

1 Simulation Study

Methods

We compare the performance of our quantile trend filtering method with the three previously published methods using designs proposed by Racine and Li (2017). The methods compared are

- **npqw**: Racine and Li (2017) constrain the response to follow a smooth location scale model of the form $Y_i = a(X_i) + b(X_i)\epsilon_i$. They estimate the τ_{th} conditional quantile given $X_i = x$ using a kernel estimator

$$q_\tau(x) = \frac{\sum_{i=1}^n \Phi_{(Y_i, b(X_i))}^{-1}(\delta_0) K_h(X_i, x)}{\sum_{i=1}^n K_h(X_i, x)} \quad (1)$$

defining $\Phi_{(Y_i, b(X_i))}^{-1}(\delta_0)$ as the quantile function of the Normal distribution with mean Y_i and standard deviation $b(X_i)$ evaluated at τ . δ_0 is a function of τ and chosen empirically, h is a tuning parameter and K is a kernel function. Code was obtained from the author for the **quantile-ll** method.

- **qsreg**: Oh et al. (2011) proposed a pseudo-data algorithm for a quantile spline estimator of the form

$$\sum_i \rho_\tau(y_i - g(x_i)) + \lambda \int (g''(x))^2 dx. \quad (2)$$

If $\rho_\tau(\cdot)$ were differentiable, the solution to this equation would take a form similar to that of the squared loss smoothing spline with weights equal to $\frac{\rho'_\tau(y_i - g(x_i))}{2(y_i - g(x_i))}$. Relying on this idea, Nychka proposed to solve the problem by iteratively solving the weighted smoothing spline. To address the non-differentiability they propose an approximation

$$\rho_{\tau, \delta}(u) = [\tau I(u > 0) + (1 - \alpha) I(u < 0)] u^2 / \delta \quad (3)$$

The function **qsreg** in the **fields** R package was used. The smoothing parameter is chosen automatically using generalized cross validation on the pseudo data.

- **rqss**: Koenker et al. (1994) Koenker proposed smoothing splines using trend filtering with the second order differencing matrix which results in linear splines. The function **rqss** in the **quantreg** package implements this method. The smoothing parameter λ is chosen using a grid search and minimizing

$$SIC(p_\lambda) = \log[n^{-1} \sum \rho_\tau(y_i - \hat{g}(x_i))] + \frac{1}{2n} p_\lambda \log n \quad (4)$$

where $p_\lambda = \sum I(y_i = \hat{g}_i(x_i))$, which can be thought of as active knots.

- **detrendr_SIC**: Our method where we minimize $\sum_i \rho_\tau(y_i - \theta_i) + \lambda \|D\theta\|_1$ and λ is chosen using SIC from above. A single value of λ was chosen by scaling and summing SIC values across all quantiles.
- **detrendr_valid**: Our method where lambda is chosen by leaving out every 5th observation as a validation data set and evaluating the check loss function on the validation data.
- **detrendr_eBIC**: The traditional BIC is given by

$$\text{BIC}(s) = -2\log(L\{\hat{\theta}(s)\}) + \nu(s)\log n \quad (5)$$

where $\theta(s)$ is the parameter θ with those components outside s being set to 0, and $\nu(s)$ is the number of components in s . If we assume an asymmetric Laplace likelihood $L(y|\theta) = \left(\frac{\tau^n(1-\tau)}{\sigma}\right)^n \exp\left\{-\sum_i \rho_\tau\left(\frac{y_i - \theta_i}{\sigma}\right)\right\}$ and the number of non-zero elements of $D\theta$ as df

$$\text{BIC}(df) = 2 \sum_i \frac{1}{\sigma} \rho_\tau(y_i - \theta_i) + df \log n \quad (6)$$

We can choose and $\sigma > 0$ and have found empirically that $\sigma = \frac{1-|1-2\tau|}{2}$ produces stable estimates. Chen and Chen (2008) proposed the extended BIC for large parameter spaces

$$\text{BIC}_\gamma(s) = -2\log(L\{\hat{\theta}(s)\}) + \nu(s)\log n + 2\gamma \log \binom{P}{j}, \quad \gamma \in [0, 1] \quad (7)$$

where P is the total number of possible parameters and j is the number of parameters included in given model. We used this criteria with $\gamma = 1$, $P = n - k$ where k is the order of the differencing matrix and $j = \nu(s)$ is the number of non-zero entries in $D^{(k)}\theta$.

BIC examples

We used a single dataset to illustrate the difference between the scaled, unscaled and extended BIC criteria.

Figure 1: Degrees of freedom (number of non-zero elements of $D\theta$) by $\log(\lambda)$.

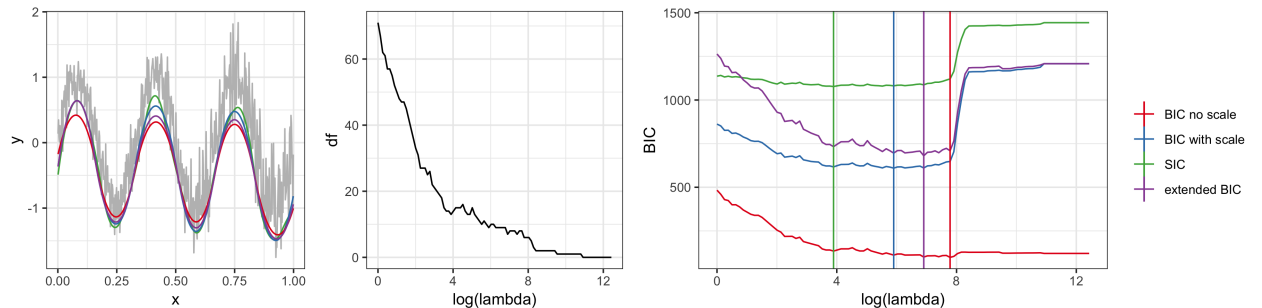
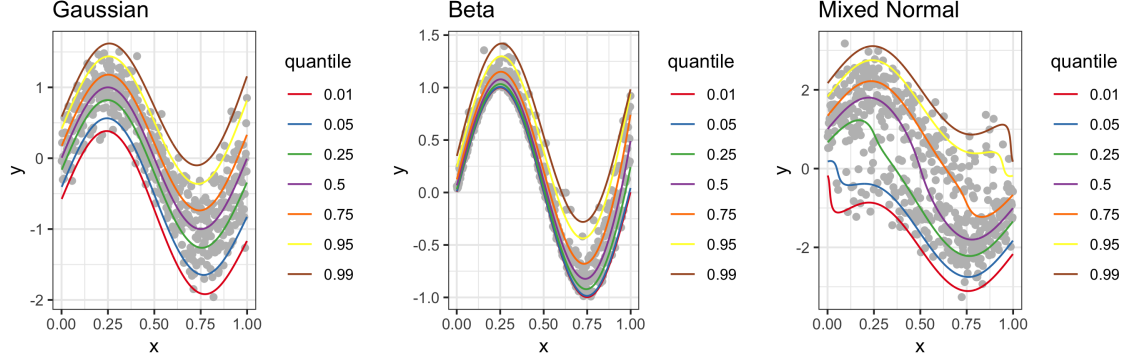


Figure 2: Simulated data with true quantiles $\tau \in \{0.01, 0.05, 0.25, 0.5, .75, 0.95, 0.99\}$



Design

Three simulation designs from Racine and Li (2017) were considered. For all designs X_i was generated as a uniformly spaced sequence in $[0, 1]$ and the response Y was generated as

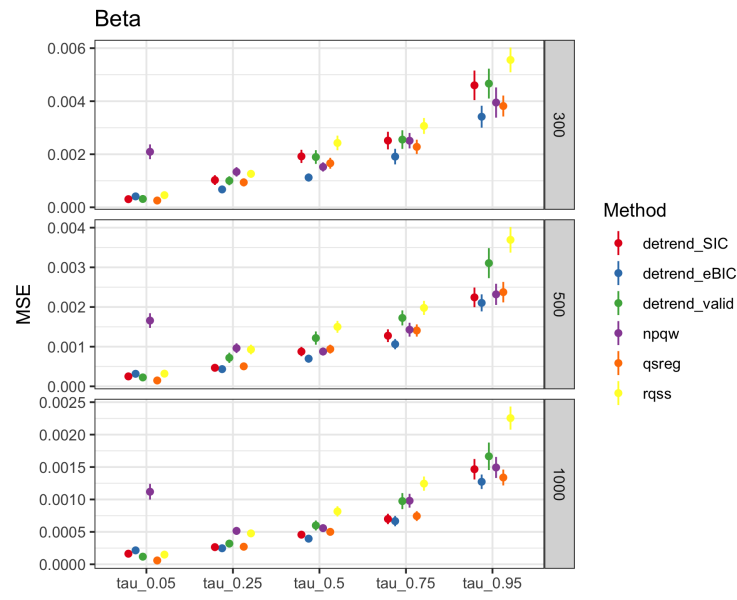
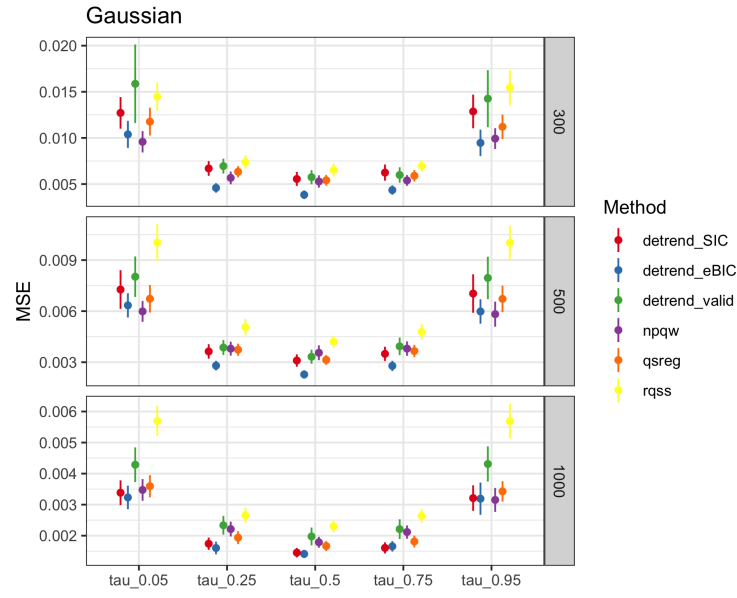
$$Y_i = \sin(2\pi x_i) + \epsilon_i(x_i)$$

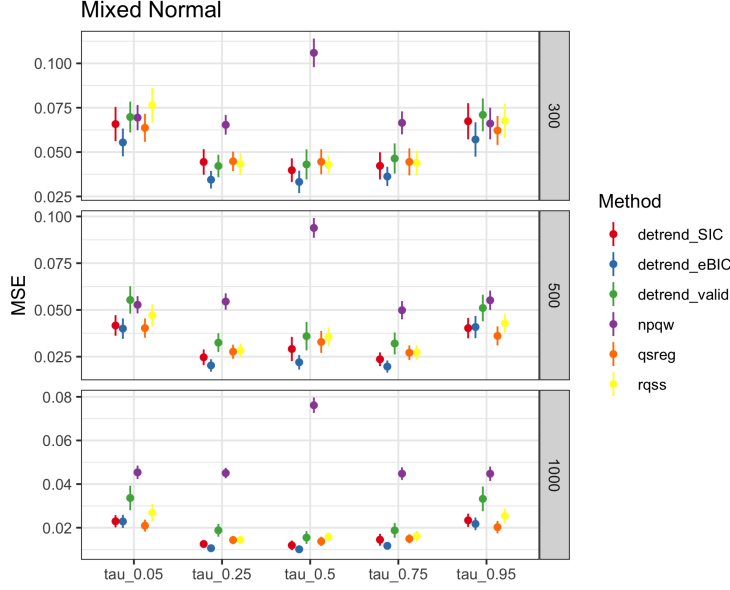
The three error distributions considered were

- Gaussian: $\epsilon_i(x_i) \sim N\left(0, \left(\frac{1+x_i^2}{4}\right)^2\right)$
- Beta: $\epsilon_i \sim \text{Beta}(1, 11 - 10x_i)$
- Mixed normal: ϵ_i is simulated from a mixture of $N(-1, 1)$ and $N(1, 1)$ with mixing probability x_i .

100 datasets were generated of sizes 300, 500 and 1000. The MSE was calculated as $\frac{1}{n} \sum_i (\hat{q}_\tau(x_i) - q_\tau(x_i))^2$.

The plots below show the mean MSE \pm twice the standard error by method, quantile level and sample size.





2 Peaks Simulation

We use another simulation design based on the applied problem we aim to solve. We assume that the measured data can be represented by

$$Y(t) = s(t) + b(t) + \epsilon \quad (8)$$

where $s(t)$ is the true signal at time t , $b(t)$ is the drift component that varies smoothly over time and $\epsilon \sim N(0, \sigma^2)$ is an error component. We assume t is a uniformly spaced sequence between 0 and 1. We generate $b(t)$ using a cubic natural spline basis function with degrees of freedom sampled from 2 to 10 with equal probability, and coefficients drawn from a normal distribution with mean and variance equal to 1. The true signal function is assumed to be zero with Gaussian peaks. The number of peaks is sampled from 5 to 15 with equal probability with centers uniformly distributed between 0.1 and 0.9 and bandwidths uniformly distributed between $2/n$ and $2/n + .01$. One hundred datasets were generated for $n = \{300, 500, 1000\}$.

We compare the methods ability to estimate the baseline using a low quantile, $\tau \in \{0.05, 0.1\}$ and calculate the MSE using the simulated baseline value as the standard. The npqw method performs significantly worse than the other methods and is not included in the figures.

Figure 3: Example of simulated peaks, baseline, and observed measurements.

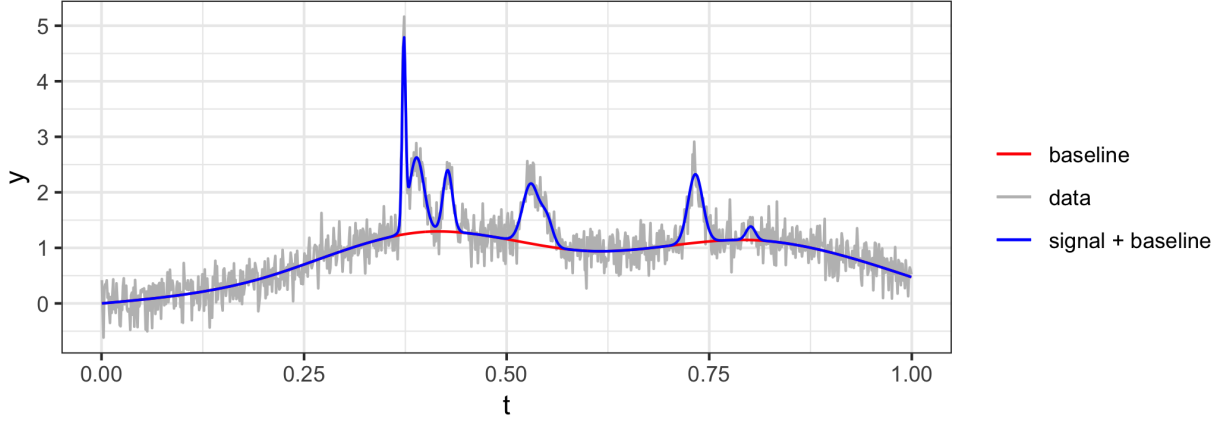
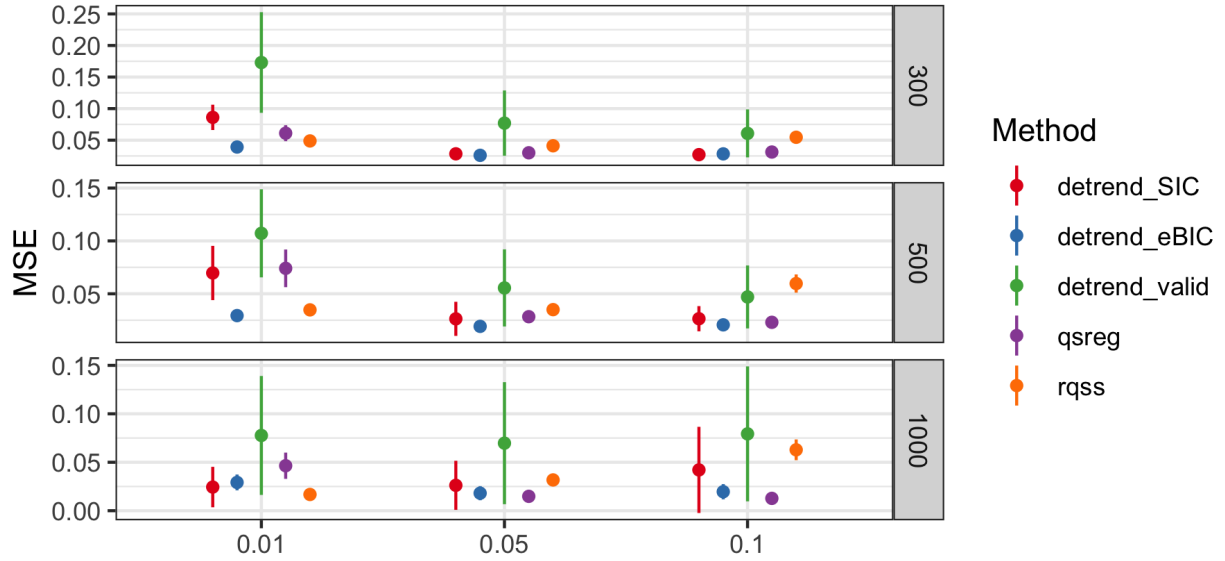


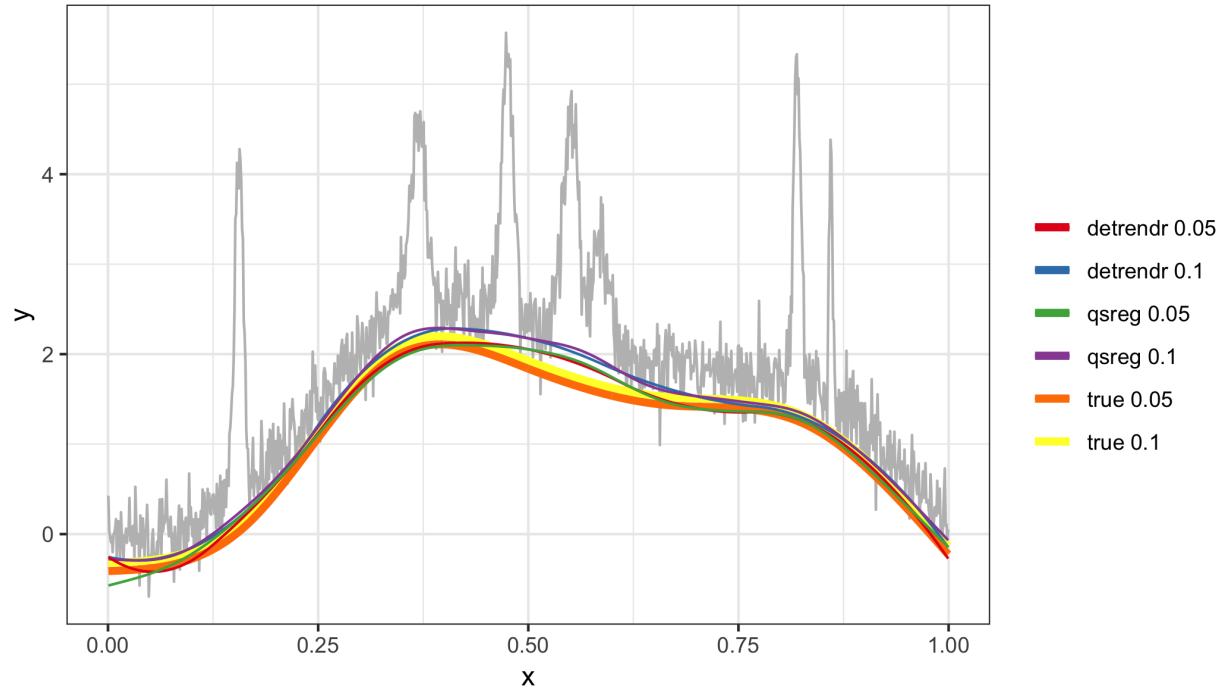
Figure 4: MSEs compared to the simulated baseline function.



References

- Chen, J. and Chen, Z. (2008), “Extended Bayesian information criteria for model selection with large model spaces,” *Biometrika*, 95, 759–771.
- Koenker, R., Ng, P., and Portnoy, S. (1994), “Quantile smoothing splines,” *Biometrika*, 81, 673–680.
- Oh, H.-S., Lee, T. C. M., and Nychka, D. W. (2011), “Fast Nonparametric Quantile Regression With Arbitrary Smoothing Methods,” *Journal of Computational and Graphical Statistics*, 20, 510–526.
- Racine, J. S. and Li, K. (2017), “Nonparametric conditional quantile estimation: A locally weighted quantile

Figure 5: Example baseline fit.



kernel approach,” *Journal of Econometrics*, 201, 72–94.