Supplemental material for the paper: Penalized complexity priors for degrees of freedom in Bayesian P-splines

Massimo Ventrucci [1] and Håvard Rue [2]

[1] Department of Statistical Sciences, University of Bologna, Italy.

[2] Department of Mathematical Sciences, NTNU, Norway.

Address for correspondence: Massimo Ventrucci, Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126, Bologna, Italy.

E-mail: `massimo.ventrucci@unibo.it`.

Phone: (+39) 051 209 4657.

Fax: (+39) 051 208 6242.

Abstract:

Key words:

1 Simulation study: all results

Figures from 1 to 4 show results from all scenarios of the simulation study described in section 6 of our paper.

2 Details on implementing the PC prior for τ_β in INLA

The PC prior for the precision of a GMRF (Simpson et al., 2014) can be implemented in INLA Rue et al. (2009). The SD prior by Klein and Kneib (2015) is the PC prior for the IGMRF precision τ_β (see equation 2.1 in the paper), thus it can also be implemented within INLA. The function `get_theta()` in the `sdPrior` `R` package (Klein, 2015) computes the scaling parameter (here denoted as θ') of the PC prior for the variance $\sigma^2_\beta = \tau^{-1}_\beta$. Recall, the PC prior for σ^2_β is a Weibull with shape $\frac{1}{2}$ and scaling parameter θ' (Klein and Kneib, 2015),

$$\pi_{PC}(\sigma^2_\beta) = \frac{1}{2\theta'}\left(\frac{\sigma^2_\beta}{\theta'}\right)^{-\frac{1}{2}} \exp\left(-\left(\frac{\sigma^2_\beta}{\theta'}\right)^{\frac{1}{2}}\right)$$

As a note, in our simulation setting, we computed θ' corresponding to α and c (the parameter of the SD prior) for both design $K = 20$ and $K = 30$, using $n = 20$. In our experience, using $n = 50$ and $K = 30$, computation with `get_theta()` was somehow slow, thus we decided to compute θ' for both designs ($K = 20$ and $K = 30$) by setting $n = 20$. As pointed out by Klein and Kneib (2015), computation of θ' is stable for changing n (where n is the number of evaluation points over the covariate domain), meaning that θ' does not vary substantially for $n > 1$ and it would be sufficient to use a B-spline basis matrix with only one row, i.e. $n = 1$, as an input for `get_theta()`. Klein and Kneib (2015) also report exemplary values for $c = 3$ computed with $n = 1$, using a B-spline design with 22 inner knots and cubic B-splines.

The PC prior for the precision τ_β is a Gumbel$(1/2, \theta)$ type 2 distribution:

$$\pi_{PC}(\tau_\beta) \quad = \quad \frac{\theta}{2}\tau_\beta^{-3/2} \exp\left(-\theta/\sqrt{\tau_\beta}\right), \tag{2.1}$$

Once θ' is computed (using `get_theta()`), then the scaling parameter of the PC prior (2.1) is obtained as $\theta = \theta'^{-1/2}$. In INLA, the PC prior (2.1) is specified through parameters $\alpha \in (0, 1)$ and $u = -\log(\alpha)/\theta$. The interpretation of (u, α) is that $Pr(\sigma_\beta > u) = \alpha$, $u > 0$, $\alpha \in (0, 1)$.

Therefore, the SD prior with scaling parameter θ' can be implemented in INLA by setting $u = -\log(\alpha)/\theta'^{-1/2}$.

The PC prior for degrees of freedom, $\pi_{PC}(d)$, with parameters α and U (see section 4.1 in our paper) can be implemented in INLA only when τ_ϵ is known. Given $d(\cdot|\tau_\epsilon)$, the mapping conditional on the noise precision τ_ϵ, to implement the induced PC prior for degrees of freedom we need to set $u = 1/\sqrt{d^{-1}(U|\tau_\epsilon)}$. Finally, given an SD prior with parameters c and α (and given the associated scaling parameter θ'), one can compute the implied PC prior for degrees of freedom, conditional on τ_ϵ. This prior will have upper bound equal to $U = d(1/u^2|\tau_\epsilon)$, where $u = -\log(\alpha)/\theta'^{-1/2}$ (entries in table 1 were calculated in this way).

3 The joint prior in action with simulated data

We illustrate the joint prior in action with a simulated dataset, from each of the following three scenarios:

- *Scenario 1*: $y_i = 5+x_i+0.2x_i^2+0.3x_i^3+\epsilon_i$, with $\epsilon_i \sim N(0, 0.01^{-1})$ and $x_i \in (0, 5)$;

B-spline design	High noise ($\tau_\epsilon = 0.25$)			Medium noise ($\tau_\epsilon = 1$)			Low noise ($\tau_\epsilon = 5$)		
	$c = 1.5$	$c = 2$	$c = 3$	$c = 1.5$	$c = 2$	$c = 3$	$c = 1.5$	$c = 2$	$c = 3$
$n = 20$; $K = 20$	2.72	2.99	3.41	3.41	3.78	4.36	4.54	5.08	5.89
$n = 50$; $K = 20$	3.11	3.44	3.95	3.95	4.40	5.10	5.31	5.95	6.90
$n = 20$; $K = 30$	2.70	2.95	3.44	3.39	3.75	4.43	4.54	5.06	6.04
$n = 50$; $K = 30$	3.09	3.40	4.00	3.93	4.37	5.20	5.34	5.98	7.16

Table 1: Implied degrees of freedom, d, for the SD prior. The entries in the table refer to the upper bound, U, for d, obtained by assuming an SD prior with parameters c and $\alpha = 0.01$, in the different simulation scenarios. The computation of U involves the use of the `sdPrior` `R` package (Klein, 2015).

- *Scenario 2*: $y_i = \cos(x_i) + \epsilon_i$, with $\epsilon_i \sim N(0, 1)$ and $x_i \in (0, 2\pi)$;
- *Scenario 3*: $y_i = 2.5 + \sin(x_i) + \epsilon_i$, with $\epsilon_i \sim N(0, 10^{-1})$ and $x_i \in (0, 0.3) \cup (0.7, 1)$

In all scenarios we use a relatively small sample size $n = 50$. In the first and second scenario, covariate values are uniformly scattered over the covariate domain. In the third scenario, the covariate values are not uniform, with a gap of observations in the middle of the domain; the aim is to verify whether the PC prior works as expected in sparse data cases. The simulated datasets and the true curves are reported in Figure 5.

We fit model (4) to (7), see the paper section 5, to each of the three datasets, with $K = 30$ cubic B-splines and an IGMRF of order 2 on $\boldsymbol{\beta}$, using algorithm 1 (see the appendix of the paper). We are interested in assessing the fit for different priors using different upper bounds $U = \{2, 4, 6\}$, while keeping a small probability $\alpha = 0.01$ for the tail event.

Results are reported in Figure 5, with colours indicating the fit associated to the different upper bounds. Setting $U = 2$, with small probability $\alpha = 0.01$, means that any deviation from the base model is hugely penalized at prior. Indeed, this choice forces the fit towards a linear trend (red line) in all three scenarios. This demonstrates that the PC prior works as expected: when the base model is over-weighted and no chance is given to more flexible models, the posterior is dragged towards the base model. Setting $U = 4$ and $U = 6$ gives a more flexible prior, allowing for larger deviations from the base model. In all three panels in Figure 5, the shape of the fitted smooth functions obtained by setting $U = 4$ or $U = 6$ (green and blue lines) resembles the true curves (black dashed lines). In general, if the true model depicts a complete cycle, such as in scenarios 2 and 3, a PC prior with an upper bound of 6 or more degrees of freedom should roughly capture it. Finally, in scenario 3 the smooth function is estimated at unobserved covariate values. The shape of the fit inside the gap depends on the order of the IGMRF prior on $\boldsymbol{\beta}$ (Eilers and Marx, 2010); for instance, a first order IGMRF would attempt to fit a constant inside the gap, while a second order one would fit a line; indeed, in the lower panel of Figure 5, we see a red line across the gap when setting $U = 2$. Basically, the PC prior works as expected, regardless of the presence of large gaps of observations over the covariate domain. The only issue we need to take care about in scenario 3 is that some B-splines might be poorly supported resulting in singularity of $\boldsymbol{B}^{\top}\boldsymbol{B}$; this can be solved by adding a small ϵ to the diagonal of $\boldsymbol{B}^{\top}\boldsymbol{B}$ when computing the degrees of freedom using formula (2.2) in the paper.

References

Eilers, P. and Marx, B. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 637–653.

Klein, N. and Kneib, T. (2015). Scale-dependent priors for variance parameters in structured additive distributional regression. *Bayesian Analysis*. doi: 10.1214/15-BA983. URL http://projecteuclid.org/euclid.ba/1448323525.

Klein, N. (2015). *sdPrior: Scale-Dependent Hyperpriors in Structured Additive Distributional Regression*. URL http://CRAN.R-project.org/package=sdPrior. R package version 0.3.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using inte- grated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71(2)**, 319–392.

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., and Sørbye, S. H. (2014). Penalising model component complexity: A principled, practical approach to constructing priors. *ArXiv e-prints*.

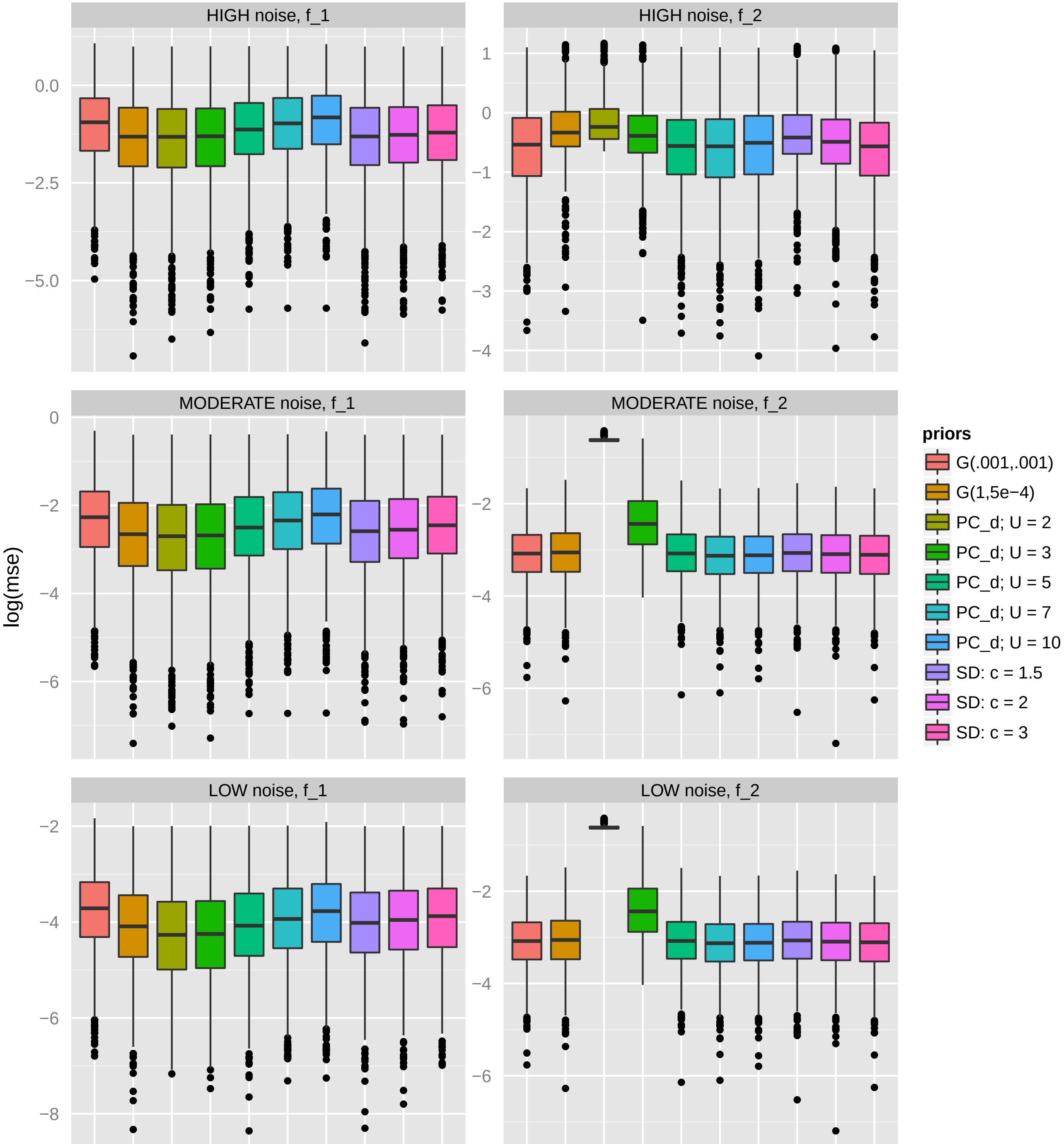

Figure 1: Simulation results: $\log(MSE)$ for f_1 (left panels) and f_2 (right panels), in presence of high noise ($\tau_\epsilon = 0.25$, top panels), moderate noise ($\tau_\epsilon = 1$, middle panels), low noise ($\tau_\epsilon = 5$, bottom panels), sample size **n=20**, **K=20**. In the legend on the right, label "G" indicates the Gamma prior; "PC_d" indicates our PC prior for degrees of freedom (joint prior), with $\alpha = 0.01$ and $U = \{2, 3, 5, 7, 10\}$; "SD" denotes scale dependent prior with $\alpha = 0.01$ and $c = \{1.5, 2, 3\}$.

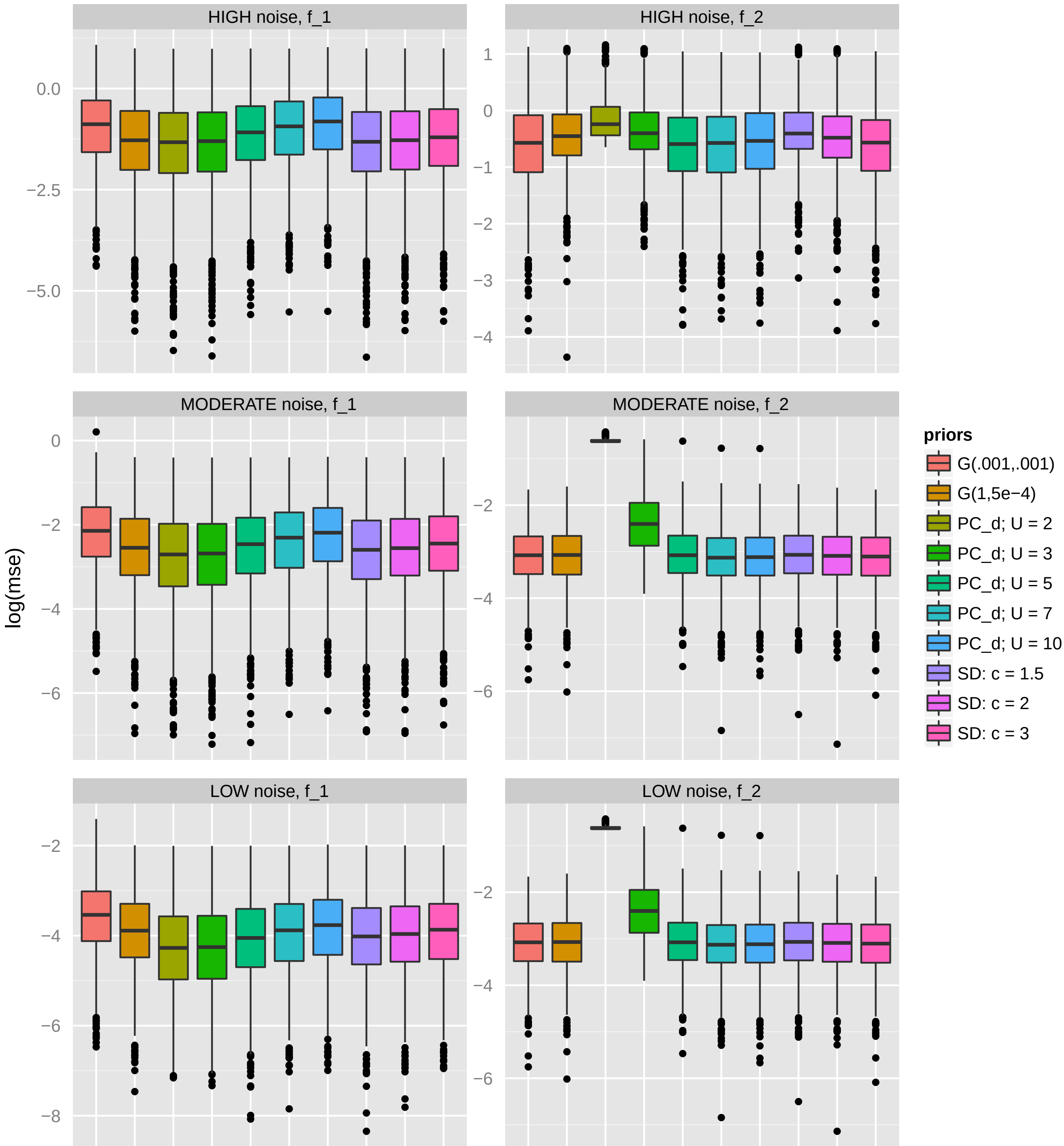

Figure 2: Simulation results: $\log(MSE)$ for f_1 (left panels) and f_2 (right panels), in presence of high noise ($\tau_\epsilon = 0.25$, top panels), moderate noise ($\tau_\epsilon = 1$, middle panels), low noise ($\tau_\epsilon = 5$, bottom panels), sample size **n=20**, **K=30**. In the legend on the right, label "G" indicates the Gamma prior; "PC_d" indicates our PC prior for degrees of freedom (joint prior), with $\alpha = 0.01$ and $U = \{2, 3, 5, 7, 10\}$; "SD" denotes scale dependent prior with $\alpha = 0.01$ and $c = \{1.5, 2, 3\}$.

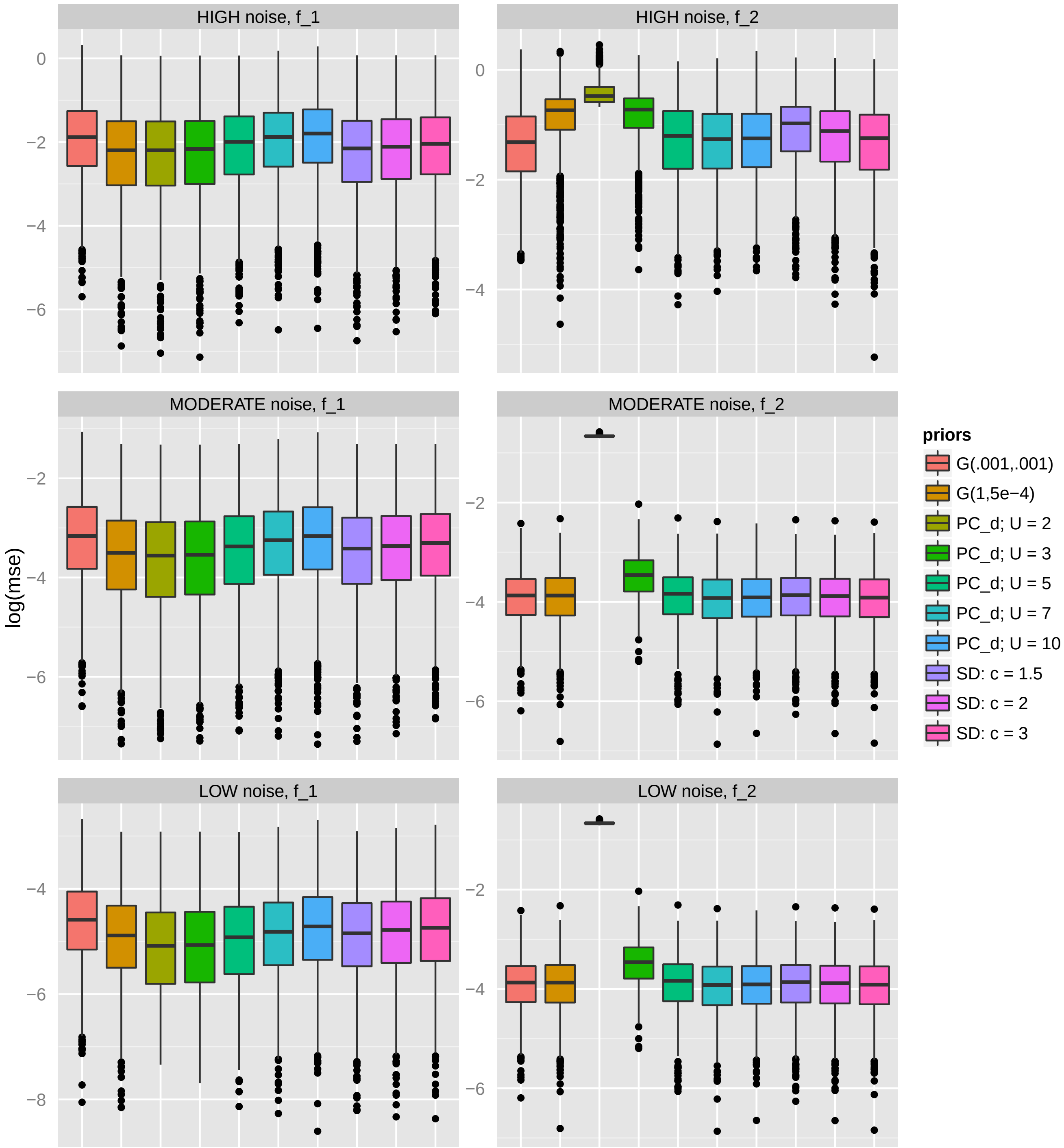

Figure 3: Simulation results: $\log(MSE)$ for f_1 (left panels) and f_2 (right panels), in presence of high noise ($\tau_\epsilon = 0.25$, top panels), moderate noise ($\tau_\epsilon = 1$, middle panels), low noise ($\tau_\epsilon = 5$, bottom panels), sample size **n=50**, **K=20**. In the legend on the right, label "G" indicates the Gamma prior; "PC_d" indicates our PC prior for degrees of freedom (joint prior), with $\alpha = 0.01$ and $U = \{2, 3, 5, 7, 10\}$; "SD" denotes scale dependent prior with $\alpha = 0.01$ and $c = \{1.5, 2, 3\}$.

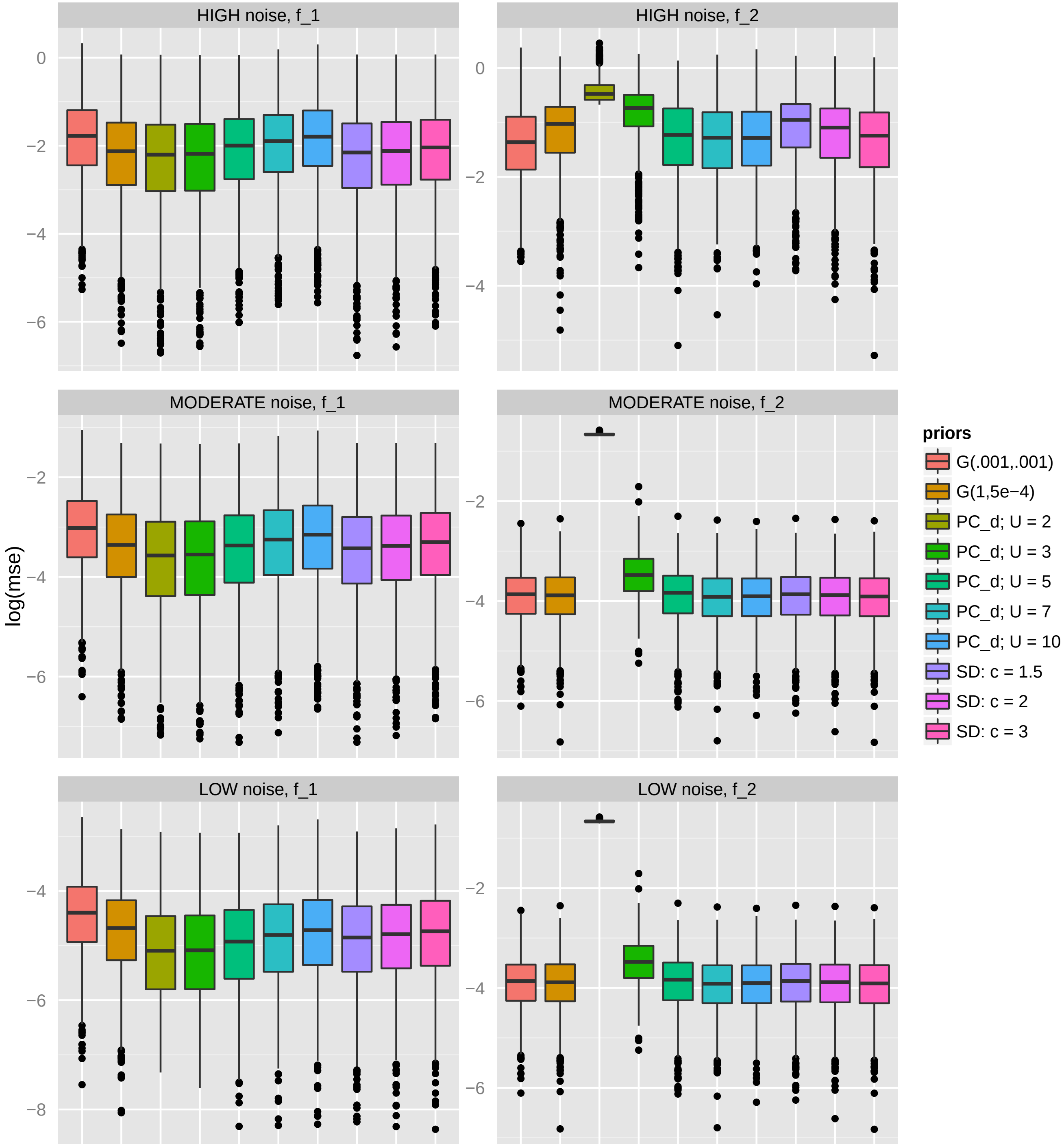

Figure 4: Simulation results: $\log(MSE)$ for f_1 (left panels) and f_2 (right panels), in presence of high noise ($\tau_\epsilon = 0.25$, top panels), moderate noise ($\tau_\epsilon = 1$, middle panels), low noise ($\tau_\epsilon = 5$, bottom panels), sample size **n=50**, **K=30**. In the legend on the right, label "G" indicates the Gamma prior; "PC_d" indicates our PC prior for degrees of freedom (joint prior), with $\alpha = 0.01$ and $U = \{2, 3, 5, 7, 10\}$; "SD" denotes scale dependent prior with $\alpha = 0.01$ and $c = \{1.5, 2, 3\}$.

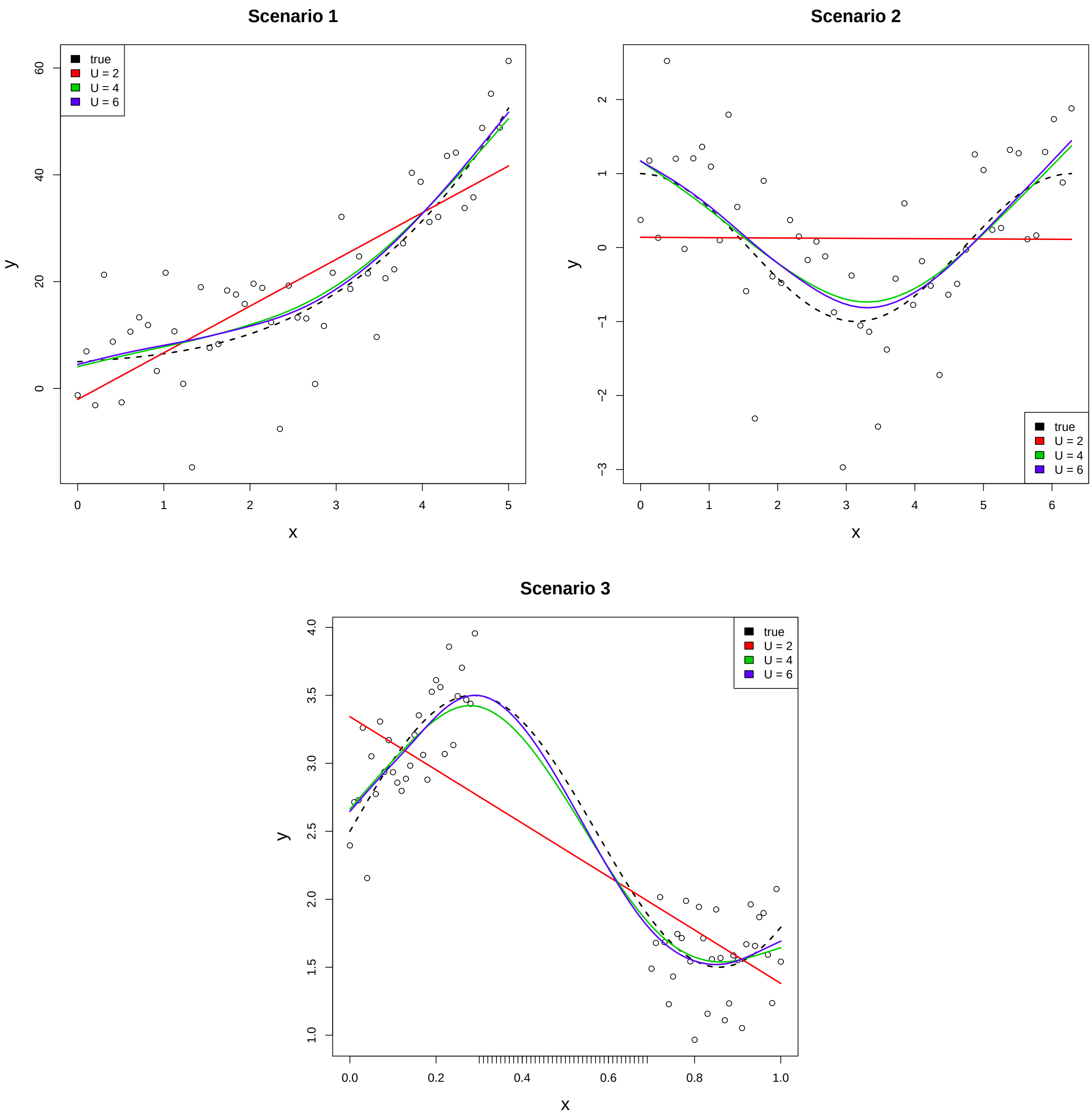

Figure 5: The joint prior in action with simulated data. The three panels report the fit of the P-spline model with the joint prior (as described in section 5 of the paper), with $K = 30$ cubic B-splines and an IGMRF prior of order 2 on $\boldsymbol{\beta}$; the PC prior for degrees of freedom is scaled according to $U = \{2, 4, 6\}$ and $\alpha = 0.01$. In each panel, the fit corresponds to the linear base model when $U = 2$. When the PC prior defines larger upper bounds (e.g., $U = 4$, $U = 6$) the fit resembles quite well the true curve.