Supplementary Material for: "An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails"

M. de Carvalho[†], S. Pereira[‡], P. Pereira[*], and P. de Zea Bermudez[‡]

[†] *School of Mathematics, University of Edinburgh, UK (miguel.decarvalho@ed.ac.uk)*

[‡] *Faculdade de Ciências and CEAUL, Universidade de Lisboa, Portugal (sapereira@fc.ul.pt)*

[*] *ESTSetúbal/IPS and CEAUL, Portugal (paula.pereira@estsetubal.ips.pt)*

1 JAGS CODE

```
model{
  # Priors
   betak[1]~dnorm(0,0.001)
   betanu[1]~dnorm(0,0.001)
   betaxi[1]~dnorm(0,0.001)
   for(j in 2:p) {
      betak[j]~ddexp(0,lambdak)
      betanu[j]~ddexp(0,lambdanu)
      betaxi[j]~ddexp(0,lambdaxi)
   }
   lambdak ~ dgamma(0.1,0.1)
   lambdanu ~ dgamma(0.1,0.1)
   lambdaxi ~ dgamma(0.1,0.1)
   # Likelihood
   for(i in 1:n){
      spy[i] <-((1/sigma[i])*((1+xi[i]*y[i]/sigma[i])^(-1 /xi[i]-1)) *
                k[i]*(1-(1+xi[i]*y[i]/sigma[i])^(-1/ xi[i]))^(k[i] - 1))/C
      ones[i]~dbern(spy[i])
      log(k[i]) <-inprod(X[i,],betak[])
      log(nu[i]) <-inprod(X[i,],betanu[])
      log(xi[i]) <-inprod(X[i,], betaxi[])
      sigma[i] <- nu[i]/(1+xi[i])
   }
}
```

2 SUPPLEMENTARY NUMERICAL RESULTS

2.1 ADDITIONAL REPORTS ON SIMULATION STUDY

In this section we report further evidence supporting the Monte Carlo simulation study in the paper. Specifically, we report the conditional quantiles (0.025, 0.975) (see Figs. 1–4) as well as a table on the frequency of variable selection (cf Table 1). In addition, we also report here the mean integrated squared error (MISE) for Scenarios 2–4; see Fig. 5. A similar chart for Scenario 1 is already provided in the main paper (see Fig. 4).

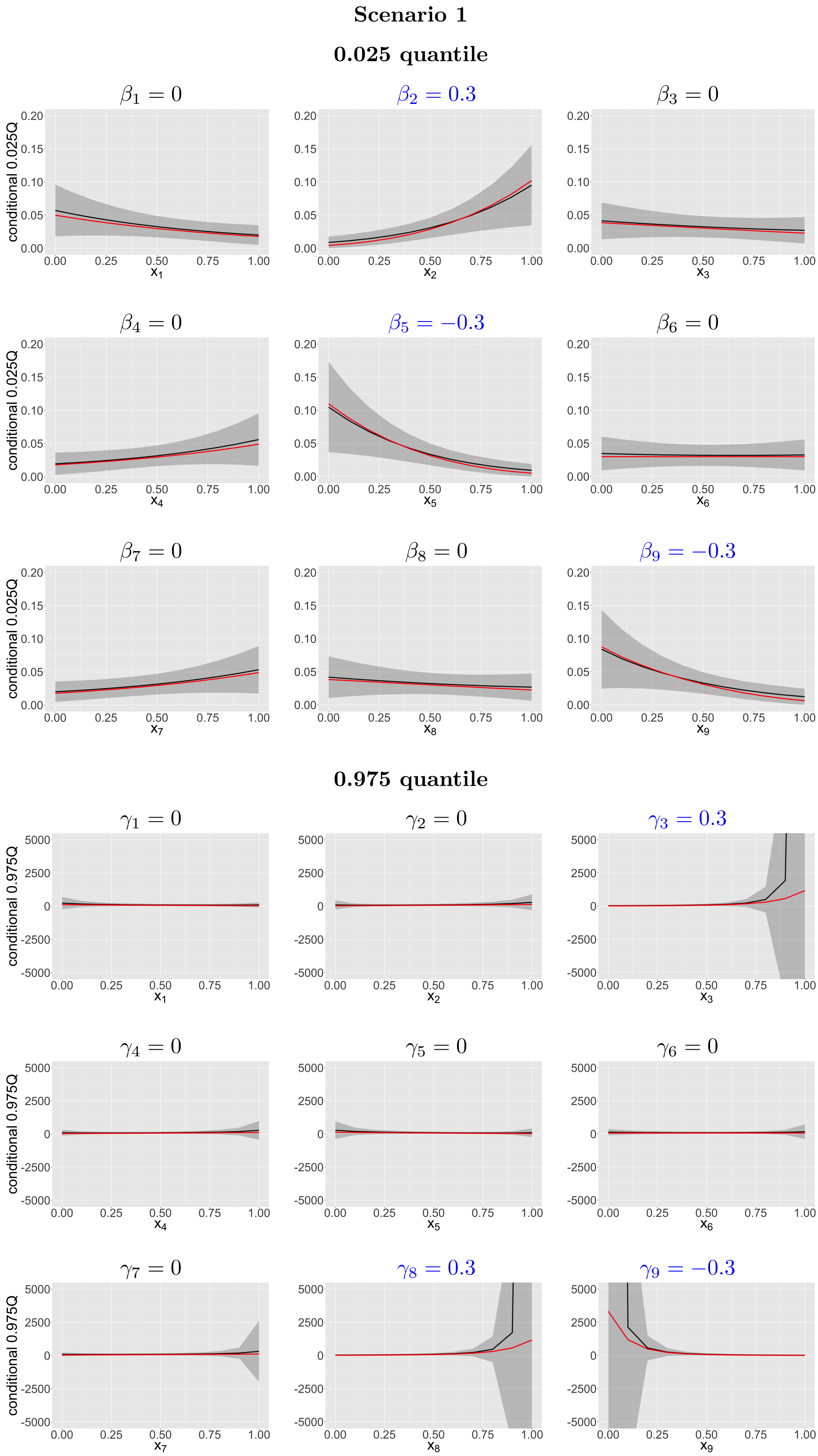

Figure 1: Monte Carlo mean of posterior mean (0.025, 0.975) quantile estimates (black) along with pointwise 95% Monte Carlo bands, conditioning on a vector whose components are zeros except at the first and dth components (i.e. $(1, x_1, 0, \ldots, 0)^\top$), $(1, 0, x_2, \ldots, 0)^\top$, and so on). The true lines are depicted in red.

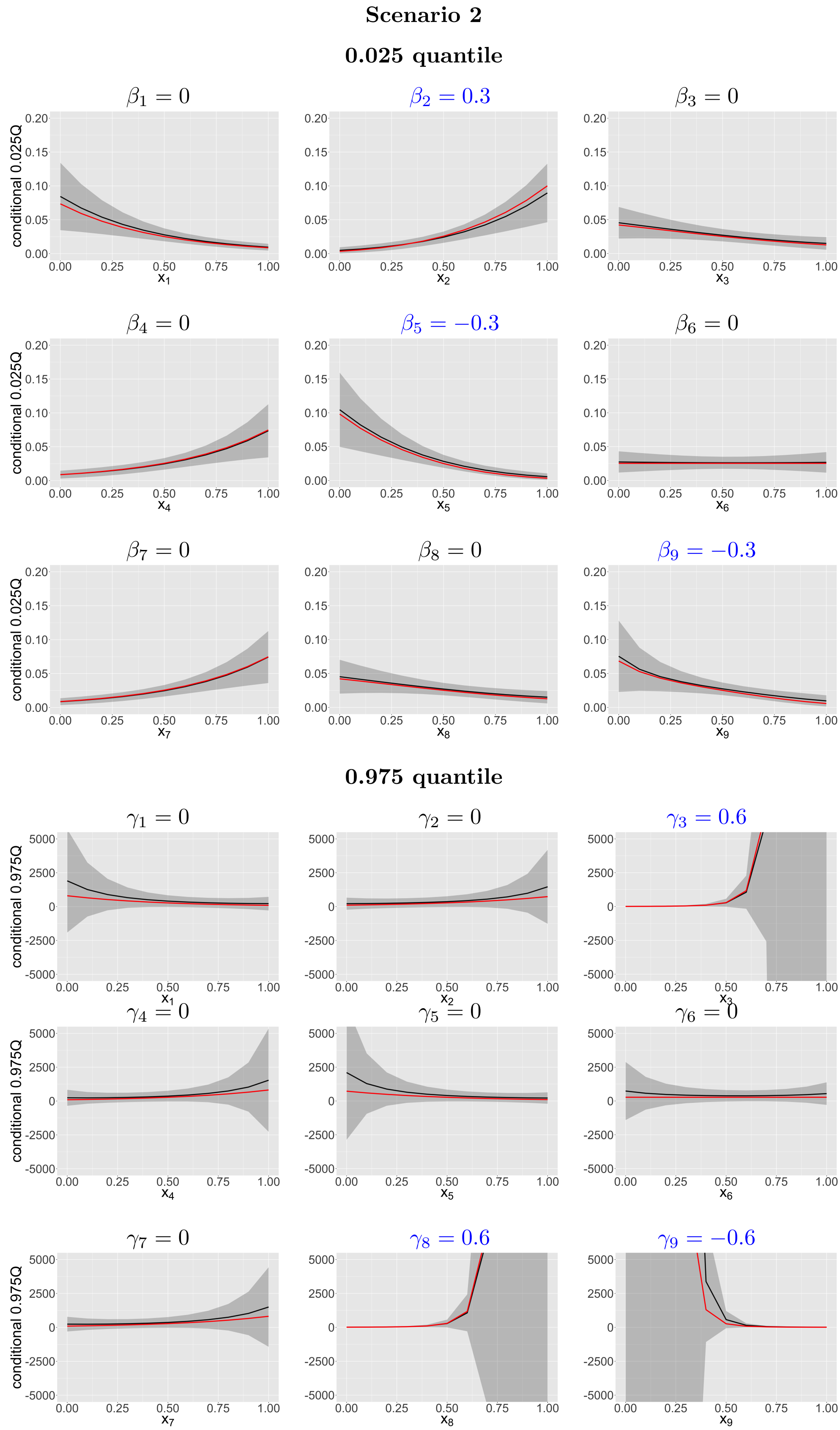

Figure 2: Monte Carlo mean of posterior mean (0.025, 0.975) quantile estimates (black) along with pointwise 95% Monte Carlo bands, conditioning on a vector whose components are zeros except at the first and dth components (i.e. $(1, x_1, 0, \ldots, 0)^\top$), $(1, 0, x_2, \ldots, 0)^\top$, and so on). The true lines are depicted in red.

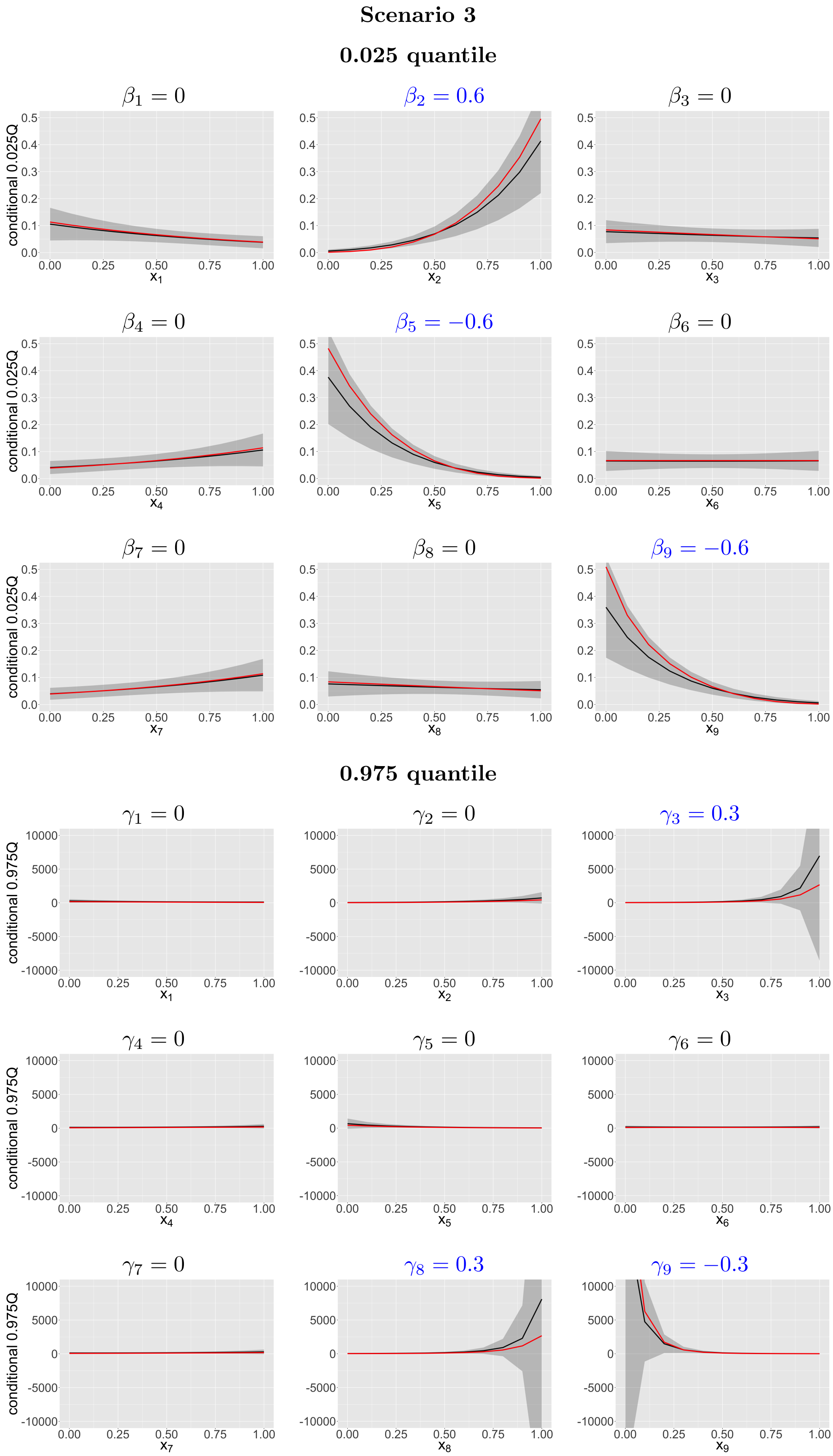

Figure 3: Monte Carlo mean of posterior mean (0.025, 0.975) quantile estimates (black) along with pointwise 95% Monte Carlo bands, conditioning on a vector whose components are zeros except at the first and dth components (i.e. $(1, x_1, 0, \ldots, 0)^{\top}$), $(1, 0, x_2, \ldots, 0)^{\top}$, and so on). The true lines are depicted in red.

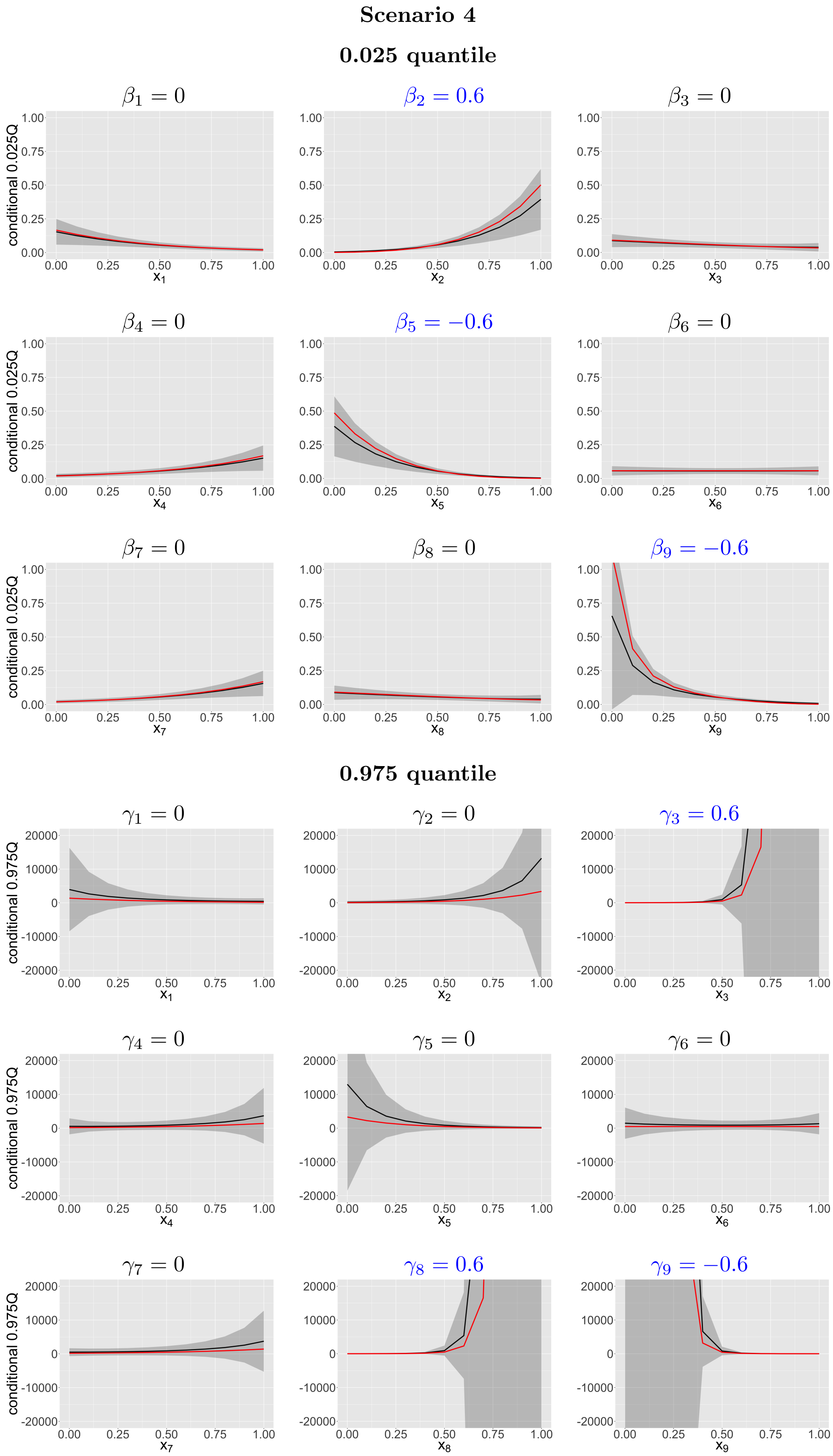

Figure 4: Monte Carlo mean of posterior mean (0.025, 0.975) quantile estimates (black) along with pointwise 95% Monte Carlo bands, conditioning on a vector whose components are zeros except at the first and dth components (i.e. $(1, x_1, 0, \ldots, 0)^\top$), $(1, 0, x_2, \ldots, 0)^\top$, and so on). The true lines are depicted in red.

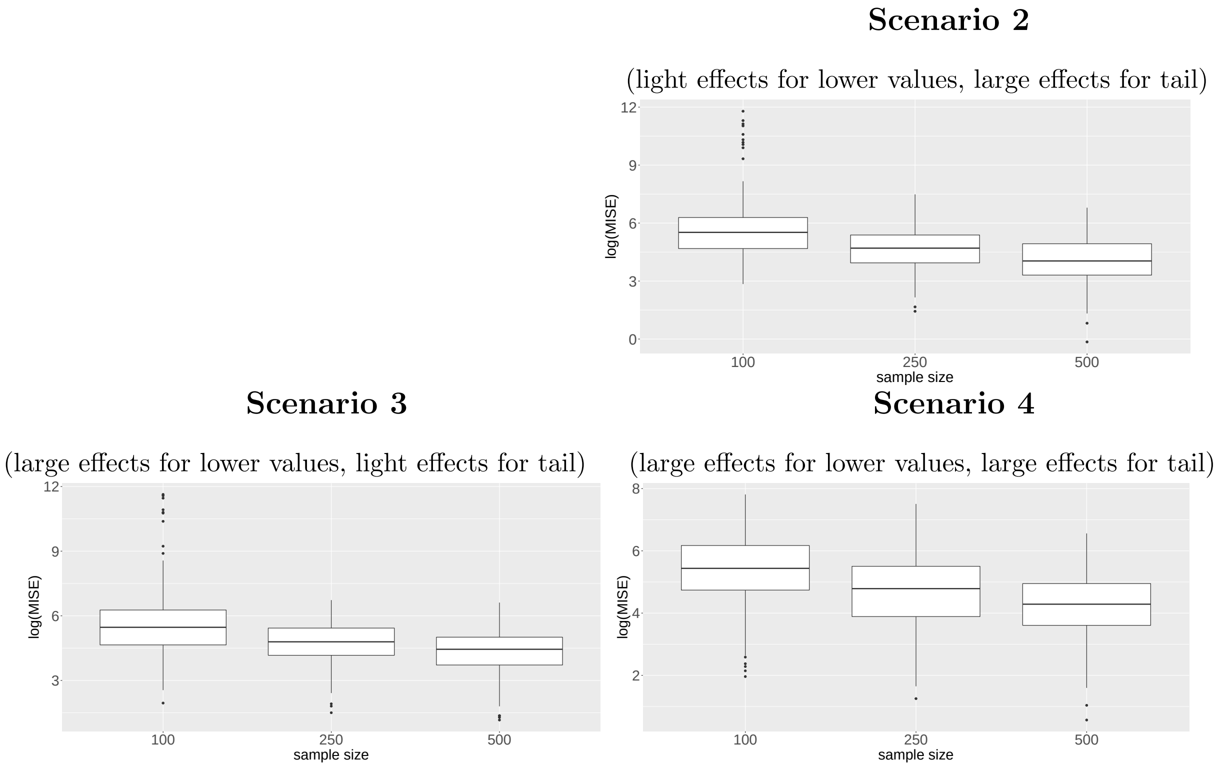

Figure 5: Side-by-side boxplots of MISE on the log scale for Monte Carlo simulation study for Scenarios 2–4.

Scenario 1	$\beta_1 = 0$	$\beta_2 = 0.3$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = -0.3$	$\beta_6 = 0$	$\beta_7 = 0$	$\beta_8 = 0$	$\beta_9 = -0.3$
	6	127	1	6	116	0	6	0	108
	$\alpha_1 = -0.3$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0.3$	$\alpha_5 = 0$	$\alpha_6 = 0$	$\alpha_7 = 0.3$	$\alpha_8 = 0$	$\alpha_9 = 0$
	7	4	0	15	2	0	6	0	6
	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_3 = 0.3$	$\gamma_4 = 0$	$\gamma_5 = 0$	$\gamma_6 = 0$	$\gamma_7 = 0$	$\gamma_8 = 0.3$	$\gamma_9 = -0.3$
	2	4	184	8	8	4	3	183	151
Scenario 2	$\beta_1 = 0$	$\beta_2 = 0.3$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = -0.3$	$\beta_6 = 0$	$\beta_7 = 0$	$\beta_8 = 0$	$\beta_9 = -0.3$
	14	233	6	5	234	0	6	3	235
	$\alpha_1 = -0.6$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0.6$	$\alpha_5 = 0$	$\alpha_6 = 0$	$\alpha_7 = 0.6$	$\alpha_8 = 0$	$\alpha_9 = 0$
	229	12	5	241	5	2	242	1	6
	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_3 = 0.6$	$\gamma_4 = 0$	$\gamma_5 = 0$	$\gamma_6 = 0$	$\gamma_7 = 0$	$\gamma_8 = 0.6$	$\gamma_9 = -0.6$
	3	5	250	11	15	10	13	250	250
Scenario 3	$\beta_1 = 0$	$\beta_2 = 0.6$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = -0.6$	$\beta_6 = 0$	$\beta_7 = 0$	$\beta_8 = 0$	$\beta_9 = -0.6$
	52	239	30	38	238	32	49	27	237
	$\alpha_1 = -0.3$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0.3$	$\alpha_5 = 0$	$\alpha_6 = 0$	$\alpha_7 = 0.3$	$\alpha_8 = 0$	$\alpha_9 = 0$
	84	72	11	88	60	17	74	22	60
	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_3 = 0.3$	$\gamma_4 = 0$	$\gamma_5 = 0$	$\gamma_6 = 0$	$\gamma_7 = 0$	$\gamma_8 = 0.3$	$\gamma_9 = -0.3$
	2	5	249	8	11	6	9	250	245
Scenario 4	$\beta_1 = 0$	$\beta_2 = 0.6$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = -0.6$	$\beta_6 = 0$	$\beta_7 = 0$	$\beta_8 = 0$	$\beta_9 = -0.6$
	50	244	46	42	244	36	49	46	244
	$\alpha_1 = -0.6$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0.6$	$\alpha_5 = 0$	$\alpha_6 = 0$	$\alpha_7 = 0.6$	$\alpha_8 = 0$	$\alpha_9 = 0$
	235	78	23	240	72	21	245	26	40
	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_3 = 0.6$	$\gamma_4 = 0$	$\gamma_5 = 0$	$\gamma_6 = 0$	$\gamma_7 = 0$	$\gamma_8 = 0.6$	$\gamma_9 = -0.6$
	6	15	250	11	22	11	16	250	250

Table 1: Frequency of variable selection for $M = 500$ Monte Carlo repetitions and $n = 500$.

2.2 SIMULATION STUDY UNDER MISSPECIFICATION

In this section we assess the performance of the proposed methods under misspecification. Data ($n = 250$) are simulated from

$$Y \mid \mathbf{X} = \mathbf{x} \sim \text{Burr}(c(\mathbf{x}), k), \quad c(\mathbf{x}) = \exp(-\mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma}), \quad k = 1,$$

with $\gamma = (0.6, 0, 0.6, 0, 0, -0.6, 0, 0, 0, -0.6)^{\mathrm{T}}$, for all $\mathbf{x}$. Reference limiting 'true values' of $\kappa(\mathbf{x})$ and $\xi(\mathbf{x})$ can be respectively derived through the study of the limits $\lim_{y\to 0} F(y|\mathbf{x})/y^{\kappa(\mathbf{x})}$ and $\lim_{y\to\infty}\{1 - F(y|\mathbf{x})\}/y^{-1/\xi(\mathbf{x})}$, where $F(y|\mathbf{x})$ is the conditional distribution function of the Burr$(c(\mathbf{x}), k)$ model; such reference true values are respectively given by $\kappa(\mathbf{x}) = c(\mathbf{x}) = \exp(-\mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma})$ and $\xi(\mathbf{x}) = 1/\{c(\mathbf{x})k\} = \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma})$. We consider again $M = 500$ Monte Carlo replicates; we fit the following version of the proposed model, with power carrier $G_{\kappa(\mathbf{x})}(v) = v^{\kappa(\mathbf{x})}$, and with link functions

$$\kappa_{\mathbf{x}} = \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}), \quad \nu_{\mathbf{x}} = \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}), \quad \xi_{\mathbf{x}} = \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\gamma}).$$

This misspecified setting still allows for direct comparison of regression coefficients in the left and right tails, and as it can be seen from Fig. 6 the proposed methods are able to learn about the coefficients for $\kappa(\mathbf{x})$ and $\xi(\mathbf{x})$ under this misspecified setting. We now show that the regression coefficients of $\kappa(\mathbf{x})$ still track the effect of covariates on lower values of the response. The results of the conditional 0.025 quantile function are reported in Fig. 7, whereas the corresponding estimates of the regression coefficients of $\kappa(\mathbf{x})$ are reported in Fig. 6. The joint analysis of the latter figures, reveals that the effect of covariates on lower values is well tracked by the regression coefficients of $\kappa(\mathbf{x})$, despite the fact that misspecification induces some bias on the conditional quantile function. In addition, it can be seen from Figs. 6–7 that the effect of covariates for the lower values is actually the opposite of that for the tail in terms of sign.

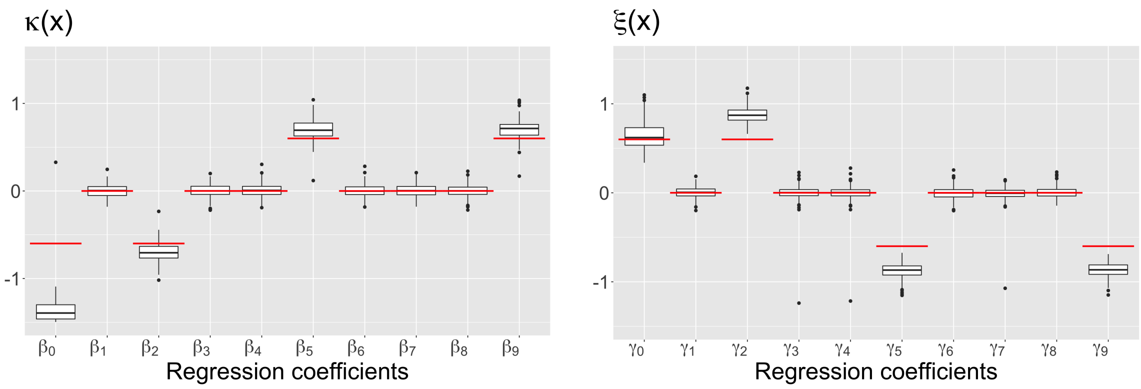

Figure 6: Side-by-side boxplots with regression coefficient estimates for Monte Carlo simulation study plotted against the true values (—).

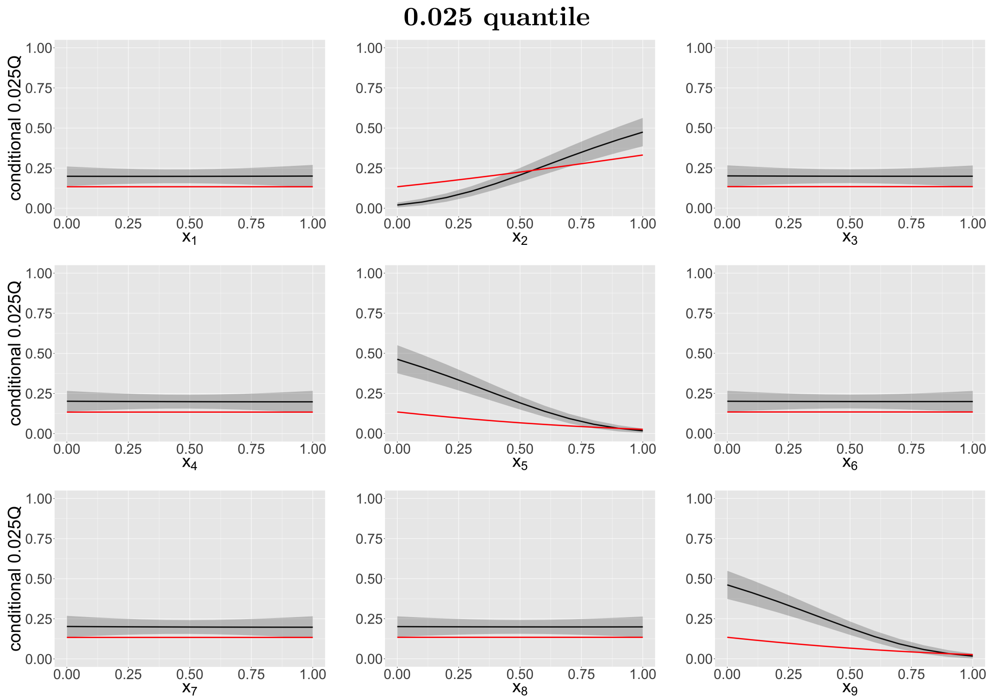

Figure 7: Monte Carlo posterior mean 0.025 quantile estimates (black) along with 95% Monte Carlo bands plotted against the true (red), conditioning on a vector whose components are zeros except at first and dth components (i.e. $(1, x_1, 0, \ldots, 0)^\top$), $(1, 0, x_2, \ldots, 0)^\top$, and so on).

2.3 SIMULATION RESULTS FOR A SIMPLIFIED MODEL

In this section we analyze numerically a simplified version of the proposed model where $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$ all have the same global shrinkage parameter. Figs. 8–10 should thus be compared with Figs. 1–3 in the paper. The upshot from the simulation experiments reported below, for this version of the model, is essentially the same as that from Section 3 of the paper.

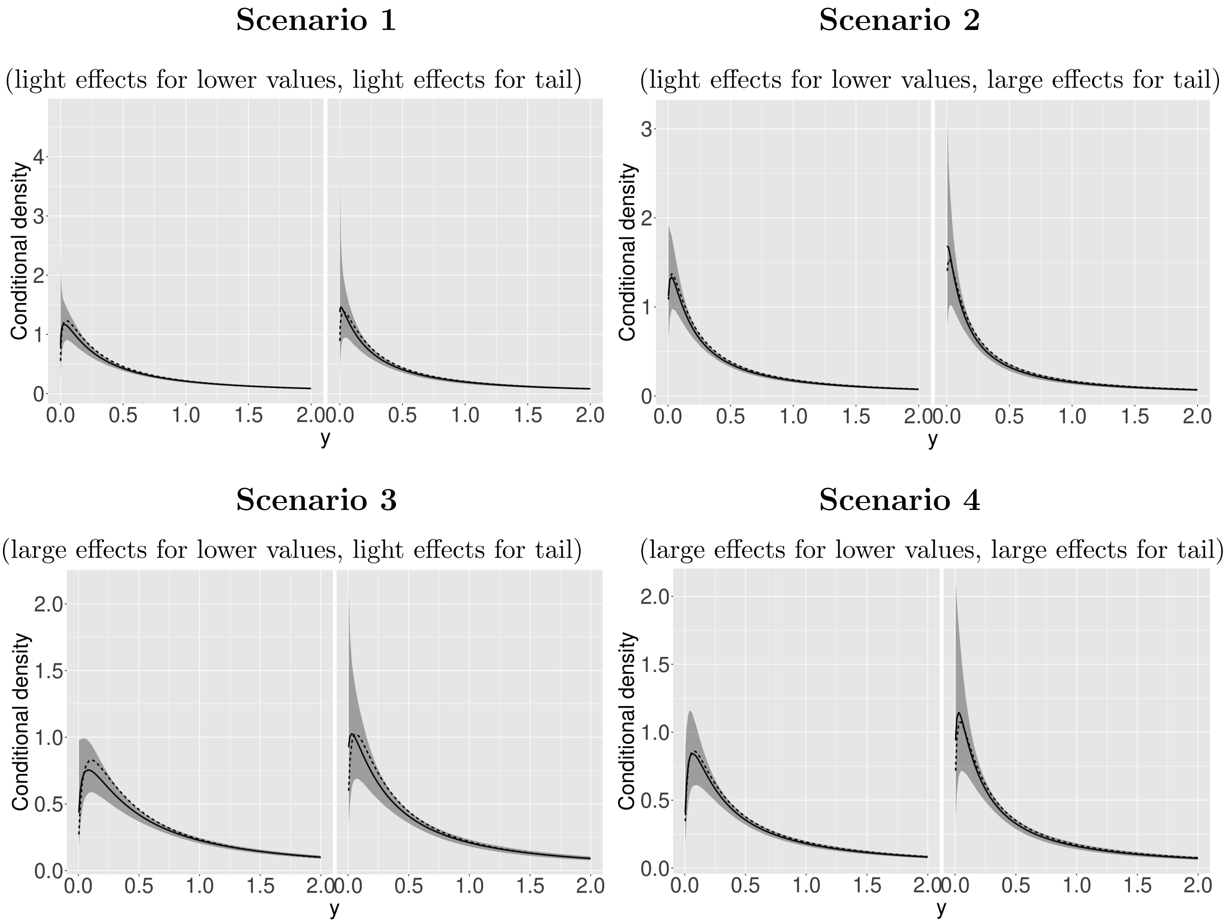

Figure 8: Cross sections of posterior mean conditional density (solid) along with pointwise credible bands against true (dashed) for a one-shot experiment with $n = 500$; the cross sections result from conditioning on $\mathbf{x} = (0.25, \ldots, 0.25)$ (left) and $\mathbf{x} = (0.50, \ldots, 0.50)$ (right).

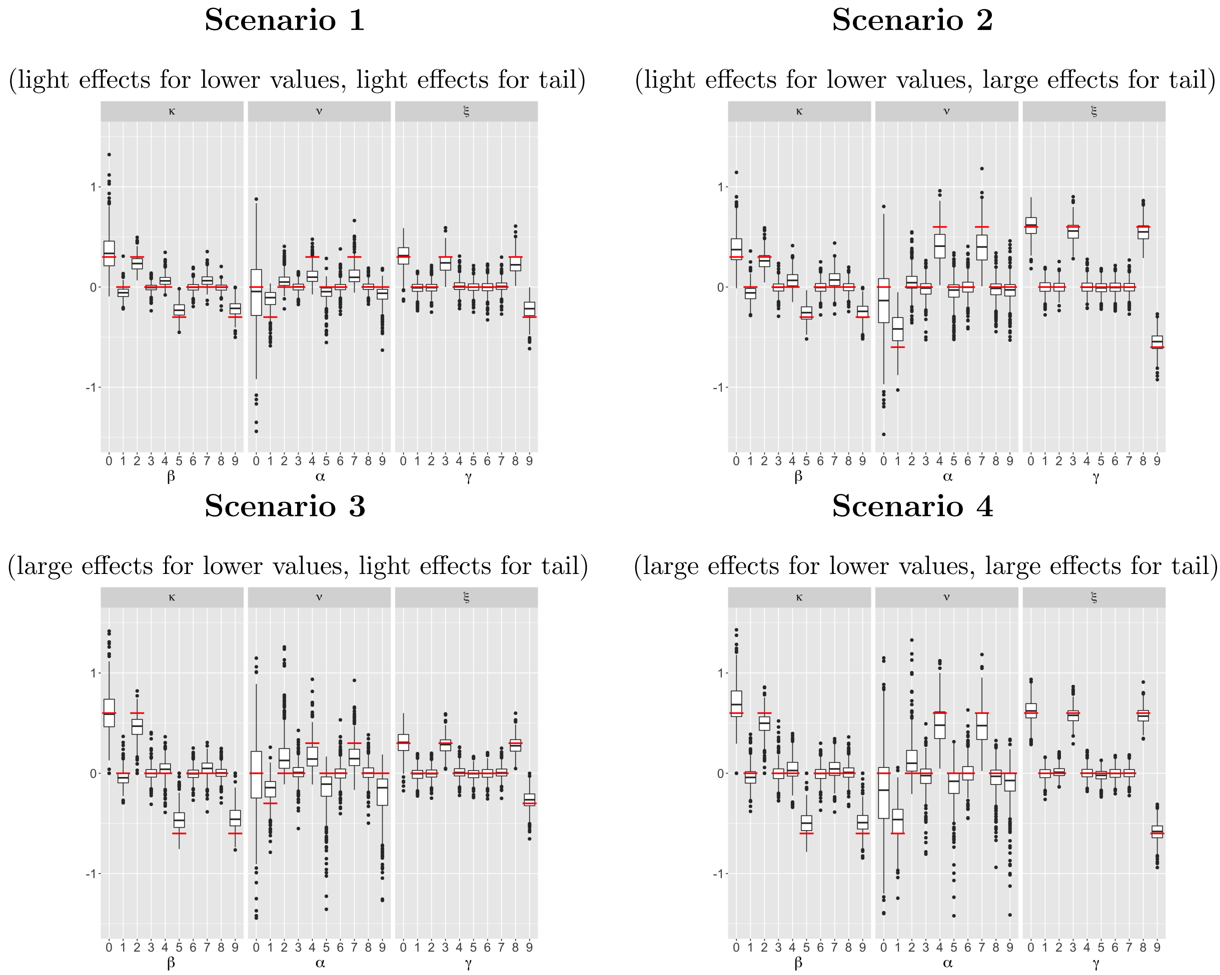

Figure 9: Side-by-side boxplots with regression coefficient estimates for Monte Carlo simulation study ($n = 250$) plotted against the true values (—). The values 0–9 represent the coefficient indices.

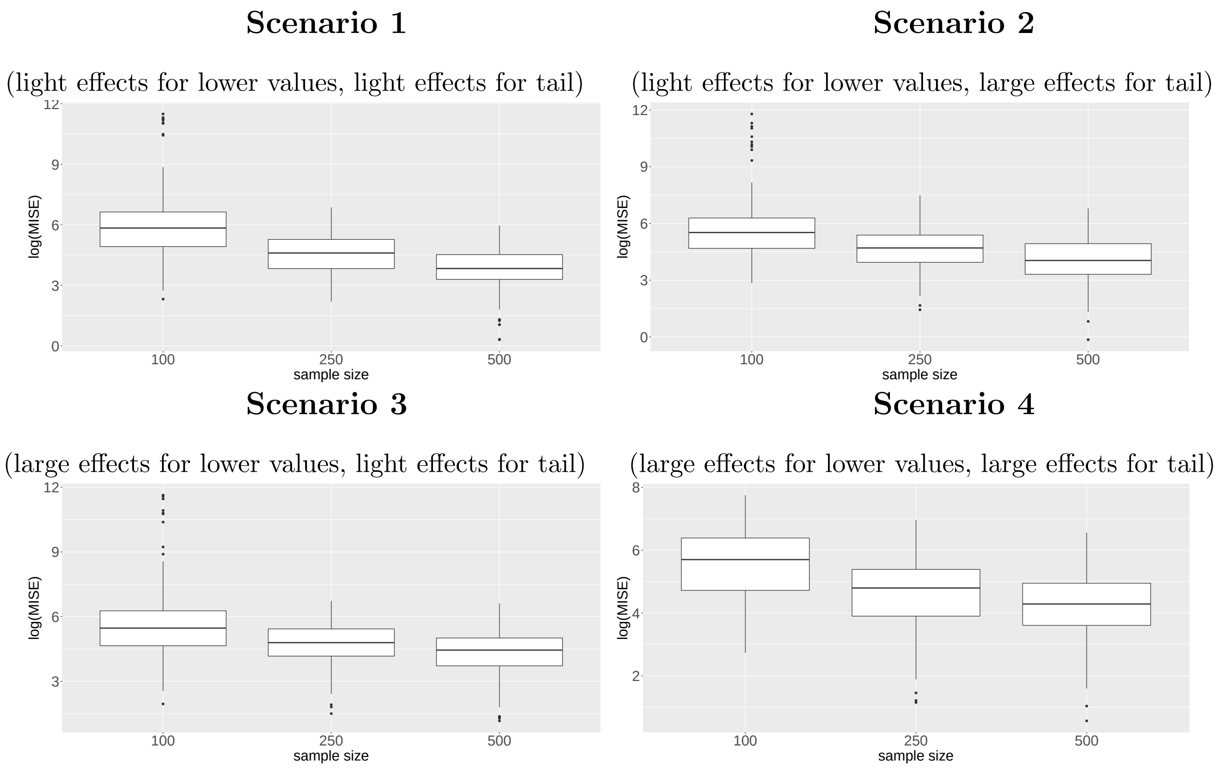

Figure 10: Side-by-side boxplots of MISE on the log scale for Monte Carlo simulation study for Scenarios 1–4.

3 SUPPLEMENTARY EMPIRICAL RESULTS

This section presents some supplements for the data analysis from the paper. In Fig. 11 we present conditional quantiles (0.025, 0.50, 0.975) so to directly assess how rainfall itself is impacted by all potential drivers examined in the paper, namely: Atlantic multi-decadal Oscillation (AMO), El Niño-Southern Oscillation (expressed by NINO34 index) (ENSO), North Pacific Index (NP), Pacific Decadal Oscillation (PDO), Southern Oscillation Index (SOI), and North Atlantic Oscillation (NAO).

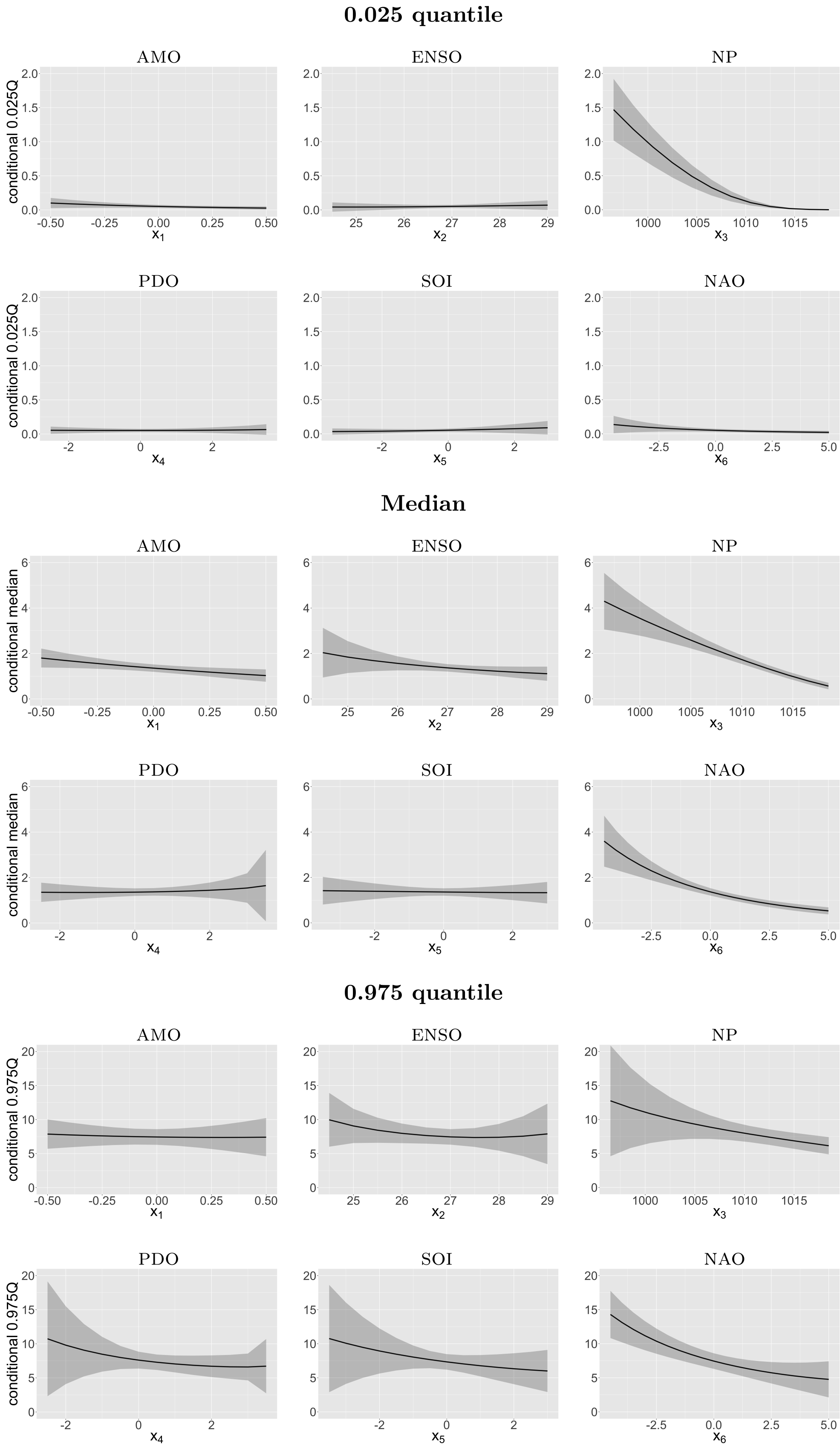

Figure 11: Posterior mean conditional (0.025, 0.5, 0.975) quantile estimates (black) along with credible bands, conditioning on a vector whose components are zeros except at the first and dth components (i.e. $(1, x_1, 0, \dots, 0)^\top), (1, 0, x_2, \dots, 0)^\top$, and so on).