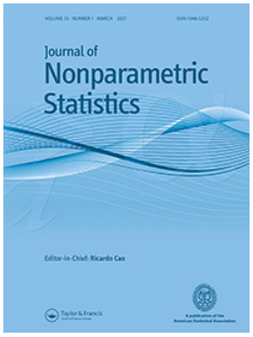


Taylor & Francis
Taylor & Francis Group

Journal of Nonparametric Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Regularised rank quasi-likelihood estimation for generalised additive models

Hannah E. Correia & Asheber Abebe

To cite this article: Hannah E. Correia & Asheber Abebe (2021) Regularised rank quasi-likelihood estimation for generalised additive models, Journal of Nonparametric Statistics, 33:1, 101-117, DOI: 10.1080/10485252.2021.1921176

To link to this article: https://doi.org/10.1080/10485252.2021.1921176

JOURNAL OF NONPARAMETRIC STATISTICS
2021, VOL. 33, NO. 1, 101–117
https://doi.org/10.1080/10485252.2021.1921176

Taylor & Francis
Taylor & Francis Group

Regularised rank quasi-likelihood estimation for generalised additive models

Hannah E. Correia [a,b,c] and Asheber Abebe [a]

[a]Department of Mathematics and Statistics, Auburn University, Auburn, AL, USA; [b]Harvard Data Science Initiative, Harvard University, Cambridge, MA, USA; [c]Department of Biostatistics, Harvard University, Boston, MA, USA

ABSTRACT

Generalised additive models (GAMs) provide flexible models for a wide array of data sources. In the past, improvements of GAM estimation have focused on the smoothers used in the local scoring algorithm used for estimation, but poor prediction for non-Gaussian data motivates the need for robust estimation of GAMs. In this paper, rank-based estimation, as a robust and efficient alternative to the likelihood-based estimation of GAMs, is proposed. It is shown that rank GAM estimators can be obtained through iteratively reweighted likelihood-based GAM estimation which we call the iterated regularised rank quasi-likelihood (IRRQL). Simulation experiments support the use of rank-based GAM estimation for heavy-tailed or contaminated sources of data.

ARTICLE HISTORY
Received 9 May 2020
Accepted 18 April 2021

KEYWORDS
Wilcoxon score function; iteratively reweighted least squares; smoothing spline; robust estimation

MSC2010 CODES
62G05; 62G30; 62G35; 62J12

1. Introduction

Generalized additive models (GAMs) model the mean of a response variable μ via a monotonic link function $g(\mu) = \eta$ using smooth nonparametric coefficient functions $f_j(\cdot)$ and taking the form

$$\eta = f_0 + \sum_{j=1}^{p} f_j(X_j) \,, \tag{1}$$

where $X_1, \ldots, X_p$ are covariates contributing to $\mu = E(Y)$ and Y has an exponential family distribution with variance $V(Y) = \phi^2 \nu(\mu)$. Here $\phi > 0$ is an unknown dispersion parameter and the function $\nu(\cdot)$ is assumed to be twice continuously differentiable. GAMs provide flexible models that are useful for applications in business, economics, medicine, ecology, and environmental health, among other disciplines.

Estimation of the smooth functions $f_1(\cdot), \ldots, f_p(\cdot)$ in GAM required to estimate the model was done via local scoring by Hastie and Tibshirani (1990) in which the smooth

CONTACT Hannah E. Correia hcorreia@hsph.harvard.edu Department of Mathematics and Statistics, Auburn University, Auburn, 36849 AL, USA; Harvard Data Science Initiative, Harvard University, Cambridge, 02138 MA, USA; Department of Biostatistics, Harvard University, 655 Huntington Ave., Boston, 02115 MA, USA

Supplemental data for this article can be accessed here. https://doi.org/10.1080/10485252.2021.1921176

functions $f_j(\cdot)$ were estimated individually using a local scoring algorithm. A major focus of improvements to GAM estimation has been on the representation and estimation of the smoother functions. Hastie and Tibshirani (1990) focus on the use of linear smoothers, because many guarantee convergence via their convenient properties of symmetry and shrinking (Buja, Hastie, and Tibshirani 1989). However, linear smoothers pose a difficulty in model selection and inference due to the problem of calculating the effective degrees of freedom of the smoothing terms. This issue in turn affects the criteria typically applied in model selection, such as Akaike information criterion (AIC) or generalised cross-validation (GCV). Wahba (1990) elegantly solved this problem with generalised smoothing splines and developed an algorithm for estimating multiple smoothing parameters using GCV scores for GAMs constructed with smoothing splines (Gu and Wahba 1991); however, model selection using this approach is computationally expensive. Wood and Augustin (2002) proposed a compromise using penalised regression splines which reduced the class of usable smoothers but were computationally efficient. An issue with regression splines mentioned by Hastie and Tibshirani (1990) is the complicated choice of the placement of knots. Wood and Augustin (2002) opted for penalised regression splines as a solution: use a large number of knots and apply a penalty to avoid overfitting. The estimation of the smooth functions is then obtained by the least squares procedure subject to a roughness penalty. The choice of the smoothing parameter used in the roughness penalty is discussed in detail in Wood and Augustin (2002). Model (1) is fit by maximising the penalised log-likelihood through iteratively reweighted least squares in the algorithm given by Hastie and Tibshirani (1990).

Later, Wood (2004) proposed an improved method for multiple smoothing parameter estimation to cope with fixed penalties and suggested GAMs with a ridge penalty to optimise numerical stability and deal with issues of model identifiability. This involves performing Newton or steepest descent updates of the log of the smoothing parameters λ_j within the iteratively reweighted least squares (IRLS) backfitting algorithm. The approach we propose in this paper will take advantage of this formulation.

Classical fitting of GAMs may not be adequate to fit heavy-tailed or contaminated data, since the approach is sensitive to data contamination. Hence, there is a need for developing robust methodology for fitting GAMs. Robust fitting strategies for GAMs have been considered previously. Croux, Gijbels, and Prosdocimi (2012) used the extended quasi-likelihood (EQL) approach to obtain an M-estimator instead of the least squares estimator of Wood and Augustin (2002) and proposed a robust form of GCV to select the smoothing parameters λ_j. Alimadad and Salibian-Barrera (2011) built on this work by replacing the maximum likelihood-based weights in the IRLS algorithm with robust quasi-likelihood weights using M-estimators. Both contributed to fitting robust GAMs resistant to the presence of outliers but at a computational cost. Wong, Yao, and Lee (2014) used an iterated least-squares procedure to develop an efficient algorithm for M-estimation in GAMs. This paper instead applies R-estimators to GAM estimation and devises a natural extension of the iterative rank estimation in GLMs by Miakonkana and Abebe (2014) to construct robust and efficient GAMs.

The paper is structured as follows: Section 2 provides our rank-based estimation approach and Section 3 outlines the iterative algorithm used to fit rank-based GAMs. Simulations are given in Section 4. We offer concluding remarks in Section 5.

2. Estimation

To accommodate a wide range of data generating phenomena when fitting GAMs, robust approaches have been recently proposed. These include Alimadad and Salibian-Barrera (2011), Croux et al. (2012), and Wong et al. (2014). These are all M-type estimators. We are interested in R-estimators as defined in Jaeckel (1972) and Hettmansperger and McKean (2011). As discussed in Draper (1988), both M- and R-estimators provide robust fits with no clear winner. While both M and R are location invariant, only R-estimators are automatically scale invariant. M-estimators can be made scale invariant with the addition of a preliminarily estimated scale parameter (Rousseeuw and Leroy 2005). This makes R-estimators very attractive for estimation in complex model settings. A drawback of both M- and R-estimation is that they are computationally expensive. For the linear regression model, Sievers and Abebe (2004) gave an approach that uses iterative least squares fitting to obtain R-estimators. More recently, Miakonkana and Abebe (2014) extended this to the fitting of generalised linear models. Wong et al. (2014) derived a computationally efficient M-estimator for GAMs, again using iterative fitting of GAMs via penalised least squares. In this chapter, we propose a rank estimator of GAMs and develop an efficient iterative computational algorithm. Our method, which we call the *iterated regularised rank quasi-likelihood (IRRQL)* procedure, depends on the ranking of Pearson residuals to account for the mean-variance dependence in GAMs.

2.1. Rank-based estimation

To facilitate the introduction of our rank-based approach for estimation of GAMs, we will start with the linear regression model $Y_i = \alpha + \mathbf{x}_i^T\boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$. The vector of model residuals is given by $\mathbf{z}(\boldsymbol{\beta})$ with ith component $z_i(\boldsymbol{\beta}) = Y_i - \mathbf{x}_i^T\boldsymbol{\beta}$. Jaeckel (1972) proposed to estimate the regression slope parameter $\boldsymbol{\beta}$ by minimising

$$\|\mathbf{z}(\boldsymbol{\beta})\|_\varphi \equiv \sum_{i=1}^{n} \varphi\left(\frac{R(z_i(\boldsymbol{\beta}))}{n+1}\right) z_i(\boldsymbol{\beta}), \tag{2}$$

where $\varphi : (0,1) \to \mathbb{R}$ is a nondecreasing score function such that $\int \varphi = 0$ and $\int \varphi^2 = 1$ and $R(\cdot)$ is the rank function. He showed that this produces a regression estimator that is equivalent to the rank score estimator given by Jurečková (1971). He also showed that the quantity $\|\cdot\|_\varphi$ is a convex pseudo-norm on $\mathbb{R}^n$. Because $\|\cdot\|_\varphi$ is a pseudo-norm, it is invariant to constant translations; hence, it cannot be used to estimate the intercept α. There are several choices for φ with the simplest one given by the linear score function $\varphi(u) = \sqrt{12}(u - 1/2)$ resulting in the so-called Wilcoxon pseudo-norm. For a general discussion regarding the use of (2) in the linear model, one may refer to the monograph (Hettmansperger and McKean 2011).

For linear regression, it has been shown that the estimator resulting from Wilcoxon estimation is robust in the presence of outliers and heavy-tailed error distributions. It is also very efficient. For instance, it achieves 95.5% relative efficiency versus the least squares estimator when the underlying error distribution is normal, and the relative efficiency is much higher for distributions with tails heavier than the normal. The worst-case relative efficiency is 86.4% for symmetric error distributions. So there is much appeal to using $\|\cdot\|_\varphi$ for estimation purposes. The inference also extends to hypothesis testing (Hettmansperger

and McKean 2011). For example, we can easily define drop, Wald, or score tests for testing the significance of model parameters.

In recent years, the method has been employed for models other than the linear model. Bindele and Abebe (2015) studied rank estimation of semiparametric models with responses missing at random. They showed that the rank estimator remains robust and efficient, with efficiency improving relative to standard imputation methods, when a large proportion of the responses are missing. The approach has been used for estimation of general nonlinear regression (Abebe and McKean 2013), generalised linear models (Miakonkana and Abebe 2014), varying coefficient models (Wang et al. 2009), and functional regression (Denhere and Bindele 2016), among others. Some of the development has been facilitated by the iterative reweighted least squares procedure given by Sievers and Abebe (2004). This greatly simplifies the computation of rank regression coefficients even for complex models (Abebe et al. 2016).

2.2. Rank-based GAM estimation

For obtaining the rank estimator of GAMs, we will use a penalised version of the rank quasi-score function given in Miakonkana and Abebe (2014). The responses $\{Y_i\}_{i=1}^n$ are assumed to be independent and follow a distribution from the exponential family with expectation μ_i and variance $\phi^2 v(\mu_i)$. To simplify our discussion and theoretical development, we will consider the simple $p = 1$ version of the GAM (1) given by

$$g(\mu_i) = f(x_i)$$

as well as the linear (Wilcoxon) score function $\sqrt{12}(u - 1/2)$. Our approach extends directly to $p > 1$ and other score functions φ.

Taking a set of prespecified basis functions $\mathbf{b} = (b_1(\cdot), \ldots, b_m(\cdot))'$, the function f is assumed to have a representation

$$f(x_i; \boldsymbol{\theta}) = \sum_{j=1}^{m} b_j(x_i)\theta_j \equiv \mathbf{b}_i^T \boldsymbol{\theta} \tag{3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is a vector of basis coefficients and we suppress x_i in $\mathbf{b}$. Define the Pearson residuals as

$$z_i(\boldsymbol{\theta}) = \frac{Y_i - \mu_i}{\sqrt{v(\mu_i)}}.$$

Note that we are ignoring the extra dispersion parameter ϕ since the ranks are invariant to scale transformations. The rank quasi-likelihood function as defined in Miakonkana and Abebe (2014) is then

$$\boldsymbol{\ell}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{R(z_i(\boldsymbol{\theta}))}{n+1} - \frac{1}{2} \right\} \frac{\partial \mu_i / \partial \boldsymbol{\theta}}{\sqrt{v(\mu_i)}} .$$

By taking $h \equiv g^{-1}$, we have $\mu_i = h(\mathbf{b}_i^T \boldsymbol{\theta})$ and $\partial \mu_i / \partial \boldsymbol{\theta} = h'(\mathbf{b}_i^T \boldsymbol{\theta})\mathbf{b}(x_i)$ and therefore

$$\boldsymbol{\ell}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{R(z_i(\boldsymbol{\theta}))}{n+1} - \frac{1}{2} \right\} \frac{h'(\mathbf{b}_i^T \boldsymbol{\theta})\mathbf{b}_i}{\sqrt{v(h(\mathbf{b}_i^T \boldsymbol{\theta}))}} .$$

Theoretically, the rank estimator of $\boldsymbol{\theta}$ is found by solving $\boldsymbol{\ell}(\boldsymbol{\theta}) = \mathbf{0}$. However, for the estimation of GAMs, we will need to impose a smoothness penalty. Thus we define the *regularised rank quasi-likelihood (RRQL)* function and solve

$$\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}) \equiv \boldsymbol{\ell}(\boldsymbol{\theta}) - \mathbf{S}_{\lambda_n}\boldsymbol{\theta} = \mathbf{0}\,,$$

where $\mathbf{S}_{\lambda_n} = 2\lambda_n\mathbf{D}$, $\lambda_n > 0$ is a smoothing parameter and $\mathbf{D}$ is an $m \times m$ penalty matrix. We let $\widetilde{\boldsymbol{\theta}}_n$ represent the zero of the RRQL function; that is $\widetilde{\boldsymbol{\theta}}_n$ solves $\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}) = \mathbf{0}$.

However, finding a direct solution of $\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}) = \mathbf{0}$ is difficult. Below, we will define an iterative scheme to approximate $\widetilde{\boldsymbol{\theta}}_n$. To that end, define the pseudo-Pearson 'residuals'

$$z_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \frac{Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta})}{\sqrt{v(h(\mathbf{b}_i^T\boldsymbol{\theta}^*))}}$$

and define the corresponding rank estimator as the minimiser of $\|\mathbf{z}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\|_w$, where

$$\|\mathbf{z}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\|_w = \sum_{i=1}^{n}\left\{\frac{R(z_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*))}{n+1} - \frac{1}{2}\right\} z_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$$

is as given in (2) with $\varphi(u) = \sqrt{12}(u - 12)$. Using the IRLS scheme of Sievers and Abebe (2004), this can be represented as

$$\|\mathbf{z}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\|_w = \sum_{i=1}^{n} w_i(\boldsymbol{\theta})\frac{(Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta}))^2}{v(h(\mathbf{b}_i^T\boldsymbol{\theta}^*)}$$

where, letting $m = \text{med}\{Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta})\}$, the weights are defines as

$$w_i(\boldsymbol{\theta}) = \begin{cases} \dfrac{\dfrac{R(z_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*))}{n+1}}{Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta}) - m}^{-\frac{1}{2}} & \text{if } Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta}) - m \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Since $\varphi(u)$ is odd about 1/2, these weights are all non-negative. To obtain our iteration, we take the weights w_i at a different value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}'$. Now taking the derivative of $\|\cdot\|_w$ with respect to $\boldsymbol{\theta}$ we obtain the approximate rank score function

$$2\sum_{i=1}^{n} w_i(\boldsymbol{\theta}')\frac{(Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta}))}{v(h(\mathbf{b}_i^T\boldsymbol{\theta}^*)}h'(\mathbf{b}_i^T\boldsymbol{\theta})\mathbf{b}_i$$

Following Wedderburn (1974), if we now take $\boldsymbol{\theta}^* = \boldsymbol{\theta}$, then we get the weighted quasi-likelihood function

$$\boldsymbol{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{i=1}^{n} w_i(\boldsymbol{\theta}')\left\{\frac{Y_i - h(\mathbf{b}_i^T\boldsymbol{\theta})}{v(h(\mathbf{b}_i^T\boldsymbol{\theta}))}\right\}h'(\mathbf{b}_i^T\boldsymbol{\theta})\mathbf{b}_i = \sum_{i=1}^{n} w_i(\boldsymbol{\theta}')\frac{(Y_i - \mu_i)}{v(\mu_i)}\frac{\partial\mu_i}{\partial\boldsymbol{\theta}}\,,$$

which is exactly a weighted form of the classical GLM quasi-likelihood function.

We can now define the penalised quasi-likelihood for GAM estimation as

$$\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \boldsymbol{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}') - \mathbf{S}_{\lambda_n}\boldsymbol{\theta}\,.$$

Now, suppose we have a suitable initial estimator of $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}_n^{(0)}$. This can be the classical penalised likelihood estimator. For $k = 1, 2, \ldots$, we define $\widehat{\boldsymbol{\theta}}_n^{(k)}$ as a solution of the *iterated regularised rank quasi-likelihood (IRRQL)* function

$$\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n^{(k-1)}) = \mathbf{0},$$

which can be computed by iteratively solving a weighted GAM estimating equation. For the unpenalised version $\boldsymbol{\ell}_0(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n^{(k-1)})$, Miakonkana and Abebe (2014) showed that the iteration converges to the solution of $\boldsymbol{\ell}_0(\boldsymbol{\theta})$.

Let $\mathbf{f} = (f(x_1), \ldots, f(x_n))^T$ and the $n \times m$ coefficient matrix be $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n]^T$. Note that we have two rank estimators of $\boldsymbol{\theta}$: $\widetilde{\boldsymbol{\theta}}_n$ which solves the RRQL and $\widehat{\boldsymbol{\theta}}_n^{(k)}$, $k = 1, 2, \ldots$ given $\widehat{\boldsymbol{\theta}}_n^{(0)}$ which solves the IRRQL. We can correspondingly define two rank-based GAM estimators of $\mathbf{f}$ using equation (3). We define these as

$$\widetilde{\mathbf{f}}_n = \mathbf{B}\widetilde{\boldsymbol{\theta}}_n$$

and

$$\widehat{\mathbf{f}}_n^{(k)} = \mathbf{B}\widehat{\boldsymbol{\theta}}_n^{(k)}, \quad k = 1, 2, \ldots.$$

Note that for a given k, $\widehat{\mathbf{f}}_n^{(k)}$ is just a regular weighted GAM estimator. Thus, its asymptotic properties are well understood and are part of the standard GAM literature (cf. Hastie and Tibshirani 1990; Wood 2006). The conditions needed for consistency and the asymptotic results given in Theorems 2.1 and 2.2 are the same as those given in Wong et al. (2014). The results are given here for completeness and detailed proofs are found in the online appendix of Wong et al. (2014). Theorem 2.1 gives consistency of $\widehat{\mathbf{f}}_n^{(k)}$. However, we need to understand whether $\widehat{\mathbf{f}}_n^{(k)}$ provides a good approximation of the rank estimator $\widetilde{\mathbf{f}}_n$. (A1) – (A4) below give conditions under which $\widehat{\mathbf{f}}_n^{(k)}$ gives a valid approximation of $\widetilde{\mathbf{f}}_n$. The theorem following the conditions gives the asymptotic equivalence of $\widehat{\mathbf{f}}_n^{(k)}$ and $\widetilde{\mathbf{f}}_n$. Before giving the conditions, we note that when using the iteratively reweighted least squares (IRLS) approach of fitting GAMs, there is a reproducing kernel Hilbert space (RKHS) representation $\mathbf{f}^T\Gamma^{1/2}\mathbf{R}\Gamma^{1/2}\mathbf{f}$ of the penalty function $\lambda_n\boldsymbol{\beta}^T\mathbf{D}\boldsymbol{\beta}$, where Γ is the IRLS weight (Wong et al. 2014). In this set up, the residual smoother matrix for GAM estimation is $\mathbf{H}_{\lambda_n} = (\mathbf{I} + 2\lambda_n\mathbf{R})^{-1}$, where $\mathbf{R}$ is the reproducing kernel. The conditions needed for consistency are

(A1) The function f is bounded; that is, $\sup_{-\infty<t<\infty} |f(t)| < \infty$.
(A2) Let $\mathscr{F}$ be the space of all f's. We assume that $\mathscr{F}$ is a reproducing kernel Hilbert space.
(A3) Let $\mathscr{C}_\alpha = \{f \in \mathscr{F} : \|f\|_{\mathscr{F}} \leq \alpha\}$ for some constant α. We assume that $\mathscr{C}_\alpha$ is compact with respect to L_2 norm.
(A4) Let d_n be the maximum diagonal element of $\mathbf{H}_{\lambda_n}$. We assume that $\lambda_n/n \to 0$ and $d_n \to 0$ as $n \to \infty$. Moreover, $tr(\mathbf{H}_{\lambda_n})/\lambda_n < K < \infty$.

Theorem 2.1: *Under (A1)–(A4), for* $k \in \mathbb{N}$, $n^{-1}E\{\|\widehat{\mathbf{f}}_n^{(k)} - \mathbf{f}\|^2\} \to 0$ *as* $n \to \infty$.

Theorem 2.2: *Under (A1)–(A4), for* $k \in \mathbb{N}$, $\|\widehat{\mathbf{f}}_n^{(k)} - \widetilde{\mathbf{f}}_n\| / E\{\|\widehat{\mathbf{f}}_n^{(k)} - \mathbf{f}\|\} \xrightarrow{\mathcal{P}} 0$ *as* $n \to \infty$.

There are certain practical considerations that need attention. The first is the degrees of freedom of the estimation problem. The RKHS literature defines the effective degrees of freedom as $tr(\mathbf{H}_{\lambda_n})$. So, (A4) specifies a balance between the effective degrees of freedom and the smoothing parameter. We still need a way to estimate the smoothing parameter λ_n. In this paper, we employ generalised cross-validation to select the parameter λ_n. This is the most common approach in the literature (Wood 2006). Finally, one may question the value of fully iterating k. If the one step estimator gives us consistent estimators (Theorems 2.1 and 2.2), then why do we need to iterate more than once? This was answered in Sievers and Abebe (2004) and Miakonkana and Abebe (2014) where, using fixed-point theory, it was established that as $k \to \infty$ the IRLS rank estimator converges to the true rank estimator for finite samples. In our notation, this is $\lim_{k\to\infty} \widehat{\mathbf{f}}_n^{(k)} = \widetilde{\mathbf{f}}_n$ for n fixed.

2.3. Robustness

We will focus on the influence function (Hampel et al. 2005) as a measure of robustness. The influence function (IF) measures the sensitivity of the estimator to infinitesimal changes in data. As we will demonstrate below, the IF of our rank-based estimator can be found in a straightforward manner.

The score function for the classical GAM estimation is

$$\boldsymbol{\psi}(y, \theta) = \frac{y - \mu}{v(\mu)} \frac{\partial \mu}{\partial \boldsymbol{\theta}} - \frac{1}{n} S_\lambda \boldsymbol{\theta}.$$

Letting $F(y, x, y)$ be the joint distribution function of (y, x) and $\hat{F}$ the corresponding joint empirical distribution function, Wong et al. (2014) give the IF of the GAM estimator $\hat{\hat{\boldsymbol{\theta}}}_n = T(\hat{F})$ as

$$IF(y; \boldsymbol{\psi}, F) = -\left\{ \int \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(u, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=T(F)} \mathrm{d}F(u, x) \right\}^{-1} \boldsymbol{\psi}(y, T(F)),$$

which is unbounded since $\boldsymbol{\psi}$ is unbounded in y. Similarly, the score function for the rank-based GAM estimation is equivalent to

$$\boldsymbol{\psi}_r(z, \theta) = G\left(\frac{y - \mu}{\sqrt{v(\mu)}} - \frac{1}{2} \right) \frac{\partial \mu}{\partial \boldsymbol{\theta}} - \frac{1}{n} S_\lambda \boldsymbol{\theta}\ ,$$

where G is the distribution function of the Pearson residual $z = (y - \mu)/\sqrt{v(\mu)}$. Let H be the joint distribution of (z, x) with a corresponding empirical distribution function $\hat{H}(z, x) = n^{-1} \sum I(\{z_i \le z\} \cap \{x_i \le x\})$. Then the generic functional representation of the rank-based GAM estimator can be given as $\tilde{\boldsymbol{\theta}}_n = T(\hat{H})$ and its IF is

$$IF(z; \boldsymbol{\psi}_r, H) = -\left\{ \int \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}_r(v, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=T(H)} \mathrm{d}H(v, x) \right\}^{-1} \boldsymbol{\psi}_r(z, T(H))\ .$$

Since the dependence of $\boldsymbol{\psi}_r$ on y is only through G, it is bounded in y. As a result, the IF of the rank estimator $IF(z; \boldsymbol{\psi}_r, H)$ is also bounded in y, thus establishing the robustness

of our procedure. We emphasise that this boundedness is true in the residual space for each fixed x, but not in the factor space, since the IF depends on $\partial\mu/\partial\boldsymbol{\theta} = h'(\mathbf{b}(x)^T\boldsymbol{\theta})\mathbf{b}(x)$ unless x has bounded domain (e.g. controlled designs). This is similar to the performance of rank estimators for the linear model (Hettmansperger and McKean 2011) as well as generalised linear models (Miakonkana and Abebe 2014). The unboundedness of the IF in factor space means the breakdown point of this estimator is 0 (Hettmansperger and McKean 2011). However, efficient high breakdown estimators can be constructed using the weighting scheme proposed by Chang, McKean, Naranjo, and Sheather (1999) and Abebe and McKean (2013).

3. Computational algorithm

In this section, we give a brief outline of the iterative algorithm for solving the IRRQL used to obtain the rank estimates $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\mathbf{f}}_n$.

Step 0: Obtain initial estimates $\widehat{\boldsymbol{\theta}}_n^{(0)}$, $\widehat{\mathbf{f}}_n^{(0)} = \mathbf{B}\widehat{\boldsymbol{\theta}}_n^{(0)}$, Pearson residuals $z(\widehat{\boldsymbol{\theta}}_n^{(0)})$, and weights $w(\widehat{\boldsymbol{\theta}}_n^{(0)})$. Set $k = 1$ and let $\epsilon > 0$ be a given tolerance.

Step 1: Obtain a weighted GAM (Hastie and Tibshirani 1990) estimate $\widehat{\boldsymbol{\theta}}_n^{(k)}$ by solving

$$\boldsymbol{\ell}_{\lambda_n}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k-1)}) = \mathbf{0},$$

where λ_n is chosen by generalised cross-validation (Wood 2004).

Step 2: Calculate $\widehat{\mathbf{f}}_n^{(k)} = \mathbf{B}\widehat{\boldsymbol{\theta}}_n^{(k)}$.

Step 3: If $\|\widehat{\mathbf{f}}_n^{(k)} - \widehat{\mathbf{f}}_n^{(k-1)}\| \geq \epsilon\|\widehat{\mathbf{f}}_n^{(k-1)}\|$, then set $k = k+1$ and return to **Step 1**. Otherwise, take $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n^{(k)}, \widehat{\mathbf{f}}_n = \widehat{\mathbf{f}}_n^{(k)}$ and STOP.

Thin plate spline approximation, the default approach in the `mgcv` package in R (Wood 2006), requires $O(n^3)$ operations, which is prohibitive especially for large n. We employed thin plate regression splines, which can also become expensive to calculate for large datasets. For this reason, we used the strategy in Wood (2003) where data are subsampled up to a specific threshold. This threshold is the number of unique data points up to 2000, and for data with more than 2000 unique data points, it is set at 2000. The computation time using a MacBook Pro with 2.8 GHz Quad-Core Intel Core i7 processor and 16GB RAM for Gaussian error and one function to estimate with $n = 500$ took about 0.02 s for the likelihood-based GAM and 0.1 s for the IRRQL based GAM, while for $n = 8000$ took about 0.6 s for the likelihood-based GAM and 1.6 s for the IRRQL based GAM.

4. Simulations and real data studies

4.1. Simulations

We performed simulation experiments to compare our proposed procedure (labeled as R) to the classical likelihood-based GAM estimator (labeled as L) and an M-estimator of Croux et al. (2012) using robust Huber weights (labeled as M). The R code implementation of the M-estimation procedures of Alimadad and Salibian-Barrera (2011) and Wong et al. (2014) included only the binomial and Poisson distributions, so we did not make

direct comparisons of our method with theirs. In our simulation, we considered two additive models: one with only one function and a second one with four functions. We also considered a Gamma generalised additive model with four functions.

With p functions $f_1, \ldots, f_p$, the relative efficiencies of R and M versus L for the ith function were obtained as

$$\mathrm{RE}_i(R) = \frac{\sum_{j=1}^{n}(f_{ij} - \widehat{f}_{\mathrm{L},i,j})^2}{\sum_{j=1}^{n}(f_{ij} - \widehat{f}_{\mathrm{R},i,j})^2} \quad \text{and} \quad \mathrm{RE}_i(M) = \frac{\sum_{j=1}^{n}(f_{ij} - \widehat{f}_{\mathrm{L},i,j})^2}{\sum_{j=1}^{n}(f_{ij} - \widehat{f}_{\mathrm{M},i,j})^2},$$

respectively, for each estimated function. When there was more than one function we also computed the relative efficiencies for the overall fit as

$$\mathrm{RE}(R) = \frac{\sum_{j=1}^{n}(y_j - \widehat{f}_{\mathrm{L},j})^2}{\sum_{j=1}^{n}(y_j - \widehat{f}_{\mathrm{R},j})^2} \quad \text{and} \quad \mathrm{RE}(M) = \frac{\sum_{j=1}^{n}(y_j - \widehat{f}_{\mathrm{L},j})^2}{\sum_{j=1}^{n}(y_j - \widehat{f}_{\mathrm{M},j})^2},$$

where $\widehat{f}_{*,j} = \sum_{i=1}^{p}\widehat{f}_{*,i,j}$ and y_j are the responses for $j = 1, \ldots, n$. For simplicity, we only use the Wilcoxon score function $\varphi(u) = \sqrt{12}(u - 1/2)$ for IRRQL estimators labelled by R. All our simulations used 500 iterations.

4.1.1. One regressor

The first simulation example considered $n = 100$ samples generated according to the model

$$y_j = f(x_j) + \varepsilon_j, \quad j = 1, \ldots, n,$$

where f was

$$f(x_j) = 0.2(10^6 x_j^{11})(1 - x_j)^6 + 10^4 x_j^3 (1 - x_j)^{10}$$

which is given as f_2 in Gu and Wahba (1991) and x_j are uniformly distributed on $[0, 1]$.

We considered two error distribution scenarios: one to simulate the effect of heavy tails and the second to simulate the effect of contamination. To study the effect of tail thickness on the estimators, the errors ε were randomly generated from Student's t distributions with increasing degrees of freedom e^k, where k is taken from 1 to 5 in steps of 0.5. To study the effect of contamination in the measured response on the estimators, the errors ε were drawn from a contaminated normal distribution. The contaminated normal distribution is defined by creating a normal-normal Huber contaminated distribution as

$$CN(\delta, \sigma) = (1 - \delta)N(0, 1) + \delta N(0, \sigma^2),$$

where $\delta \in [0, 1]$ and $\sigma > 0$. This means the errors are drawn from the $N(0, 1)$ distribution with probability $1 - \delta$ and from the $N(0, \sigma^2)$ distribution with probability δ. For our simulation experiment, we took $\sigma = 3$ and δ taken from 0 to 0.35 in steps of 0.05. The results are shown in Figures 1 and 2, respectively. Figure 1 shows that both M and R are much more efficient than L for heavy-tailed distributions, and M is consistently less efficient than R. Figure 2 shows that the rank estimator is more efficient than M across all levels of contamination considered, with M losing efficiency when there is more than 20% contamination. Both M and R are more efficient than L for contamination over 5%.

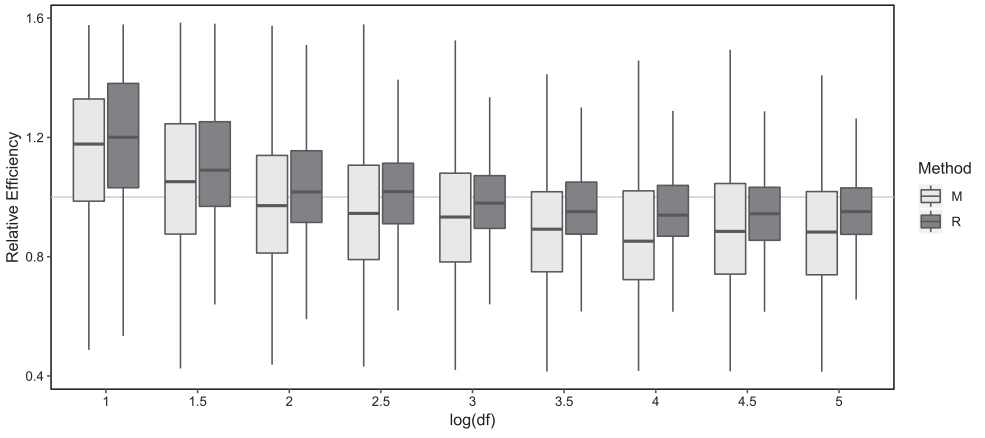


Figure 1. Relative efficiencies vs *L* for simulation with increasing *df* of the *t* distribution.

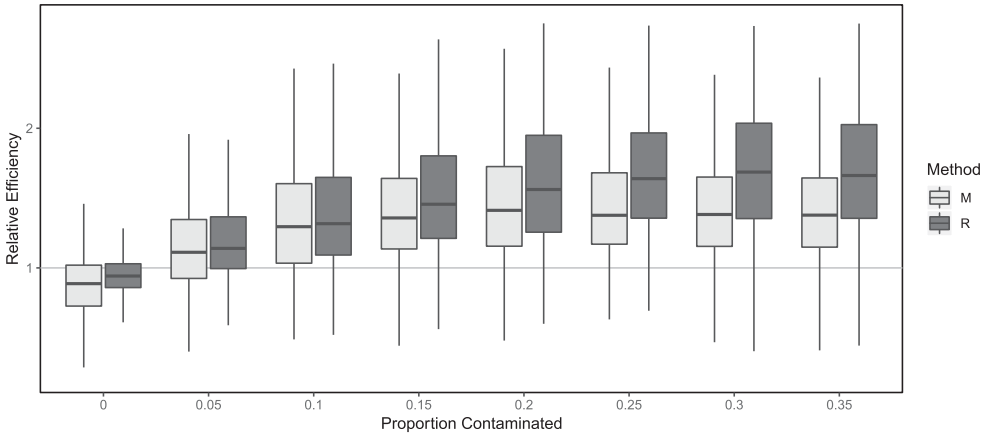


Figure 2. Relative efficiencies vs *L* for simulation with increasing proportions of contamination of the normal distribution.

We compared the computation time of R and M using the same simulation parameters for one case of the contaminated normal setting but only 50 iterations. We found that the average computation time for R was around 0.1 s (min = 0.05 s, max = 0.26 s), whereas M took an average of 0.75 s (min = 0.50 s, max = 1.65 s). This demonstrates the computational efficiency of our algorithm. M-estimators would need to calculate an additional preliminary scale estimator to become scale invariant which increases their computation time. This highlights the advantages of the proposed estimator for complex problems where computational time efficiency is crucial.

4.1.2. Four regressors

The second simulation study considered the following four-term GAM with $n = 500$ samples:

$$y_j = x_{1j}^2 + 5\sqrt{|x_{2j} + 1|} + 2\sin(\pi x_{3j}) + x_{4j} + \varepsilon_j,$$

where each of the four functions was centred and scaled.

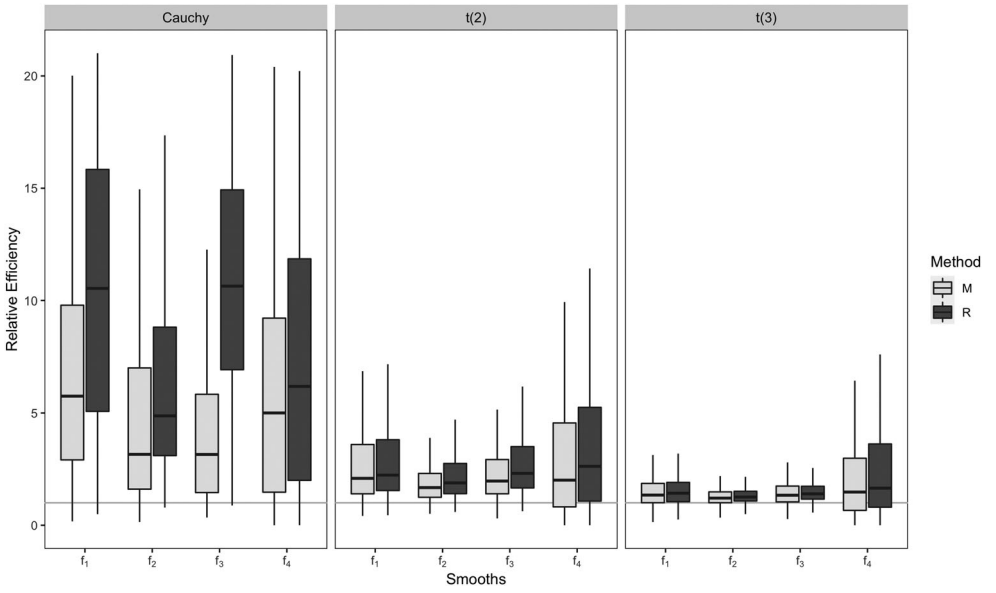


Figure 3. Relative efficiencies vs *L* for simulation with errors drawn from Cauchy, $t(2)$, and $t(3)$ distributions with a Gaussian-distributed response.

For this simulation experiment, we drew errors from six distribution types: a standard Cauchy distribution; a Student's t distribution with 2 degrees of freedom; a Student's t distribution with 3 degrees of freedom; a $N(0, \sigma^2)$ distribution with $\sigma = 1/2$; a Laplace$(0, 1)$ distribution; and a contaminated normal distribution defined as

$$CN(\delta, \sigma) = (1 - \delta)N(5, \sigma^2) + \delta N(0, \sigma^2),$$

where $\delta = 0.95$ and $\sigma = 1/2$. The simulation was performed 500 times. Relative efficiencies compared to L estimation were calculated for M and R procedures, and MSE were calculated for all three estimation procedures.

R was more efficient than M for errors from the Cauchy, Student's t(2) and t(3), and Laplace distributions, and both R and M were both more efficient than L for errors from contaminated normal distributions (Figures 3 and 4). MSE for R-estimated functions were consistently lower than L- and M-estimated functions with errors from Cauchy, Student's t(3) and t(3), Laplace, and contaminated normal distributions (Figures S9–S13). MSE for R-estimated functions with normal errors were not lower than L- and M-estimated functions for more complex second and third functions (Figure S14).

4.1.3. Gamma regression with four regressors

To examine the performance of our IRRQL estimation method for non-normal distributions, we conducted a third simulation study with $n = 500$ samples according to the model

$$f(\mathbf{x}) = 2\sin(\pi x_1) + \exp(2x_2) + \left[0.2(10^6 x_3^{11})(1 - x_3)^6 + 10^4 x_3^3 (1 - x_3)^{10}\right] + 0x_4\ ,$$

which is given in Gu and Wahba (1991) and $x_i,\ i = 1, \ldots, 4$, are uniformly distributed on $[0, 1]$. Each of the four functions were centred and scaled.

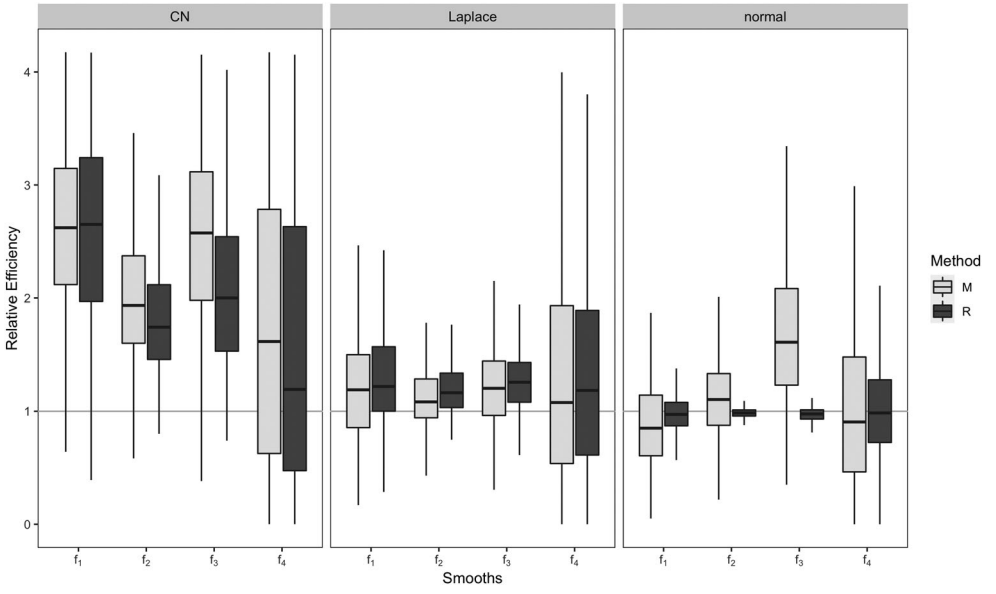


Figure 4. Relative efficiencies vs *L* for simulation with errors drawn from contaminated normal (CN), Laplace, and normal distributions with a Gaussian-distributed response.

A vector of responses was drawn from a Gamma distribution $y \sim \Gamma(4, 0.25\exp(f(\mathbf{x})/4))$, where $f(\mathbf{x}) = f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4)$ is the true function as given above. The response vector was then contaminated by taking a proportion $\delta \in [0, 1]$ of the responses in each case and replacing them by $N(max(y), 0.2^2)$. In our simulation study, we took $\delta \in \{0, 0.1, 0.2\}$. *L*, *M*, and *R* were compared by assuming either a Gaussian family with identity link or a Gamma family with log link. The simulation was run 500 times. *R* was consistently more efficient than *L* for 10% and 20% contamination, and *R* was also as efficient as *M* at all contamination levels. *R* generally reduced variance in relative efficiencies compared to *M* for higher contamination levels (Figure 5). *R* had lower MSE than *L* for all levels of contamination when using the Gaussian with identity link assumption, and *R* regularly had lower MSE than *M* for most of the smooths at all contamination levels (Figures S15–S17). When using the Gamma with log link assumption, *R* had lower MSE values than *L* and *M* for 10% and 20% contamination levels, while MSE for *L*, *M*, and *R* were similar for 0% contamination (Figures S18–S20).

4.2. Real data application

The IRRQL method was applied to data from a large-scale study which accumulated crime and sociodemographic data from census tracts in a representative sample of large cities and metropolitan areas throughout the United States in 2000 (Peterson and Krivo 2010). For computational efficiency, only data from the state of Ohio was used for this application. A GAM of the form given in (1) was fit, with $\mathbf{X} = \{\text{pop00}, \text{femhed}, \text{percap}\}$ being the predictor variables tract population, percent of female-headed households, and per capita income, respectively. The response $Y = \text{crimrt}$ is the three-year average total crime rate

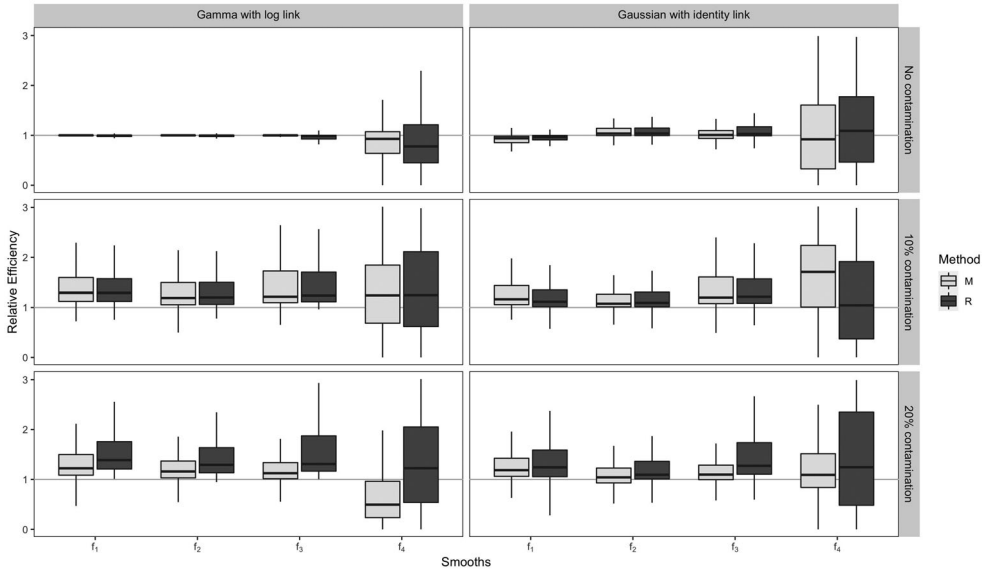


Figure 5. Relative efficiencies vs L for simulation with response drawn from the contaminated Gamma distribution with contamination $\delta = \{0, 0.1, 0.2\}$.

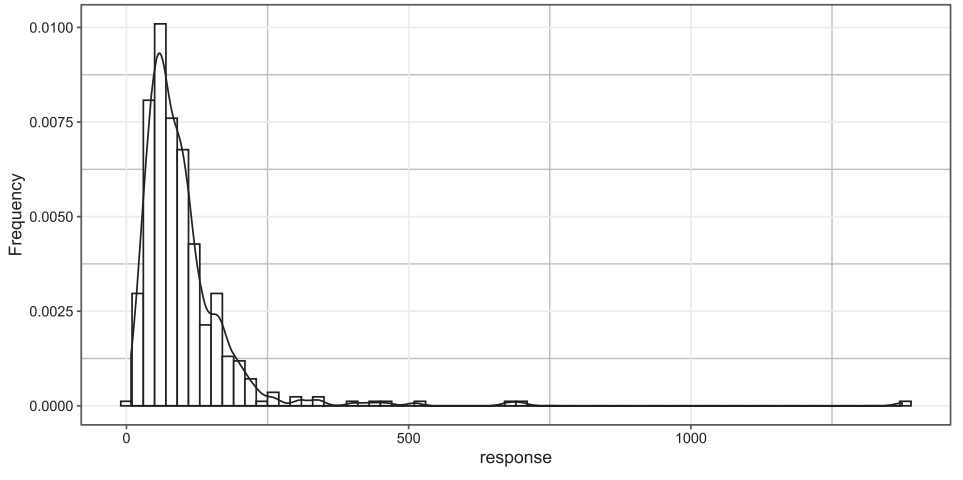
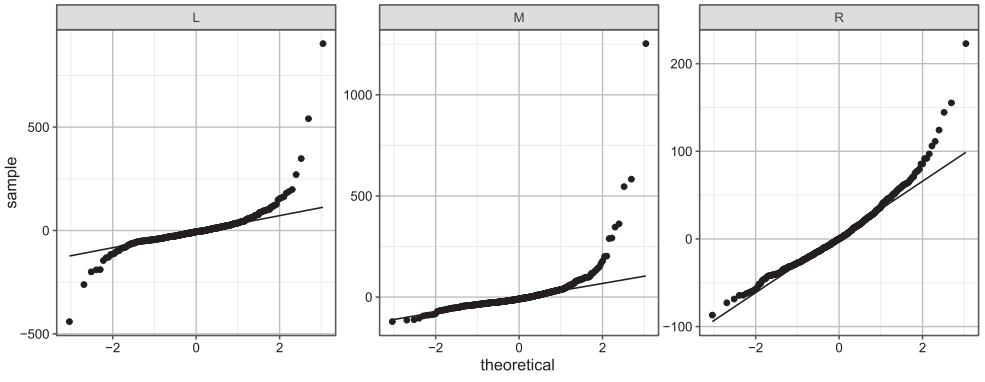


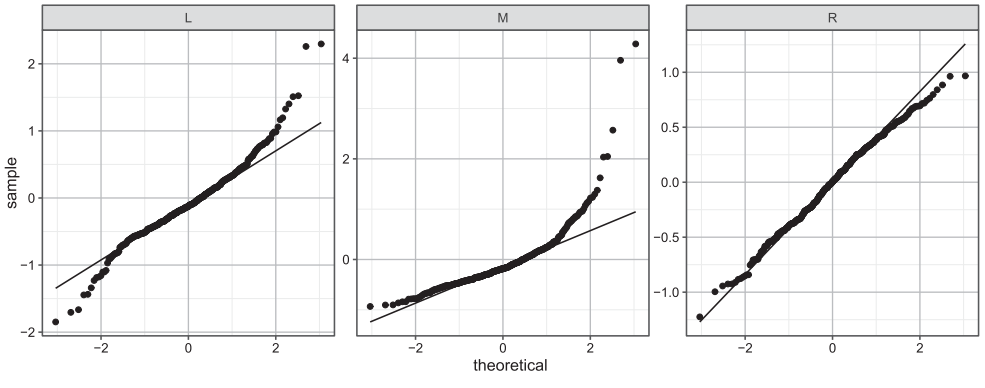
Figure 6. Histogram with density overlay of the untransformed response crime rate.

for each tract and is heavily right-skewed (Figure 6). A 10-fold cross-validation was performed, comparing L, M, and R. Two alternative family assumptions were used for each estimation procedure: a Gaussian with identity link and a Gamma with log link.

Assuming a Gamma distribution with log link improved fit for L, however R had good fit even when the distribution was misspecified (Figure 7). The R method outperformed both L and M in prediction under both family assumptions in the 10-fold cross-validation (Table 1). We therefore utilised the GAM fit with the R method using a Gamma family assumption with log link for inference.



(a) Gaussian with identity link assumption.



(b) Gamma with log link assumption.

Figure 7. Normal Q–Q plots for crime rate model residuals for *L*, *M*, and *R* using two alternative family assumptions for each estimation procedure: (a) Gaussian with identity link and (b) Gamma with log link. (a) Gaussian with identity link assumption and (b) Gamma with log link assumption.

Table 1. Means and standard deviations of median absolute deviation (MAD) from 10-fold cross-validation for each estimation method when using either Gaussian with identity link or Gamma with log link assumptions.

	Method	MAD (Std Dev)
Gaussian	L	27.694 (5.119)
	M	83.174 (10.906)
	R	23.060 (4.722)
Gamma	L	24.327 (5.337)
	M	80.224 (10.706)
	R	23.021 (4.889)

The estimated smooths from the GAM fit with the *R* method using the Gamma family assumption are illustrated in Figure 8. As tract population increased up to 6000, three-year average total crime rate decreased. Additionally, as the percent of female-headed households increased, crime rate decreased. However, there was no change in crime rate when

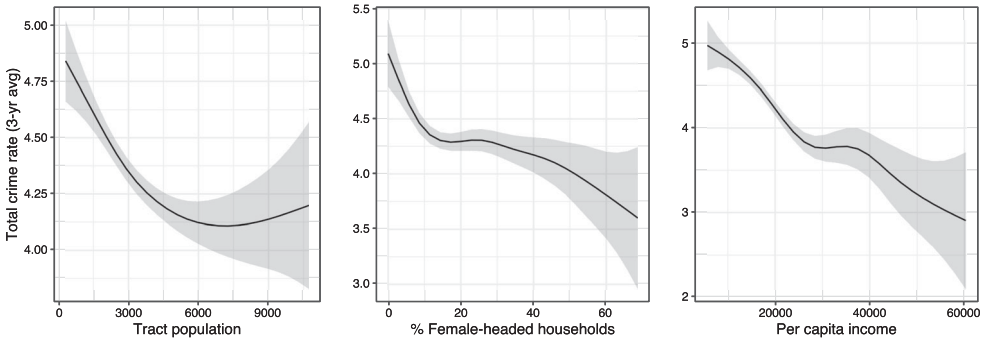


Figure 8. Smooths of predictor variables from the GAM fit with the *R* method using a Gamma family assumption with log link on crime data.

the percent of female-headed households was approximately between 15% and 25%. Similarly, crime rate decreased as per capita income increased, but per capita income between $28,000 and $35,000 was estimated to produce no change in crime rate.

5. Discussion

This paper proposes and studies rank-based estimators of generalised additive models. This provides a viable alternative to the usual likelihood-based estimator of GAMs. Our estimation algorithm is simple. In our reformulation of rank estimators of GAMs as iteratively reweighted penalised least squares estimators, we manage to (1) take advantage of existing weighted GAM theory to establish the theoretical properties of the proposed estimator and (2) take advantage of existing software (eg. `mgcv` in R) to fit the models. In particular, our estimation procedure proceeds by performing repeated weighted GAM fits until convergence conditions are met. We evaluated the relative change in fits to establish convergence.

Pointwise confidence intervals as well as simultaneous confidence bands used for inference can be constructed using the fitted model following the bootstrap procedure described in Section 6.5 of Ruppert et al. (2003). This involves estimating each function over a grid of x values and then using bootstrap to approximate the standard error at each of the grid points. For simultaneous confidence intervals, one may employ percentile bootstrap confidence intervals (Efron and Tibshirani 1994) based on the maximum absolute standardised deviation of the estimated function, as opposed to using quantiles based on data simulated from the multivariate normal distribution.

Our simulation experiments show that the proposed IRRQL estimation method outperforms GAM and LAD for data derived from processes that are heavy-tailed or contaminated. This is fairly common in climate studies, and investigators often depend on simplifying the problem so that they can apply basic nonparametric tests such as the Wilcoxon rank-sum test. However, such approaches are not easy to apply for high-dimensional data with complex underlying structure. Thus, the proposed method gives a practical approach for studying problems where classical fitting of GAMs is inefficient.

Data availability statement

The data that support the findings of this study are openly available in the Inter-university Consortium for Political and Social Research at https://doi.org/10.3886/ICPSR27501.v1, reference number 27501.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the NSF Graduate Research Fellowship (Division of Graduate Education) [grant number DGE-1414475] and NSF under Division of Mathematical Sciences [grant number DMS-1343651].

ORCID

Hannah E. Correia http://orcid.org/0000-0003-3476-3674
Asheber Abebe http://orcid.org/0000-0001-5759-2383

References

Abebe, A., and McKean, J.W. (2013), 'Weighted Wilcoxon Estimators in Nonlinear Regression', *Australian & New Zealand Journal of Statistics*, 55, 401–420.
Abebe, A., McKean, J.W., Kloke, J.D., and Bilgic, Y.K. (2016), *Iterated Reweighted Rank-Based Estimates for GEE Models*, Cham: Springer International Publishing. 61–79.
Alimadad, A., and Salibian-Barrera, M. (2011), 'An Outlier-robust Fit for Generalized Additive Models with Applications to Disease Outbreak Detection', *Journal of the American Statistical Association*, 106, 719–731.
Bindele, H.F., and Abebe, A. (2015), 'Semi-parametric Rank Regression with Missing Responses', *The Journal of Multivariate*, 142, 117–132.
Buja, A., Hastie, T., and Tibshirani, R. (1989), 'Linear Smoothers and Additive Models', *The Annals of Statistics*, 17, 453–510.
Chang, W.H., McKean, J.W., Naranjo, J.D., and Sheather, S.J. (1999), 'High-Breakdown Rank Regression', *Journal of the American Statistical Association*, 94, 205–219.
Croux, C., Gijbels, I., and Prosdocimi, I. (2012), 'Robust Estimation of Mean and Dispersion Functions in Extended Generalized Additive Models', *Biometrics*, 68, 31–44.
Denhere, M., and Bindele, H.F. (2016), 'Rank Estimation for the Functional Linear Model', *Journal of Applied Statistics*, 43 (10) 1–17.
Draper, D. (1988), 'Rank Based Robust Analysis of Linear Models. I. Exposition and Review', *Statistical Science*, 3, 239–271. with comments and a rejoinder by the author.
Efron, B., and Tibshirani, R. (1994), *An Introduction to the Bootstrap*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.
Gu, C., and Wahba, G. (1991), 'Minimizing GCV/GML Scores with Multiple Smoothing Parameters Via the Newton Method', *SIAM Journal on Scientific and Statistical Computing*, 12, 383–398.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P., and Stahel, W.A. (2005), *Robust Statistics: The Approach Based on Influence Functions*, New York: John Wiley & Sons, Inc.
Hastie, T.J., and Tibshirani, R.J. (1990), *Generalized additive models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, London: Chapman and Hall, Ltd.
Hettmansperger, T.P., and McKean, J.W. (2011), *Robust Nonparametric Statistical Methods*, Vol. 119 of *Monographs on Statistics and Applied Probability* (2nd ed.), Boca Raton, FL: CRC Press.
Jaeckel, L.A. (1972), 'Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals', *Annals of Mathematical Statistics*, 43, 1449–1458.

Jurečková, J. (1971), 'Nonparametric Estimate of Regression Coefficients', *Annals of Mathematical Statistics*, 42, 1328–1338.
Miakonkana, G.-v. M., and Abebe, A. (2014), 'Iterative Rank Estimation for Generalized Linear Models', *Journal of Statistical Planning and Inference*, 151/152, 60–72.
Peterson, R.D., and Krivo, L.J. (2010), 'National Neighborhood Crime Study (NNCS), 2000 (ICPSR 27501),' Tech. rep., Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], distributed 2010-05-05; Accessed 2019-04-23.
Rousseeuw, P.J., and Leroy, A.M. (2005), *Robust Regression and Outlier Detection* (Vol. 589), New York: John Wiley & Sons, Inc.
Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
Sievers, G.L., and Abebe, A. (2004), 'Rank Estimation of Regression Coefficients Using Iterated Reweighted Least Squares', *Journal of Statistical Computation and Simulation*, 74, 821–831.
Wahba, G. (1990), *Spline models for observational data*, Vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).
Wang, L., Kai, B., and Li, R. (2009), 'Local Rank Inference for Varying Coefficient Models', *Journal of the American Statistical Association*, 104, 1631–1645.
Wedderburn, R.W.M. (1974), 'Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method', *Biometrika*, 61, 439–447.
Wong, R.K.W., Yao, F., and Lee, T.C.M. (2014), 'Robust Estimation for Generalized Additive Models', *J. Comput. Graph. Statist.*, 23, 270–289.
Wood, S.N. (2003), 'Thin Plate Regression Splines', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 95–114.
Wood, S.N. (2004), 'Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models', *Journal of the American Statistical Association*, 99, 673–686.
Wood, S.N. (2006), *Generalized Additive Models*, Texts in Statistical Science Series, Boca Raton, FL: Chapman & Hall/CRC, an Introduction with *R*.
Wood, S.N., and Augustin, N.H. (2002), 'GAMs with Integrated Model Selection Using Penalized Regression Splines and Applications to Environmental Modelling', *Ecological Modelling*, 157, 157–177.