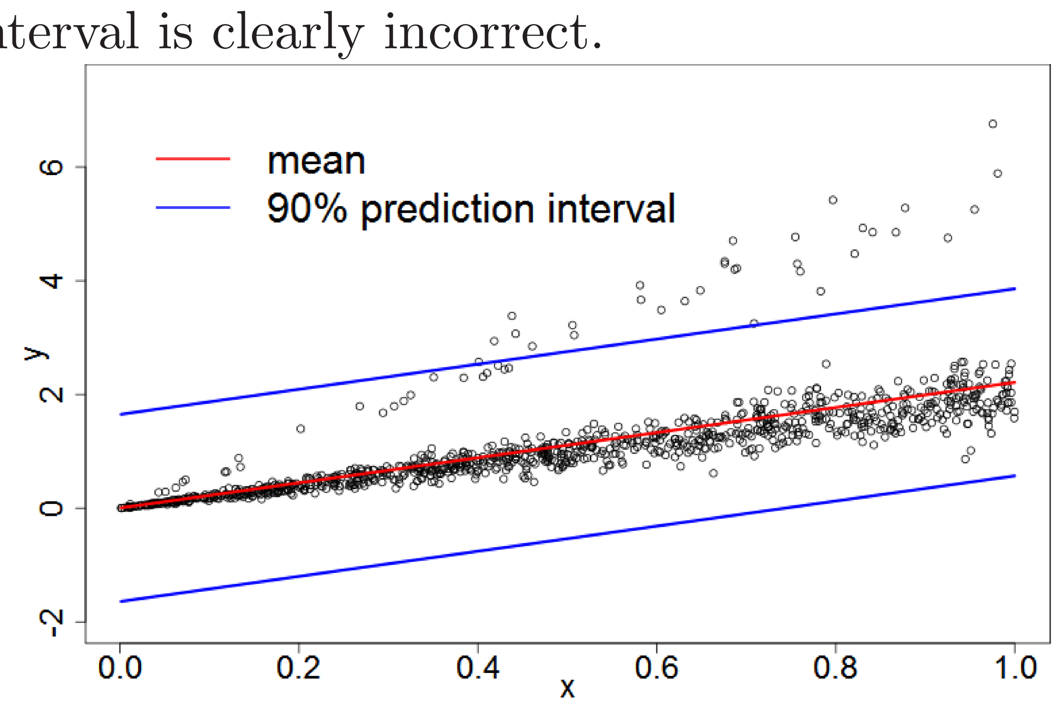
Automated smoothing parameter selection for Quantile Additive Models

University of BRISTOL

Bertrand Nortier and Simon N. Wood , Univ. of Bristol, Mathematics.

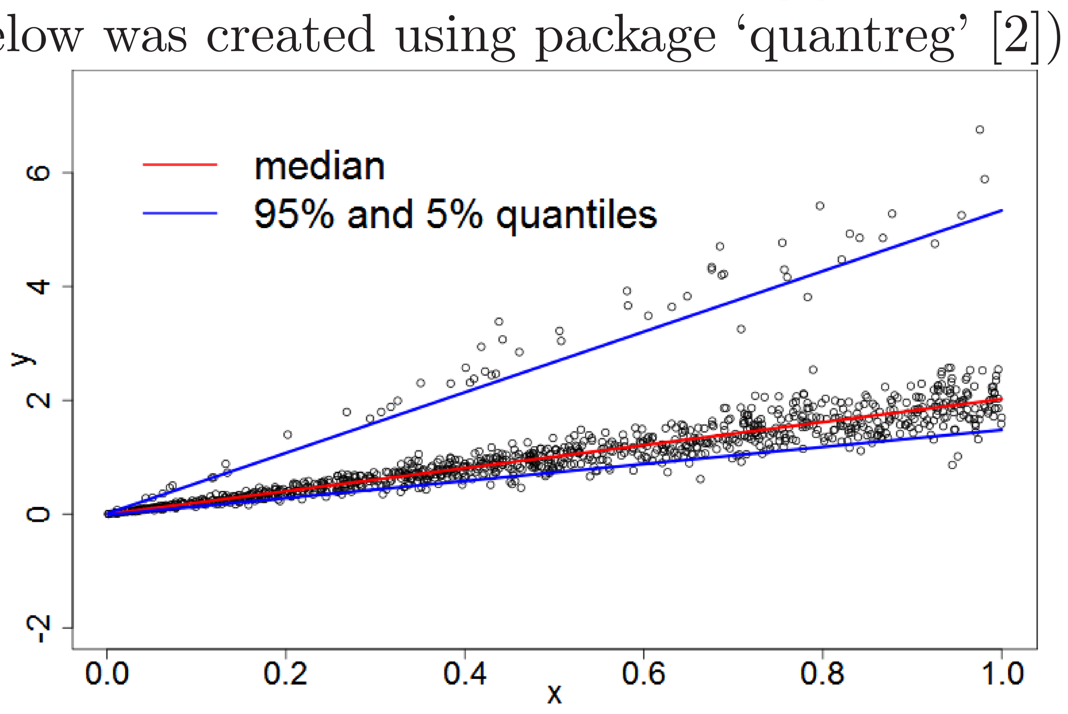
Issue with Linear Regression

Prediction intervals based on standard linear model assumptions may not be appropriate when the data displays some of the following characteristics: **1) heteroschedasticity 2) non-normality 3) has outliers**. In the picture below, we see that: 1) the mean is highly influenced by the outliers and 2) the 90% prediction interval is clearly incorrect.



Quantile Regression

For data as described previously, it may be more appropriate to use **Quantile Regression** developed by Koenker and Bassett [1] (the chart below was created using package 'quantreg' [2]).



Quantile regression is based on loss function:

$$\mathcal{L}_\tau(u) = \begin{cases} (\tau - 1)\,u & (u < 0) \\ \tau u & (u \geq 0) \end{cases} = u\left(\tau - \mathcal{I}(u < 0)\right)$$

and the vector of parameters can be estimated using:

$$\hat{\beta}^{QR} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_\tau\left(y_i - x_i^T \beta\right)$$

QAM

Quantile Additive Models (QAMs) ([3] chap. 5, [4]) are non-linear multivariate extensions of Quantile Regression.

$$\hat{\mathbf{g}}^{QAM} = \arg\min_{\mathbf{g}} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_\tau\left(y_i - \mathbf{g}(x_i)\right)$$

$$\begin{aligned} \mathbf{g}(\mathbf{x}_i) &= g_1(x_{1i}) + g_2(x_{2i}) + \ldots \\ &+ g_p(x_{pi}) + g_{12}(x_{1i}, x_{2i}) + g_{13}(x_{1i}, x_{3i}) \\ &+ \ldots + g_{123}(x_{1i}, x_{2i}, x_{3i}) + \ldots \end{aligned}$$

First, function $\mathbf{g}$ can be represented as a **weighted sum of basis functions** leading to a representation as: $\mathbf{g}(x_i) = \mathbf{X}_i^T \boldsymbol{\beta}$. Second, as the model is overparametrized, we introduce **smoothing penalties on the g_j to control the bias-variance tradeoff**. For a smooth curve this might be: $\int (g''(x))^2 dx$ which, given a basis can be re-written as $\boldsymbol{\beta}^T S_j \boldsymbol{\beta}$:

$$\hat{\beta}^{QAM-Penalty} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_\tau\left(y_i - \mathbf{X}_i^T \boldsymbol{\beta}\right) + \frac{1}{2} \sum_{j=1}^{p} \lambda_j \boldsymbol{\beta}^T S_j \boldsymbol{\beta}$$

GACV for Hyperparameters

Yuan [5] based on the work of Craven and Wahba [6] and Nychka et al. [7] propose an approximation of Leave One Out Cross Validation (LOOCV) where only the knowledge of $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}$ is required to determine the vector of hyperparameters / smoothing parameters $\boldsymbol{\Lambda} = (\lambda_1, \ldots \lambda_p)^T$.

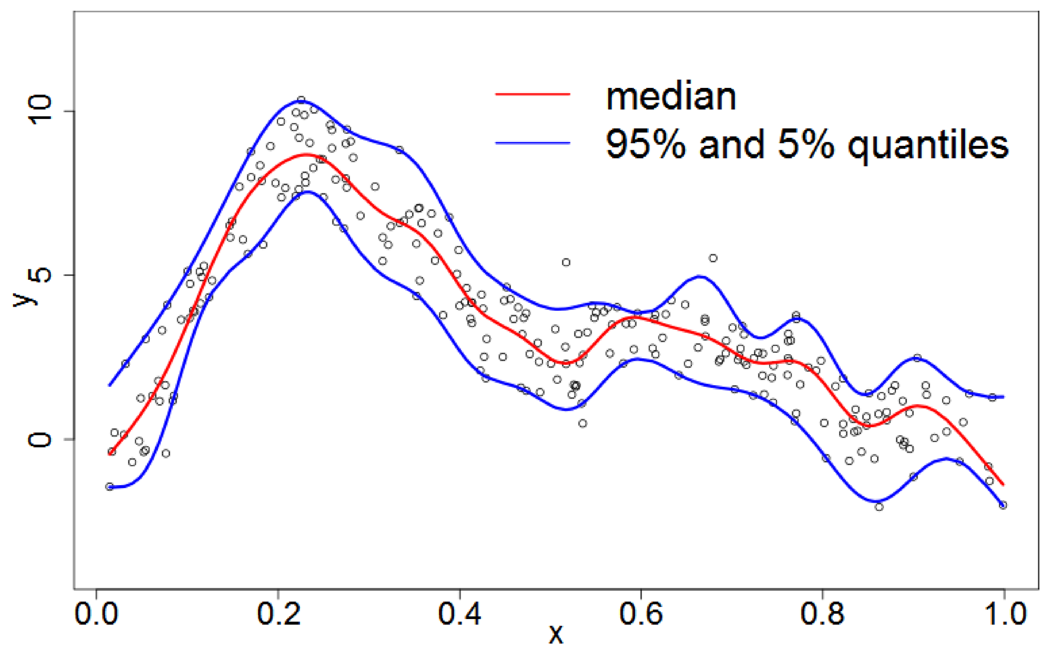
$$GACV_{\alpha,\tau}(\boldsymbol{\Lambda}) = \frac{\sum_{i=1}^{n} \mathcal{L}_{\alpha,\tau}\left(y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}{n - edf_{\alpha,\tau}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}$$

$\mathcal{L}_{\alpha,\tau}(u)$ is a smooth approximation of $\mathcal{L}_\tau(u)$

$edf_{\alpha,\tau}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)$ is the effective degrees of freedom

As $\alpha \to 0$ $\mathcal{L}_{\alpha,\tau}(u) \to \mathcal{L}_\tau(u)$

However, there are two issues: **a) Λ is optimized using a grid approach** and **b) Reiss and Huang** [8] **note that the GACV tends to undersmooth for non-central quantiles.** We also note that although the central quantile is often fitted correctly, it is sometimes undersmoothed. Below is an example were extreme quantiles and the median are undersmoothed.



Prop. 1: QGACV

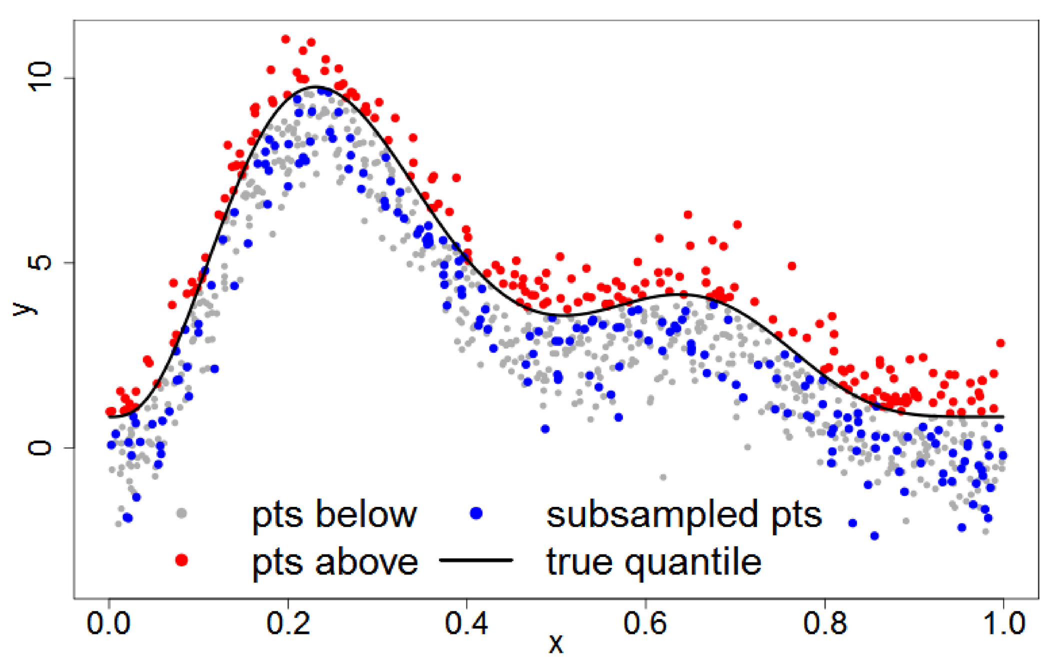
The derivation of the GACV is based on averaging weights of an ACV criterion ([7], [5]) ($h_{\alpha,\tau,ii}$ are diag. elements of the 'hat' matrix).

$$ACV_{\alpha,\tau}(\boldsymbol{\Lambda}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathcal{L}_{\alpha,\tau}\left(y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}{1 - h_{\alpha,\tau,ii}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}$$

As pointed out by Reiss and Huang [8], this averaging does not take into account that the weights on each side of the regression curve are vastly different. However, we note that **the same quantile regression curve could be obtained if we subsampled from the side with the highest number of points and we applied a symmetric loss function resulting in more even weights**. We then derive:

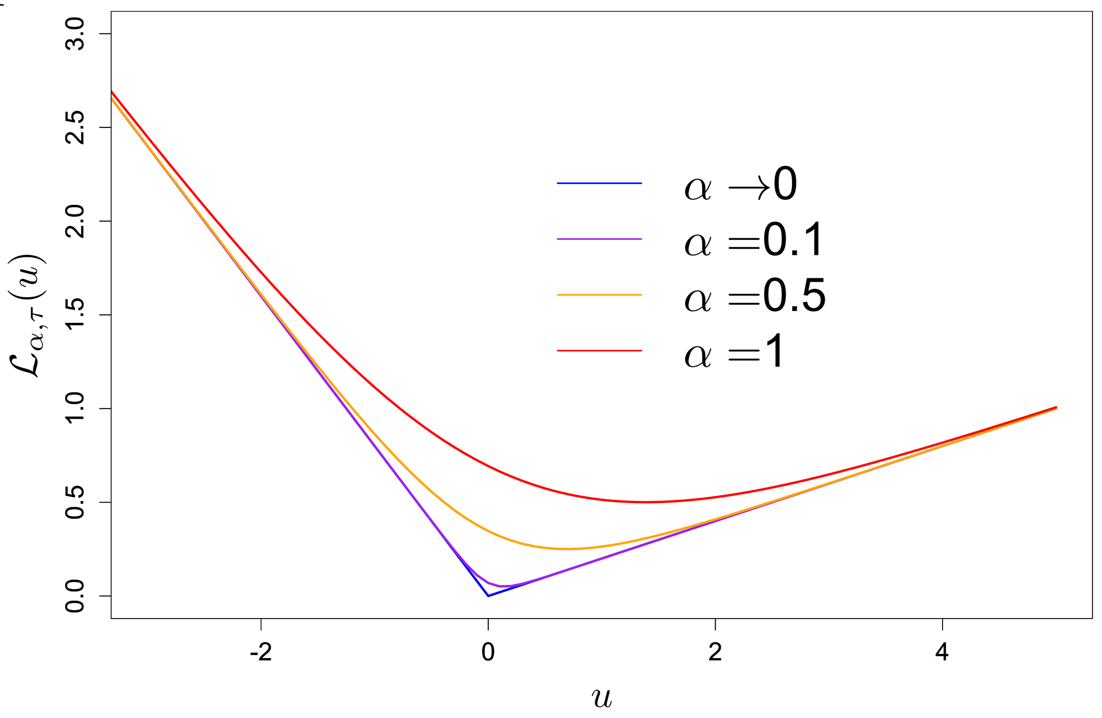
$$QGACV_{\alpha,\tau}(\boldsymbol{\Lambda}) = \frac{\sum_{i=1}^{n} \mathcal{L}_{\alpha,\tau}\left(y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}{2n\phi - (2\phi + 1)\, edf_{\alpha,\tau}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{\Lambda}}\right)}$$

$$\phi = \min(\tau, 1 - \tau)$$

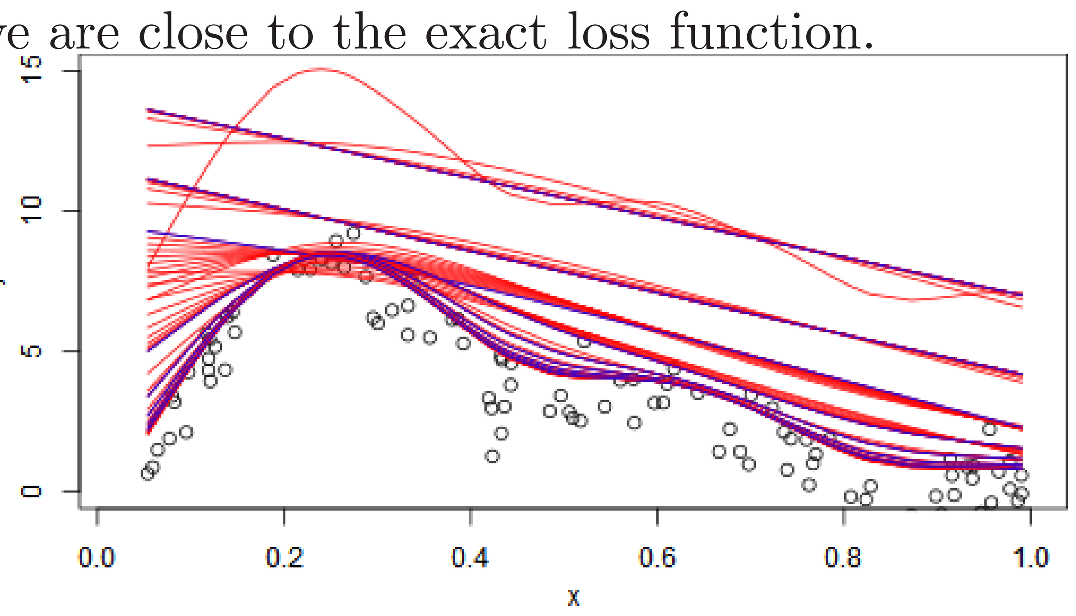


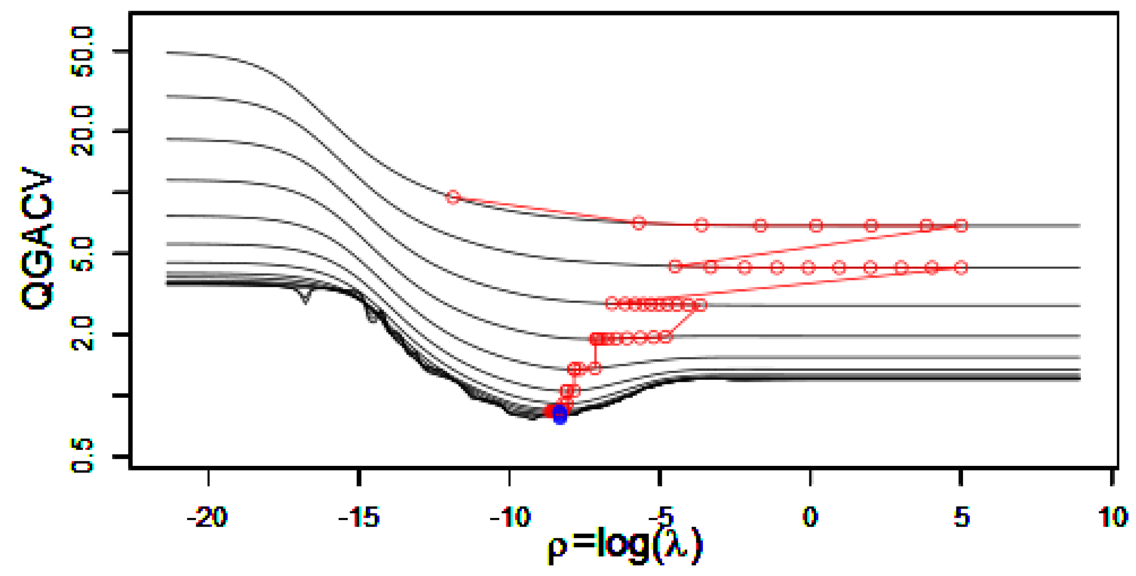
Prop. 2: graduated optim.

Parameter α determines the degree of approximation of the loss function.



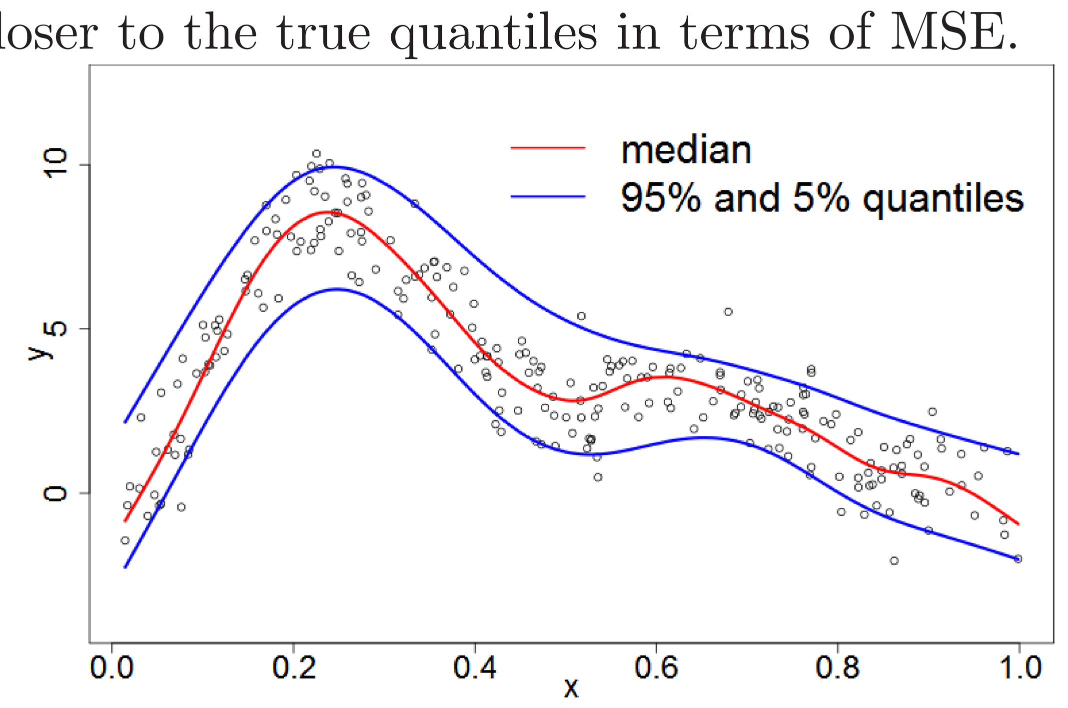
Although the QGACV criterion is a non-convex function of $\boldsymbol{\Lambda}$, we note that **as α increases, the QGACV function becomes closer to a quasi-convex function**. We then propose to use **Graduated Optimization/Non-Convexity (Blake and Zisserman [9], chap. 7)** to determine the vector of smoothing parameters $\boldsymbol{\Lambda}$. It consists in solving the optimization problem at decreasing values of α (i.e. we start at a large value of $\alpha = \alpha_{start}$ optimize, then reduce α and optimize, starting from the previous minimum and so on until we reach a value α_{goal} that is 'optimal'). Once the vector of smoothing parameters $\boldsymbol{\Lambda}$ is known, we continue to decrease α to determine the vector of parameters $\boldsymbol{\beta}$ until we are close to the exact loss function.





Results/Conclusion

Using QGACV/Graduated Optimization both the central and extreme quantile fits are closer to the true quantiles in terms of MSE.



References/Acknowledgements

[1] Roger Koenker, Gilbert Bassett Jr., 'Regression Quantiles', *Econometrica: Journal of the Econometric Society*, pp.33–50, 1978.

[2] Roger Koenker, Stephen Portnoy, Pin Tian Ng, Achim Zeileis, Philip Grosjean, Brian D. Ripley, 'R Package quantreg', 2018.

[3] Peter Bloomfield, William L. Steiger, 'Least Absolute Deviations: Theory, Applications and Algorithms', *Springer*, 1984.

[4] Matteo Fasiolo, Yannig Goude, Raphael Nedellec, Simon N. Wood, 'Fast Calibrated Additive Quantile Regression', *arXiv preprint*, arXiv:1707.03307, 2017.

[5] Ming Yuan, 'GACV for Quantile Smoothing Splines', *Computational Statistics & Data Analysis*, Volume 50, Issue 3, pp.813-829, 2006.

[6] Peter Craven, Grace Wahba, 'Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation.', *Numerische Mathematik*, Volume 31, Issue 4, pp.377-404, 1978.

[7] Doug Nychka, Gerry Gray, Perry Haaland, David Martin, Michael O'connell, 'A Nonparametric Regression Approach to Syringe Grading for Quality Improvement', *Journal of the American Statistical Association*, Volume 90, No 432, pp.1171-1178, 1995.

[8] Philip T. Reiss, Lei Huang, 'Smoothness Selection for Penalized Quantile Regression Splines', *The International Journal of Biostatistics*, Volume 8, Issue 1, Article 10, 2012.

[9] Andrew Blake, Andrew Zisserman, 'Visual Reconstruction', *MIT Press*, 1987.

This poster uses class 'baposter' and is based on poster examples created by Brian Amberg. **http://www.brian-amberg.de/uni/poster/**