Supplementary Material to 'Multiple smoothing parameters selection in additive regression quantiles'

Muggeo V.M.R.,* Torretta F., Eilers P., Sciandra M., Attanasio M.

Dipartimento di Scienze Econom, Az. e Statistiche – Università di Palermo

Erasmus University Medical Centre, Rotterdam, the Netherlands

A Additional simulation results

In this supplementary material we provide further simulation evidence. More specifically, here we report differences in terms of the MIAE (Mean Integrated Absolute Error) for one and two smooth terms from the simulation studies described in section 4 of the main manuscript.

Moreover we discuss further simulation results about comparisons involving the L-curve, 10-folds CV, and boosting. Also we report some evidences about varying coefficient models, choice of the difference order in the penalty, and the extra tuning parameter γ.

A.1 Performances in terms of MIAE

Figures 2 and 3 in the main manuscript contrast the three competitors in terms of Mean Integrated Square Errors (MISE) for homoscedastic and heteroscedastic errors, respectively. Figure 1 below portrays the same comparisons by considering the differences between fitted and true quantile curves in absolute value: Mean Integrated Absolute Error, MIAE. The simulation settings are those described in the main manuscript (section 4).

Focussing on the proposed approach, 'pspline+HFS' performs slightly better with truly linear signal and worse at middle quantiles with strongly nonlinear signal $g_0(x)$. However, as for the MISE, no systematic pattern comes out with overall comparable performances across the scenarios.

*E-mail: vito.muggeo@unipa.it

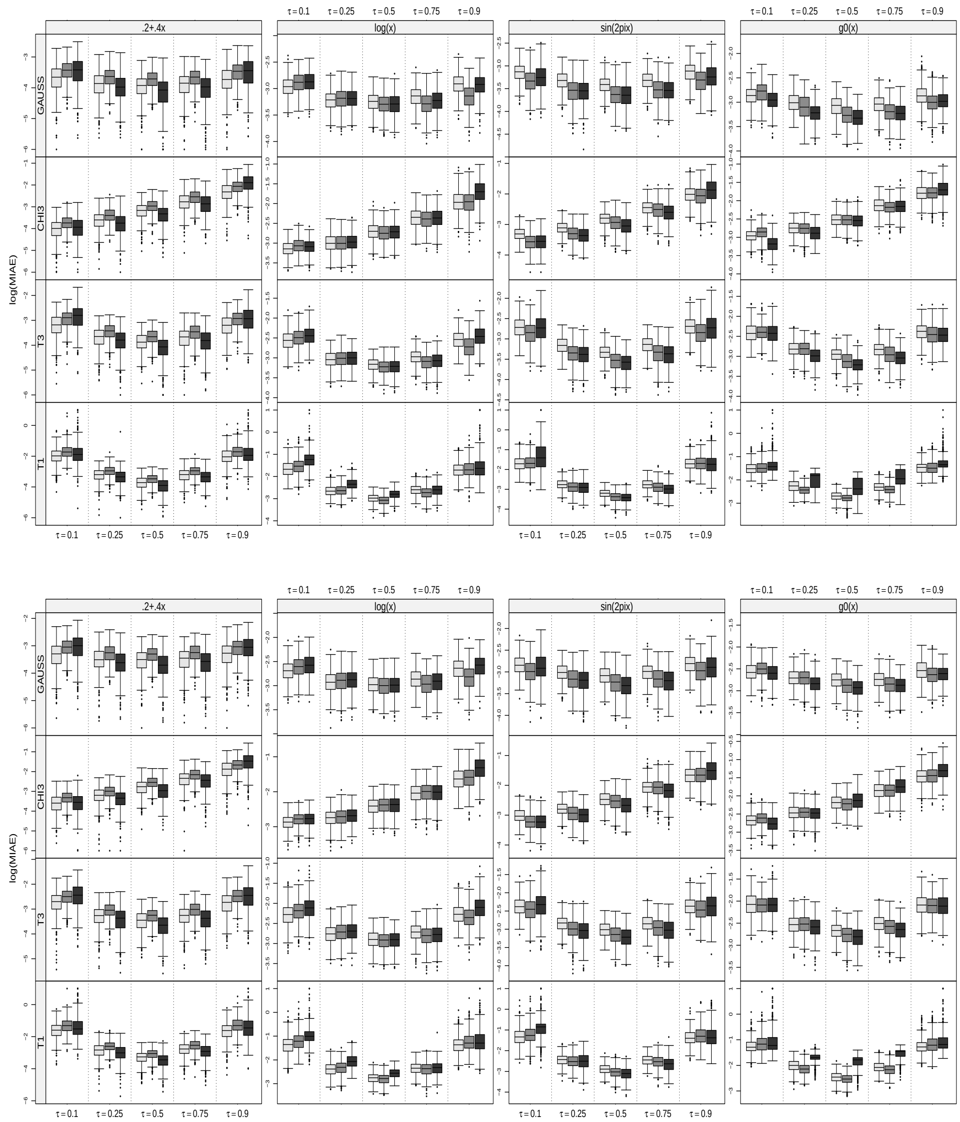

Figure 1: Contrasting the three main competitors for smoothing parameter selection in terms of MIAE (on log scale) by different error distributions and signals: 'sspline+SIC' (light grey box), 'pspline+SIC' (medium grey box) and 'pspline+HFS' (dark grey box). The top plots refer to errors with constant scale $\sigma(x) = 0.2$, the bottom ones to non-constant scale function $\sigma(x) = 0.2(1 + x)$.

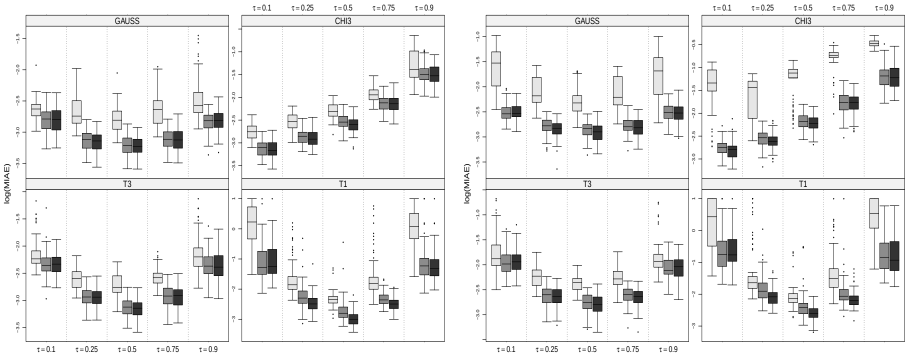

Figure 2: Contrasting the three main competitors (in terms of log(MIAE)) for smoothing parameter selection with 2 additive terms by different error distributions and quantile curves: sspline+SIC (light grey box), pspline+SIC (medium grey box) and pspline+HFS (dark grey box). Left panels: constant scale function; Right panels: non-constant scale function. The true signal is $1 + 2\cos(x_{1i}) + \sin(2\pi x_{2i})$.

Figure 2 depicts the MIAE differences for two smooth covariates: as in the above single smooth term, the patterns are substantially undistinguishable with respect to the MISE with 'sspline+SIC' the worst, and the proposed 'pspline+HFS' performing similarly or slightly better than 'pspline+SIC' in the t_1-errors scenarios.

Due to such strong similarities between MISE and MIAE, even observed in scenarios being discussed afterwards, hereafter we will present results only in terms of square errors.

Also, we have considered the same simulations settings with $n = 100$: results are not shown since no important and particular patterns emerge with respect to results obtained with $n = 400$.

A.2 Comparisons with L-curve approach

As discussed in the paper, the L-curve constitutes an possible alternative approach to select the smoothing parameter. Even if it could not be generalized to multiple smooth terms, we report some simulation evidence comparing it (pspline+L) to our proposal (pspline+HFS) when the QR equation involves just a single smooth-term. Figure 3 portrays the usual boxplots for the log(MISE) values across all aforementioned scenarios.

Here differences are quite clear cut: 'pspline+HFS' exhibits higher MISE values only with t_1 errors at middle quantile curves ($\tau = 0.25, 0.50, 0.75$) and signals $\log(x)$ and $g_0(x)$, while it performs better at all percentile values for every considered distribution when the signal is linear or sinusoidal.

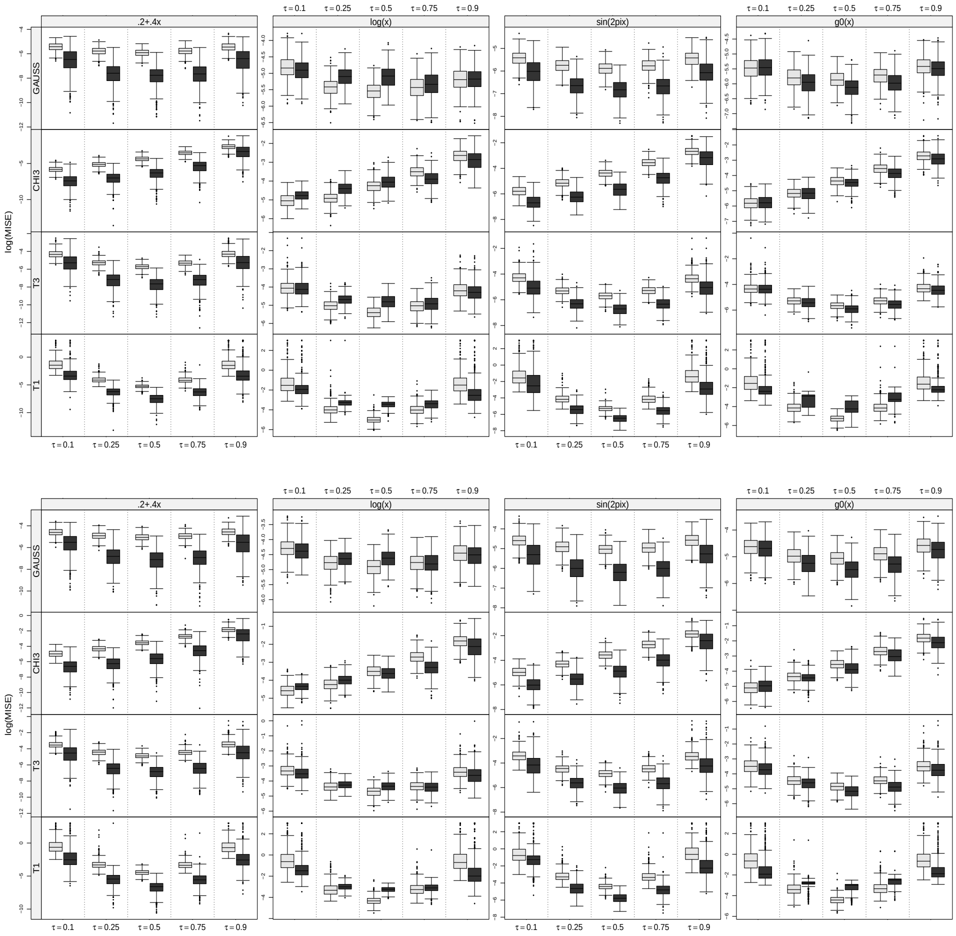

Figure 3: Contrasting ‘pspline+L’ (light grey box) and ‘pspline+HFS’ (dark grey box) for smoothing parameter selection in terms of MISE (on log scale) by different error distributions and signals. The top plots refer to errors with constant scale $\sigma(x) = 0.2$, the bottom ones to non-constant scale function $\sigma(x) = 0.2(1 + x)$.

A.3 Comparisons with cross-validation and boosting

Cross validation (CV) and boosting represent alternative options to select the smoothing parameter in QR. As remarked in section 4.1, CV is very taxing from a computational perspective, as for each candidate value of the smoothing parameter, the dataset has to be split into 10, say, groups, the model fitted to one group and tested on the remaining 9 groups. All groups are employed as training data one at a time and the prediction error coming from the testing part is averaged and saved. Therefore for a 15-dimensional grid of lambda values, the ten folds CV requires

150 fits. Clearly the computation burden gets prohibitive when multiple smoothing parameters have to be selected, and thus we assess CV only with one smooth term. Smoothing parameter selection via boosting is recent proposal as discussed in relevant papers reported in bibliography of the manuscript. It is computational efficient and thus it can be used even in additive QR. We use the gamboost() function with default values from the R package mboot to run boosting QR with smooth terms.

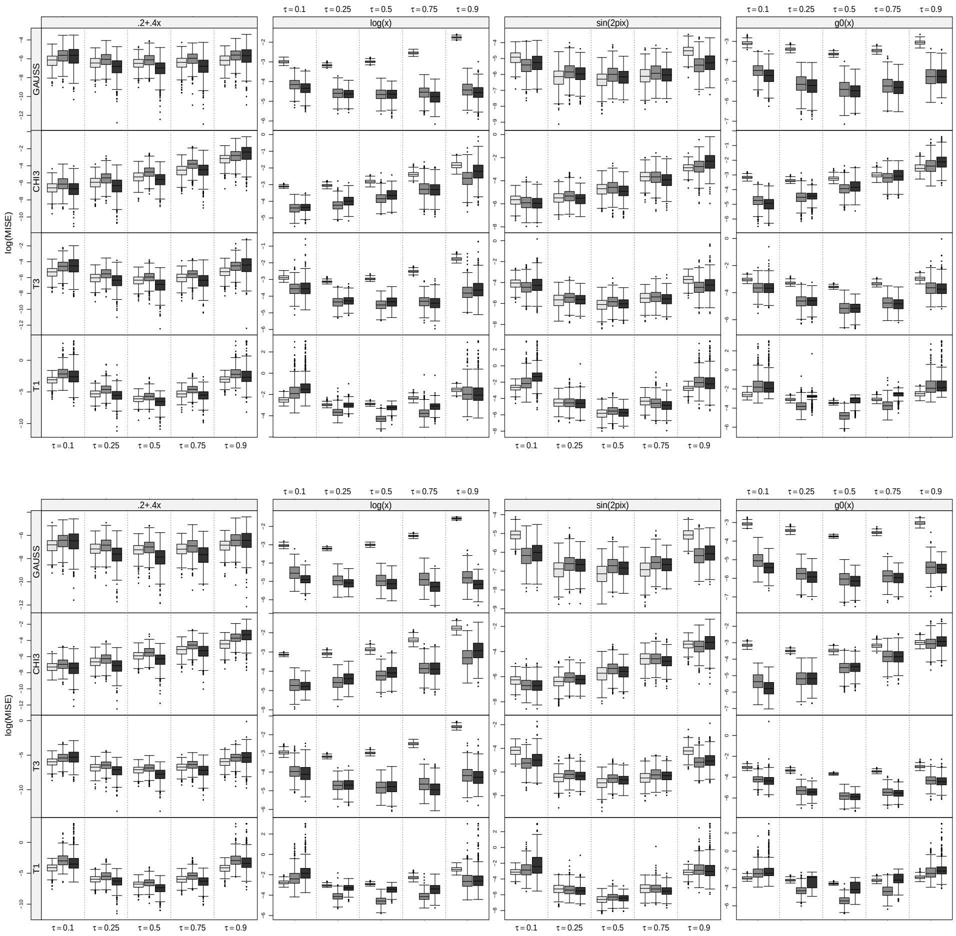

Figure 4: Contrasting boosting (light grey box), 10-folds CV (middle grey box) and the proposed 'pspline+HFS' (dark grey box) for smoothing parameter selection (1 smooth covariate) in terms of MISE (on log scale) by different error distributions and signals. The top plots refer to errors with constant scale $\sigma(x) = 0.2$, the bottom ones to non-constant scale function $\sigma(x) = 0.2(1 + x)$. Boosting has been implemented with default values in gamboost() function by the mboost package for R

Figure 4 and 5 report results, respectively for one and two smoothing parameter selection.

Boosting with default values return higher MISE values when the signal is $\log(\cdot)$, and $g_0(\cdot)$. On the other hand, CV (10-folds) and the proposed iterative approach provide quite similar MISE values across the scenarios with some differences at middle quantiles for signals linear, $\log(\cdot)$, and $g_0(\cdot)$ with gaussian and t_1 errors.

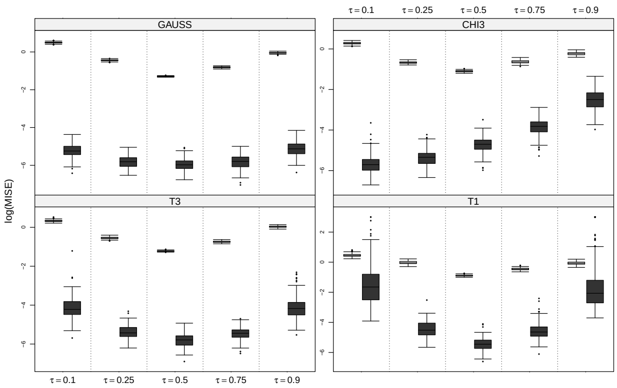

Figure 5: Contrasting boosting (light grey box) and the proposed pspline+HFS (dark grey box) for smoothing parameter selection in QR with 2 additive terms by different distributions (iid errors) and quantile curves. Boosting has been implemented with default values in gamboost() function by the mboost package for R. The true signal is $1 + 2\cos(x_{1i}) + \sin(2\pi x_{2i})$.

In the additive model with 2 smooth terms (Figure 5), our proposal still outperforms the boosting, although it should be recognized that different, likely better, performances could be expected by modifying properly some settings in the boosting algorithm.

A.4 Performances in VC models

As remarked in section 3.2, vary coefficients (VC) models can be estimated within the proposed framework. To assess empirically the performances, Figure 6 illustrates the MISE differences among pspline+SIC, pspline+L, and pspline+HFS. The true quantile function is $\sin(2\pi x_i)z_i$ where $x_i \sim U(0,1)$ and $z_i \sim U(2,4)$ and errors come from the same distributions Normal, χ^2_3, t_3 and t_1.

Perhaps surprisingly, the L-curve approach performs always worse than 'pspline+SIC' and 'pspline+HFS' competitors which show very similar mean square errors.

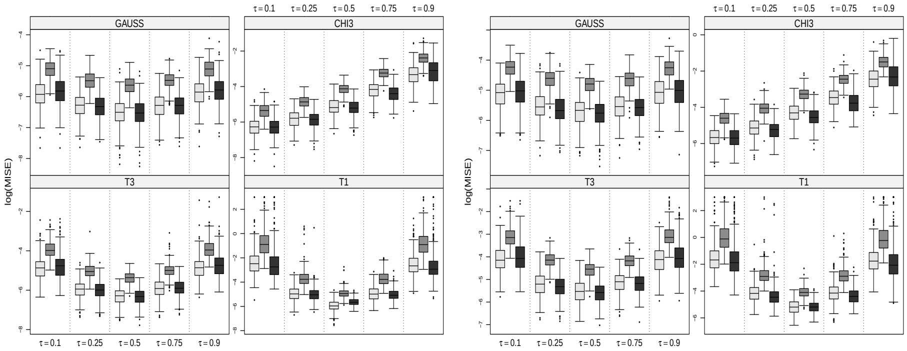

Figure 6: Contrasting the three competitors (in terms of log(MISE)) for smoothing parameter selection in a varying coefficient model by different error distributions and quantile curves: pspline+SIC (light grey box), pspline+L (medium grey box) and pspline+HFS (dark grey box). Left panels: constant scale function; Right panels: non-constant scale function.

A.5 The role of difference order in the penalty

The penalized objective to be minimized (eq. (2.2) in the manuscript) includes penalties on the d-order differenced coefficients. Like for mean regression, when the smoothing parameter λ gets larger the fitted curve approaches a $d-1$ degree polynomial, while $\lambda = 0$ indicates no penalization, resulting in a potentially wiggly curve: at intermediate values of λ, the fit is a piecewise polynomial of degree $d-1$. $d = 1$ and $d = 2$ lead to a fitted quantile curve which is, respectively, a piecewise-constant or piecewise-linear curve which are unreasonable for most of applications. Hence $d \geq 3$ is requested if a smooth fitted curve is requested. but as a referee has pointed out, one could wonder if the difference order influences the performance of the proposed algorithm.

Figure 7 portrays the MISE values of the aforementioned scenarios coming from the proposed algorithm when the penalty specifies $d = 1, 2, 3, 4$.

Unsurprisingly $d = 2$ performs somewhat better for truly linear quantile curve, while $d = 1$ performs generally worse, especially in the sinusoidal case. However focussing on the d values returning a smooth fit ($d = 3$ or $d = 4$) we observe quite similar MISE values with somewhat lower MISE values of $d = 3$ at $\tau = 0.75$ and $\tau = 0.90$ for logarithm signal.

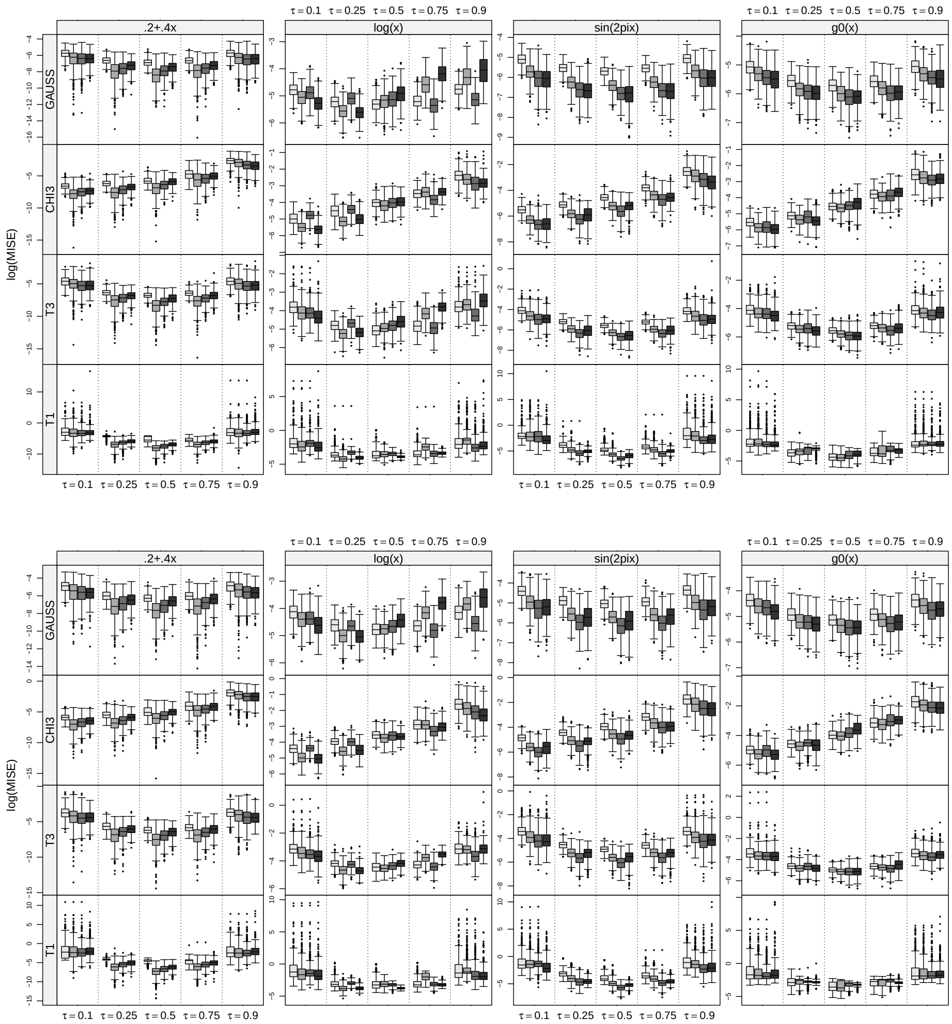

Figure 7: Performance of the proposed 'pspline+HFS' algorithm using difference penalties $d = 1, 2, 3, 4$ (from lighter to darker grey box). The top plots refer to errors with constant scale $\sigma(x) = 0.2$, the bottom ones to non-constant scale function $\sigma(x) = 0.2(1 + x)$.

A.6 The role of the extra-penalty factor γ

Formulas of the scale parameter estimates in step 3. of the algorithm involve a tuning parameter γ aimed to provide an extra penalty for the model complexity. As remarked in section 3.2, the idea is sometimes discussed in mean regression when smoothing parameter selection is carried out via generalized cross validation, for instance. Evidence suggests that $\gamma = 1$ leads to undersmoothing and the larger γ, the smoother the fit.

To assess possible effects of γ on the fitted quantile curves, Figure 8 reports the MISE values when four different values are employed in the algorithm: $\gamma = 1, 1.5, 2, 2.5$.

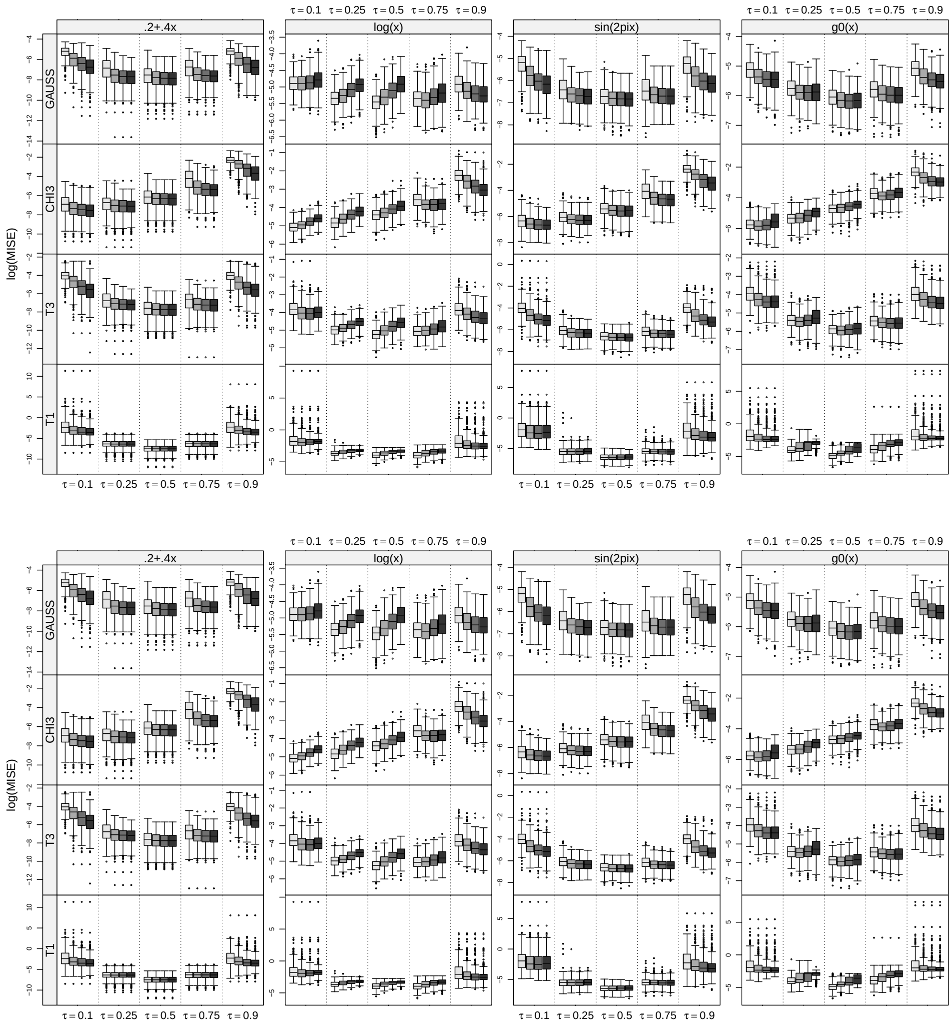

Figure 8: Performance of the proposed 'pspline+HFS' algorithm using different γ values ($\gamma = 1, 1.5, 2, 2.5$ from lighter to darker grey box). The top plots refer to errors with constant scale $\sigma(x) = 0.2$, the bottom ones to non-constant scale function $\sigma(x) = 0.2(1 + x)$.

Apart form some sporadic case (middle τ values with logarithm signal), $\gamma = 1$ is associated with higher MISE values, while differences when $\gamma \geq 1.5$ are quite minor, especially at $\gamma = 2$ or 2.5. We recommend $\gamma = 2$ which is the value set in all simulation experiments.