Smoothers, Diagnostics, Prior Sensitivity and Model Comparison

1 Splines

All univariate smooth terms **s** use covariates in the zn form derived from population proportions of Census 2021 variables, with the brms default "thin plate regression splines" as in mgcv, with basis dimension 7.[1] The interaction term with the zn forms of IMD and VRx uses the **t2** tensor product of splines with marginal basis dimension 5.[2] Univariate and tensor product splines can be viewed from within the file codeM1.

A univariate smooth term s(x,k=7) is interpreted as the sum of a linear term and 5 splines, each with associated parameters (in mgcv when k=7 the penalty matrix has rank 5 and a 2-dimensional subspace of parameters is unpenalized). Each of the 5 penalized spline parameters is obtained as the product of a single positive sd pertaining to the smooth term, and a standardized coefficient for the particular penalized spline. The default prior for the linear term is flat. The default prior for the sd is a half Student-t, constrained to > 0, with 3 degrees of freedom and scaled by 2.5. Default priors for the standardized coefficients are standard normal.

An interaction term t2(x,y,k=5) is handled similarly, after decomposition into components of dimension 9, 6, and 6.

2 Random Effects, Parametric terms, Special parameter

Likewise, the parameters for the random effects for REG are obtained as the product of a single positive sd and standardized coefficients, with default priors as above. Parameters and default priors for the random effects for LTLA are similar. The parametric terms in the model are transformed by a QR decomposition and given default flat priors. The default prior for the special parameter ϕ is Gamma(0.01,0.01).

3 Sampling and Diagnostics

Sampling used the default settings with default priors, 4 chains of length 1000 after a warmup of 1000, but maximum treedepth was reset from 10 to 13. The output was checked with standard Bayesplot tests including divergence, treedepth, $\hat{R}$ (modified Gelman-Rubin statistic split-$\hat{R}$), ebfmi (estimated bayesian fraction of missing information), posterior predictive density.[3] Bayesian R2 and Mean Absolute Deviation MAD = mean(abs(observed-fitted)) were computed for the fitted model.[4]

4 Prior sensitivity

The default priors in brms could affect the results, however these are completely non-informative for all parametric terms, standard normal for terms which are so defined, and have support on the positive half line $\mathbf{R}^+$ for all the intrinsically positive components: the standard deviations and the special parameter ϕ. Despite the infinite range of support, the priors might put too much weight on too narrow or too wide a range. Sensitivity to the choice of prior was tested first with `priorsense` and then by fitting a new model $\mathcal{M}$1pr which retains the formula of $\mathcal{M}$1 but resets the sd and sds priors to a half-normal with sd=5.

The programme `priorsense` raises the informative priors (individually or in groups) to a power and examines the effect on the posterior distribution of selected parameters.[5, 6] Problems are flagged when the gradient of the Cumulative Jensen-Shannon distance between the base posterior and posteriors resulting from power-scaling exceeds a threshold, set at the default value 0.05. The commented file code_priortest applies `priorsense` using the output of $\mathcal{M}$1.

If all informative priors are scaled simultaneously, no problems are detected in the pointwise fitted values nor in the centred regional and local random effects. The centred regional effects are virtually unchanged and the differences between selected effects remain credible. The same REG:LTLA with the top and bottom 10 centred local effects are identified after simultaneous powerscaling, their numerical effects are virtually unchanged and the differences between top and bottom 10 effects remain credible.

Some other parameters are affected by the simultaneous powerscaling. Considering each of the 19 informative priors separately, problems arise only from powerscaling the prior on the t2 spline interaction term. However, no problems are detected in the pointwise fitted values nor in the centred regional and local random effects when powerscaling the t2 prior, and their numerical values are virtually unchanged and differences remain credible.

A separate test replaced the default scaled half-student (df=3) priors used by $\mathcal{M}$1 for all sd and sds terms which govern the random effects and splines. These defaults were replaced by half-normal priors with sd=5, a wider option over the range (0,15). The file code_M1pr fits the new model $\mathcal{M}$1pr and generates fitted values and Figures 1 to 7 and 8a, 8b and 8c. Whereas $\mathcal{M}$1 showed East Lindsey and Blackpool respectively as the 10th and 11th ranked centred local effects, $\mathcal{M}$1pr shows Blackpool ranked 10th and East Lindsey ranked 11th. The remaining top and bottom 10 centred local effects are unchanged. For the outliers, Salford Central & University is above the chosen probability threshold (0.01) in `lookout` and also appears unexceptional with fit.gpd in `mev`, but the other six outliers are detected by both methods. The various Figures are otherwise unchanged.

5 Model comparison

The MixIS method of model comparison used to generate Table 2 adapts the approach for LOO comparison which is given theoretical justification on the assumption that the true generating process can be modelled with a Dirichlet process incorporating the model covariates and the pointwise elpd values.[7]

The file code_mixcompare conducts a limited simulation with synthetic beta-binomial data (N=1000) including outliers, a factor variable, a parametric term and a smooth term. The true generating process includes a separate term not included in either model used in the comparison, giving rise to the outliers. The true process is used to generate 200 new datasets and compute directly (without MixIS) the predicted pointwise elpd for each new dataset using the two models fitted to the original data. In turn, this gives 200 simulated values of $\tilde{\Delta}$, the difference in the summed predicted elpd for the two models. The sd of these simulated $\tilde{\Delta}$ values is then compared to the sd obtained from the pointwise difference of elpd_mixis from the original data. The simulated $\tilde{\Delta}$ is approximately normal with a mean close to the Δ obtained from the original data and its sd is similar to the sd obtained from the original data, though smaller. This gives some confirmation of Table 2 and suggests that the reported p-values may be conservative.

References

[1] Wood SN. Generalized Additive Models: An Introduction with R. 2nd ed. New York: Chapman and Hall/CRC; 2017.

[2] Wood SN, Scheipl F, Faraway JJ. Straightforward intermediate rank tensor product smoothing in mixed models. Stat Comput. 2012;https://doi.org/10.1007/s11222-012-9314-z.

[3] Gabry J, Mahr T.: bayesplot: Plotting for Bayesian Models. R package version 1.13.0. 2025. https://mc-stan.org/bayesplot/. Accessed 14 Nov 2025.

[4] Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for Bayesian Regression Models. Am Stat. 2019;https://doi.org/10.1080/00031305.2018.1549100.

[5] Kallioinen N, Paananen T, Bürkner PC, Vehtari A. Detecting and diagnosing prior and likelihood sensitivity with power-scaling perturbations. Statistics and Computing. 2023;https://doi.org/10.1007/s11222-023-10366-5.

[6] Kallioinen N, Paananen T, Bürkner PC, Vehtari A, Weber F.: priorsense: Prior Diagnostics and Sensitivity Analysis. R package version 1.2.0 https://cran.r-project.org/package=priorsense. Accessed 13 Apr 2026.

[7] Sivula T, Magnusson M, Matamoros AA, Vehtari A. Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. Bayesian Analysis. 2025;https://doi.org/10.1214/25-BA1569.