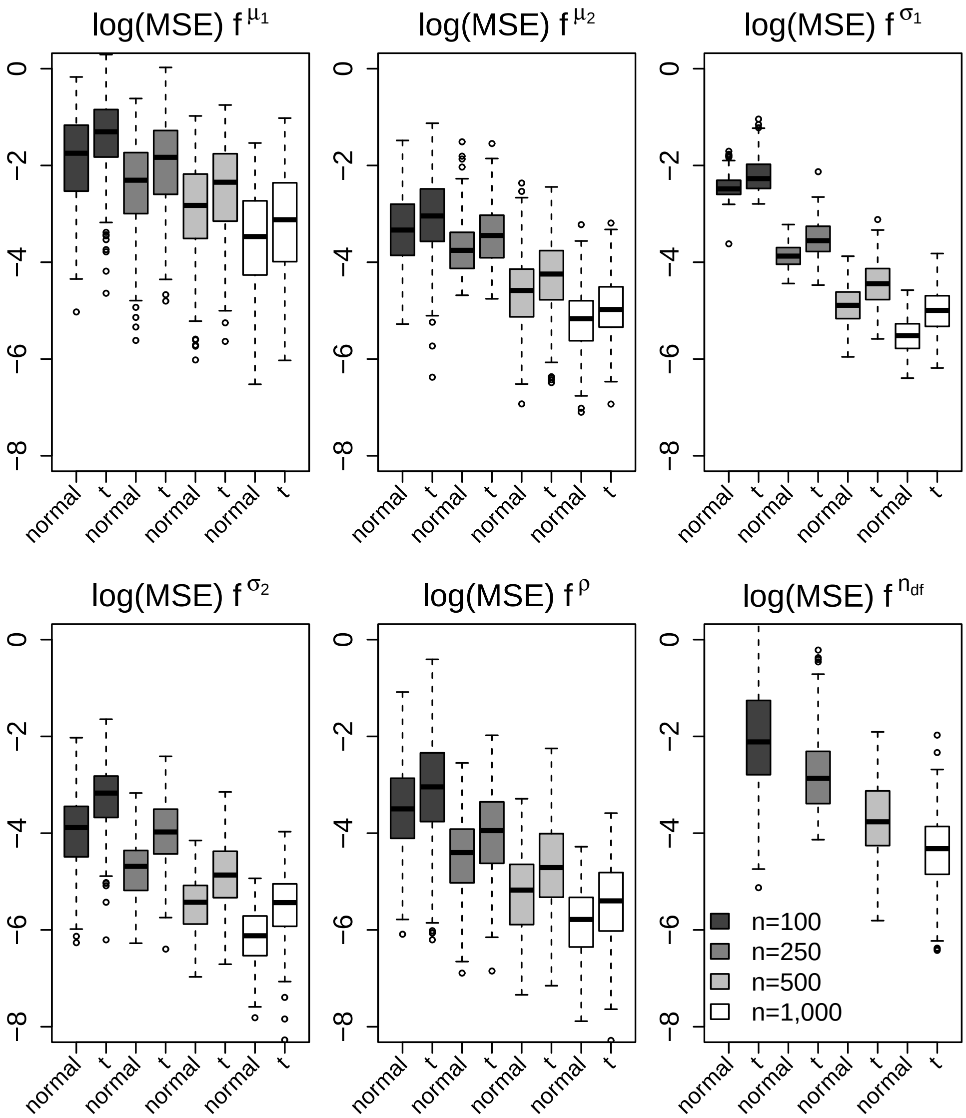

Figure C1: Simulation study. Boxplots of logarithmic mean squared errors of nonlinear effects in the bivariate normal distribution (left panels) and the bivariate t distribution (right panels).

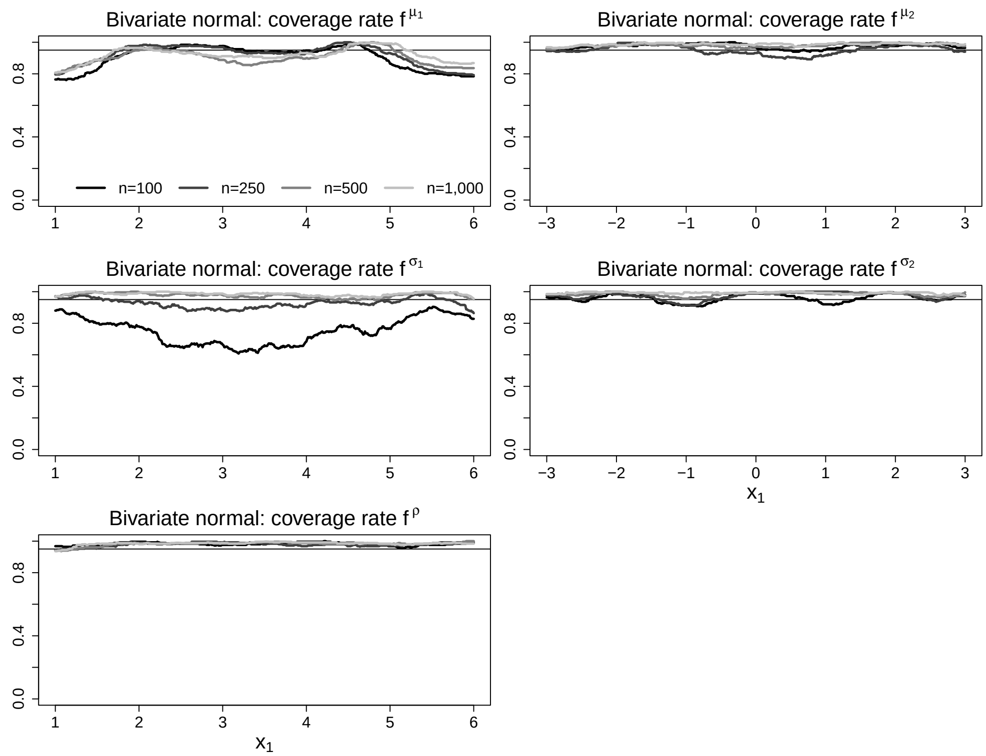

Figure C2: Simulation study, bivariate normal distribution. Coverage rates of pointwise 95% credible intervals (from black to light grey the sample size is increasing from $n = 100$ to $n = 1,000$). The horizontal line displays the 95% mark.

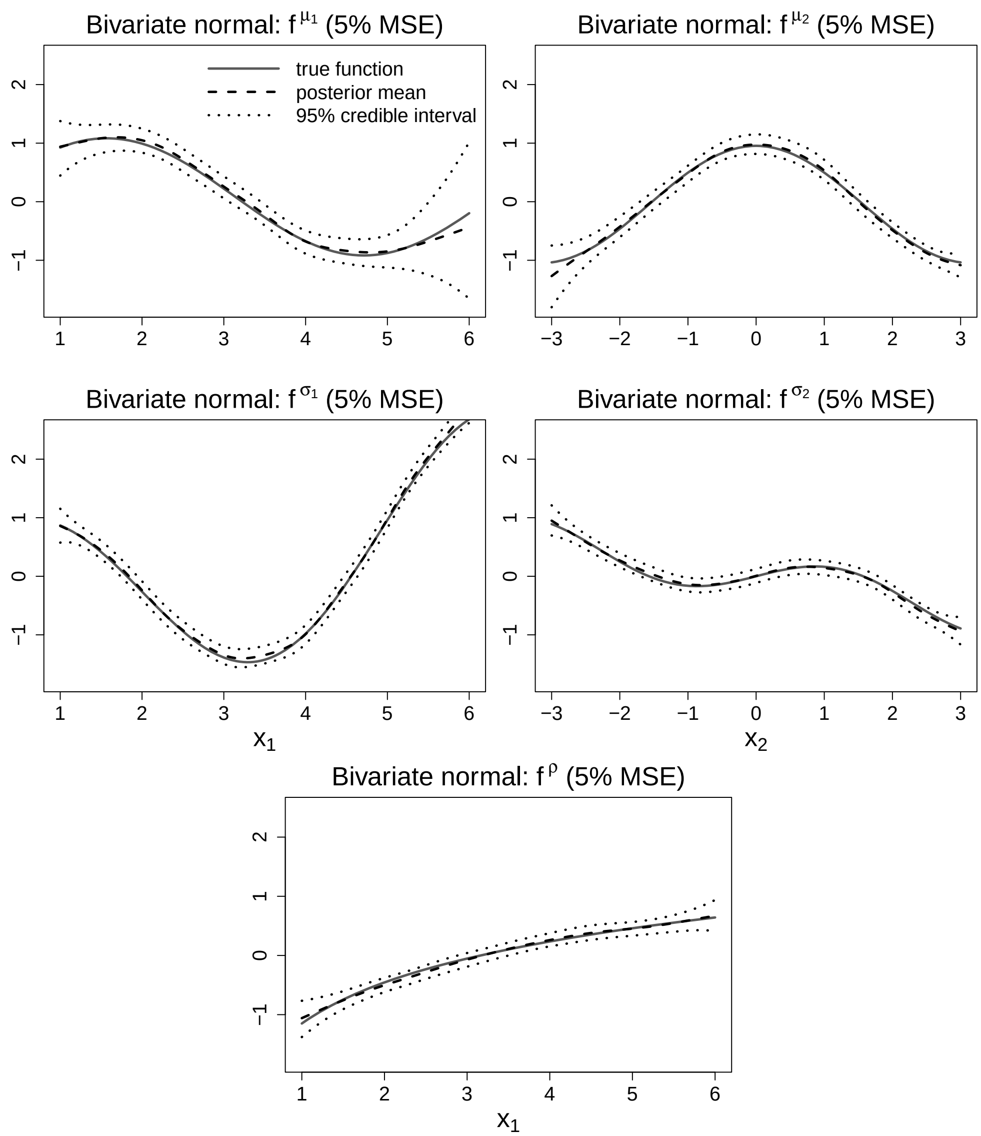

Figure C3: Simulation study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 5% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

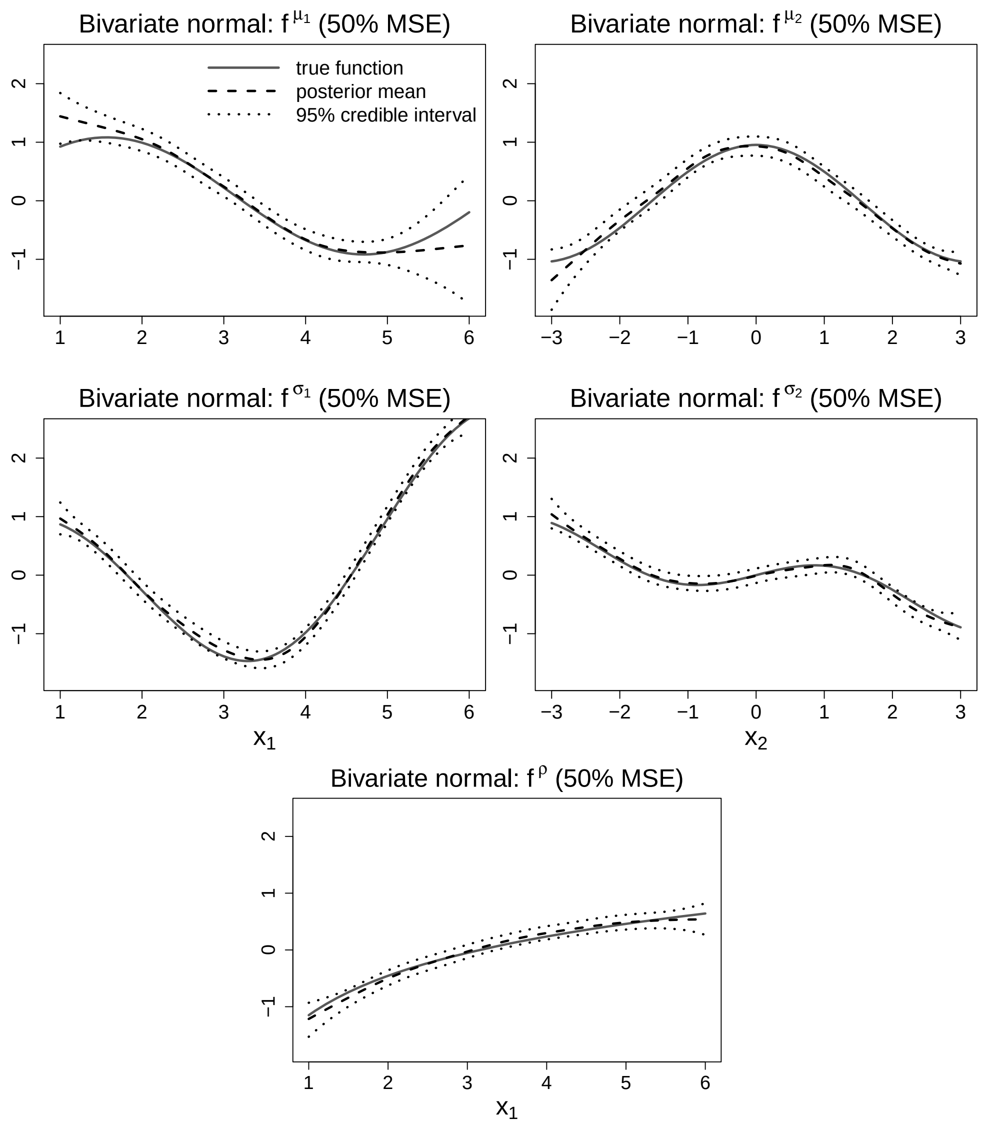

Figure C4: Simulation study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 50% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

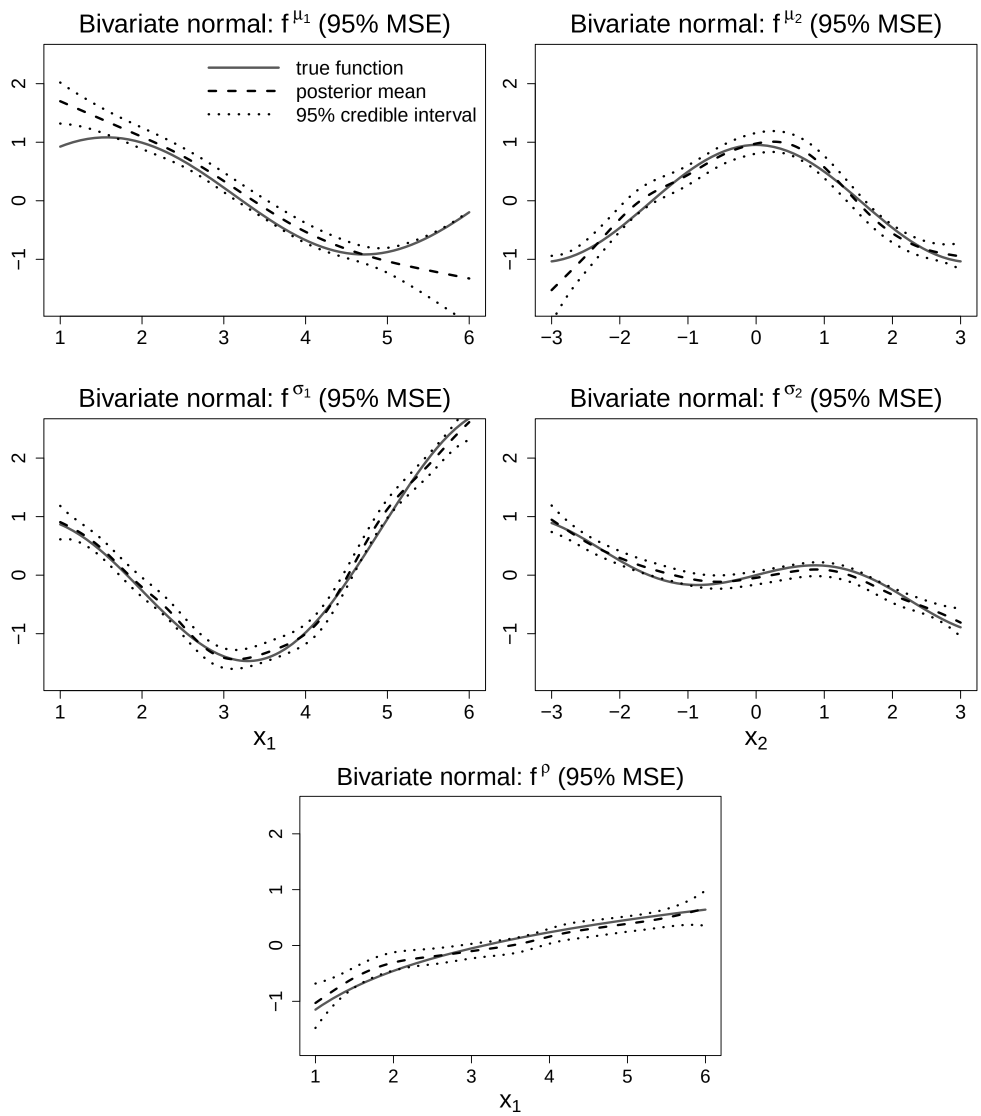

Figure C5: Simulation study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 95% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

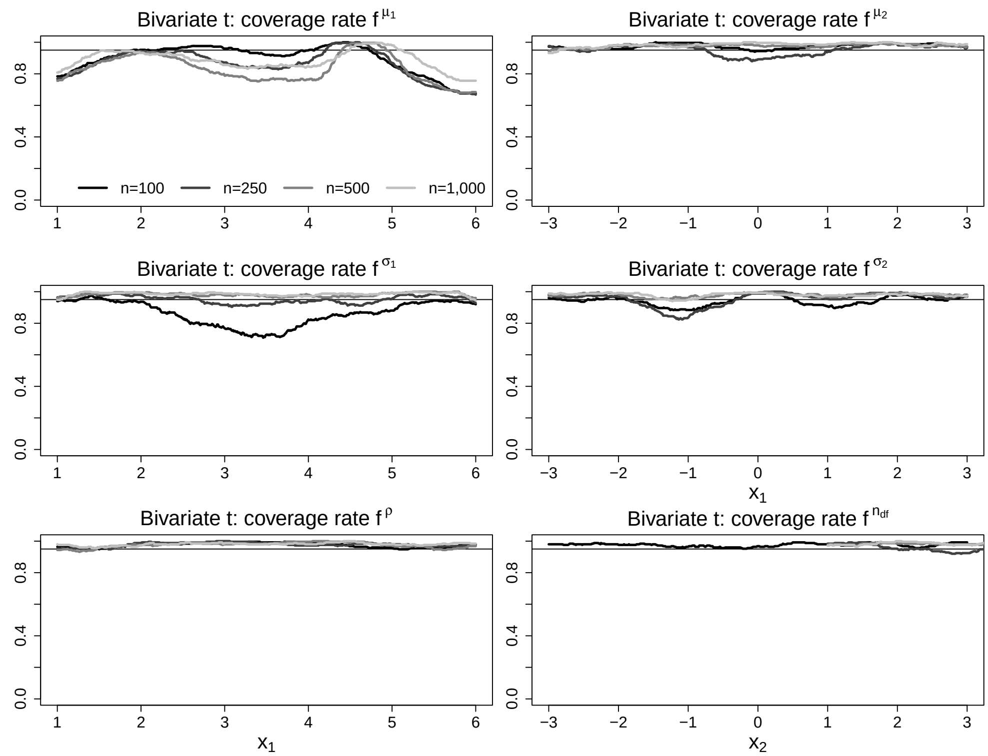

Figure C6: Simulation study, bivariate t distribution. Coverage rates of pointwise 95% credible intervals (from black to light grey the sample size is increasing from $n = 100$ to $n = 1,000$). The horizontal line displays the 95% mark.

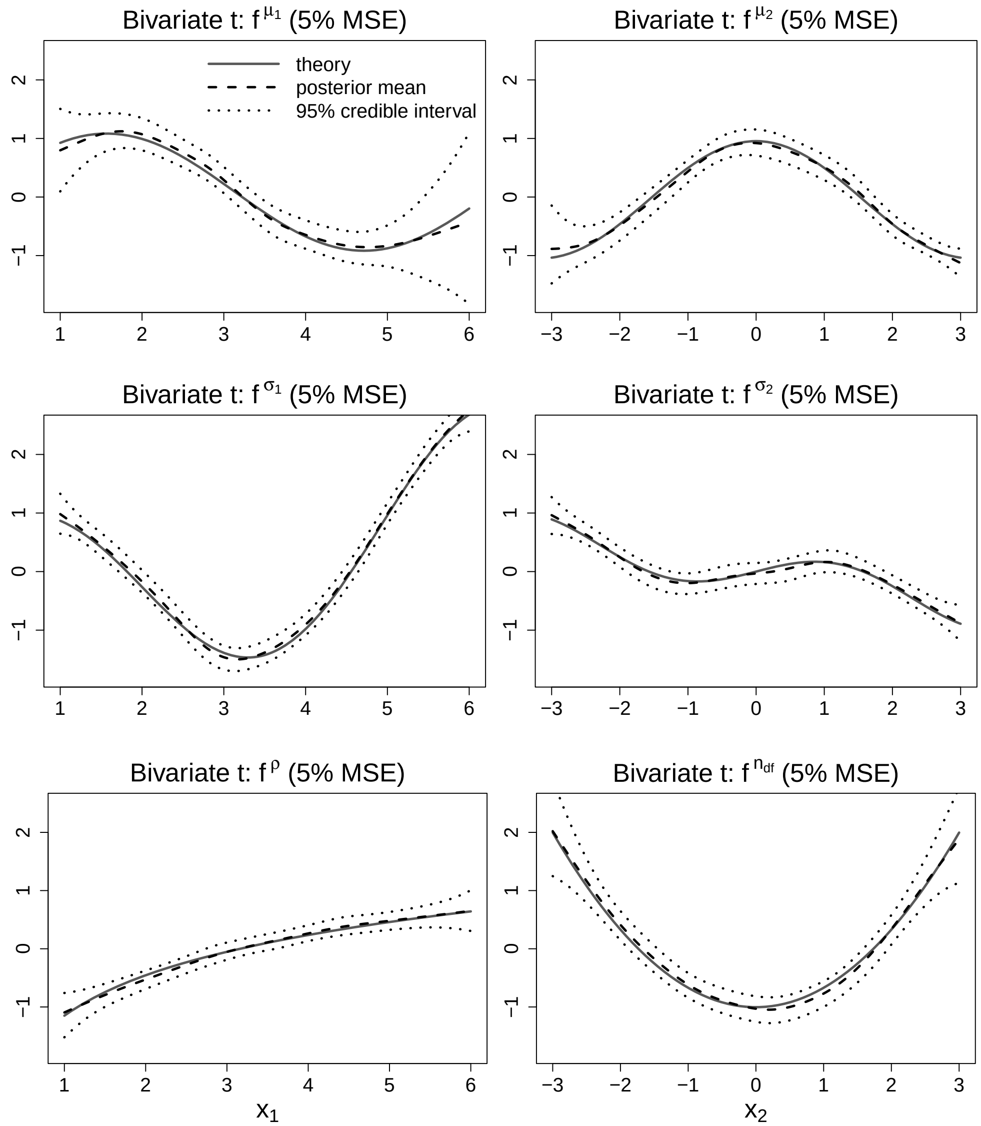

Figure C7: Simulation study, bivariate t distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 5% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

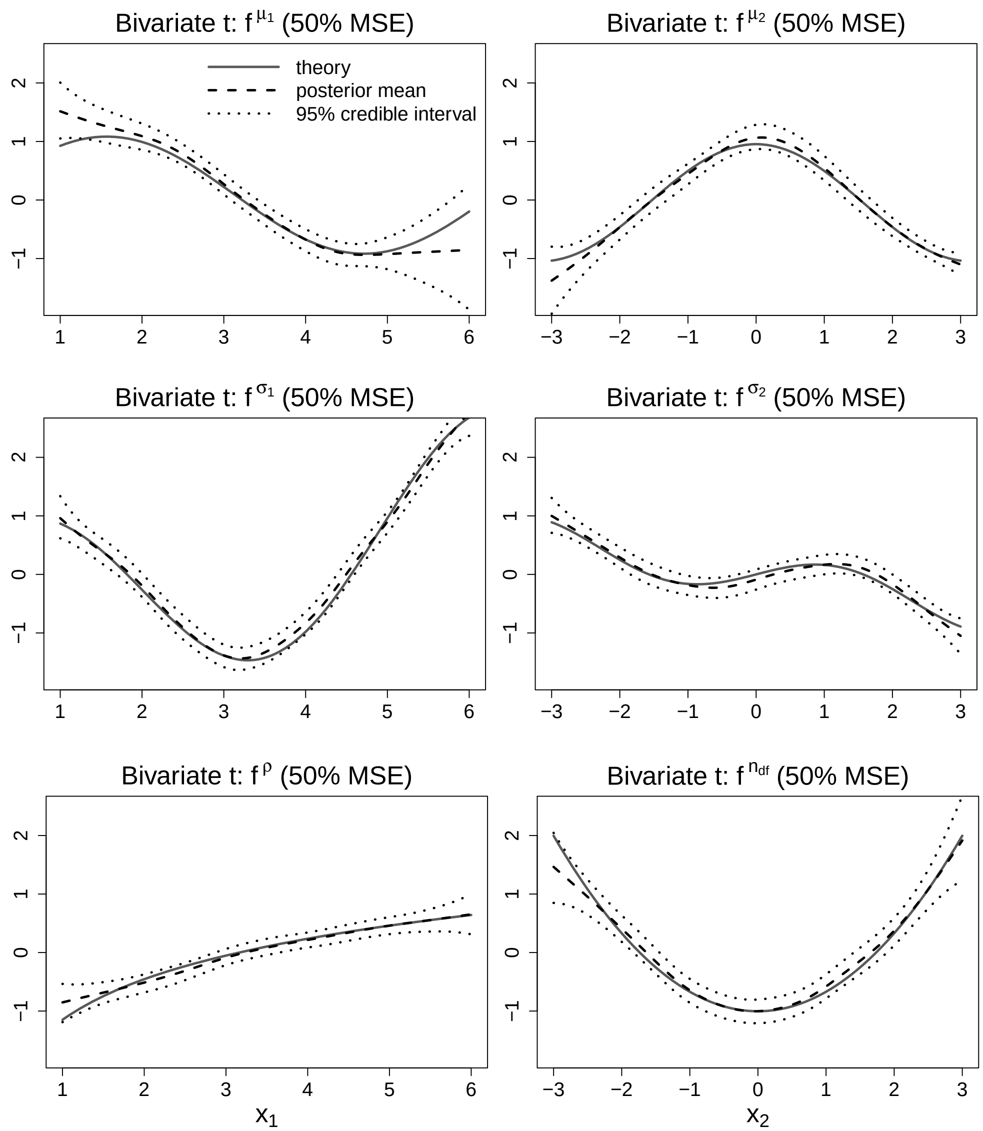

Figure C8: Simulation study, bivariate t distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 50% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

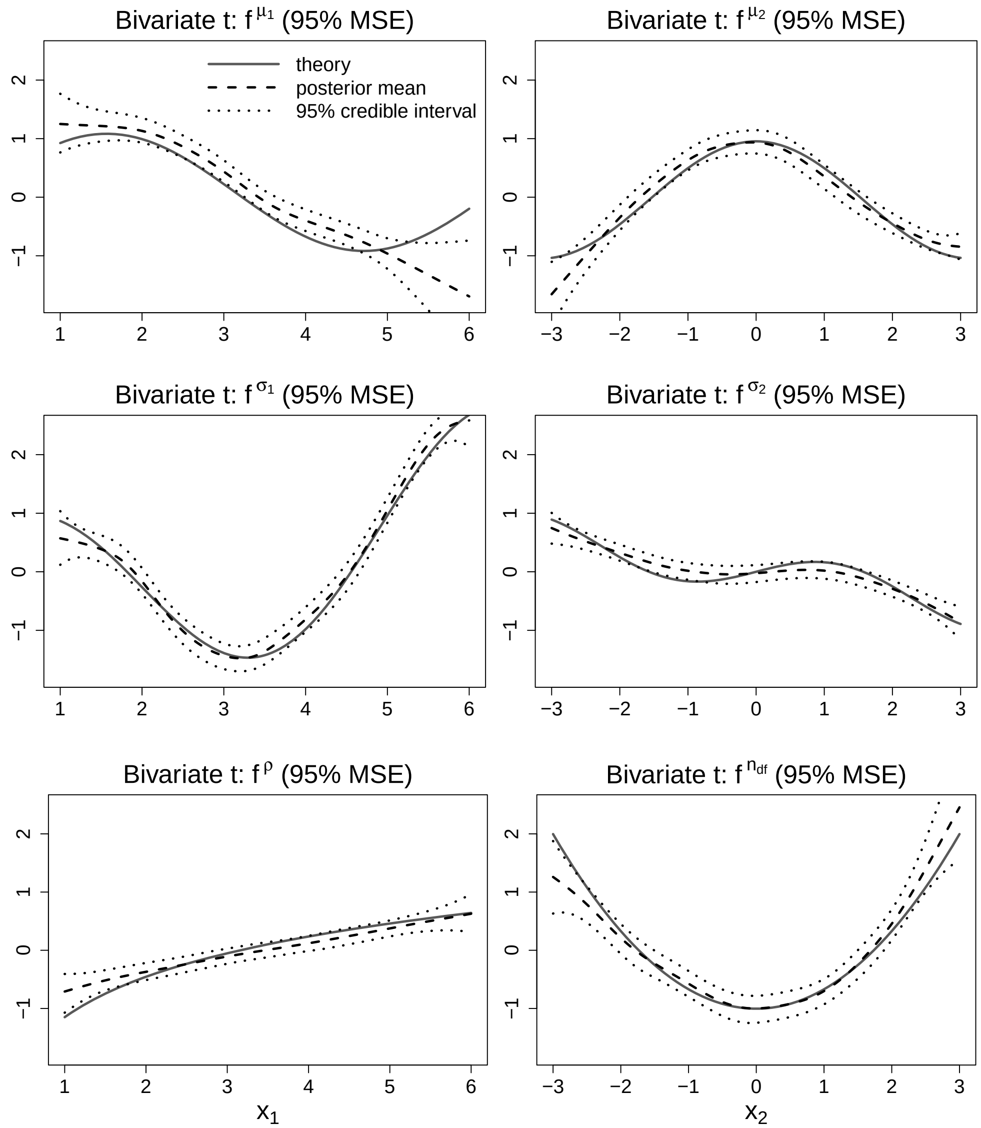

Figure C9: Simulation study, bivariate t distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines) for sample size $n = 1,000$. The solid lines are the true functions. The realisations shown here correspond to the 95% quantile of the MSE distribution of the corresponding function obtained from 250 simulation replications.

D Simulation Study for Models with Concurvity

In this section, we investigate the robustness of multivariate distributional regression to concurvity. Concurvity thereby refers to the situation that for at least one distribution parameter ϑ_k at least one of the functions $f_j^{\vartheta_k}$ from (1) is highly correlated with the rest of the predictor $\sum_{l \neq j} f_l^{\vartheta_k}$, i.e. if $f_j^{\vartheta_k} \approx \sum_{l \neq j} f_l^{\vartheta_k}$. The data generating process we consider to create concurvity is as follows:

(1) As response distribution we assume a bivariate normal distribution.

(2) Three types of models are defined in which four distribution parameters are constant and concurvity is either simulated in μ_1, σ_1 or ρ. The constant parameters are $\mu_2 = 0$, $\sigma_2 = 1$ and depending on which type of model is under consideration also $\mu_1 = 0$, $\sigma_1 = 1$ or $\rho = 0.5$. The models are called μ_1, σ_1 and ρ model in the discussion and the captions of Figures D10 to D18.

(3) We use three covariates x_1, x_2 and x_3 with $x_1, x_2 \sim \mathrm{U}[-3, 3]$ and

$$x_3 = \frac{c \cdot (-\sin(x_1) + \cos(x_2)) + (1 - c) \cdot \epsilon}{\sqrt{c^2 + (1 - c)^2}}$$

The value $c \in \{0, 0.25, 0.5, 0.75, 1\}$ controls the amount of concurvity: $c = 1$ for a perfectly deterministic relationship and $c = 0$ for independence. The standardisation with $\sqrt{c^2 + (1 - c)^2}$ guarantees comparability for different values of c. Then, we assume $\epsilon \sim \mathrm{N}(0, \tau_\epsilon^2)$ and τ_ϵ^2 is the empirical variance of $-\sin(x_1) + \cos(x_2)$.

(4) In addition to the three types of models, we investigate two scenarios:

- *Scenario 1:* Covariate x_3 is functionally related to the predictor with $\tilde{f}_3(x_3) = \frac{x_3}{3}$ and x_1, x_2 enter with

$$\tilde{f}_1(x_1) = 0.1 \exp(x_1), \quad \tilde{f}_2(x_2) = 0.1 \exp(-x_2),$$

  i.e. f_3 is implicitly correlated to $\tilde{f}_1$ and $\tilde{f}_2$ only via the relation of the covariates.

- *Scenario 2:* Covariate x_3 is functionally related to the predictor with $\tilde{f}_3(x_3) = \frac{x_3}{3}$ and x_1, x_2 enter with

$$\tilde{f}_1(x_1) = \sin(x_1), \quad \tilde{f}_2(x_2) = \cos(x_2),$$

  i.e. $\tilde{f}_3$ is explicitly correlated to $\tilde{f}_1$ and $\tilde{f}_2$ since for $c = 1$ the effect $\tilde{f}_3$ is a linear combination of $\tilde{f}_1$ and $\tilde{f}_2$.

(5) The predictor is given by

$$\eta = f_1(x_1) + f_2(x_2) + f_3(x_3)$$

where $f_j(x_{ij}) = \tilde{f}_j(x_{ij}) - \sum_{k=1} \tilde{f}_j(x_{kj})$, $i = 1, \dots, n$, $j = 1, 2, 3$. Depending on the type of the model, we then link this predictor to the parameter via $\mu_1 = \eta$, $\sigma_1 = \exp(\eta)$ or $\rho = \frac{\eta}{\sqrt{1+\eta^2}}$.

(6) For each scenario and each type of model, we use sample sizes $n = 100$ and $n = 1,000$. The number of simulation replications is chosen to be 250.

(7) The functions f_1, f_2 and f_3 are approximated with cubic Bayesian P-splines with 20 inner knots on an equidistant grid within the range of the covariates and second order random walk prior. Inverse gamma priors with $a = b = 0.001$ are assigned for the hyperparameters τ^2.

(8) We use 12,000 iterations with a burnin phase of 2,000 iterations and a thinning parameter of size 10.

Results of this study are summarized in terms of boxplots of logarithmic MSEs in Figures D10 to D12 as well as sampling paths and autocorrelation plots of one coefficient β_j for each of the functions f_1, f_2 and f_3 in scenario 1 and 2 of the μ_1, σ_1 and ρ model for $n = 100$ (Figures D13 to D18).

(1) *MSE.* Overall estimation of posterior means of the functions seem to be fairly robust against increasing concurvity.

(2) *Bias (not shown).* The bias is rather small in all 12 models and we do not recognize a systematic deterioration of the bias with increasing value of c. For scenario 2, we observe slightly higher biases compared to scenario 1 which is in line with our findings of the MSEs. In general, the bias seem to be fairly robust to concurvity (complying with results obtained for collinearity in linear models).

(3) *MCMC samples.* In all 12 models, acceptance rates are between 70% and 100% and are not impaired by the induced concurvity. Furthermore, in all six models under scenario 1, the positive mixing behaviour and small autocorrelations are preserved also for $c > 0$ (with the exception of $c = 1$ with partly higher autocorrelations). The same observation can be made for $c = 0, 0.25, 0.5, 0.75$ under scenario 2. In the worst situation, which is the perfectly deterministic

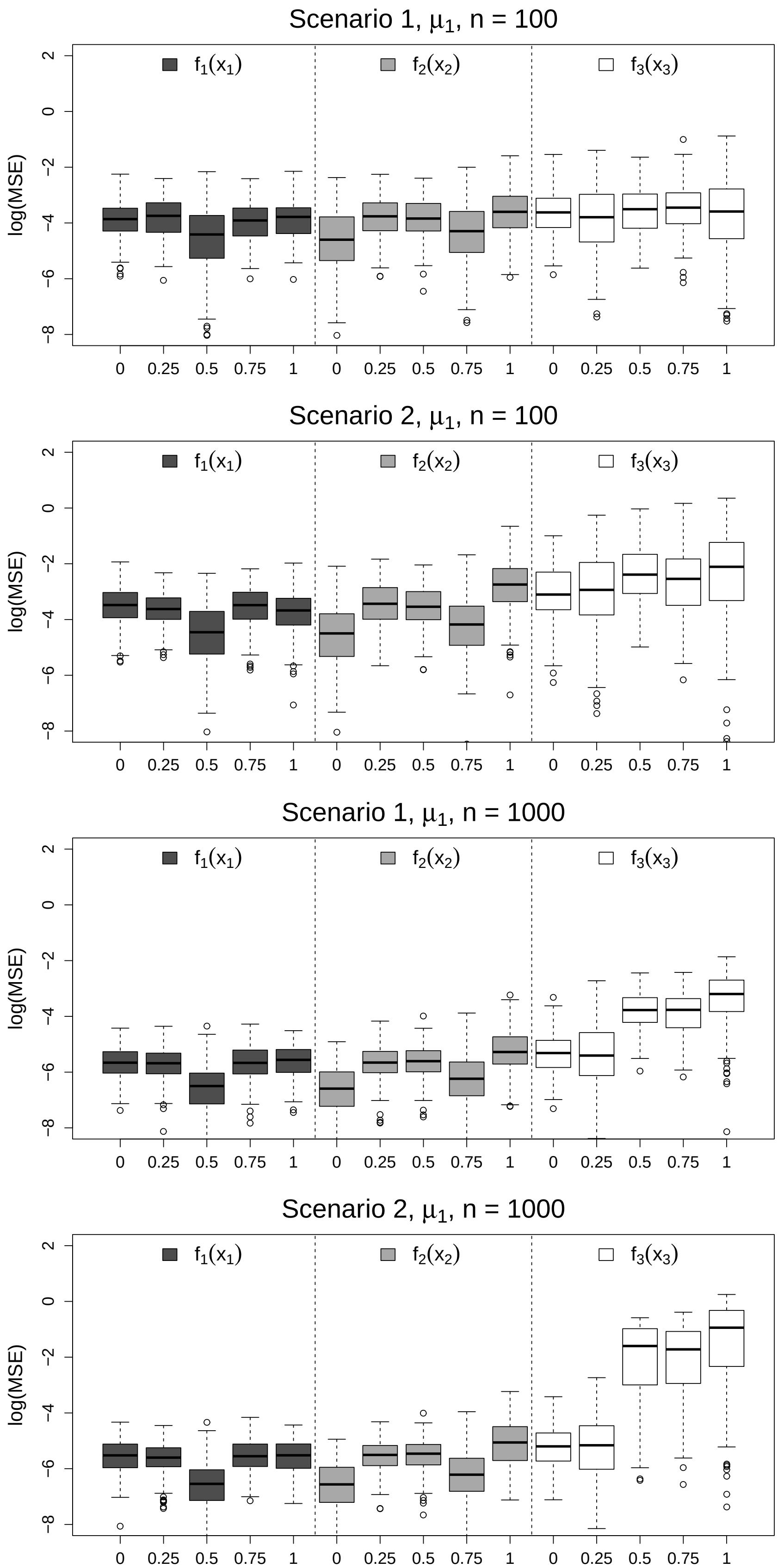

Figure D10: Simulation study, concurvity. Boxplots of logarithmic mean squared errors of the effects f_1, f_2 and f_3 in the μ_1 model for the different amounts of concurvity c on the x-axes. The upper two boxes correspond to scenario 1 and 2 for sample size $n = 100$ and the lower two boxes to scenario 1 and 2 for $n = 1,000$.

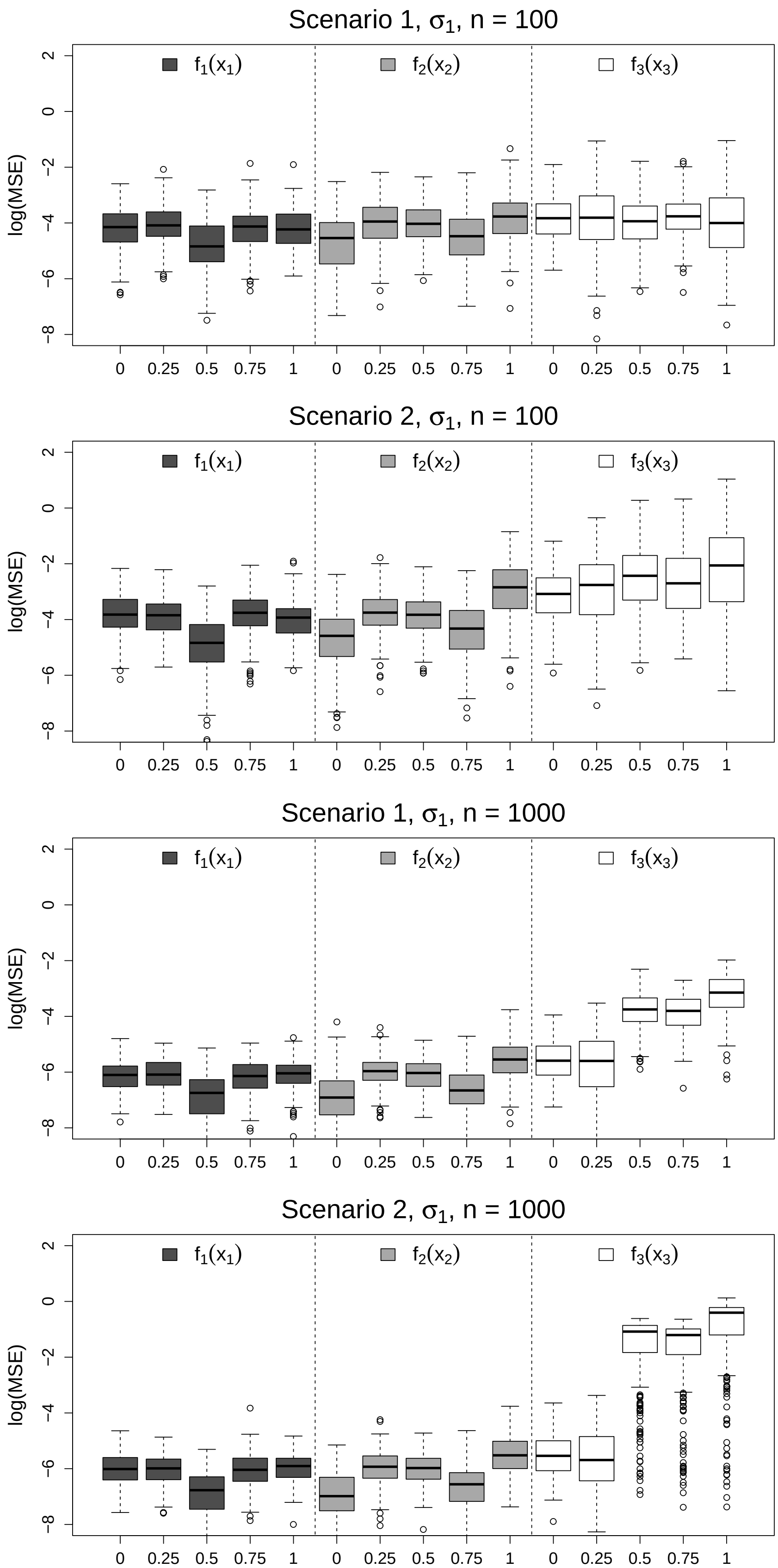

Figure D11: Simulation study, concurvity. Boxplots of logarithmic mean squared errors of the effects f_1, f_2 and f_3 in the σ_1 model for the different amounts of concurvity c on the x-axes. The upper two boxes correspond to scenario 1 and 2 for sample size $n = 100$ and the lower two boxes to scenario 1 and 2 for $n = 1,000$.

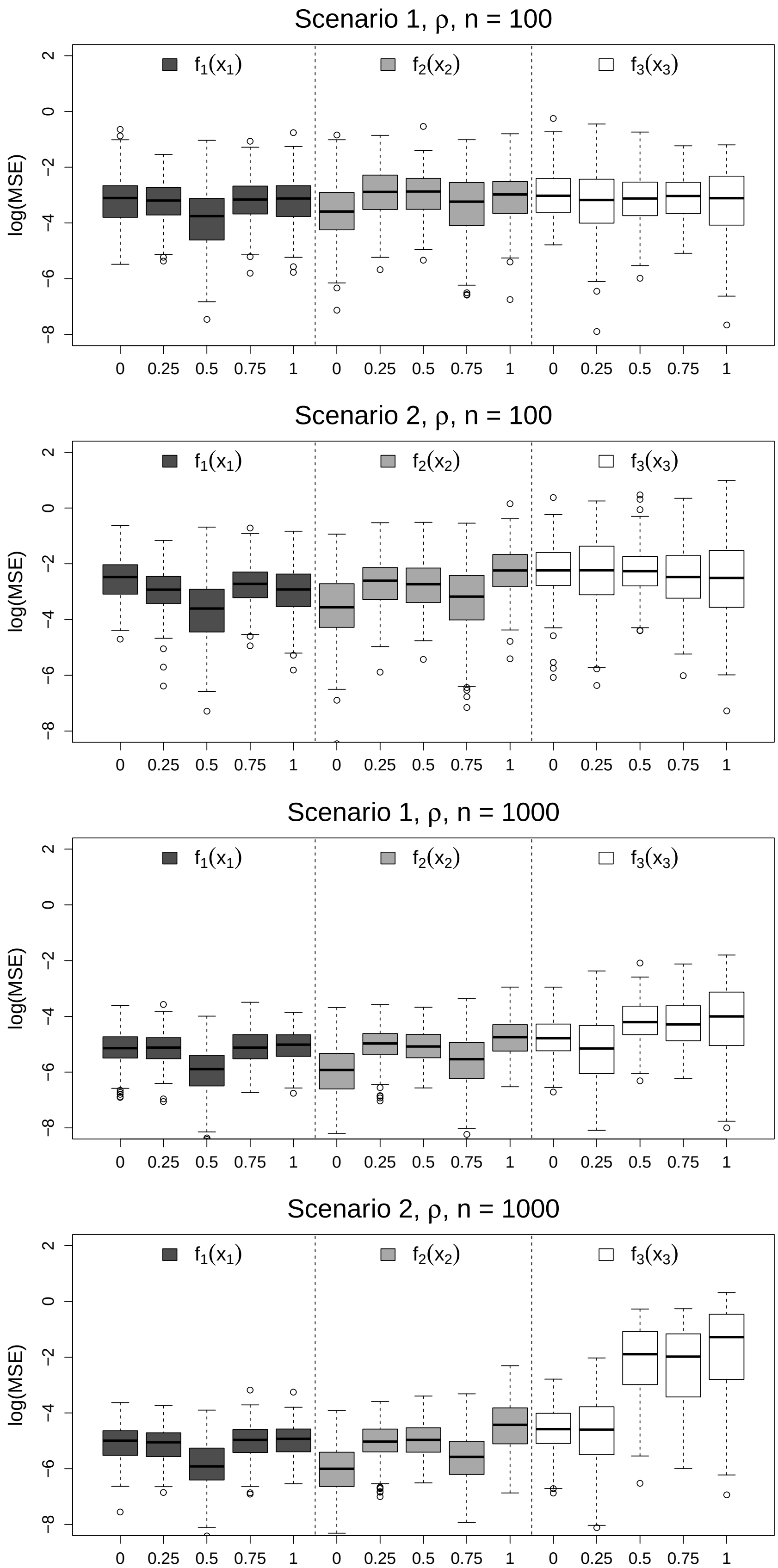

Figure D12: Simulation study, concurvity. Boxplots of logarithmic mean squared errors of the effects f_1, f_2 and f_3 in the ρ model for the different amounts of concurvity c on the x-axes. The upper two boxes correspond to scenario 1 and 2 for sample size $n = 100$ and the lower two boxes to scenario 1 and 2 for $n = 1,000$.

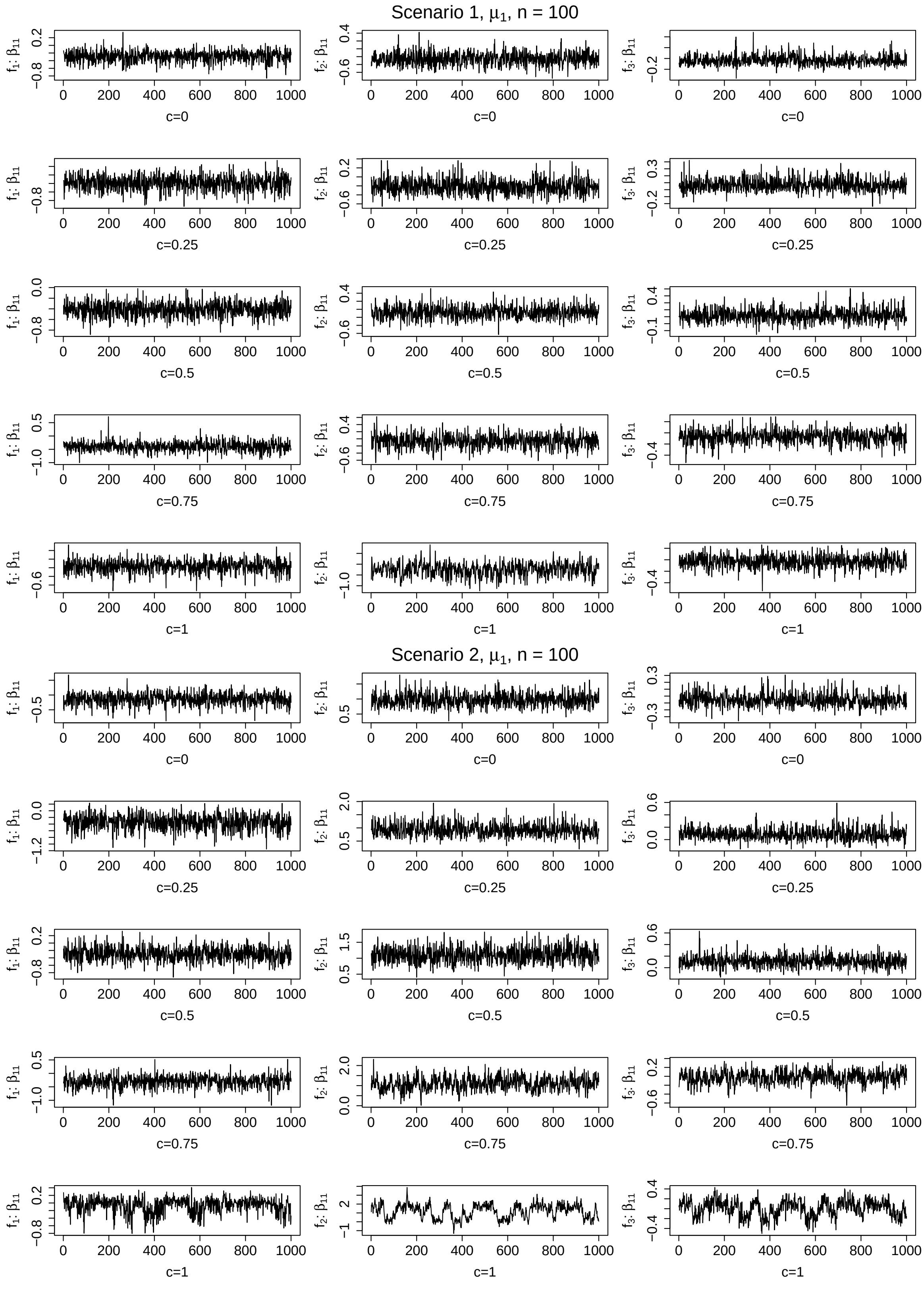

Figure D13: Simulation study, concurvity. Samplepaths of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the μ_1 model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

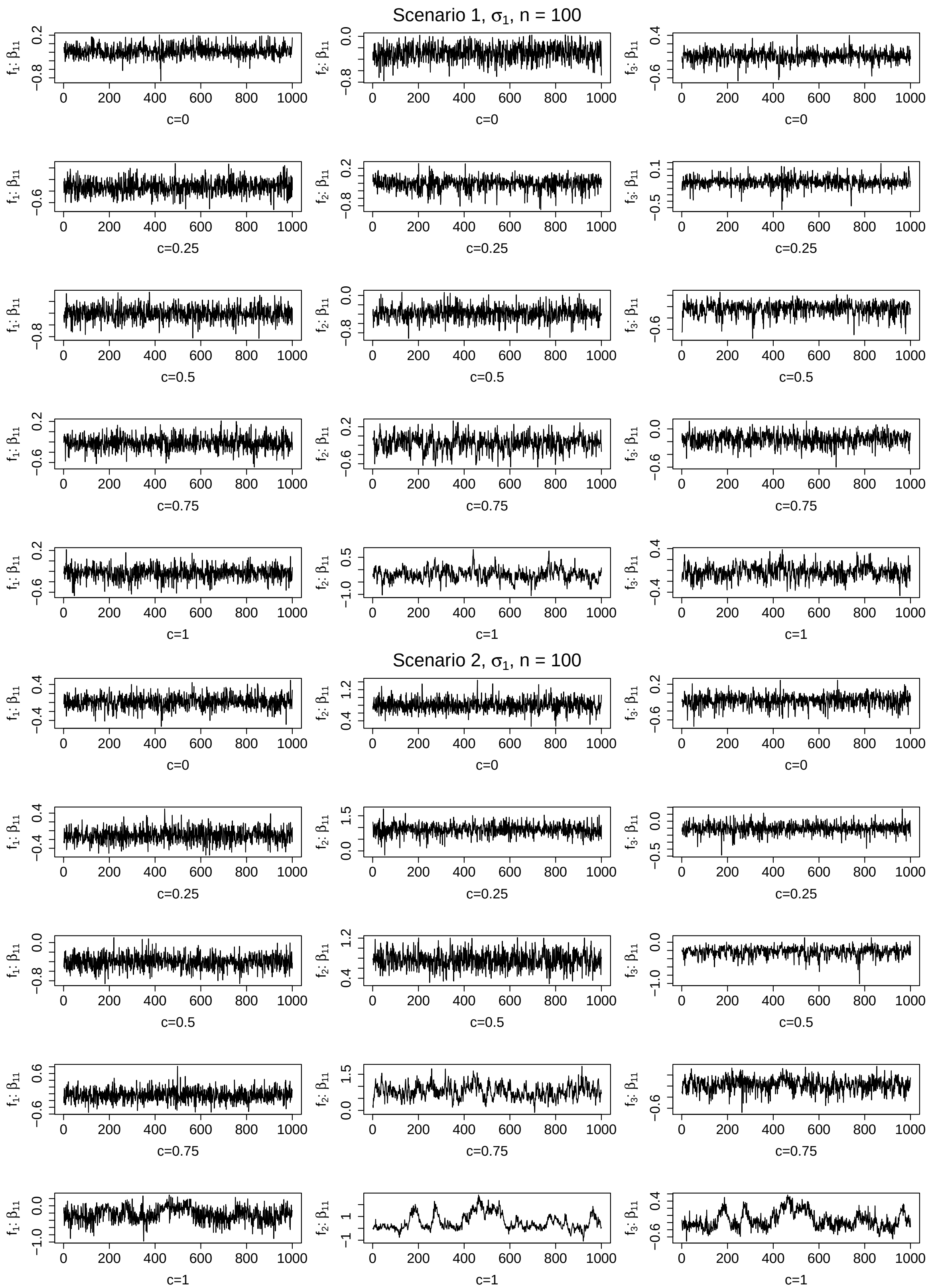

Figure D14: Simulation study, concurvity. Samplepaths of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the σ_1 model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

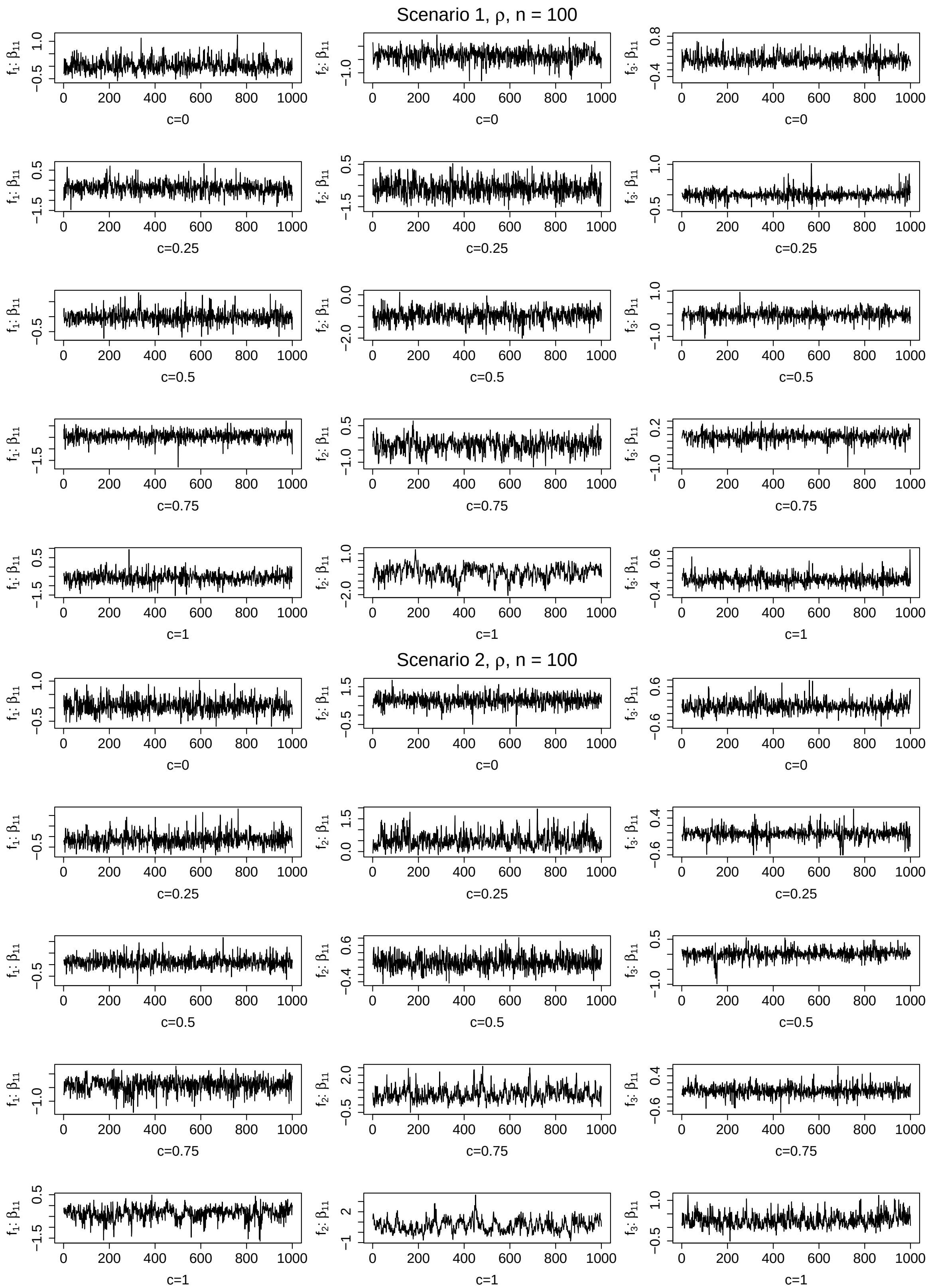

Figure D15: Simulation study, concurvity. Samplepaths of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the ρ model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

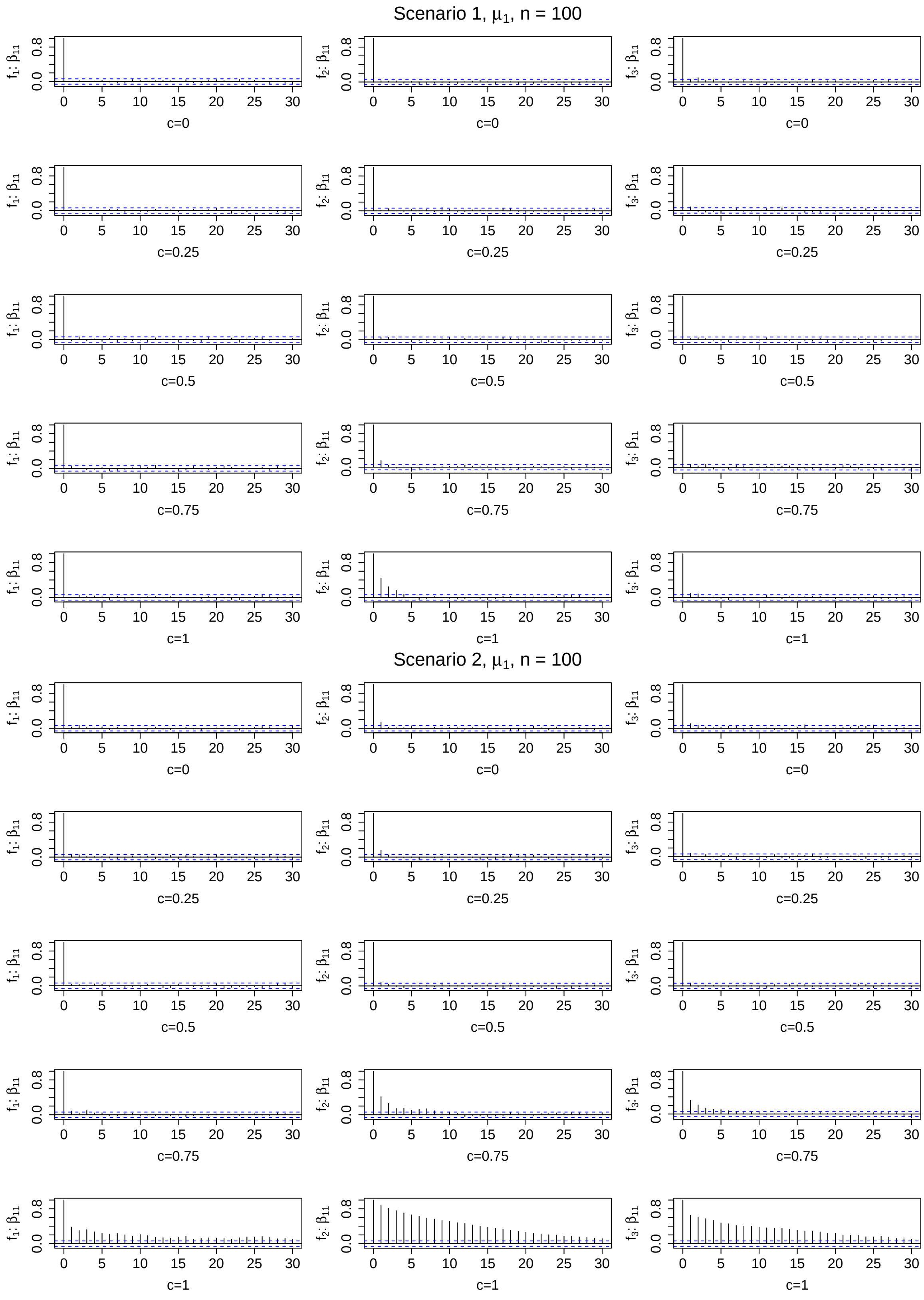

Figure D16: Simulation study, concurvity. Autocorrelation plots of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the μ_1 model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

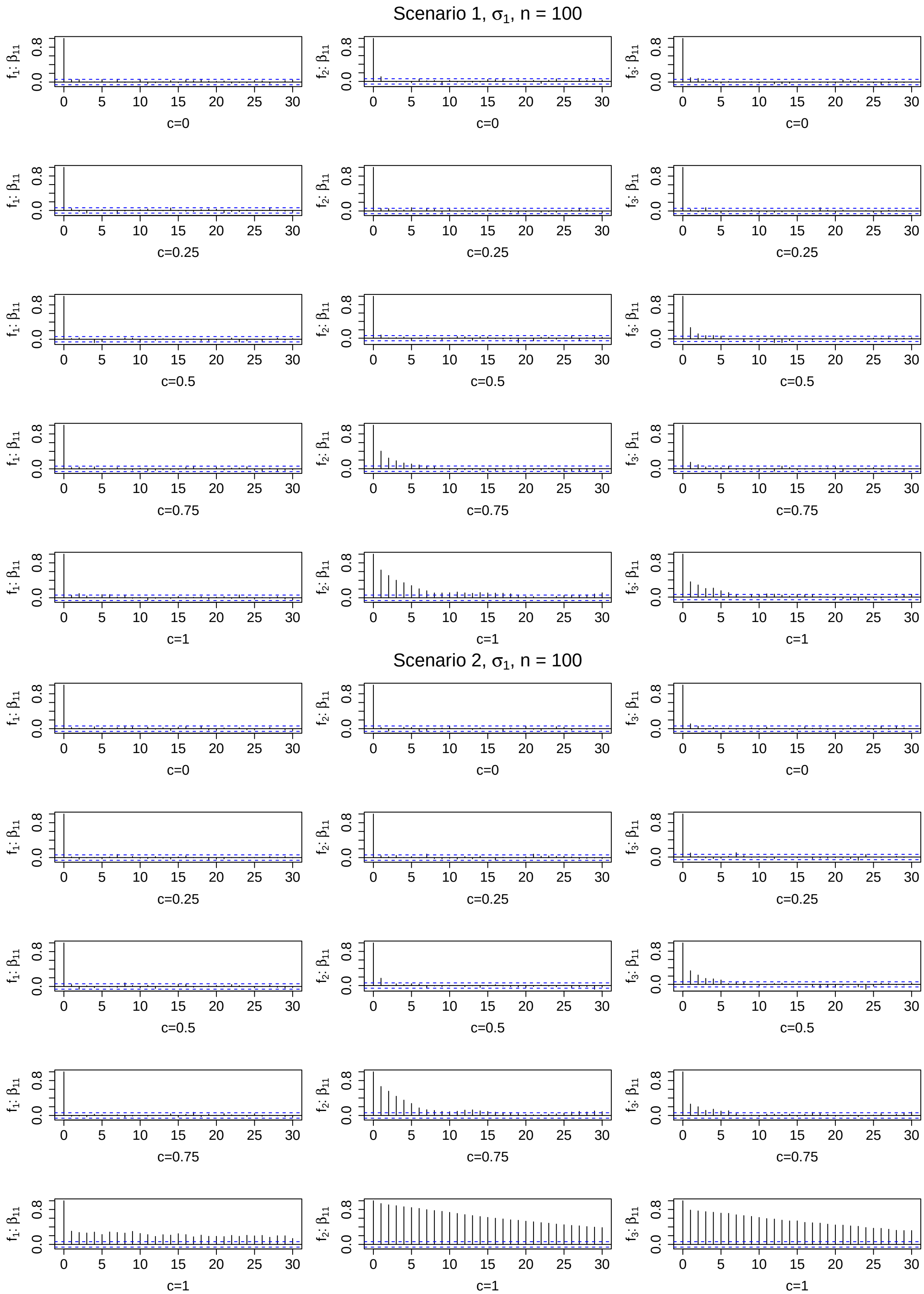

Figure D17: Simulation study, concurvity. Autocorrelation plots of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the σ_1 model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

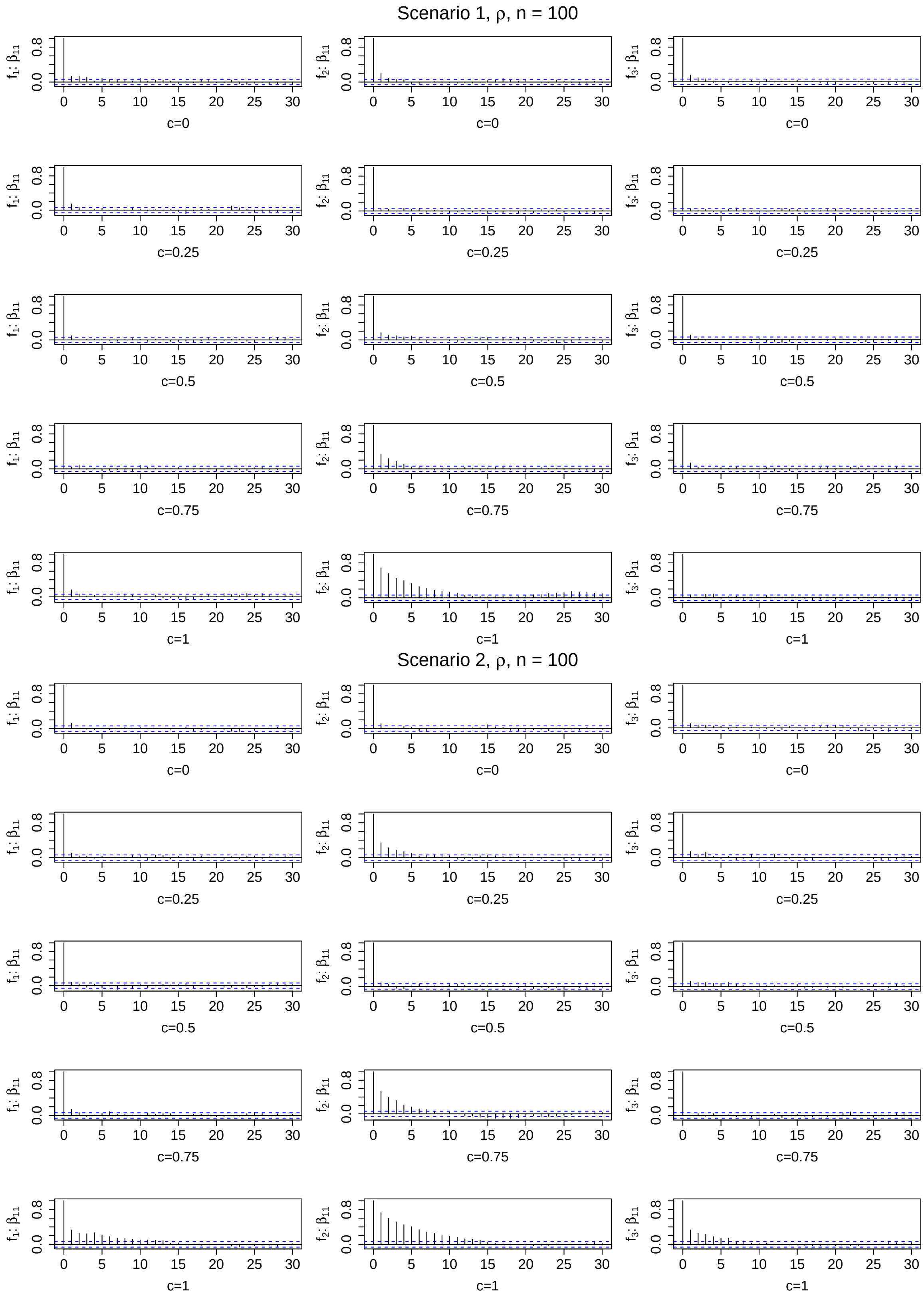

Figure D18: Simulation study, concurvity. Autocorrelation plots of coefficient β_{11} of the function estimates f_1, f_2 and f_3 (column-wise) in the ρ model of one randomly selected replication for the different amounts of concurvity c (row-wise). The upper four rows correspond to scenario 1 and the lower four to scenario 2 for sample size $n = 100$.

Covariate	Description
stunting	stunting (height-for-age) according to the WHO definition (continuous, $-5.99 \leq$ stunting ≤ 5.97)
wasting	wasting (weight-for-height) according to the WHO definition (continuous, $-5.97 \leq$ wasting ≤ 5.95)
cage	age of the child in months (continuous, $0 \leq$ cage ≤ 35)
breastfeeding	months of breastfeeding (continuous, $0 \leq$ breastfeeding ≤ 35)
csex	sex of the child (categorical, 1=male, 2=female)
ctwin	child is twin (categorical, 0=no, 1=yes)
mbmi	BMI of the mother in kg/m^2 (continuous, $12.45 \leq$ mbmi ≤ 39.85)
mweight	weight of the mother in kg (continuous, $27.9 \leq$ mweight ≤ 106.9)
mage	age of the mother in years (continuous, $13 \leq$ mage ≤ 49)
medu	education of the mother in single years (continuous, $0 \leq$ medu ≤ 22)
edupartner	education of the mother's partner in single years (continuous, $0 \leq$ edupartner ≤ 22)
munemployed	mother is unemployed (categorical, 0=yes, 1=no)
mreligion	religion of the mother (categorical, 1=Hindu, 2=Muslim, 3= Christian, 4=Sikh, 5=Other)
electricity	availability of electricity (categorical, 0=no, 1=yes)
radio	ownership of a radio (categorical, 0=no, 1=yes)
television	ownership of a television (categorical, 0=no, 1=yes)
refrigerator	ownership of a refrigerator (categorical, 0=no, 1=yes)
bicycle	ownership of a bicycle (categorical, 0=no, 1=yes)
motorcycle	ownership of a motorcycle (categorical, 0=no, 1=yes)
car	ownership of a car (categorical, 0=no, 1=yes)
dist	district (categorical, $5 \leq$ dist ≤ 482, 438 districts)

Table E1: Childhood undernutrition. Description of covariates included in the full models.

($c = 1$) scenario 2, we however have to deal with several replications for which the MCMC chains between f_1, f_2 and f_3 are not independent resulting in partly non-stationary paths and higher autocorrelations.

E Supplementary Material to Section 3

Table E1 summarises the covariates and dependent variables included in the full models.

Figures E22 to E29 show nonlinear and spatial effects for the parameters $\mu_{stunting}$, $\mu_{wasting}$, $\sigma_{stunting}$ and $\sigma_{wasting}$, n_{df} of the bivariate t model obtained from the stepwise

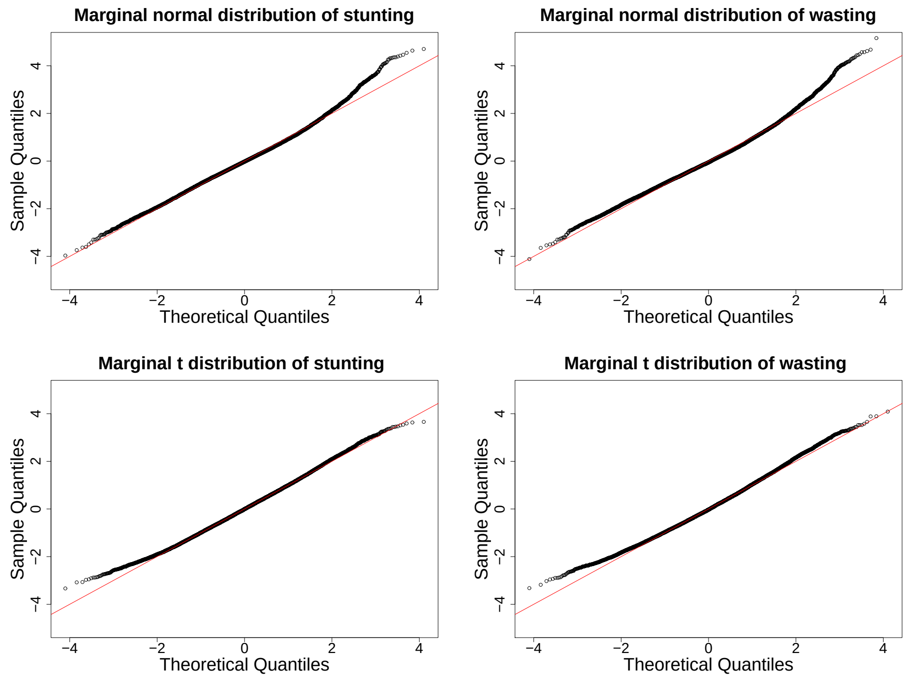

Figure E19: Childhood undernutrition. Quantile residuals of the marginal distributions in the selected bivariate normal model (first row) and the selected bivariate t model (second row).

model selection. All shown effects are centred around zero. The predictors for these parameters are of the form (4) given in the main paper. Estimates of the parametric effects can be found in Tables E2 to E7. Figure E19 shows the marginal residual plots for the optimal bivariate normal and t model and Figure E20 illustrates the positive mixing behaviour of MCMC in this study.

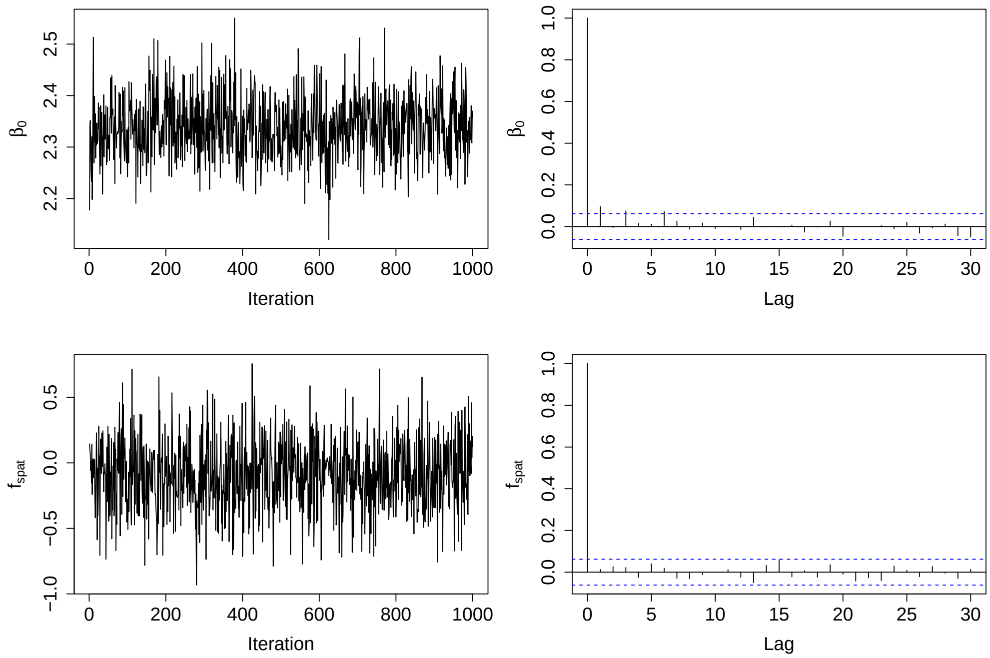

Figure E20: Childhood undernutrition, selected bivariate t model. Sampling-paths and autocorrelation plots for the intercept $\beta_0^{n_{df}}$ (first row) and for one coefficient of $f_{spat}^{n_{df}}$ of a randomly selected district in India (second row).

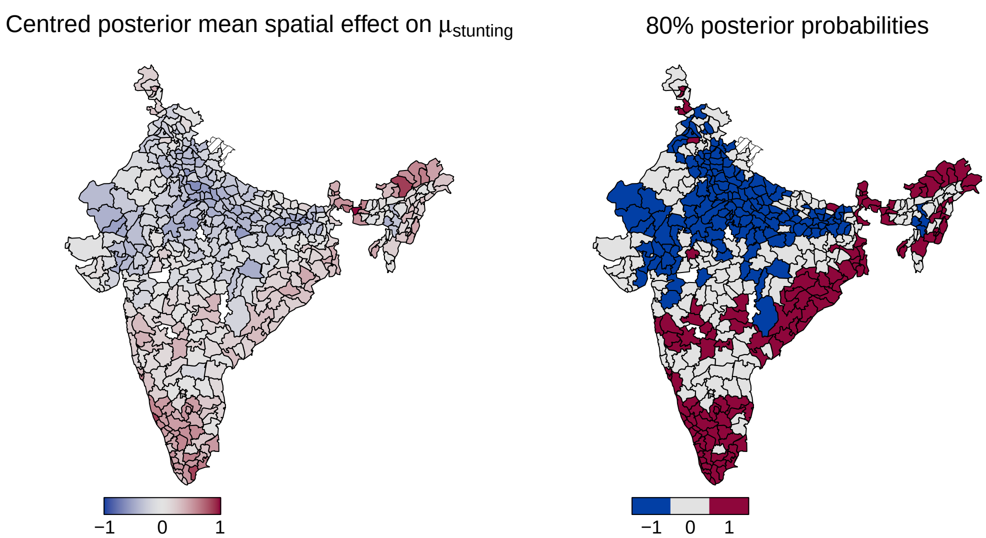

Figure E21: Childhood undernutrition, bivariate t distribution. Posterior mean estimates of the complete spatial effects on $\mu_{stunting}$ (left) and 80% posterior probabilities (right). In the right panel a value of 1 corresponds to a strictly positive 80% credible interval and a value of -1 to a strictly negative credible interval. A value of 0 indicates that the corresponding credible interval contains zero.

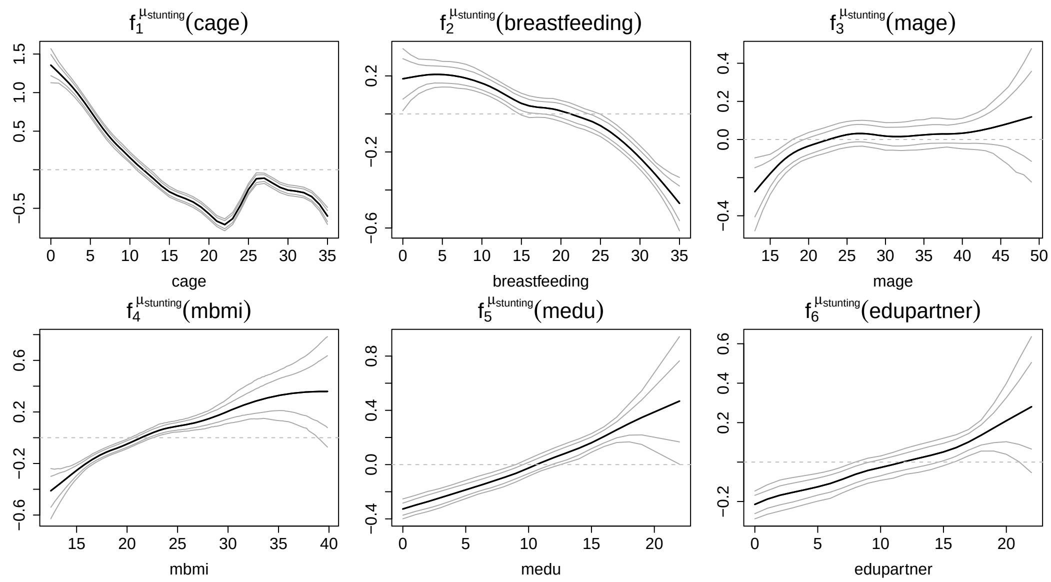

Figure E22: Childhood undernutrition, selected bivariate t model. Posterior mean estimates of nonlinear effects on $\mu_{stunting}$ together with 80% and 95% pointwise credible intervals. The effects are centred around zero.

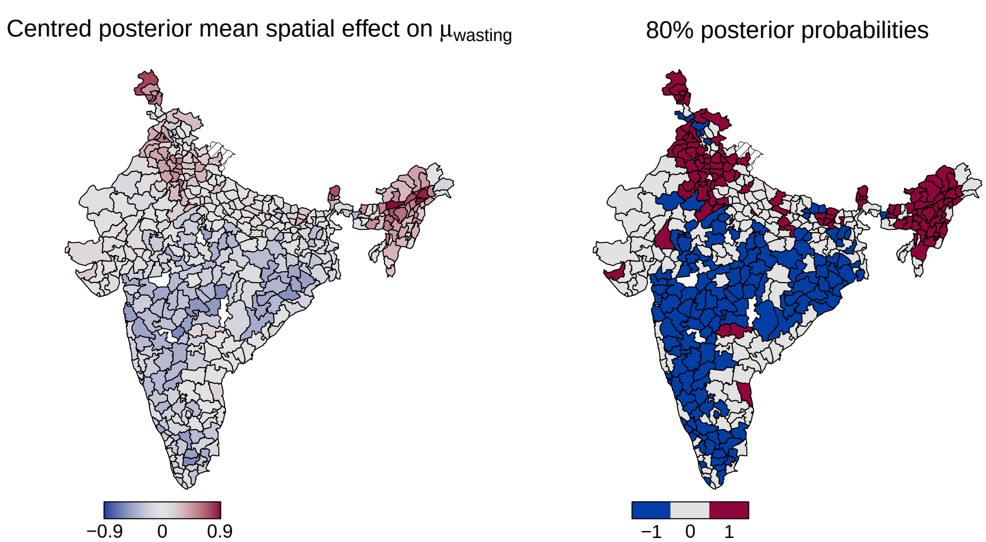

Figure E23: Childhood undernutrition, bivariate t distribution. Posterior mean estimates of the complete spatial effects on $\mu_{wasting}$ (left) and 80% posterior probabilities (right). In the right panel a value of 1 corresponds to a strictly positive 80% credible interval and a value of -1 to a strictly negative credible interval. A value of 0 indicates that the corresponding credible interval contains zero.

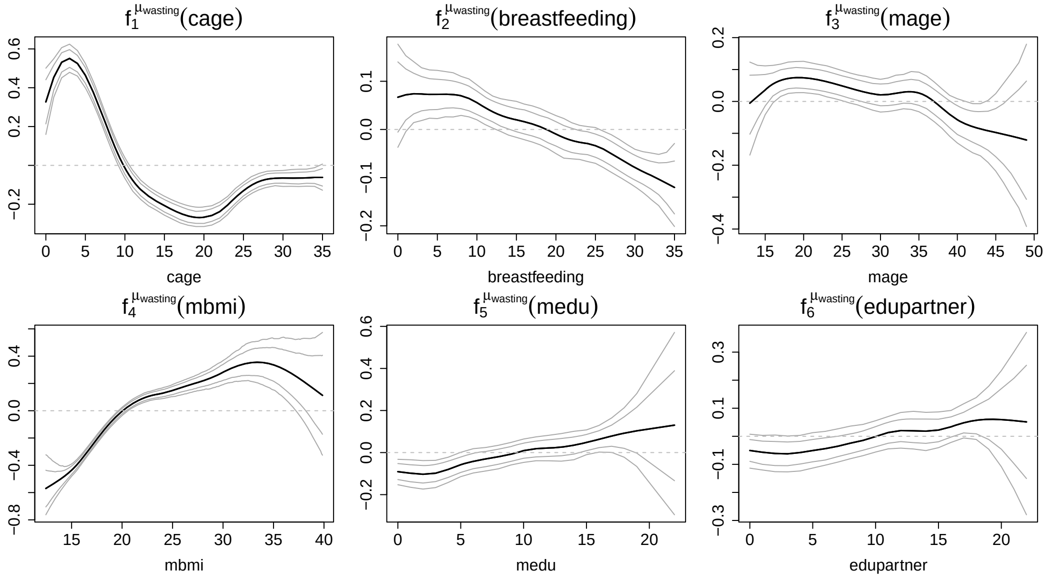

Figure E24: Childhood undernutrition, selected bivariate t model. Posterior mean estimates of selected nonlinear effects on $\mu_{wasting}$ together with 80% and 95% pointwise credible intervals. The effects are centred around zero.

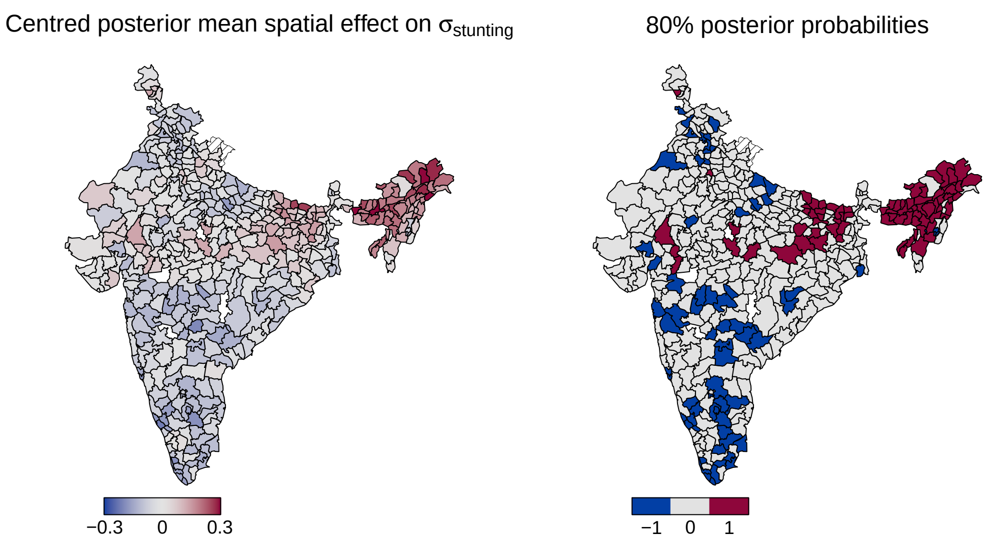

Figure E25: Childhood undernutrition, bivariate t distribution. Posterior mean estimates of the complete spatial effects on $\sigma_{stunting}$ (left) and 80% posterior probabilities (right). In the right panel a value of 1 corresponds to a strictly positive 80% credible interval and a value of -1 to a strictly negative credible interval. A value of 0 indicates that the corresponding credible interval contains zero.

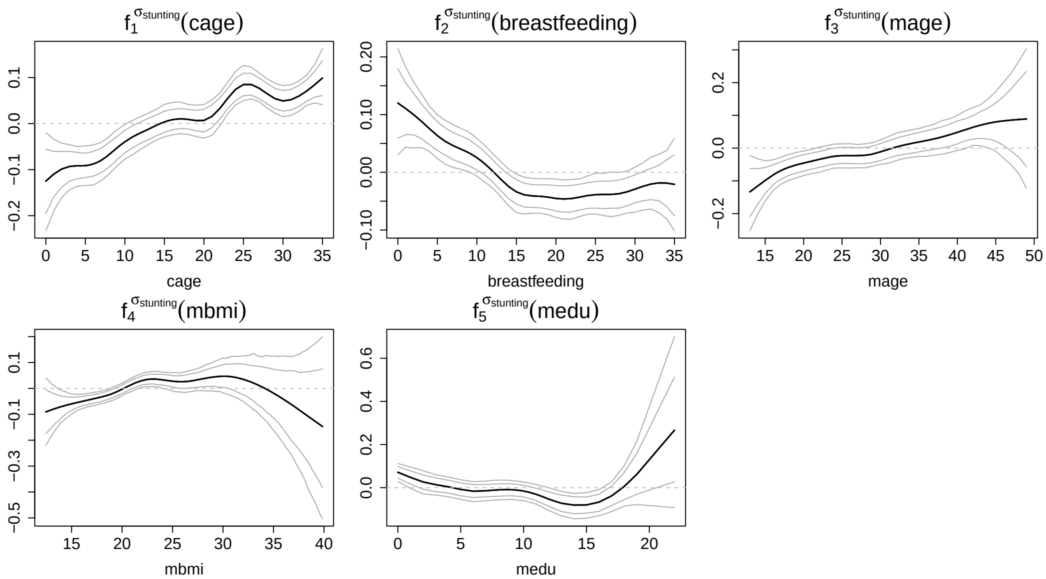

Figure E26: Childhood undernutrition, selected bivariate t model. Posterior mean estimates of selected nonlinear effects on $\sigma_{stunting}$ together with 80% and 95% pointwise credible intervals. The effects are centred around zero.

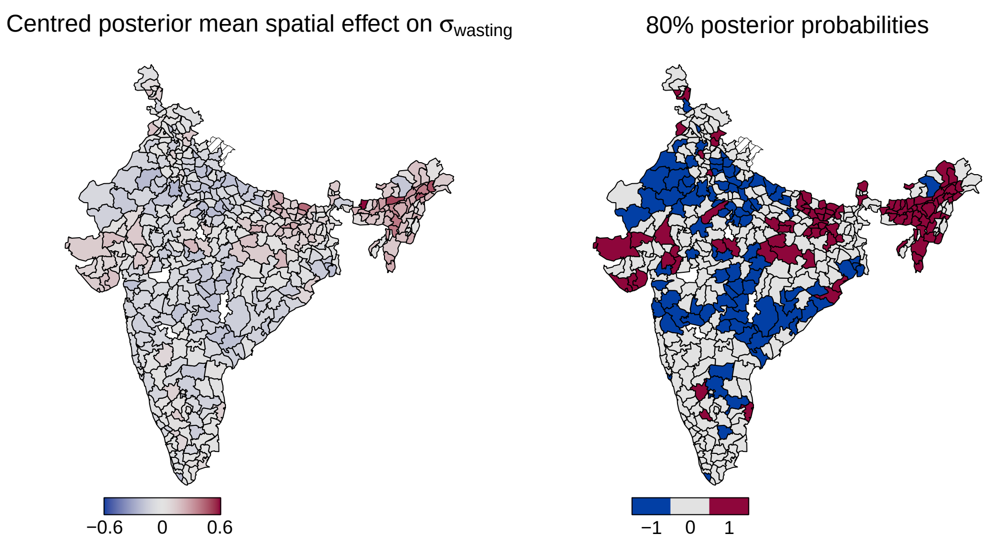

Figure E27: Childhood undernutrition, bivariate t distribution. Posterior mean estimates of the complete spatial effects on $\sigma_{wasting}$ (left) and 80% posterior probabilities (right). In the right panel a value of 1 corresponds to a strictly positive 80% credible interval and a value of -1 to a strictly negative credible interval. A value of 0 indicates that the corresponding credible interval contains zero.

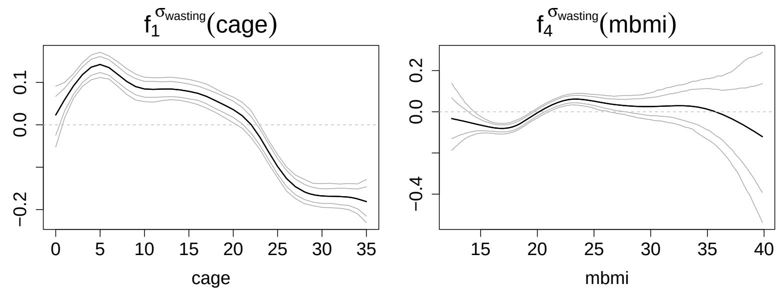

Figure E28: Childhood undernutrition, selected bivariate t model. Posterior mean estimates of nonlinear effects on $\sigma_{wasting}$ together with 80% and 95% pointwise credible intervals. The effects are centred around zero.

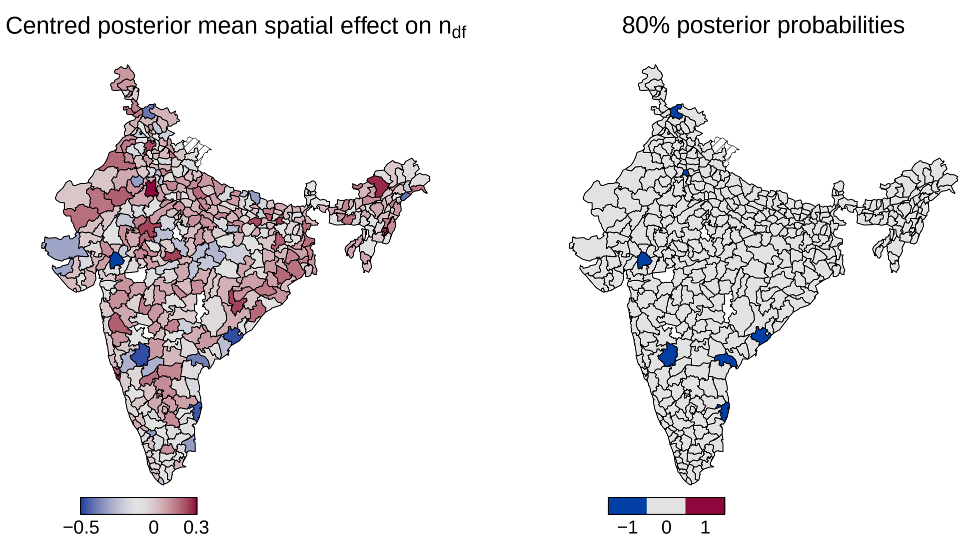

Figure E29: Childhood undernutrition, selected bivariate t model. Posterior mean estimates of the complete spatial effects on n_{df} (left) and 80% posterior probabilities (right). In the right panel a value of 1 corresponds to a strictly positive 80% credible interval and a value of -1 to a strictly negative credible interval. A value of 0 indicates that the corresponding credible interval contains zero. Note that in the left panel the plot range is not centred around zero for reasons of visibility.

Covariate	mean	std	2.5%	10%	50%	90	97.5%
const	-1.59	0.05	-1.69	-1.65	-1.59	-1.53	-1.49
csex	-0.02	0.02	-0.06	-0.04	-0.02	-0.00	0.01
ctwin	-0.75	0.08	-0.92	-0.86	-0.76	-0.64	-0.58
munemployed	-0.06	0.02	-0.10	-0.09	-0.06	-0.03	-0.02
motorcycle	0.10	0.03	0.04	0.06	0.10	0.14	0.16
mreligion2	-0.05	0.03	-0.11	-0.09	-0.05	-0.02	0.00
mreligion3	0.12	0.04	0.03	0.06	0.12	0.18	0.21
mreligion4	0.13	0.07	-0.01	0.04	0.13	0.22	0.26
mreligion5	0.10	0.06	-0.02	0.02	0.10	0.17	0.20
electricity	0.10	0.02	0.05	0.07	0.10	0.13	0.14
radio	0.05	0.02	0.01	0.02	0.05	0.07	0.08
television	0.14	0.02	0.10	0.12	0.14	0.17	0.19
refrigerator	0.11	0.03	0.05	0.07	0.11	0.15	0.17
bicycle	-0.02	0.02	-0.06	-0.05	-0.02	-0.00	0.01
car	0.03	0.06	-0.08	-0.05	0.03	0.11	0.15

Table E2: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on $\mu_{stunting}$.

Covariate	mean	std	2.5%	10%	50%	90%	97.5%
const	-0.79	0.04	-0.87	-0.84	-0.79	-0.75	-0.72
csex	0.01	0.01	-0.02	-0.01	0.01	0.03	0.04
ctwin	-0.13	0.07	-0.27	-0.22	-0.13	-0.04	0.01
munemployed	-0.03	0.01	-0.06	-0.05	-0.03	-0.01	0.00
motorcycle	0.02	0.02	-0.02	-0.01	0.02	0.05	0.07
mreligion2	-0.00	0.02	-0.04	-0.03	-0.00	0.02	0.03
mreligion3	0.05	0.04	-0.03	-0.00	0.05	0.10	0.12
mreligion4	0.06	0.06	-0.05	-0.01	0.06	0.13	0.16
mreligion5	0.02	0.04	-0.07	-0.03	0.02	0.08	0.11
electricity	0.03	0.02	-0.01	0.01	0.03	0.05	0.06
radio	0.00	0.01	-0.03	-0.02	0.00	0.02	0.03
television	0.00	0.02	-0.03	-0.02	0.00	0.03	0.04
refrigerator	0.07	0.03	0.01	0.03	0.07	0.10	0.12
bicycle	0.03	0.01	-0.00	0.01	0.03	0.05	0.06
car	0.02	0.05	-0.07	-0.05	0.02	0.08	0.11

Table E3: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on $\mu_{wasting}$.

Covariate	mean	std	2.5%	10%	50%	90	97.5%
const	0.25	0.03	0.19	0.21	0.25	0.29	0.31
csex	-0.02	0.01	-0.04	-0.03	-0.02	-0.01	0.00
ctwin	-0.00	0.06	-0.11	-0.07	-0.00	0.07	0.11
munemployed	0.01	0.01	-0.01	-0.00	0.01	0.03	0.04
motorcycle	-0.02	0.02	-0.06	-0.04	-0.02	0.01	0.02
mreligion2	0.03	0.02	-0.01	0.00	0.03	0.05	0.06
mreligion3	-0.07	0.03	-0.13	-0.11	-0.07	-0.03	-0.01
mreligion4	0.04	0.05	-0.05	-0.02	0.04	0.10	0.13
mreligion5	-0.14	0.04	-0.21	-0.19	-0.14	-0.09	-0.07
electricity	-0.03	0.01	-0.06	-0.05	-0.03	-0.01	-0.00
radio	-0.01	0.01	-0.03	-0.02	-0.01	0.01	0.02
television	0.00	0.02	-0.03	-0.02	0.00	0.02	0.03
refrigerator	0.03	0.02	-0.02	-0.00	0.03	0.05	0.07
bicycle	-0.02	0.01	-0.04	-0.03	-0.02	-0.00	0.01
car	-0.04	0.04	-0.13	-0.10	-0.04	0.01	0.05

Table E4: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on $\sigma_{stunting}$.

Covariate	mean	std	2.5%	10%	50%	90	97.5%
const	-0.01	0.02	-0.04	-0.03	-0.01	0.01	0.02
csex	-0.03	0.01	-0.05	-0.04	-0.03	-0.01	-0.01
ctwin	0.07	0.06	-0.05	-0.01	0.06	0.14	0.18
munemployed	0.01	0.01	-0.02	-0.01	0.01	0.03	0.04
motorcycle	0.01	0.02	-0.03	-0.02	0.01	0.04	0.05
mreligion2	-0.00	0.02	-0.03	-0.02	-0.00	0.02	0.03
mreligion3	-0.02	0.03	-0.08	-0.06	-0.02	0.02	0.03
mreligion4	0.01	0.04	-0.08	-0.04	0.01	0.07	0.10
mreligion5	-0.12	0.04	-0.20	-0.17	-0.12	-0.07	-0.04
electricity	-0.02	0.02	-0.05	-0.04	-0.02	0.00	0.01
radio	-0.00	0.01	-0.02	-0.02	-0.00	0.02	0.02
television	-0.02	0.01	-0.05	-0.04	-0.02	-0.00	0.01
refrigerator	0.04	0.02	-0.00	0.01	0.04	0.07	0.09
bicycle	-0.01	0.01	-0.03	-0.02	-0.01	0.01	0.02
car	-0.04	0.04	-0.13	-0.10	-0.04	0.01	0.04

Table E5: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on $\sigma_{wasting}$.

Covariate	mean	std	2.5%	10%	50%	90	97.5%
const	-0.21	0.02	-0.24	-0.23	-0.21	-0.18	-0.17
csex	0.00	0.01	-0.03	-0.02	0.00	0.02	0.03
ctwin	0.00	0.07	-0.14	-0.09	0.01	0.10	0.15
munemployed	-0.04	0.02	-0.08	-0.06	-0.04	-0.02	-0.01
motorcycle	-0.03	0.03	-0.08	-0.06	-0.03	0.01	0.03
mreligion2	-0.03	0.02	-0.08	-0.06	-0.03	-0.00	0.01
mreligion3	-0.01	0.04	-0.09	-0.06	-0.01	0.03	0.06
mreligion4	0.01	0.06	-0.10	-0.07	0.01	0.08	0.12
mreligion5	0.08	0.05	-0.01	0.02	0.08	0.14	0.18
electricity	0.02	0.02	-0.02	-0.01	0.02	0.04	0.06
radio	-0.00	0.02	-0.03	-0.02	-0.00	0.02	0.03
television	0.01	0.02	-0.03	-0.02	0.01	0.04	0.05
refrigerator	0.06	0.03	0.01	0.03	0.06	0.10	0.12
bicycle	-0.00	0.02	-0.03	-0.02	-0.00	0.02	0.03
car	0.11	0.06	-0.00	0.04	0.12	0.19	0.22

Table E6: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on ρ.

Covariate	mean	std	2.5%	10%	50%	90	97.5%
const	2.34	0.06	2.23	2.27	2.34	2.42	2.46
csex	-0.22	0.04	-0.31	-0.28	-0.22	-0.17	-0.15
ctwin	-0.08	0.23	-0.51	-0.35	-0.10	0.22	0.37
munemployed	-0.00	0.05	-0.10	-0.07	-0.00	0.06	0.09
motorcycle	-0.12	0.08	-0.27	-0.22	-0.12	-0.02	0.04
mreligion2	-0.01	0.07	-0.13	-0.09	-0.01	0.08	0.13
mreligion3	-0.13	0.11	-0.35	-0.27	-0.13	0.01	0.08
mreligion4	0.02	0.16	-0.31	-0.19	0.03	0.23	0.33
mreligion5	-0.21	0.13	-0.46	-0.37	-0.21	-0.04	0.06
electricity	-0.03	0.05	-0.13	-0.09	-0.03	0.04	0.09
radio	-0.06	0.05	-0.15	-0.12	-0.06	0.01	0.04
television	-0.03	0.06	-0.14	-0.10	-0.03	0.04	0.08
refrigerator	0.25	0.09	0.08	0.14	0.25	0.37	0.44
bicycle	-0.01	0.05	-0.09	-0.07	-0.01	0.05	0.08
car	-0.41	0.14	-0.70	-0.59	-0.41	-0.23	-0.11

Table E7: Childhood undernutrition, selected bivariate t model. Estimates of linear effects on n_{df}.

Party	mean	std	2.5%	10%	50%	90	97.5%
CDU/CSU	0.04	0.02	0.01	0.02	0.03	0.06	0.08
SPD	0.15	0.05	0.08	0.10	0.15	0.22	0.26
FDP	0.07	0.02	0.03	0.04	0.06	0.10	0.12
The Greens	0.11	0.04	0.06	0.07	0.11	0.16	0.19
The Left	0.43	0.11	0.25	0.30	0.41	0.58	0.70
Others	0.11	0.04	0.05	0.07	0.10	0.15	0.19

Table F8: Federal election. Estimated hyperparameters τ^2 for the spatial effects.

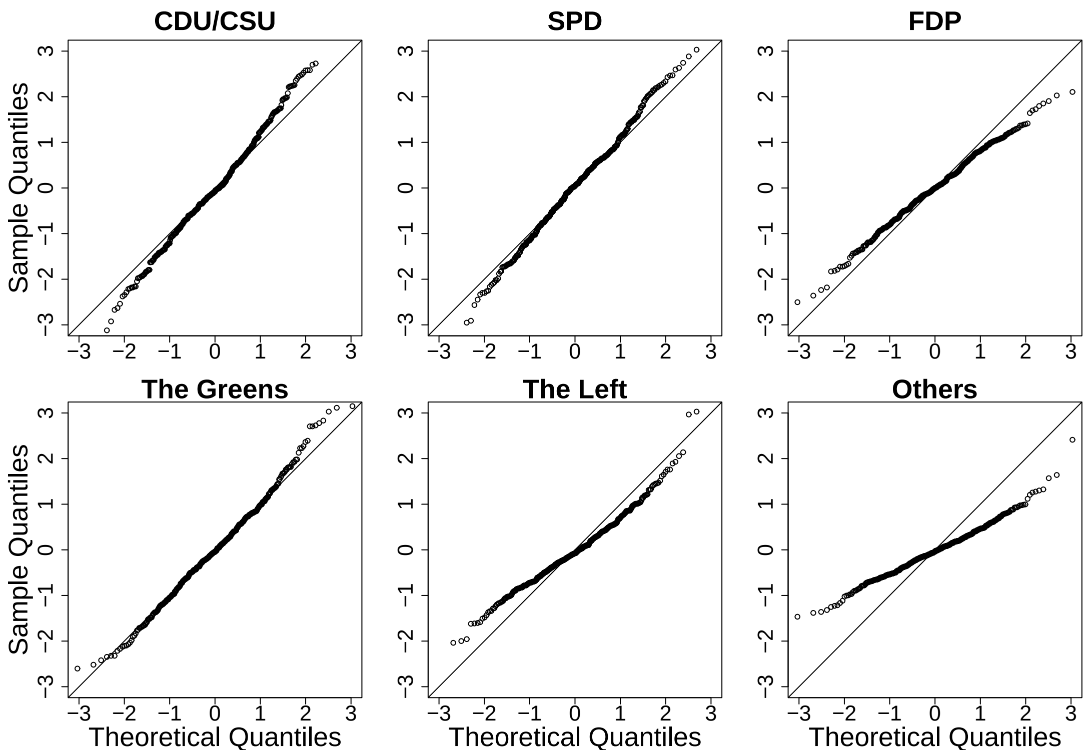

Figure F30: Federal election. Quantile residuals of marginal distributions.

F Supplementary Material to Section 4

Table F8 summarises the posterior distribution of the smoothing variances τ^2 of the spatial effects. In Figures F32 and F33 we show posterior mean estimates of the covariates *region*, *PoE*, *GDPpc* and *unemployment*. In Figure F33 80% and 95% pointwise credible intervals indicate uncertainties of the estimated effects. In each graph effects are centred around zero. Marginal residual plots are shown in Figure F30. Two selected sampling-paths and autocorrelation plots depicted in Figure F31 shall indicate some empirical evidence for the convergence of the Markov chains.

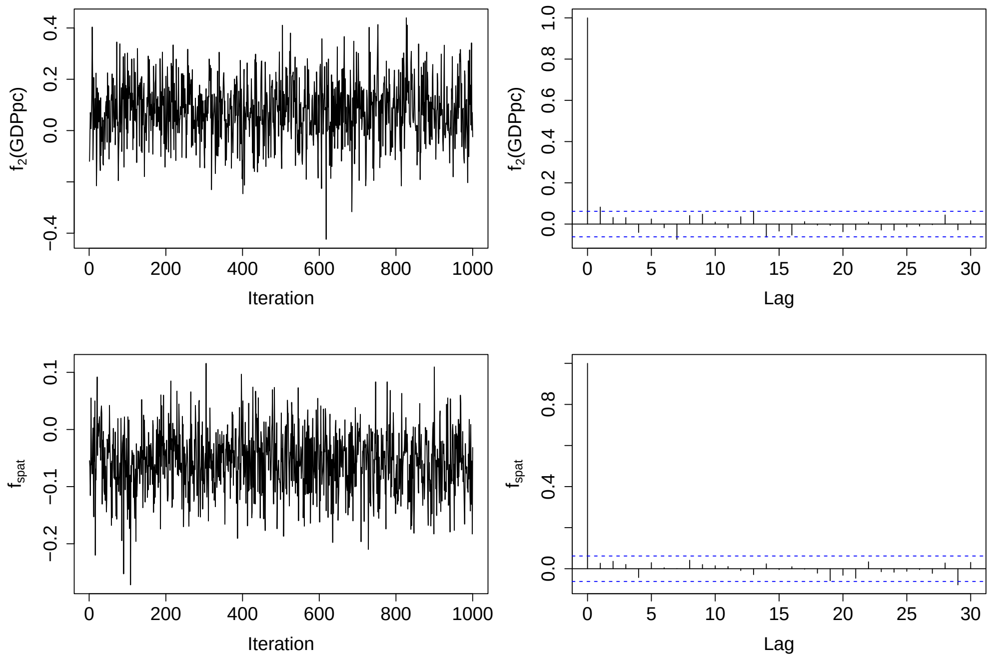

Figure F31: Federal election. Sampling-paths and autocorrelation plots for the coefficient β_1 of the nonlinear function $f_2(GDPpc)$ (first row) and for one coefficient of $f_{spat}^{n_{df}}$ of a randomly selected region (second row). The paths belong to the effects on the predictor for the party CDU/CSU

References

Bundesinstitut fuer Bau, S.-u. R. B. (2002). Regierungsbezirke, GeoBasis-BKG.

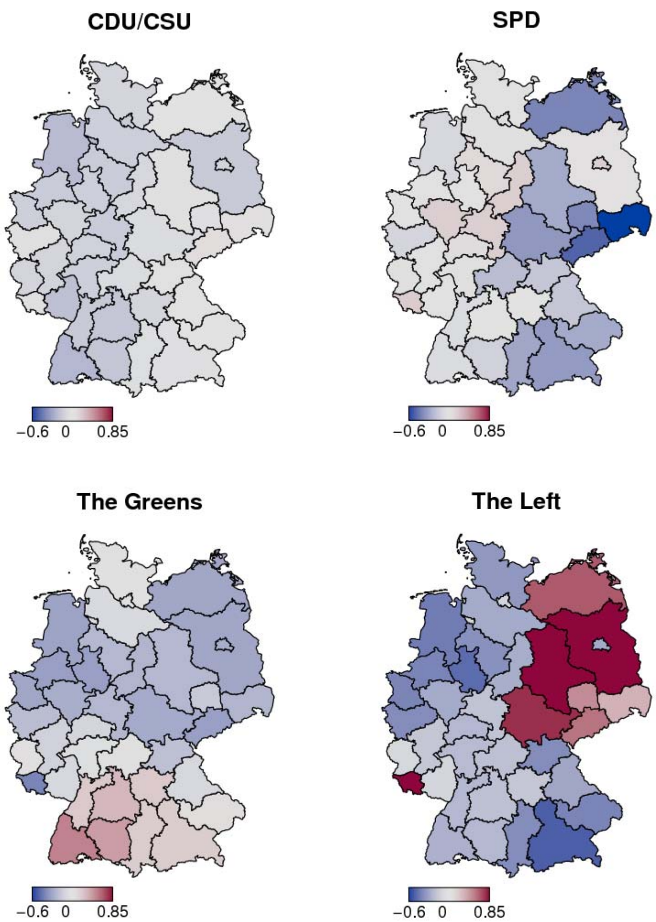

Figure F32: Federal election. Posterior mean estimates of spatial effects. The effects are centred around zero. Note that the plot range is not centred around zero for reasons of visibility. Geometric information provided by Bundesinstitut fuer Bau (2002)

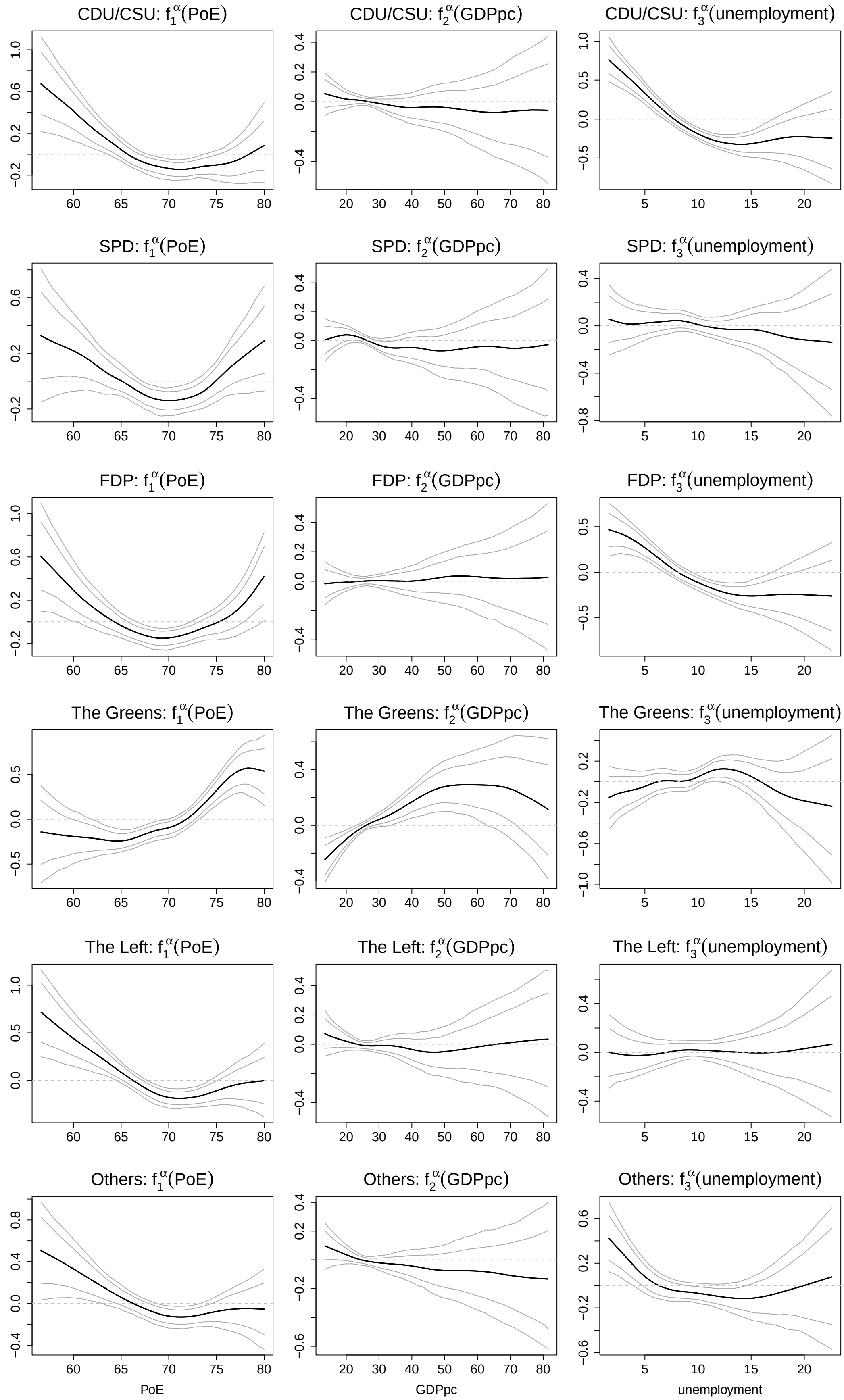

Figure F33: Federal election. Posterior mean estimates of nonparametric effects together with 80% and 95% pointwise credible intervals. The effects are centred around zero.