Multiple smoothing parameters selection in additive regression quantiles

Vito M.R. Muggeo [1], Federico Torretta [1], Paul H. C. Eilers [2] Mariangela Sciandra [1] and Massimo Attanasio [1]

[1] Dipartimento di Scienze Economiche, Aziendali e Statistiche, University of Palermo, Palermo, Italy

[2] Erasmus University Medical Centre, Rotterdam, the Netherlands

---

Address for correspondence: Vito Muggeo, Dipartimento di Scienze Economiche, Aziendali e Statistiche, viale delle Scienze, Edificio 13, 90128 Palermo, Italy.

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

Phone: (+39) 091 23895240.

Fax: (+39) 091 485612.

Abstract: We propose an iterative algorithm to select the smoothing parameters in additive quantile regression, wherein the functional forms of the covariate effects are unspecified and expressed via B-spline bases with difference penalties on the spline coefficients. The proposed algorithm relies on viewing the penalized coefficients as random effects from the symmetric Laplace distribution and it turns out to be very efficient and particularly attractive with multiple smooth terms. Through simulations we compare our proposal with some alternative approaches, including the traditional ones based on minimization of the Schwarz Information Criterion. A real-data analysis is presented to illustrate the method in practice.

---

1   Introduction

Quantile regression (QR) is nowadays a well established framework in observational studies when interest is in modelling the quantiles of the response variable as a function of one or more covariates (Austin et al., 2005; Li et al., 2010). Implementation of QR was first discussed by Koenker and Bassett (1978), and since then many tutorials and books have appeared in the literature, the most notable being the one by Koenker (2005). Some advantages of QR with respect to the more usual mean regression include robustness to possible outliers and influential observations, and the ability to provide a complete picture of the response conditional distribution, see Austin et al. (2005) and Waldmann (2018) for a gentle introduction and discussion.

The linear QR model belongs to the mainstream of statistical methodology now, and the availability of specialized software makes it as simple to use as the familiar mean regression model. See the *quantreg* package in R, the *quantreg* procedure in SAS, and the *qreg* command in Stata, all of them focussing mainly on linear QR. However, when linearity is questionable, alternative and more flexible approaches should be considered. Nonparametric or flexible modelling in QR is crucial in data analysis with important and noteworthy applications in many fields, the most famous one being growth charts (Cole and Green, 1992; Wei et al., 2006; Li et al., 2010; Muggeo et al., 2013). Here the anthropometric variable of main interest, e.g. weight or height, is regressed on age, and a flexible relationship has to be fitted to obtain 'reference' values; the ultimate goal could be to identify observations out of such reference intervals. Unfortunately flexible modelling of nonlinear effects within QR appears to be somewhat limited, especially when multiple smooth, yet unspecified, relationships have to be included in the same regression equation. In this context, the gap with the corresponding counterparts for mean regression, such as the generalized additive models (Wood, 2006), is somewhat substantial. A possible reason limiting the widespread usage of *additive* QR is the lack of efficient algorithms able to fit additive QR with automatic choice of the smoothing parameters.

Nonparametric smoothing in QR has been discussed using different techniques, such as kernel (Xiang, 1996; Yu and Jones, 1998; Liu et al., 2019), smoothing splines from both a frequentist as well as Bayesian paradigm (Thompson et al., 2010), and low rank splines with or without penalty (Wei et al., 2006; He and Shi, 1994; Ng and Maechler, 2007). Unpenalized splines strongly depend on the number and the positions of knots and therefore these are not recommended, especially when there exist

regions in the covariate range with few observations. Penalized splines are effective to deal with such 'unlucky' data configurations: Bosch et al. (1995) express the problem in terms of cubic smoothing splines with an associated quadratic penalty, Koenker et al. (1994) also use smoothing splines but with a total variation penalty on the first derivative, and Ng and Maechler (2007) use splines with L_1 or L_∞ roughness measures. Bollaerts et al. (2006) employ B-splines with a L_1 penalty on the first or second order differences of the coefficients. All the aforementioned papers proposing the penalized approach, rely on a smoothing parameter which has to be selected for producing the final fit. Methods to select the smoothing parameter include the traditional cross validation (Bollaerts et al., 2006) and the Schwarz Information Criterion (Koenker et al., 1994; Koenker, 2011). Additional less-known approaches to select the smoothing parameter in QR are the generalized approximate cross validation understood as an approximation of the generalized comparative Kullback-Leiber distance for quantile smoothing splines (Yuan, 2006), and the relatively new criterion based on the L-curve originally proposed by Hansen (1992), discussed by Frasso and Eilers (2015), and employed in the nonparametric QR by Andriyana et al. (2014). All these methods are far from being efficient because they typically rely on grid search. Namely one fixes a pre-specified grid of smoothing parameter values, and fits the model for every candidate value in the grid. Then, the final fit is selected according to the best value of the criterion. There are two possible pitfalls with this approach: the optimal smoothing parameter value may depend on the number and position of the "candidate" values and, more importantly, the computational burden becomes particularly expensive when the regression equation involves multiple additive components with the consequent multidimensional grid of smoothing parameters. To avoid grid search, alternatively one could employ some numerical optimization methods to optimize the objective as a function of the smoothing parameters only, for instance via the simulated annealing option of `optim` in R, as in Koenker (2011). However the computational load still remains an issue.

To the best of our knowledge there are some R packages implementing nonparametric QR. Without the presumption of being exhaustive we cite some options. The well-known `quantreg` package (Koenker, 2016) includes the function `rqss()` which allows multiple smooth terms via `qss`, but it does not include automatic selection of the smoothing parameters. The R package `cobs` (Ng and Maechler, 2016) has a built-in function to perform automatic selection of smoothing parameter via the SIC, but it works just with a single covariate. For additive QR models we mention the R packages `mboost` (Hothorn et al., 2018; Mayr and Hofner, 2018) for a boosting-based approach, and the recent `qgam` (Fasiolo et al., 2018) relying on the extended log-F distribution

for the response. Both packages make use of P-splines, but not of L_1 penalty.

The goal of this paper is to set up an efficient algorithm to fit additive QR models with L_1 penalties where multiple smooth terms are included with corresponding smoothing parameters to be estimated. The approach is similar to Geraci and Bottai (2007), in that some coefficients are viewed as random effects coming from a Laplace distribution. However, unlike Geraci and Bottai (2007), we deal with smoothing methods and end up with a rather plain and efficient algorithm without performing Gibbs sampling to solve intractable integrals. We presented our idea in a statistical workshop (Torretta et al., 2015); here we provide details, justification of the approach and report results from extensive simulation studies.

The rest of the paper is structured as follows. In section 2 we briefly review the P-spline framework for smoothing in QR; in section 3 we describe the proposed algorithm in detail, and in section 4 we present results from some simulation experiments. Section 5 is devoted to the analysis of a real dataset and finally section 6 reports conclusions and some discussions.

2   P-spline quantile regression framework

Let Y be the quantitative response variable, $Q_Y(\tau|z, x)$ the τ-th quantile of the distribution of Y conditional to covariates $z_1, z_2, \ldots, z_p$ and $x_1, x_2, \ldots, x_K$. The xs are quantitative and understood to affect the response quantile through flexible relationships; the zs are further covariates, possibly categorical, entering the model linearly. The additive QR equation can be expressed via

$$Q_{Y_i}(\tau|z_i, x_i) = z_i^T \beta_\tau + s_{1\tau}(x_{1i}) + \ldots + s_{K\tau}(x_{Ki}), \qquad (2.1)$$

where β_τ quantifies the linear effect of covariates z, and the $\{s_{k\tau}(x_k)\}_{k=1,\ldots,K}$ are smooth but otherwise unspecified functions. Following Bollaerts et al. (2006), Ng and Maechler (2007), and Muggeo et al. (2013) among others, we use a low-rank B-spline basis as smoother, thus $s_{k\tau}(x_{ki}) = \sum_j^{J_k} b_{k\tau j} B_{kj}(x_{ki})$, where $B_{kj}(x_i)$ is the j-th spline function of the basis $B_k(x_{ki}) = (B_{k1}(x_{ki}), B_{k2}(x_{ki}), \ldots, B_{kJ_k}(x_{ki}))$ evaluated at x_{ki}, and the $b_{k\tau j}$s are J_k spline coefficients of the vector $b_{k\tau}$ (Eilers and Marx, 1996). J_k is usually taken as a large number to build a generous basis guaranteeing potentially flexible fits, for instance $J_k = min\{40, n/4\}$, as suggested by Ruppert

et al. (2003). When the regression equation involves multiple smooth terms as in (2.1) each basis has to be made identifiable. Among the different approaches, we employ the sum-to-zero constraints such that the estimated smooth curve sums to zero over the observed covariate space (Wood, 2006). Hereafter we assume each basis is identifiable and the model intercept is included among the linear parameters β_τ.

Denoting the response quantile in (2.1) by Q_i, the penalized objective function to be minimized can be written as

$$\mathcal{L}_\lambda(\beta, b_1, \dots, b_K) = \sum_{i=1}^{n} \rho_\tau(y_i - Q_i) + \sum_{k=1}^{K} \left\{ \lambda_{k\tau} \sum_{j=1}^{J_k - d_k} |\Delta^{d_k} b_{k\tau}|_j \right\}, \qquad (2.2)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, and the K penalties $\lambda_{k\tau} \sum_{j=1}^{J_k - d_k} |\Delta^{d_k} b_{k\tau}|_j$ control the wiggliness of the corresponding fitted curves $\hat{s}_{k\tau}(\cdot)$ by shrinking the relevant spline coefficients differences. More specifically Δ^{d_k} is the d_k order difference operator applied to the spline coefficients vector b_k; for instance if $d_k = 1$ it is $\sum_{j}^{J_k - d_k} |\Delta^{d_k} b_{k\tau}|_j = |b_{k\tau 1} - b_{k\tau 2}| + |b_{k\tau 2} - b_{k\tau 3}| + \dots + |b_{k\tau J_k - 1} - b_{k\tau J_k}|$. $\lambda_{k\tau}$ is the positive smoothing parameter regulating under- or over-smoothing of the fitted curve. Like for mean regression, the fitted curve $\hat{s}_k(\cdot)$ approaches a $d_k - 1$ degree polynomial when $\lambda_{k\tau}$ gets larger, while $\lambda_{k\tau} = 0$ indicates no penalization, resulting in a potentially wiggly curve.

For fixed smoothing parameter, minimization of (2.2) is a relatively simple task since standard linear programming techniques may be used (He and Ng, 1999; Bollaerts et al., 2006; Koenker, 2011). Unfortunately the smoothing parameter is not fixed in practice, and as briefly discussed in the Introduction, all of the current methods working with L_1 penalties, select it via grid-search or numerical optimization. This is sustainable for a single λ, but it can become time-consuming and possibly infeasible with additive models where multiple smoothing parameters λ_k have to be estimated. For instance, to alleviate the computational burden, Bang and Jhun (2012) and Andriyana et al. (2014) propose to search for a unique λ, and compute the term specific smoothing parameters via the 'ad hoc' adjustment $\lambda_k = \lambda / max\{|\hat{b}_{jk}|\}_{j=1,\dots,J_k}$ or $\lambda_k = \lambda / sdev\{|\hat{b}_{jk}|\}_{j=1,\dots,J_k}$, where the $\hat{b}_{jk}$s are estimates from a preliminary unpenalized fit. However such approach can clearly produce non-optimal solutions. Alternatively two recent approaches to smoothing in QR deal with L_2 penalty, via the boosting (Fenske et al., 2013; Mayr and Hofner, 2018) or by means of a smooth approximation of the check loss (Fasiolo et al., 2018). In the next section we discuss an iterative algorithm to select the λ_ks in an additive QR (2.2) with L_1 objective and penalties.

3   The proposal: an iterative algorithm for smoothing parameter selection in QR

To estimate the multiple lambda parameters of objective (2.2), we propose to use an iterative algorithm which is sometimes referred in literature as 'Schall algorithm'. However the approach was discussed by Fellner (1986) for robust estimation of linear mixed models based on equations of Harville (1977), and extended to generalized mixed models by Schall (1991). The algorithm has also been employed for expectile smoothing by Schnabel and Eilers (2009). The underlying idea of such 'Harville-Fellner-Schall' algorithm exploits the link between penalized smoothing methods and random effects models: the penalized coefficients are viewed as random effects having relevant variance parameters, see Currie and Durban (2002) and Wand (2003) for details. Viewing the penalized coefficients as random effects from a known distribution allows us to estimate the smoothing parameter as the ratio of the error variance divided by the variance of the 'random effects', namely of the penalized coefficients. We apply a similar idea to estimate the smoothing parameters in nonparametric QR via minimization of (2.2). We first outline the algorithm and then postpone discussion.

The proposed algorithm for additive QR (2.1) at fixed τ, with p linear parameters and K smooth terms, is summarized as follows.

1. Fix a (small) value for all smoothing parameters $\lambda_{k\tau}^{(0)}$, $\quad k = 1, \ldots, K$;

2. Fit the QR (2.1) by minimizing the objective (2.2) at the fixed $\lambda_{k\tau}^{(0)}$ such that parameter estimate $\hat{b}_{k\tau}$ and fitted quantiles $\widehat{Q}_i$ are obtained;

3. Compute $\hat{\phi}_\tau = \frac{\sum_i \rho_\tau(y_i - \widehat{Q}_i)}{(n - p - \sum_k \gamma_k edf_k)}$ and $\hat{\psi}_{k\tau} = \frac{\sum_j |\Delta^{d_k} \hat{b}_{k\tau}|_j}{\gamma_k edf_k}$ $k = 1, \ldots, K$, where edf_k are the term-specific degrees of freedom to be defined later and $\gamma_k \geq 1$ is a fixed factor to further penalize for the complexity;

4. Compute $\hat{\lambda}_{k\tau} = \frac{\hat{\phi}_\tau}{\hat{\psi}_{k\tau}}$, $\quad k = 1, \ldots, K$;

5. Set $\hat{\lambda}_{k\tau} \to \lambda_{k\tau}^{(0)}$ $(k = 1, \ldots, K)$ and repeat steps 2. to 4. till convergence.

Convergence in step 5 is established throughout relative changes in the smoothing parameter values. Namely when the variation in absolute value between the current and the previous iteration of each λ_k is less than a fixed tolerance, the algorithm stops by returning the QR model fitted at the last λ value.

3.1   A theoretical justification of the iterative algorithm

Here we illustrate the rationale behind the proposed algorithm. For the sake of simplicity we consider just one set of coefficients being penalized with a single tuning parameter λ, but extension to multiple λ is straightforward.

It is known that the objective function in QR may be obtained assuming an asymmetric Laplace ($\mathcal{AL}$) distribution for the response conditional distribution (Yu and Moyeed, 2001; Geraci and Bottai, 2007). Formally $Y_i|b \sim \mathcal{AL}(Q_i, \phi, \tau)$ with density function $f(y_i|b; Q_i, \phi, \tau) = \frac{\tau(1-\tau)}{\phi} \exp\{-\frac{\rho_\tau(y_i - Q_i)}{\phi}\}$, where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the check function like in (2.2). Q_i is the location parameter, $\phi > 0$ is the scale parameter, and $\tau \in (0,1)$ the skewness parameter assumed known and fixed, see Yu and Zhang (2005). Let the location parameters Q_i depend on the values x_{ij}s via the equation $Q_i = \sum_j^J b_j x_{ij}$, and let the coefficients $(b_1, \ldots, b_J)$ assumed independent random effects from a zero-mean symmetric Laplace distribution with scale parameter ψ, whose joint density is $f(b_1, \ldots, b_J; \psi) = (2\psi)^{-J} \exp\{\sum_j |b_j|/\psi\}$.

Hence, given the aforementioned assumptions and n independent observations, the joint density $f(y_1, \ldots, y_n, b_1, \ldots, b_J)$ representing the likelihood is

$$\left\{ \frac{\tau(1-\tau)}{\phi} \right\}^n (2\psi)^{-J} \exp\left\{ -\frac{\sum_i \rho_\tau(y_i - Q_i)}{\phi} - \frac{\sum_j |b_j|}{\psi} \right\}.$$

By using the simple re-parameterization $\psi = \frac{\phi}{\lambda}$ and taking the log, yields the log likelihood, apart from a constant,

$$\ell = -n \log \phi - J \log \phi + J \log \lambda - \frac{\sum_i \rho_\tau(y_i - Q_i)}{\phi} - \frac{\lambda \sum_j |b_j|}{\phi}. \qquad (3.1)$$

If ϕ and λ are considered as nuisance, then maximization of (3.1) with respect to the b_js justifies minimization of (2.2). Actually (2.2) is more general as it includes multiple λ_k and penalty on the differenced coefficients, however the rationale is clearly the same.

The partial derivatives of (3.1) are easily obtained

$$\frac{\partial \ell}{\partial \lambda} = \frac{J}{\lambda} - \frac{\sum_j |b_j|}{\phi}$$

$$\frac{\partial \ell}{\partial \phi} = \frac{1}{\phi^2} \left\{ -\phi(n + J) + \sum_i \rho_\tau(y_i - Q_i) + \lambda \sum_j |b_j| \right\}$$

The root of $\frac{\partial \ell}{\partial \lambda} = 0$ is $\hat{\lambda} = \frac{\phi}{\sum_j |b_j|/J}$, while by the estimating equation $\frac{\partial \ell}{\partial \phi} = 0$, after plugging in $\hat{\lambda}$, we get $\hat{\phi} = \sum_i \rho_\tau(y_i - Q_i)/n$. Namely the maximum likelihood (ML) estimate of scale parameters ratio λ is

$$\hat{\lambda} = \frac{\sum_i \rho_\tau(y_i - \widehat{Q}_i)/n}{\sum_j |\hat{b}_j|/J} = \frac{\hat{\phi}}{\hat{\psi}}, \tag{3.2}$$

which justifies step 4. in the aforementioned algorithm, apart from the denominators n and J. The rationale of replacing them with the effective degrees of freedom and details about their computation are discussed in next subsections.

3.2 Remarks

As illustrated above, parameter estimation is justified via maximization of the joint likelihood depending on fixed and random, i.e. penalized, parameters. Unlike Geraci and Bottai (2007), it should be stressed that no tentative to integrate out the random effects is carried out, and thus the objective to be optimized is the joint, rather than the marginal, likelihood. It is worth noting that in the usual mixed models framework, the variances ratio expression, i.e. the counterpart of (3.2) for mean regressions, typically comes from the marginal likelihood maximization. However as discussed by McCulloch (1997), it can be motivated under different likelihoods, including the marginal, penalized and joint likelihood of fixed and random parameters as taken in this paper for smoothing QR.

The scale parameter estimate $\hat{\psi}$ in (3.2) depends on the numerator $\sum_j |\hat{b}_j|$ which, in turn, relies on independence of random effects as shown in subsection 3.1. Actually the coefficients of a B-spline basis with a difference penalty are not associated with independent random effects, but it is straightforward to transform the B-spline functions in order to take independence of the b_js (Currie and Durban, 2002). As an alternative (Eilers and Marx, 2010), it is also possible to keep the B-spline basis and to take the differenced coefficients $\sum_j^{J-d} |\Delta^d b|_j$ as depicted in step 3. of the above algorithm.

As illustrated by (3.2) in the previous section, $\hat{\lambda}$ is the joint ML estimate under the Laplace assumptions for the y_is and b_js. In linear mixed models, a restricted ML (REML)-like approach is well-known to attenuate bias of the variance parameter estimators and to improve greatly their performance (e.g. Fellner, 1986; Schall, 1991; Wand, 2003; Wood, 2006). Therefore we speculate a similar behaviour apply in QR. Unfortunately a restricted version of log likelihood (3.1) is not straightforward

to obtain and it has not been derived formally. However a natural and intuitive choice to grant such adjustment is to consider the equivalent degrees of freedom in estimation of $\hat{\psi}$ and $\hat{\phi}$, namely replacing J and n respectively by the effective and the residual degrees of freedom. Preliminary simulation experiments showing the better performance of the REML-like with respect to the ML approach have also emphasized a general tendency to under-smooth, namely to produce relatively too wiggly fitted curves. A practical modification to fix this issue is to further penalize for the model complexity in the generalized cross validation score, see Wood (2006, page 178): the idea is to increase the amount that each edf_k counts. We undertake a similar approach, and therefore in step 3. of the algorithm we use REML-like estimates for the scale parameters wherein each edf_k is increased by a factor of $\gamma_k > 1$.

The proposed algorithm straightforwardly applies to so called varying coefficient models, i.e. regression equations involving also interactions between a smooth and a linear term, such as $z_i s_\tau(x_i)$, see Andriyana et al. (2014). Detailed discussion of varying coefficients is beyond the goal of this paper, but it is worth noting that to include varying coefficient terms it suffices to modify the basis spline such that each n-dimensional column is multiplied element-wise by the n-dimensional covariate vector. Hence the ith row of the B-spline related to the varying coefficient term can be written as $\tilde{B}(x_i) = (z_i B_1(x_i), z_i B_2(x_i), \ldots, z_i B_J(x_i))^T$, and difference penalty on coefficients applies straightforwardly.

3.3 Quantifying the effective model dimension

The aforementioned algorithm in step 3. requires to quantify the term-specific degrees of freedom edf_k. To the best of our knowledge, there is no consensus in quantifying them for additive QR. Some criteria aimed at selecting the amount of smoothing, such as the cross validation (Bollaerts et al., 2006) or L-curve (Andriyana et al., 2014), do not need term-specific degrees of freedom, while other criteria, such as the Akaike or the Schwarz Information Criterion just require the overall effective model dimension. For instance in Koenker (2011), the multiple smoothing parameters in additive QR, are selected by

$$\text{SIC}_\lambda = \log\left[n^{-1}\sum_i^n \rho_\tau(y_i - \widehat{Q}_i)\right] + 0.5n^{-1}edf \log(n), \qquad (3.3)$$

where only the total edf are computed via the null residuals; see also Koenker (2005, pp 134, 234) for a general discussion, and Li and Zhu (2008) for a rigorous proof on computing the model complexity via the number of interpolated points. In a

REML-like framework, the term specific edf_k should clearly depend on the selected λ_k, namely $edf_k = J_k$ when $\lambda_k = 0$, and $edf_k \to d_k - 1$ if $\lambda_k \to \infty$. Two approaches could be undertaken to compute edf_k at intermediate values of λ_k.

The L_1 penalty leads to null estimates of some basis coefficients involved in the penalty term. Thus a natural way to quantify complexity of the fitted curve is via the corresponding non-zero penalized estimates. More specifically, for the smooth term k expressed by a J_k-rank basis and $J_k - d_k$ penalized coefficient differences, we count the number of non-zero difference estimates and the number of underlying $d_k - 1$ unpenalized coefficients; formally $edf_k = \#\{\Delta^{d_k}\hat{b}_k \neq 0\} + (d_k - 1)$. For instance if we set $d_k = 2$ and λ_k gets very large, we get all zero estimated differences, leading to $edf_k = d_k - 1 = 1$ which corresponds to the resulting linear fit.

The second viable approach relies on a smooth approximation of the L_1 norm in both the fidelity and penalty term. The goal is to build an approximate hat matrix in order to define the edf_ks accordingly. Among the several smooth approximations which could be used (e.g. Muggeo et al., 2012), we use the simple identity $|u| = u^2/\sqrt{u^2}$ (Schnabel and Eilers, 2013). However while Schnabel and Eilers (2013) use that to estimate multiple quantile curves with P-splines via iterative least squares, we exploit the smooth approximation only to build the hat matrix. Thus, given parameter estimates obtained via optimization of the L_1-norm objective (2.2), residuals $e_i = y_i - \hat{Q}_i$ and weights $w_{\tau i} = \tau - I(e_i < 0)$ arranged into the matrix $W_\tau = \text{diag}(w_{\tau 1}, \ldots, w_{\tau n})$, we note the parameter estimates can be expressed via $(\hat{\beta}, \hat{b}_\lambda)^T = (X^T W_\tau X + P_\lambda)^{-1} X^T W_\tau y$, where X is the design matrix including linear covariates and the B-spline basis functions for the smooth terms and P_λ is a block diagonal matrix including the zeroes for unpenalized coefficients and the penalty matrices relevant to the different smooth terms. For instance, the block relevant to the kth smooth term would be $\lambda_k D_k^{d_k T} V_k D_k^{d_k}$, where $D_k^{d_k}$ is the d_k-order difference matrix and V_k includes the reciprocals of the same differenced coefficients such that the quadratic form penalty equals the original L_1-norm penalty, namely $b_k^T D_k^T V_k D_k b_k = \sum_j \frac{(\Delta^{d_k} b_k)_j^2}{\sqrt{(\Delta^{d_k} b_k)_j^2}} = \sum_j |\Delta^{d_k} b_k|_j$.

The least squares formulation allows to define the hat matrix whence the model degrees of freedom may be obtained. More specifically, the elements on the main diagonal of $(X^T W_\tau X + P_\lambda)^{-1} X^T W_\tau X$ represent the degrees of freedom of each coefficient associated to the corresponding column of the design matrix. By summing the J_k elements corresponding to the coefficients $b_{k1}, \ldots, b_{kJ_k}$ relevant to the kth B-spline basis, the term-specific edf_k are obtained. Of course the trace quantifies the overall

model *edf*.

4 Simulation Studies

We assess the performance of the proposed approach via some simulation experiments. We follow some settings employed in Koenker (2011), but we also consider additional scenarios. Data were generated according to $y_i = g(x_i) + \sigma(x_i)e_i$, where the covariate $x_i \sim U(0,1)$ and the signal $g(x_i)$ is defined according to four different functions: linear $0.2 + 0.4x_i$, logarithm $\log(x_i)$, sinusoidal $\sin(2\pi x_i)$, and 'square root sinusoidal' indicated by $g_0(x_i) = \sqrt{x_i(1-x_i)}\sin((2\pi(1+2^{-7/5}))/(x_i+2^{-7/5}))$. The scale function is either constant $\sigma(x_i) = 0.2$, or depending on the covariate itself, i.e. $\sigma(x_i) = 0.2(1+x_i)$. The errors e_i are *iid* from four different distributions, Gaussian, χ_3^2, t_3, and t_1, such that the τ-th quantile equals zero. The sample size is $n = 400$ and five percentiles $\tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$ are considered. To provide an idea of the signal to noise ratio in the considered scenarios, Figure 1 portrays a set of simulated data (at $\tau = 0.5$) according to the different signal functions and error distributions.
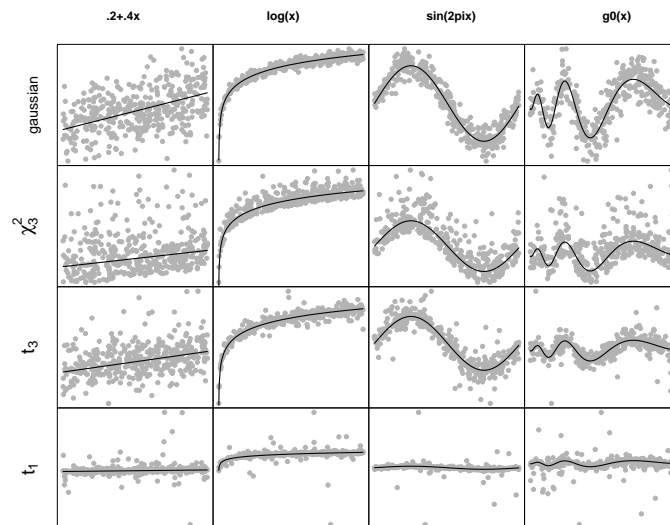

Figure 1: Some simulated data according to the signal curves and error distributions with constant scale function. The continuous line represents the true median signal.

In the manuscript we contrast three competitors: i) the *P*-spline smoother (using $min\{40, n/4\}$ cubic *B*-spline functions with difference penalty $d = 3$) and λ selected using the extra penalty factor $\gamma = 2$ in the Harville-Fellner-Schall step 3 of the proposed algorithm ('psplines+HFS'); ii) the same *P*-spline smoother with λ selected

by the SIC_λ reported in (3.3), ('psplines + SIC'); iii) the quantile smoothing splines with a total variation penalty and λ selected again by minimization of the SIC_λ, ('ssplines + SIC'). Differences of the proposed 'pspline+HFS' with respect to other possible approaches, such as cross-validation (CV) and the so-called L-curve criterion, are reported in Supplementary Material. Both CV and L-curve have been employed in literature to select just a single and not multiple, smoothing parameter in QR. Extension of ten-fold, say, CV is prohibitive in practice due to its computational load, and L-curve for bivariate smoothing has been proposed in linear regression only (Frasso and Eilers, 2013), and further studies and research are needed to assess its applicability in additive QR (2.1) . Therefore both CV and L-curve could not be considered as effective competitors in additive QR, and discussion and comparative assessment has been reported in supplementary material which also includes some comparisons with the boosting approach.
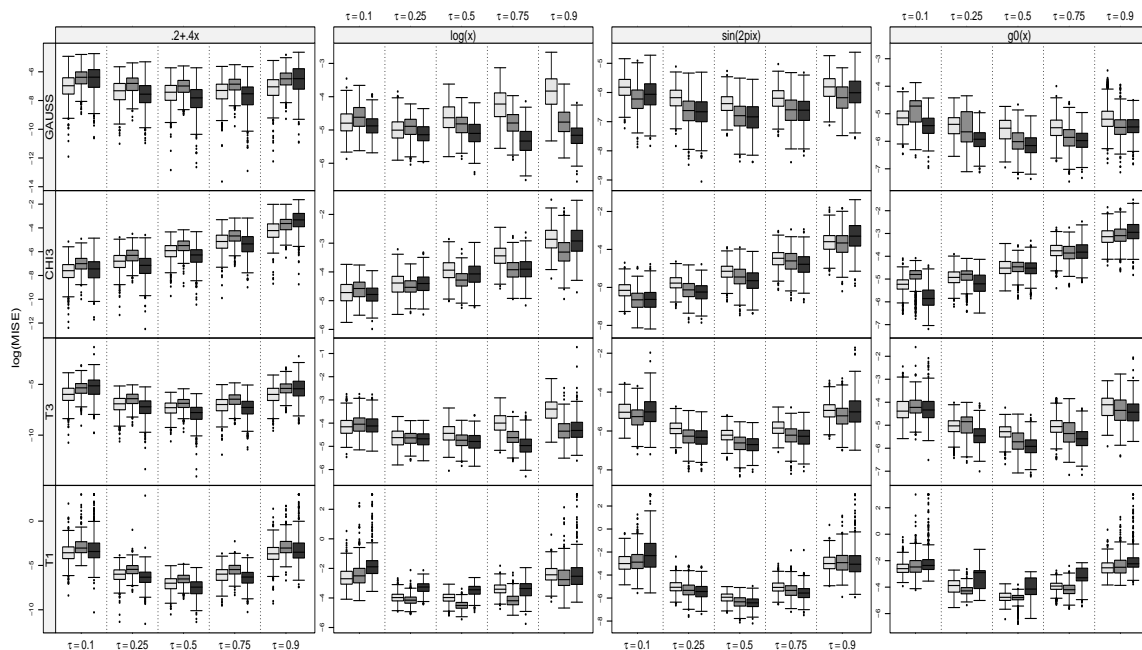

Figure 2:   Constant scale function $\sigma(x) = 0.2$. Contrasting three competitors for smoothing parameter selection in terms of MISE (on log scale) by different error distributions and signals: 'sspline+SIC' (light grey box), 'pspline+SIC' (medium grey box) and 'pspline+HFS' (dark grey box).

The three aforementioned main approaches are contrasted in terms of Mean Integrated Square Error (MISE) defined as the mean of squared differences between the true quantile curve and the corresponding fitted curve across 500 replicates, and also
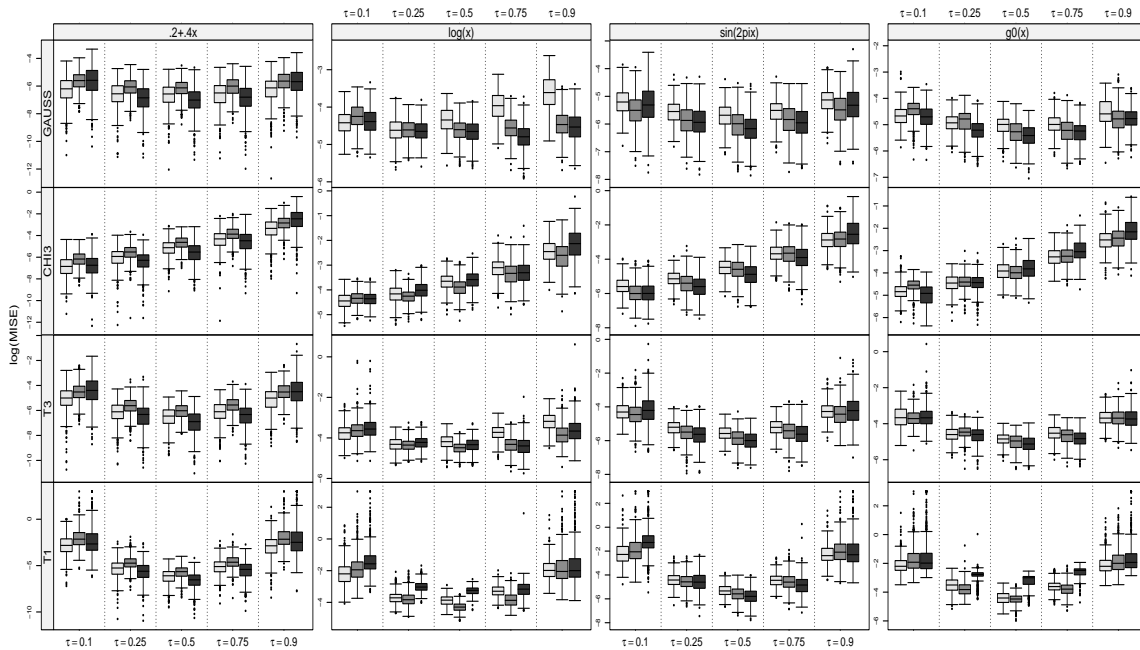
Figure 3:   Non-constant scale function $\sigma(x) = 0.2(1 + x)$. Contrasting three competitors for smoothing parameter selection in terms of MISE (on log scale) by different error distributions and signals: 'sspline+SIC' (light grey box), 'pspline+SIC' (medium grey box) and 'pspline+HFS' (dark grey box).

the Mean Integrated Absolute Error, similar to MISE but involving the absolute differences rather than the squares. Here we report results for the MISE values, while the MIAE results are reported in the Supplementary Material.

Figures 2 and 3 report the comparisons in terms of MISE: for all the error distributions considered and for every signal (linear, logarithmic, sinusoidal and sqrt+sinusoidal '$g_0(\cdot)$'), each panel reports the boxplots of the log MISE values across the replicates for the five τ values. The two figures refer to homoscedastic and heroscedastic scenarios.

Overall, no important difference emerges across the scenarios. Occasionally some criterion appears to perform slightly better/worse than the others: for instance 'sspline+SIC' returns higher MISE values at higher quantiles with $\log(x)$ signal and gaussian errors; the proposed 'pspline+HFS' exhibits slightly lower mean square errors at linear signal and middle percentiles (0.25,0.50,0.75), but at the same middle percentiles, it depicts higher values when the signal is $\log(x)$ or $g_0(x)$ and errors follow the Cauchy distribution. However no systematic patterns come out, and no difference among the homoscedastic or heteroscedastic case is noteworthy. Hence,

while performances in terms of statistical efficiency are substantially the same across the scenarios, the gain in computational efficiency is not negligible; more details are postponed to section 4.1.

The same simulation scenarios were run with an additional linear covariate to assess possible impacts on the sampling distribution of the linear coefficient estimators. No difference was observed among the MISE values and the sampling distributions and results are not shown for shortness.

We consider the more interesting scenarios of additive models, namely more smooth relationships in the QR equation. To illustrate, we consider the true quantile function $1 + 2\cos(x_{1i}) + \sin(2\pi x_{2i})$ with independent uniform covariates $x_{1i} \sim U(-4, 4)$ and $x_{2i} \sim U(0, 1)$, sample size $n = 400$, the four aforementioned error distributions and five quantile curves. Figure 4 shows the log MISE, for the aforementioned competitors 'sspline+SIC', 'pspline+SIC' and the proposed 'pspline+HFS'.

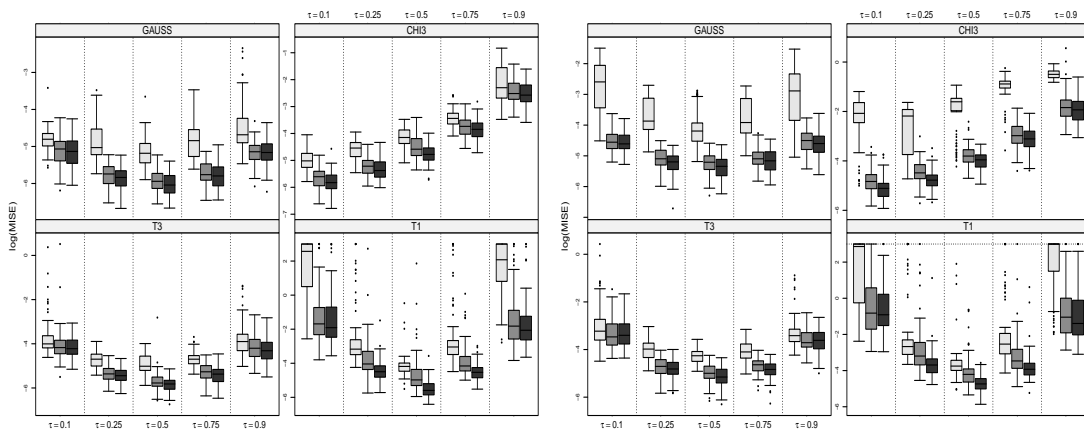Figure 4:   Contrasting the three competitors (in terms of log(MISE)) for smoothing parameter selection with 2 additive terms by different error distributions and quantile curves: 'sspline+SIC' (light grey box), 'pspline+SIC' (medium grey box) and 'pspline+HFS' (dark grey box). Right panels: constant scale function; Left panels: non-constant scale function. The true signal is $1 + 2\cos(x_{1i}) + \sin(2\pi x_{2i})$.

Unlike the single covariate case, here some differences are noteworthy. For the homoscedastic scenario the traditional 'sspline+SIC' performs bad most of times, noticeably quite worse than the others with Gaussian and Cauchy errors especially at extreme quantiles ($\tau = 0.1$ and 0.90). On the other hand, 'pspline+SIC' and 'pspline+HFS' behave about the same, with 'pspline+HFS' exhibiting lightly but constantly lower MISE values at middle quantiles ($\tau = 0.25, 0.50, 0.75$) with somewhat

remarkable differences in the t_1-distribution case. With heteroscedastic errors patterns are even more pronounced, especially with asymmetric errors where the MISE values from 'sspline+SIC' are far higher than the others.

Differences in terms of MISE exhibit approximately the same patterns and are reported in the supplementary material which also shows comparisons relevant to smoothing parameter selection in varying coefficient models.

4.1   Computational issues

As previously discussed, computational efficiency is a major feature of our proposal. The benefit with respect to crude grid search approaches could be somewhat expected, especially when multiple smoothing parameters have to be estimated: in fact grid-search approaches require multidimensional evaluation grids leading to quite taxing procedures since a large number of fits have to be obtained to pick up the best lambdas values optimizing the selected criterion, such as the SIC or L-curve. CV is even heavier, since for each candidate lambda value in the grid, the model has to be fitted and tested several times to training/testing sub-datasets.

Table 1:   Computational times to fit additive quantile regresion with 1 or 2 smooth terms for $n = 1000$ observations. Entries refer to averages (on 10 fits) of the `elapsed` components obtained via the `system.time()` function in R. Times refer to building the B-spline basis and the penalty matrices. Computations run on R 3.6.1 on Windows, Intel i7-8700 CPU 3.20 GHz, RAM 16 GB.

n. smooth terms	criterion	method	execution time (sec.)
1	sspline+SIC	grid-search (15)	0.12
		grid-search (30)	0.21
	pspline+SIC	grid-search (15)	0.12
		grid-search (30)	0.22
	pspline+HFS	iterative	0.07
2	sspline+SIC	grid-search (15×15)	57.3
		grid-search (30×30)	228.7
	pspline+SIC	grid-search (15×15)	8.61
		grid-search (30×30)	33.1
	pspline+HFS	iterative	0.33

To gain a rough assessment of the computational load, Table 1 reports the execution

time of the aforementioned approaches when fitting additive regression quantiles with 1 or 2 smooth terms at $\tau = 0.5$ for $n = 1000$ observations. For the 'grid-search' rows, the numbers in parentheses refer to the number of lambda values to be evaluated to seek the optimum: we consider 15 or 30 values, with the former leading, as it would be expected, to lower excution times but higher mean square errrors.

While for single smooth terms, all times are within reasonable ranges, differences get notable with 2 smooth terms: grid-searches show quite large times in general, but sspline+SIC exhibits larger running times, probably due to the higher number of parameters to be estimated. The proposed pspline+HFS presents times far lower than the competitors, since it is based on iterative procedure and typically less than 20 iterations are requested to get convergence. It is worth stressing that such large differences are expected to rise as the number of smoothing parameters increases, making our proposal quite attractive with multiple additive terms.

For the sake of completeness, we mention the chosen criterion could be minimized via numerical procedures rather than grid-searches: for instance by means of the Nelder-Mead algorithm with does not require gradient evaluation, or quasi-Newton methods which compute the gradient numerically. However we have experienced a somewhat strong influence of the supplied starting values on the final results, making the numerical procedures substantially unreliable. One could try to run the algorithm using different starting values or probably to rely on different algorithms such as the simulated annealing, but at the cost of increasing the computational burden.

5   Application: modelling standing long jump in children

We apply the proposed algorithm to analyse data concerning the sport performance in children. Physical fitness is a powerful marker of the health condition in childhood and adolescence; for instance, fitness is negatively associated with cardiovascular risk factors for chronic disease, high blood pressure, and total fatness (Kraemer and Häkkinen, 2008).

As an alternative to laboratory methods, physical fitness can be measured via the on-site fitness tests which are easy to be administered, especially when the population study involves schoolchildren. Among the different sport performance outcomes, the

standing long jump (SLJ) test represents a widespread yet practical, time efficient and cheap method for assessing the muscular fitness in children and adolescents. The schoolchild stands at a line marked on the ground with the feet slightly apart and jumps: the SLJ performance is the horizontal distance jumped. SLJ, sometimes also known as standing broad jump, is commonly used to assess explosive leg power, but it is also understood to be a proxy of muscular strength tests of the lower body. SLJ depends mainly on leg length, but it influenced also by neuromuscular maturation, as SLJ requires more coordination of movements and technique, for instance the so-called takeoff angle (Saint-Maurice et al., 2015).

Data analyzed here refer to wide survey carried out in years 2011 to 2013 to assess sport abilities and fitness in school children in Sicily. Data have been kindly provided by the Department of 'Scienze Psicologiche, Pedagogiche e della Formazione', University of Palermo. Measurements were gained by previously trained experts, and include measurements on performance on several physical tests along with some anthropometric measurements collected in the major Sicilian cities in the three-year period. Here we present results relevant to $n = 488$ school children in Palermo collected in 2011; further results relevant to whole Sicily will be presented elsewhere.

Beside the response SLJ, the main covariate is the child weight. The substantive research question is how the weight could affect the SLJ performance, namely in statistical terms the relationship SLJ - weight which is expected to be nonlinear. However the weight is strongly correlated with age and height, which are understood to affect the response as well, possibly in a nonlinear manner. Thus an additive QR accounting simultaneously for the three covariates flexibly appears to be the most appropriate model, namely,

$$Q_{\mathtt{slj}}(\tau) = \beta_0 + \beta_1 \mathtt{gender} + s_{1\tau}(\mathtt{age}) + s_{2\tau}(\mathtt{weight}) + s_{3\tau}(\mathtt{height}),$$

at fixed τ. Multiple quantile curves could be useful to obtain the so-called growth charts, sometimes refereed as 'normative reference values' in sport medicine (Saint-Maurice et al., 2015; Sandercock et al., 2016). However focussing only on upper quantiles, would be useful for talent selection in school age. Thus we set $\tau = 0.95$ in the above regression equation.

We fit the additive QR model using the aforementioned algorithm; the smooth functions are expressed as identifiable cubic B-spline bases and the number of basis functions was set using the empirical rule of $min\{40, n/4\}$ (Ruppert et al., 2003). A third-order difference penalty was used in the penalty term.

Figure 5 portrays the fitted curves (centred due to identifiability constraints) along partial residuals, i.e. the fitted quantile values plus the residuals from the full model. To quantify uncertainty, pointwise confidence intervals have been computed by means of the estimates covariance matrix based on bootstrap cases resampling.
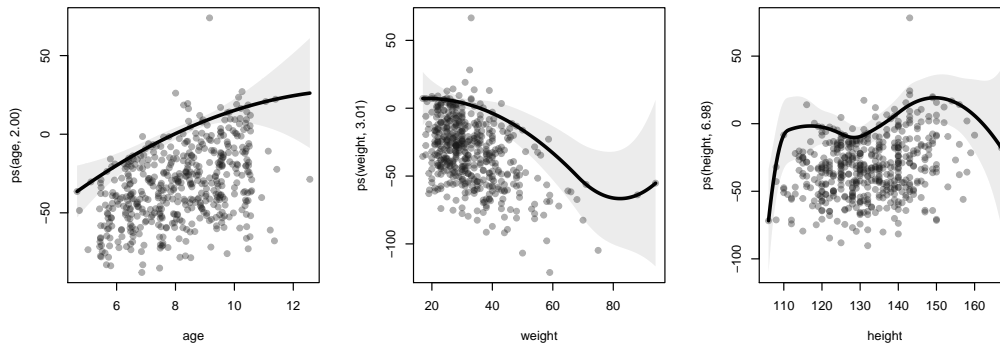

Figure 5: The smooth effect of age, weight and height on the quantile curve $\tau = 0.95$ of standing long jump. In each panel, shaded areas portray the 95% pointwise confidence intervals based on bootstrap resampling and dots represent the partial residuals.

For the age term, the smoothing parameter estimate is quite large, and thus the relevant relationship with the response corresponds to a polynomial of degree 2; on the other hand, the estimated lambdas for the weight and height terms are moderate, resulting in flexible relationships with 3.01 and 6.98 degrees of freedom, respectively. Controlling for age and height, the weight effect on the SLJ is worth discussing: till 35-40 Kg there is almost no influence on the performance, but afterwards there is an important negative effect till about 75 Kg when the relationship stabilizes. Finally the height effect shows different phases, with steeper slopes at very low (< 110cm) and at middle ($130 - 150$ cm) heights.

We have discussed estimation at single specified τ. However if the aim is to estimate multiple quantile curves at different τ values, the proposed algorithm could be applied several times at different τ values with the noncrossing constraints as discussed in Muggeo et al. (2013).

6   Conclusions

We have proposed an iterative algorithm for selecting multiple smoothing parameters in additive QR models with L_1 penalties. The idea exploits the link between smoothing and mixed modelling, a connection which is well consolidated in mean regression, but it appears not fully exploited in quantile regression. The iterative nature of the proposed algorithm makes it very attractive in presence of several lambdas where the multidimensional grid search or the derivative-free numerical optimization demand extensive computations. While the algorithm relies on assuming proper Laplace distributions for the responses and the random effects with constant scale parameters, simulations have shown that our approach performs well even when the constant-scale assumption is violated and observations exhibit heteroscedasticity. This is quite remarkable from a practical perspective, since quantile regression turns out to be very useful especially in heteroscedastic scenarios.

In addition to far better computational efficiency, simulation experiments have shown that most of the times our proposal exhibits good statistical performance as compared to the canonical approach of smoothing splines and tuning parameter selected by the Schwartz Information Criterion. The proposed approach returned slightly higher mean squared errors only in few scenarios at the lowest quantile ($\tau = 0.1$). Along with the algorithm itself, we have also discussed a strategy to quantify the equivalent degrees of freedom of the fitted model and also of each smooth term. In additive QR, quantification of the equivalent degrees of freedom for each smooth term, does not appears to have been discussed previously.

We have focused on additive models, that is multiple 'univariate' smooth terms, but generalization to bivariate smooths represent straightforward extensions where our proposal could apply. Modelling multiple quantile curves to produce growth charts for reference values can be carried out by applying the proposed algorithm at different probability values, namely with τ- specific smoothing parameter values. As an alternative option, a unique smoothing parameter could be used for each quantile curve: selecting and using a unique smoothing parameter for each τ value represents a noteworthy issue to be investigated.

Also, the presented algorithm could be employed in random effects QR for longitudinal data, where a few proposals have been discussed (Koenker, 2004; Geraci and Bottai, 2007; Lamarche, 2010): comparisons with Geraci and Bottai (2007) which use explicit marginal Laplace log likelihood and Gibbs sampler, appear particularly noteworthy.

Yet another possible extension of the proposed algorithm is about modelling quantiles from discrete distributions, for instance to study the student performance at university via the number of credits gained after the first academic year (Grilli et al., 2016): here QR is applied several times at the randomly jittered response and, if the smoothing parameter has to be estimated at each perturbated dataset, the proposed approach appears quite useful due to computational efficiency.

R code to implement the methods presented in this paper, is currently available from the corresponding author and will be also shipped in due time with the R package quantregGrowth. Detailed comparisons among the R packages to fit additive regression quantiles represent a noteworthy point to be investigated in a separate paper.

Supplementary materials

Supplementary materials for this paper, including comparisons with other approaches along with further simulation results, R code and data discussed in section 5, are available from `http://www.statmod.org/smij/archive.html`

Acknowledgements

References

Andriyana, Y., Gijbels, I., and Verhasselt, A. (2014). P-splines quantile regression estimation in varying coefficients models. *Test*, **23**, 153–194.

Austin, P. C., Tu, J., Dalyand, P. A., and Alter, D. (2005). The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. *Statistics in Medicine*, **24**, 791–816.

Bang, S. and Jhun, M. (2012). Simultaneous estimation and factor selection in quantile regression via adaptive sup-norm regularization. *Computational Statistics and Data Analysis*, **56**, 813–826.

Bollaerts, K., Eilers, P., and Aerts, M. (2006). Quantile regression with monotonicity restrictions using P-splines and the L_1-norm. *Statistical Modelling*, **6**, 189–207.

Bosch, R., Ye, Y., and Woodworth, G. (1995). A convenient algorithm for quantile regression with smoothing splines. *Computational Statistics and Data Analysis*, **19**, 613–630.

Cole, T. and Green, P. (1992). Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–1319.

Currie, I. and Durban, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, **4**, 333–349.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.

Eilers, P. and Marx, B. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 637–653.

Fasiolo, M., Goude, Y., Nedellec, R., and Wood, S. N. (2018). *Fast calibrated additive quantile regression.* URL https://arxiv.org/abs/1707.03307.

Fellner, W. H. (1986). Robust estimation of variance components. *Techonometrics*, **28**, 51–60.

Fenske, N., Fahrmeir, L., Hothorn, T., Rzehak, P., and Höhle, M. (2013). Boosting structured additive quantile regression for longitudinal childhood obesity data. *The International Journal of Biostatistics*, **9**, 1–18.

Frasso, G. and Eilers, P. H. C. (2013). L-surface and V-valley for multi-dimensional smoothing parameter selection. In V. Muggeo, V. Capursi, G. B. G. L., editor, *Proceedings of the 28th International Workshop on Statistical Modelling*, pages 151–156.

Frasso, G. and Eilers, P. H. C. (2015). L- and V-curves for optimal smoothing. *Statistical Modelling*, **15**, 91–111.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–154.

Grilli, L., Rampichini, C., and Varriale, R. (2016). Statistical modelling of gained university credits to evaluate the role of pre-enrolment assessment tests: An approach based on quantile regression for counts. *Statistical Modelling*, **16**, 47–66.

Hansen, P. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, **34**, 561–80.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of The Americal Statistical Association*, **72**, 320–338.

He, X. and Ng, P. (1999). COBS: qualitatively constrained smoothing via linear programming. *Computational Statistics*, **14**(3), 315–338.

He, X. and Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, **3**, 299–308.

Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M., and Hofner, B. (2018). *mboost: Model-Based Boosting*. URL https://CRAN.R-project.org/package=mboost. R package version 2.9-1.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**(1), 74–89.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R. (2011). Additive models for quantile regression: model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, **25**, 239–262.

Koenker, R. (2016). *quantreg: Quantile Regression*. URL https://CRAN.R-project.org/package=quantreg. R package version 5.29.

Koenker, R. and Bassett, W. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.

Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673–680.

Kraemer, W. J. and Häkkinen, K. (2008). *Handbook of Sports Medicine and Science, Strength Training for Sport*. Wiley.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, **157**, 396–408.

Li, Y. and Zhu, J. (2008). The L_1 norm quantile regression. *Journal of Computational and Graphical Statistics*, **17**, 163–185.

Li, Y., Graubard, B., and Korn, E. L. (2010). Application of nonparametric quantile regression to body mass index percentile curves from survey data. *Statistics in Medicine*, **29**, 558–572.

Liu, X., Yu, K., Xu, Q., and Tang, X. (2019). Improved local quantile regression. *Statistical Modelling*, **19**, 501–523.

Mayr, A. and Hofner, B. (2018). Boosting for statistical modelling: A non-technical introduction. *Statistical Modelling*, **18**, 365–384.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**, 162–170.

Muggeo, V. M. R., Sciandra, M., and Augugliaro, L. (2012). Quantile regression via iterative least squares computations. *Journal of Statistical Computation and Simulation*, **82**, 1557–1569.

Muggeo, V. M. R., Sciandra, M., Tomasello, A., and Calvo, S. (2013). Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environmental and Ecological Statistics*, **20**, 519–531.

Ng, P. and Maechler, M. (2007). A fast and efficient implementation of qualitatively constrained quantile smoothing splines. *Statistical Modelling*, **7**, 315–328.

Ng, P. and Maechler, M. (2016). *COBS – Constrained B-splines (Sparse matrix based)*. URL https://CRAN.R-project.org/package=cobs. R package version 1.3-1.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric regression*. Cambridge university press.

Saint-Maurice, P. F., Laurson, K. R., Kaj, M., and Csányi, T. (2015). Establishing normative reference values for standing broad jump among hungarian youth. *Research Quarterly for Exercise and Sport*, **86**, S37– S44.

Sandercock, G., Voss, C., Cohen, D., Taylor, M., and Stasinopoulos, D. M. (2016). Centile curves and normative values for the twenty metre shuttle-run test in English schoolchildren. *Journal of Sports Sciences*, **30**, 679–687.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719–727.

Schnabel, S. and Eilers, P. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis*, **53**, 4168–4177.

Schnabel, S. K. and Eilers, P. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, **97**(1), 77–87.

Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis*, **54**, 1138–1150.

Torretta, F., Muggeo, V. M. R., and Eilers, P. H. C. (2015). P-spline quantile regression with a mixed model algorithm. In Friedl, H. and Wagner, H., editors, *Proceedings of the 30th International Workshop on Statistical Modelling*, pages 372–376.

Waldmann, E. (2018). Quantile regression: A short story on how and why. *Statistical Modelling*, **18**, 203–218.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, **18**, 223–249.

Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, **25**, 1369–1382.

Wood, S. (2006). *Generalized Additive Models: an introduction with R*. Chapman & Hall.

Xiang, X. (1996). A kernel estimator of a conditional quantile. *Journal of Multivariate Analysis*, **59**, 206–216.

Yu, K. and Jones, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–237.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics–Theory and Methods*, **34**, 1867–1879.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–447.

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational statistics & data analysis*, **50**, 813–829.