Coverage Analysis of Quantile Smoothing Spline Confidence Bands and Regions

Likun Zhang and Enrique del Castillo
Dept. of Industrial Engineering and Dept. of Statistics
Penn State University
and
Andrew Berglund, Martin Tingley and Nirmal Govind
Netflix Inc.

May 10, 2017

Abstract

A coverage analysis of the confidence bands and surfaces computed by the functions in the `ConfidenceQuant` R package is presented. One covariate "bands" and two covariate "surfaces" are investigated for a variety of true functions and surfaces and different types of noise distributions. It is found that the pointwise coverage obtained is close to nominal across the experimental region in a wide variety of scenarios. The methods used in the package to obtain the bands and surfaces are briefly explained.

1 Quantile smoothing splines

The `ConfidenceQuant` R package developed for the computation of confidence bands and regions for comparative experiments fits Penalized Quantile Smoothing Splines using the `rqss` function in the `quantreg` package, developed by R. Koenker [2]. Quantile splines provide more flexible function shapes than quantile regression models and hence were adopted for this project. Given data $\{x_i\}$ and $\{y_i\}$ and a quantile number $\tau \in [0, 1]$, the type of Penalized Quantile Smoothing Spline functions fitted by `rqss` are obtained from minimizing:

$$\sum_{i=1}^{n} \rho_\tau[y_i - g(x_i)] + \lambda J(g)$$

where $\rho_\tau(u) = \tau u^+ + (1-\tau)u^-$ is the "check function" [3], $J(g)$ is a given roughness functional, and $\lambda > 0$ is a tuning factor that determines the smoothness of the fitted function g. The penalized objective trades-off the *fidelity* of the fitted function (first term) with some measure of its roughness (second term). Different types of Penalty functions have been proposed in the literature. Function `rqss` uses the Total Variation roughness penalty, defined for the case x is one dimensional (one covariate) and $g'(x)$ absolutely continuous as

$$V(g') = \int_a^b |g''(x)|dx$$

where (a, b) is the domain of the function. Koenker [3] shows that the solutions to the problem above under the Total Variation penalty can be obtained via Linear Programming (an advantage computationally) and how the solution is made of piecewise linear functions with knots at the x_i's of the form $\hat{g}(x) = \alpha_i + \beta_i(x - x_i),\ x \in [x_i, x_{i+1})$. When $g : \mathbb{R}^2 \to \mathbb{R}$ (two dimensional covariates), the total variation is defined by

$$V(\nabla g) - \int ||\nabla^2 g(x)||dx$$

where $||\cdot||$ denotes the Frobenius norm.

2 Finding the optimal smoothing constant

To fit a Penalized Quantile Smoothing Spline, the value of the smoothing parameter λ needs to be specified. Many methods have been proposed in the literature to find an "optimal" value of λ that can provide an automatic way of selecting this tuning parameter. We tried some of these methods including what is perhaps the most popular, Schwartz's Information Criterion (SIC):

$$\mathrm{SIC}(\lambda) = \log\left(\frac{1}{n}\sum_{i=1}^{n} \rho_\tau\{y_i - \hat{g}_\lambda(x_i)\right) + \frac{\log n}{2n}\mathrm{EDF}$$

where EDF stands for the expected degrees of freedom, which are equivalent to the number of 'active' knots that are interpolated by the fitted function (hence equal to the number of zero residuals). Unfortunately, neither SIC nor several modifications of it that were tried proved satisfactory for the automatic selection of λ, as the functions optimized for finding λ were too erratic for an automatic search routine.

The method we finally adopted for the selection of λ is based on work by Reiss and Huang [4], who proposed to use a Multiple Cross-validation (MCV) statistic for finding the optimal smoothing parameter in quantile smoothing splines under a *quadratic* penalty function (unlike our case). For such penalty, they show evidence that MCV provides better performance than other methods, including SIC. We used this idea for finding λ^* for total variation-penalized quantile splines fitted with `rqss` and investigated the ultimate coverage of the confidence bands generated via bootstrapping.

To choose the best smoothing parameter, Reiss and Haung propose to use K-fold cross validation to calculate the multifold crossvalidation (MCV) statistic, defined by:

$$\mathrm{MCV}(\lambda) = \frac{1}{n} \sum_{j=1}^{K} \sum_{i \in V_j} \rho_\tau [y_i - \hat{g}_\lambda^{[-V_j]}(x_i)]$$

where $V_1, \ldots, V_K$ are the K equal-sized parts (folds) of the original data, and $\hat{g}_\lambda^{[-V_j]}$ denotes the smoothing spline fitted to all folds except fold V_j. The recommended number of folds in crossvalidation ranges from 5 to 10, so we fixed $K = 10$ in the coverage analysis shown below. We observed in our numerical experiments that $\mathrm{MCV}(\lambda)$ is typically minimized for relative small values of λ which provide visually appealing smoothing. Surprisingly, we found this function better behaved in the case of two covariates than in the case of one covariate, when the function typically exhibits several local minima.

3 Confidence band computation and coverage estimation

To find a confidence band or region around a fitted quantile function without recourse to any distributional assumption, bootstrapping methods were used. Initially, given the size of the datasets to which our code will be applied, the "Bag of Little Bootstraps" (BLB) method [1] was implemented, but unfortunately, the computational advantage of the BLB method cannot be implemented with the `rqss()` function since it does not accept weighted data [1]. The final method we adopted uses a regular bootstrapping method based on B bootstraps, given it is faster and provided better coverage in our experiments than the BLB method.

In each simulation, a new x, y dataset is generated from a "known" function contaminated with different types of additive noise. First, the optimal λ that minimizes MCV is obtained. For a collection of x-values, , the penalized quantile spline evaluated at the x-values is computed for each bootstrap resample. The confidence band or region of the function is then made of the pointwise bootstrapped confidence intervals for the fitted quantile function values computed at each x. Following standard coverage simulation practice, the estimated coverage is computed as the proportion of times the true function value at each x is contained in the bootstrap intervals.

4 Confidence bands coverage (one covariate)

Given the high computing time, all coverages were estimated based on 100 simulations. In every case, the nominal confidence level was 0.95. Individual coverage of the band for particular x values in the (0,1) interval were computed without any adjustment. Data were generated from a low-frequency sinusoidal defined in the range (0,1) in one covariate with additive noise that followed different distributions (see Figure 1), as shown in the tables below. In every case in this section, $n = 3000$ data points were generated, and the optimal smoothing parameter λ^* was obtained from multi-fold cross-validation with $K = 10$ equal size folds. Standard bootstrapping was used to compute the confidence intervals at each x and provide the *pointwise* coverages shown in Tables 1-10.

[1]BLB requires 3 tuning parameters, $b < n$, the size of the partitions or "bags" obtained by sampling without replacement (values $b = n^\gamma$ with $\gamma \in (0, 5, 0.9)$ are recommended, see [1]), the number of "bags" s, and the number of resamples within each bag, r, recommended as low as 100. Since $b < n$, the bootstraped samples will have repeated data, and they can be represented by a histogram with the unique numbers observed and the counts (frequencies or "weights"), rather than using the raw data containing the repeats. If an algorithm can accept the histogram (or weighted) representation of the data, the BLB speed up will occur, but this is not the case for the `rqss` function, as its `weights` argument is not operational and will actually disappear in future versions (private communication with R. Koenker).

Tables 1 and 2 shows coverage of the 50% quantile (median) and 90% quantile confidence bands obtained using standard bootstrapping for different numbers of bootstraps B compared to using the BLB method with different parameters of the bag size b and number of resamples r. As it can be seen, regular bootstrapping with $B = 300$ resamples is enough to provide coverage close to nominal. In contrast, none of the BLB runs provided adequate coverage. Hence, in subsequent tables only one instance of BLB (with $b = n^{0.8}$ and $r = 100$, which were the best parameters) is compared against one instance of regular bootstrapping for $B = 300$.

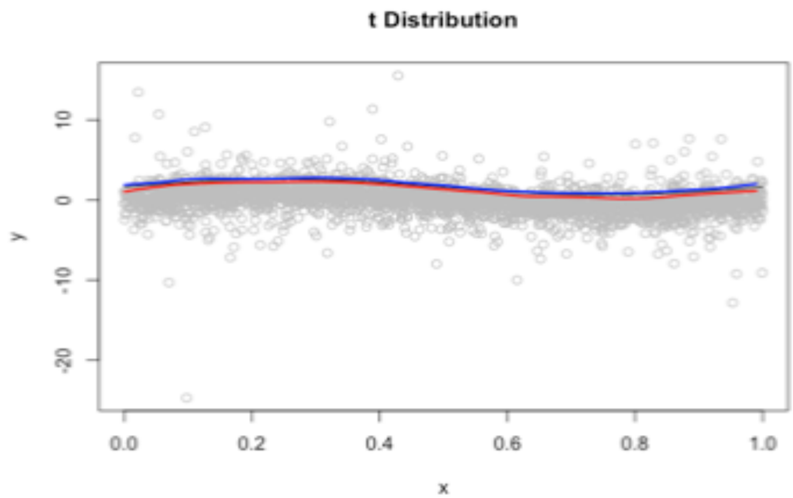

Figure 1: Simulated sinusoidal surface in one covariate with additive t_3 noise, showing the pointwise 95% confidence band for the 90% quantile.

Table 1: Pointwise coverage for t_3-distributed data, $n = 3000$. $\tau = 0.5$, $\lambda^* = 1.3797$

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 10$	79	77	86	80	85	77	80	84	80	82	75	85	88	71	85	81
$B = 50$	94	92	91	96	92	91	86	89	90	90	92	92	92	87	92	92
$B = 100$	97	93	96	94	98	94	97	98	96	94	93	91	91	95	96	92
$B = 300$	95	94	97	96	99	94	92	93	92	92	92	93	95	96	96	93
$B = 1000$	92	95	94	95	94	93	96	96	97	95	100	96	92	90	95	93
BLB ($r = 100, b = n^{0.6}$)	80	79	75	72	78	79	87	82	87	87	84	76	73	73	79	72
BLB ($r = 300, b = n^{0.6}$)	87	82	80	77	81	78	80	82	80	86	86	88	88	86	81	79
BLB ($r = 100, b = n^{0.8}$)	93	90	90	92	90	86	88	90	84	89	87	89	85	82	88	83

Table 2: *Pointwise coverage for t_3-distributed data, $n = 3000$. $\tau = 0.9$, $\lambda^* = 0.7245$*

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 10$	89	84	81	79	85	81	82	81	83	82	80	79	81	85	87	81
$B = 50$	94	95	88	91	93	91	95	90	94	97	90	88	91	93	93	91
$B = 100$	98	96	94	93	92	94	97	94	94	92	91	97	94	97	95	96
$B = 300$	96	97	97	97	93	94	89	91	93	92	94	93	93	95	97	88
$B = 1000$	97	98	96	97	99	96	93	94	95	94	93	98	95	97	96	95
BLB ($b = n^{0.8}, r = 100$)	91	90	93	87	90	92	93	93	92	93	91	94	90	85	92	84

As it can be seen in Tables 3-10, the pointwise coverage provided by the quantile smoothing spline confidence bands using regular bootstrapping is close to nominal in general, regardless of the distribution of the errors. Note how some of the distributions tested are either non-symmetric or very fat tailed.

Table 3: *Pointwise coverage for $N(0,1)$-distributed data, $n = 3000$. $\tau = 0.5$, $\lambda^* = 1.2647$*

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	95	93	97	94	95	92	79	94	87	89	96	97	98	89	88	89
BLB ($b = n^{0.8}, r = 100$)	85	83	85	91	89	85	83	91	83	84	89	86	87	83	84	74

Table 4: *Pointwise coverage for $N(0,1)$-distributed data, $n = 3000$. $\tau = 0.9$, $\lambda^* = 1.1407$*

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	94	88	91	94	98	91	78	93	90	93	97	97	94	83	93	71
BLB ($b = n^{0.8}, r = 100$)	93	83	86	93	86	76	72	94	81	79	88	89	85	77	86	63

Table 5: *Pointwise coverage for mixture $0.45N(0,1) + 0.55 * N(3,1)$ data, $n = 3000$. $\tau = 0.5$, $\lambda^* = 1.0012$*

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	91	97	96	93	93	93	95	99	98	97	95	93	90	90	96	93
BLB ($b = n^{0.8}, r = 100$)	90	89	93	89	88	86	90	91	90	89	87	87	86	86	90	93

Table 6: *Pointwise coverage for mixture $0.45N(0,1) + 0.55 * N(3,1)$ data, $n = 3000$. $\tau = 0.9$, $\lambda^* = 1.0125$*

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	96	96	93	92	94	88	87	97	94	91	95	98	93	91	92	88
BLB ($b = n^{0.8}, r = 100$)	90	86	82	80	85	85	87	87	88	92	91	88	89	90	86	76

Table 7: Pointwise coverage for double exponential (Laplace$(0,1)$) data, $n = 3000$. $\tau = 0.5$, $\lambda^* = 1.1324$

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	97	96	95	96	97	94	89	95	92	94	97	98	97	94	96	93
BLB ($b = n^{0.8}, r = 100$)	84	92	91	91	89	85	80	92	86	85	91	90	93	93	92	83

Table 8: Pointwise coverage for double exponential (Laplace$(0,1)$) data, $n = 3000$. $\tau = 0.9$, $\lambda^* = 0.8315$

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	98	94	92	93	95	92	90	96	95	94	94	92	96	94	100	94
BLB ($b = n^{0.8}, r = 100$)	89	90	88	92	93	90	87	92	90	90	89	91	95	93	93	90

Table 9: Pointwise coverage for Slash $(N(0,1)/U(0,1))$ data, $n = 3000$. $\tau = 0.5$, $\lambda^* = 0.6396$

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	97	97	93	96	92	96	94	91	90	92	96	95	93	93	97	95
BLB ($b = n^{0.8}, r = 100$)	87	87	90	87	87	96	93	92	86	91	89	92	92	92	93	94

Table 10: Pointwise coverage for Slash $(N(0,1)/U(0,1))$ data, $n = 3000$. $\tau = 0.9$, $\lambda^* = 1.5864$

x	0.07	0.13	0.19	0.25	0.31	0.37	0.43	0.49	0.55	0.61	0.67	0.73	0.79	0.85	0.91	0.97
$B = 300$	94	97	95	96	96	94	96	96	91	91	93	93	92	92	97	93
BLB ($b = n^{0.8}, r = 100$)	90	87	86	86	80	81	87	96	88	86	87	81	86	90	89	79

5 Coverage of Confidence Surfaces (two covariates)

For the case the quantile smoothing spline is a function of two covariates, three different "true" surface functions were investigated, each with additive t-noise.

The first type of surface tested for coverage analysis is a bell-shaped paraboloid surface with a maximum near the center of its domain ($(0,1) \times (0,1)$, see Figure 2) to which we added t_3 noise. $n = 4000$ data points were generated and an optimal smoothing λ value was obtained using the MCV criterion with $K = 10$ folds. Pointwise coverages for the median ($\tau = 0.5$) were computed for a grid of x-point values within the plane $(0,1) \times (0,1)$,

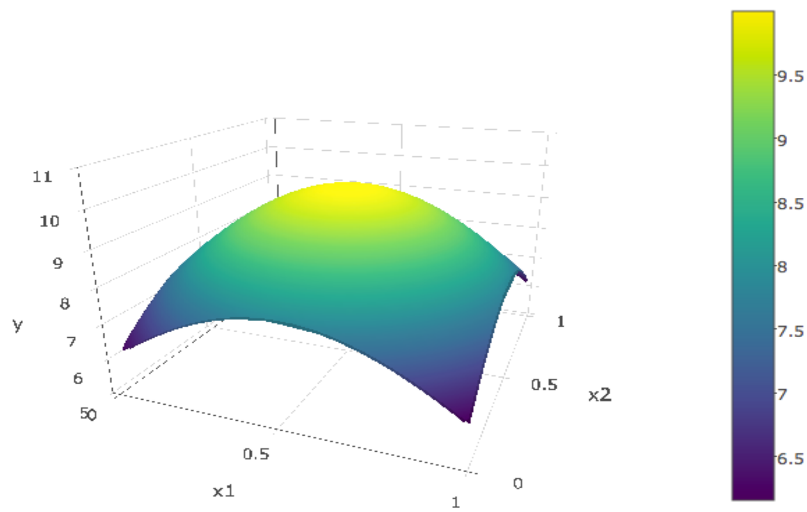

Figure 2: Bell-shaped paraboloid surface.

based on 100 simulations. As it can be seen from Table 11, the pointwise coverage is everywhere close to nominal.

To consider more complicated surfaces, a second type of surface, a sinusoidal "ridge", was used as the true surface (Figure 3) to which t-distributed (t_3) errors were added to generate the observed data. $n = 4000$ observations were generated this way and the optimal λ value was obtained using the MCV criterion with $K = 10$ folds.

Tables 12-13 show the estimated pointwise coverage for the median and 0.90 quantile for a collection of points on the plane $(0,1) \times (0,1)$, based on 100 simulations. As it can be seen, the coverage of the pointwise intervals is adequate throughout the region.

Table 11: Pointwise coverage for bell-shaped paraboloid with t_3 noise, $\tau = 0.5$, $\lambda^* = 0.8708$

x_1 \ x_2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.95	0.98	0.97	0.95	0.94	0.96	0.95	0.96	0.92
0.2	0.95	0.95	0.97	0.92	0.96	0.93	0.95	0.95	0.98
0.3	0.95	0.96	0.99	0.95	0.95	0.93	0.97	0.99	0.97
0.4	0.97	0.96	0.98	0.93	0.94	0.95	0.99	0.99	0.94
0.5	0.97	0.94	0.97	0.92	0.95	0.94	0.96	0.94	0.98
0.6	0.98	0.96	0.97	0.97	0.95	0.95	0.97	0.95	0.96
0.7	0.99	0.93	0.94	0.99	0.94	0.93	0.97	0.96	0.96
0.8	0.98	0.94	0.96	0.98	0.97	0.96	0.95	0.91	0.95
0.9	0.96	0.96	0.94	0.96	0.96	0.94	0.94	0.97	0.98

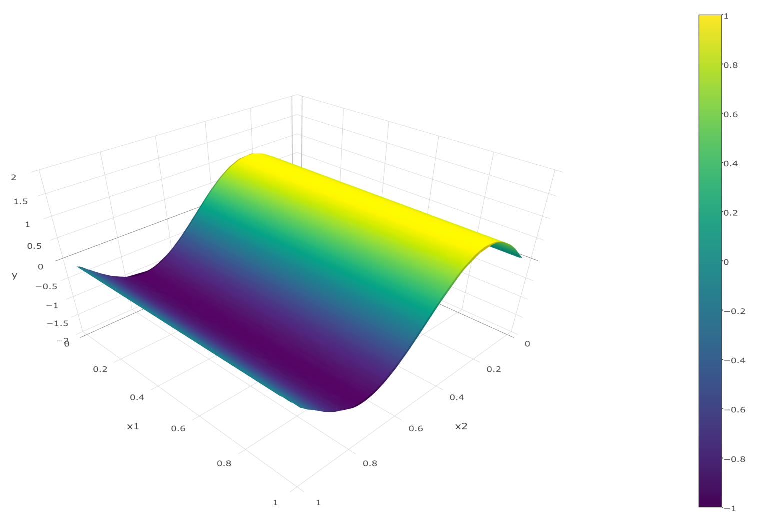

Figure 3: Sinusoidal "ridge" surface.

Table 12: Pointwise coverage for sinusoidal ridge and t_3 noise, $\tau = 0.5$, $\lambda^* = 1.1638$

x_1 \ x_2	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
0.1	0.96	0.98	0.87	0.89	0.88	0.96	0.9	0.9	0.94	0.98	0.95
0.2	0.97	0.93	0.93	0.93	0.9	0.94	0.94	0.93	0.94	0.98	0.96
0.3	0.96	0.96	0.96	0.95	0.93	0.94	0.94	0.96	0.94	0.95	0.96
0.4	0.96	0.97	0.98	0.97	0.94	0.96	0.92	0.94	0.96	0.93	0.96
0.5	0.95	0.96	0.96	0.93	0.97	0.93	0.93	0.94	0.92	0.95	0.98
0.6	0.91	0.94	0.96	0.95	0.97	0.96	0.99	0.98	0.96	0.96	0.96
0.7	0.95	0.94	0.94	0.94	0.96	0.95	0.95	0.97	0.96	0.97	0.97
0.8	0.97	0.95	0.95	0.94	0.93	0.94	0.91	0.96	0.99	0.99	0.99
0.9	0.95	0.98	0.96	0.89	0.91	0.98	0.89	0.93	0.97	0.96	0.97

Table 13: Pointwise coverage for sinusoidal ridge and t_3 noise, $\tau = 0.9$, $\lambda^* = 1.2558$

x_1 \ x_2	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
0.1	0.98	0.89	0.80	0.76	0.82	0.96	0.85	0.82	0.85	0.92	0.96
0.2	0.99	0.93	0.86	0.81	0.87	0.96	0.93	0.96	0.95	0.95	0.96
0.3	0.99	0.93	0.86	0.86	0.91	0.95	0.95	0.97	0.96	0.96	0.95
0.4	1.00	0.91	0.87	0.87	0.89	0.96	0.94	0.98	0.97	0.95	0.95
0.5	0.98	0.86	0.87	0.87	0.86	0.96	0.93	0.97	0.95	0.92	0.97
0.6	0.98	0.88	0.90	0.85	0.86	0.98	0.96	0.96	0.96	0.91	0.95
0.7	0.97	0.92	0.93	0.92	0.91	0.93	0.97	0.95	0.95	0.94	0.94
0.8	0.96	0.92	0.88	0.89	0.93	0.95	0.93	0.90	0.93	0.91	0.94
0.9	0.97	0.88	0.79	0.78	0.83	0.93	0.79	0.81	0.85	0.92	0.95

The third surface tested for the two-covariate coverage analysis is a sinusoidal surface with varying frequency to which t_3 noise was added (see Figure 4). $n = 7000$ points were generated in this case and the optimal λ was obtained using the MCV criterion with $K = 10$ folds.

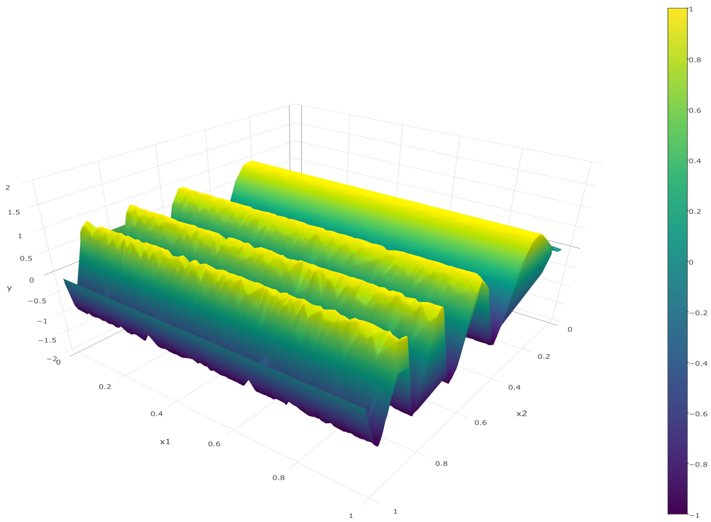

Figure 4: Sinusoidal surface of varying frequency along one coordinate.

Tables 14-15 show the estimated pointwise coverage for the median and 0.90 quantile for a collection of points on the plane $(0,1)\times(0,1)$, based on 100 simulations. In this case, the points (x_1, x_2) at which coverage was computed correspond to the lines along the ridges (peaks) of the sinusoidal function, in addition to the line $x_2 = \sqrt{8/16}$ that corresponds to the line at which the true function is zero.

Table 14: Pointwise coverage for varying frequency sinusoidal surface with t_3 noise, $\tau = 0.5$, $\lambda^* = 0.5624$

x_1 \ x_2	$\sqrt{\frac{1}{16}}$	$\sqrt{\frac{3}{16}}$	$\sqrt{\frac{5}{16}}$	$\sqrt{\frac{7}{16}}$	$\sqrt{\frac{8}{16}}$	$\sqrt{\frac{9}{16}}$	$\sqrt{\frac{11}{16}}$	$\sqrt{\frac{13}{16}}$	$\sqrt{\frac{15}{16}}$
0.1	0.98	0.96	0.92	0.92	0.93	0.78	0.77	0.74	0.68
0.2	0.99	0.97	0.97	0.94	0.97	0.83	0.78	0.79	0.60
0.3	0.96	0.95	0.93	0.94	0.99	0.89	0.77	0.76	0.49
0.4	0.93	0.97	0.94	0.91	0.95	0.79	0.80	0.75	0.56
0.5	0.91	0.96	0.91	0.89	0.92	0.84	0.81	0.75	0.50
0.6	0.94	0.93	0.94	0.89	0.94	0.90	0.84	0.76	0.56
0.7	0.97	0.93	0.96	0.88	0.96	0.88	0.74	0.76	0.51
0.8	0.96	0.97	0.92	0.91	0.94	0.89	0.84	0.67	0.55
0.9	0.97	0.97	0.96	0.84	0.99	0.88	0.78	0.68	0.65

Table 15: Pointwise coverage for varying frequency sinusoidal surface with t_3 noise, $\tau = 0.9$, $\lambda^* = 0.3310$

x_1 \ x_2	$\sqrt{\frac{1}{16}}$	$\sqrt{\frac{3}{16}}$	$\sqrt{\frac{5}{16}}$	$\sqrt{\frac{7}{16}}$	$\sqrt{\frac{8}{16}}$	$\sqrt{\frac{9}{16}}$	$\sqrt{\frac{11}{16}}$	$\sqrt{\frac{13}{16}}$	$\sqrt{\frac{15}{16}}$
0.1	0.91	0.97	0.83	0.94	0.96	0.81	0.91	0.60	0.77
0.2	0.92	0.90	0.81	0.95	0.85	0.74	0.99	0.53	0.62
0.3	0.89	0.93	0.79	0.95	0.90	0.76	0.96	0.63	0.68
0.4	0.88	0.95	0.82	0.97	0.89	0.71	0.93	0.76	0.63
0.5	0.92	0.92	0.89	0.95	0.89	0.77	0.94	0.65	0.66
0.6	0.92	0.93	0.85	0.97	0.89	0.72	0.93	0.63	0.65
0.7	0.94	0.91	0.81	0.95	0.90	0.72	0.91	0.61	0.63
0.8	0.86	0.87	0.82	0.98	0.94	0.71	0.94	0.65	0.72
0.9	0.89	0.91	0.72	0.96	0.88	0.62	0.89	0.57	0.76

As it can be seen, in the region of the $x_1 - x_2$ plane where oscillations are slower, the coverage is adequate, but in the region where the frequency of the oscillations is higher the coverage is inadequate, with coverage decreasing from left to right in each of Tables 14 and 15. Luckily, a behavior like this function is hard to find in practice.

In summary, the confidence bands and surfaces for Quantile Smoothing Spline models fitted by the `ConfidenceQuant` R programs exhibit a robust behavior with respect to several functional forms and types of distributional noise, providing close to nominal pointwise coverage. They are therefore adequate in comparative experiments where each response is modeled with this type of models.

References

[1] Kleiner, A., Talwakar, A., Sarkar, P, and Jordan, M.I., "A scalable bootstrap for massive data", J. of the Royal Society, B, 76, Part 4, pp. 795-816.

[2] Koenker, R., (2015). "Quantile Regression in R: a Vignette", (accompanies the `quantreg` R package),

`www.econ.uiuc.edu/~roger/research/rq/vig.pdf`

[3] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, 2005.

[4] Reiss, P.T. and Huang, L., (2012). "Smoothness selection for penalized quantile regression splines". Working paper, New York University, available at

`http://works.bepress.com/phil_reiss/20/`