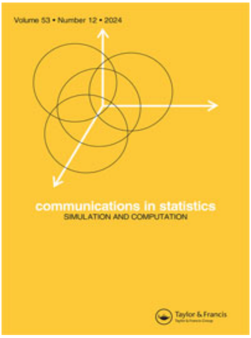

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lssp20

Quantile regression based on the skewed exponential power distribution

Lukas Arnroth & Johan Vegelius

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2024, VOL. 53, NO. 12, 6189–6205
https://doi.org/10.1080/03610918.2023.2224945

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Quantile regression based on the skewed exponential power distribution

Lukas Arnroth and Johan Vegelius

Department of Statistics, Uppsala University, Uppsala, Sweden

ABSTRACT

Bayesian quantile regression generally relies on the asymmetric Laplace distribution (ALD) as the error distribution. We consider methods for L_p-quantile regression based on the skewed exponential power distribution (SEPD). Both Bayesian and frequentist estimation procedures are outlined and compared with previous work based on the SEPD. We find that our proposed methods greatly outperform a previous method in terms of quantile estimation. Further, compared with standard quantile regression, we find that our proposed methods generally perform better in terms of root mean square error (RMSE). Empirical evidence of the statistical properties of the proposed models is provided through a simulation study. Further, a real data application illustrates their performance.

ARTICLE HISTORY
Received 26 May 2022
Accepted 6 June 2023

KEYWORDS
Lp-quantiles; Markov Chain Monte Carlo; Non-informative prior

1. Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has received increasing attention as a more comprehensive approach to the statistical analysis of linear models compared to regression on the mean. Quantile regression minimizes the sum of asymmetrically weighted absolute residuals and allows for the exploration of the relationship between the response distribution's quantiles and available covariates.

Let y_i be a response variable and $\boldsymbol{x}_i$a $k \times 1$ vector of covariates for the ith observation. Consider the linear model

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients for the explanatory variables, and ε_i is the error term. Quantile regression models the τth conditional quantile of y given x as

$$Q_{y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau, \quad i = 1, 2, ..., n, \tag{2}$$

where $Q_Y(\cdot)$ denotes the quantile function of the random variable Y. The linear model (1) with $Q_{\varepsilon_i}(\tau|\boldsymbol{x}_i) = 0$ is equivalent to (2). The τ-specific coefficient vector $\boldsymbol{\beta}_\tau$ can be estimated by minimizing the loss function

$$\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau), \tag{3}$$

where $\rho_\tau(\cdot)$ is the check function defined by

CONTACT Lukas Arnroth lukas.arnroth@statistik.uu.se Department of Statistics, Uppsala University, Uppsala, Sweden.

$$\rho_\tau(u) = u(\tau - I(u < 0)) = |\tau - I(u < 0)||u|, \tag{4}$$

and $I(\cdot)$ denotes the indicator function.

Minimization of (3) is equivalent to maximum likelihood estimation assuming the error of (1) is distributed according to the asymmetric Laplace distribution (ALD). Yu and Moyeed (2001) first proposed a Bayesian quantile regression based on the ALD. Recently, both Bayesian and frequentist methods for quantile regression based on the skewed exponential power distribution (SEPD) have been proposed (Bernardi, Bottone, and Petrella 2018; Yang, Gallagher, and McMahan 2019). The SEPD is a flexible class of distributions that includes the ALD as a special case. Other special cases of the SEPD are the normal, skewed normal, Laplace, uniform, and exponential power distribution (EPD) (Zhu and Zinde-Walsh 2009). Maximization of the SEPD is closely related to L_p-quantile regression (Chen 1996; Daouia, Girard, and Stupfler 2019), and is based on the following modification of (4) (Daouia, Girard, and Stupfler 2019)

$$\rho_{\tau,p}(u) = |\tau - I(u \leq 0)||u|^p, \quad p \geq 1. \tag{5}$$

The τ-specific vector of L_p-quantile regression estimates $\boldsymbol{\beta}_\tau(p)$ is then based on the loss function

$$\sum_{i=1}^{n} \rho_{\tau,p}(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}),$$

which includes (3) as a special case. Apart from quantile regression, L_p-quantile regression also includes asymmetric least squares as a special case, known as expectile regression (Waltrup et al. 2015). See, e.g. Xing and Qian (2017) for a Bayesian approach on expectiles. Extreme value theory has been a primary area of research where $\tau \to 1$ is of particular interest; see, e.g. Daouia, Gijbels, and Stupfler (2019); Daouia, Girard, and Stupfler (2019); Usseglio-Carleve (2018). Compared to quantile regression, the interpretation of L_p-quantile regression estimates is less intuitive, but the estimates can be transformed to have the same interpretation as quantile estimates (Jones 1994; Daouia, Girard, and Stupfler 2019). L_p-quantile regression is mainly motivated by its property of bridging between the robustness of quantile regression and sensitivity of expectile regression. Such is our motivation for quantile regression based on the SEPD.

The motivation for quantile regression based on the SEPD in the approach by Bernardi, Bottone, and Petrella (2018) is increased robustness relative to the ALD distribution. However, with our approach, based on the interpretation as L_p-quantiles, we get estimates that perform much more similar to estimates from standard quantile regression when plugging them into the empirical cumulative distribution function (CDF) of the data. We outline Bayesian and frequentist estimation methods and contrast their performance with the standard quantile regression and the method by Bernardi, Bottone, and Petrella (2018). When comparing the proposed method with that of Bernardi, Bottone, and Petrella (2018), primary focus is on the prediction performance of the methods in terms of quantiles. When comparing with standard quantile regression, we consider the performance of the developed methods in relation to standard quantile regression in terms of both quantile estimation and RMSE. Comparisons are carried out through a simulation study and with application to real data.

The paper proceeds as follows. In Sec. 2, we provide a discussion on the SEPD and its connection to quantile regression. In Sec. 3, we present new results for a non-informative prior for the SEPD and describe the estimation procedures. In Sec. 4, we present the results of a simulation study. In Sec. 5, we present a data application. Finally, in Sec. 6, we summarize the main results and draw conclusions from the study.

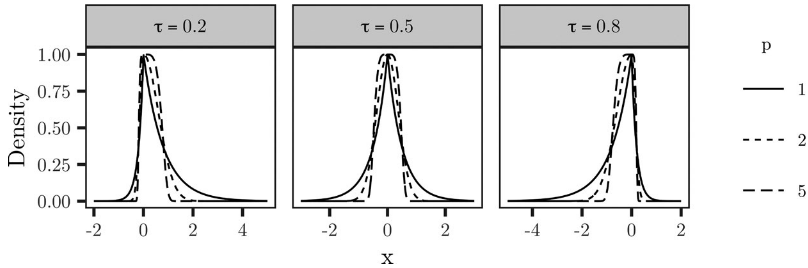

Figure 1. Density function of the SEPD with $\mu = 0$ and $\sigma = 1$.

2. Preliminaries

In this section, we will outline the density function of the SEPD and show how it relates to quantiles *via* L_p-quantiles. The density function of the SEPD is (Zhu and Zinde-Walsh 2009)

$$f(x;\mu,\sigma,p,\tau) = \begin{cases} \dfrac{1}{\sigma}\exp\left\{-\Gamma\left(1+\dfrac{1}{p}\right)^p \dfrac{|\mu-x|^p}{\sigma^p\tau^p}\right\} & \text{if } x \le \mu \\ \dfrac{1}{\sigma}\exp\left\{-\Gamma(1+\dfrac{1}{p})^p \dfrac{|x-\mu|^p}{\sigma^p(1-\tau)^p}\right\} & \text{if } x > \mu, \end{cases} \tag{6}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters respectively, $p > 0$ is the shape parameter, and $\tau \in (0,1)$ characterizes the degree of asymmetry. Special cases are ALD ($p=1$), skew normal, ($p=2$), normal, ($\tau = 0.5$, $p=2$), Laplace ($\tau = 0.5$, $p=1$) and Uniform ($p \to \infty$). The density function (6) is displayed in Figure 1 for some interesting values of τ and p.

Bernardi, Bottone, and Petrella (2018) motivates using (6) for Bayesian quantile regression by the property of μ coinciding with the τth quantile, i.e. $P(Y \le \mu) = \tau$ for $Y \sim \text{SEPD}(\mu,\sigma,p,\tau)$ (Zhu & Zinde-Walsh 2009). However, this interpretation of the parameters of the SEPD does not generalize well to estimation of the location parameter under misspecification of τ.

To elaborate on the issue of misspecification, assume that there exists some $\mu \in \mathbb{R}$ such that $P_{\theta_0}(Y \le \mu) = \tau$ where it is assumed that $Y \sim \text{SEPD}(\mu_0,\sigma_0,p_0,\tau_0)$ where $\theta_0 = (\mu_0,\sigma_0,p_0,\tau_0)^T$. Clearly, $\tau \ne \tau_0$ implies that $\mu \ne \mu_0$ by monotinicity of the CDF when evaluated under θ_0. Thus, the procedure of Bernardi, Bottone, and Petrella (2018) assumes the existence of some p such that $P_{\theta_p}(Y \le \mu) = \tau$ where $\theta_p = (\mu_0,\sigma_0,p,\tau_0)$. Without loss of generality, let $\mu_0 = 0$ and $\sigma_0 = 1$. Further, we can simplify the situation and assume $\tau < \tau_0$ so that $\mu \le 0$, then

$$P_{\theta_p}(Y \le \mu) = \int_{-\infty}^{\mu} \exp\left\{-\Gamma(1+1/p)^p \frac{(-y)^p}{\tau_0^p}\right\} = \tau_0\left[1 - G\left(|\mu|^p; \frac{1}{p}, \frac{\Gamma(1+1/p)^p}{\tau_0^p}\right)\right] = \tau, \tag{7}$$

where $G(\cdot;\alpha,\beta)$ is the CDF of a Gamma distribution with shape α and rate β. If, e.g. $\tau_0 = 0.2$, then, for $\mu < 0$, the left-hand side of (7) is a monotonically decreasing function in p with supremum τ_0. Hence, we note the existence of a p that satisfies (7) for each $\tau \in (0, 0.2)$, but condition (7) shows that the situation is not as simple as stated by Bernardi, Bottone, and Petrella (2018).

Therefore, when treating the skewness as known and fixed to some value $\tau \ne \tau_0$, where τ_0 is that of the data generating procedure, the performance of the procedure outlined in Bernardi, Bottone, and Petrella (2018), as quantile estimator, varies with the distance between τ and τ_0, which we illustrate in Sec. 4. Only for the special cases $p=1$ or $\tau = \tau_0$ does the procedure of Bernardi, Bottone, and Petrella (2018) result in location estimation that performs reliably for quantiles. Instead, the location parameter should be interpreted from the perspective of L_p-quantiles since the exponential term in (6) is a slight modification of (5) in terms of the asymmetry parameter,

$$\rho_{p,\tau}(u) = |\tau - I(u \leq 0)|^{-p}|u|^p = \begin{cases} \dfrac{|u|^p}{\tau^p}, & \text{if } u \leq 0 \\ \dfrac{|u|^p}{(1-\tau)^p}, & \text{if } u > 0. \end{cases} \tag{8}$$

The L_p-quantile based on (8) is defined as

$$q_\tau(p) = \arg\min_{q\in\mathbb{R}} E\rho_{p,\tau}(Y-q), \tag{9}$$

where $q_\tau(p) = \mu$ in the case where $Y \sim \text{SEPD}(\mu, \sigma, p, \tau)$. To understand the link to quantile regression, note that (8) is continuous and convex for $p > 1$, with derivative

$$\psi_{p,\tau}(y - q_\tau(p)) = \frac{\partial \rho_{p,\tau}(y - q_\tau(p))}{\partial q_\tau(p)} = \begin{cases} \dfrac{p|y-q_\tau(p)|^{p-1}}{\tau^p}, & \text{if } y - q_\tau(p) \leq 0 \\ -\dfrac{p|y-q_\tau(p)|^{p-1}}{(1-\tau)^p}, & \text{if } y - q_\tau(p) > 0. \end{cases}$$

The solution of (9) is a 0 of $E_Y\psi_{p,\tau}(Y - q_\tau(p))$, see Chen (1996). Hence, by the first order condition, τ can be expressed as

$$\tau = \left[\left(\frac{E|Y-q_\tau(p)|^{p-1}}{E|Y-q_\tau(p)|^{p-1}I(Y \leq q_\tau(p))} - 1\right)^{\frac{1}{p}} + 1\right]^{-1}. \tag{10}$$

When $p=1$, (10) reduces to $\tau = EI(Y \leq q_\tau(1))$, i.e. quantile regression. For $p \neq 1$, the L_p-quantile $\tilde{\tau}$ such that $q_{\tilde{\tau}}(p) = q_\tau(1)$ is given by

$$\tilde{\tau} = \left[\left(\frac{E|Y-q_\tau(1)|^{p-1}}{E|Y-q_\tau(1)|^{p-1}I(Y \leq q_\tau(1))} - 1\right)^{\frac{1}{p}} + 1\right]^{-1}, \tag{11}$$

which is a slight modification of (2) in Daouia, Girard, and Stupfler (2019). Using $\tilde{\tau}$ greatly improves the performance of the estimated coefficients as quantile estimators. As was pointed out by a reviewer, using $\tilde{\tau}$ in place of τ can be viewed as bias correction. Our proposed estimation procedure should thus result in estimators from a misspecification, or bias-correction, of the quantile in (9) as

$$\arg\min_{q\in\mathbb{R}} E\rho_{p,\tilde{\tau}}(Y-q),$$

where the expectation is taken with respect to the underlying distribution of $Y \sim \text{SEPD}(\mu, \sigma, p, \tau)$.

In Figure 2, we illustrate the relationship between $\tilde{\tau}$ in (11) and τ, for multiple p. The true values of the data generating distribution are $p = 2, \tau = 0.5$ and $\sigma = 1$. As $p < 1$ is of particular

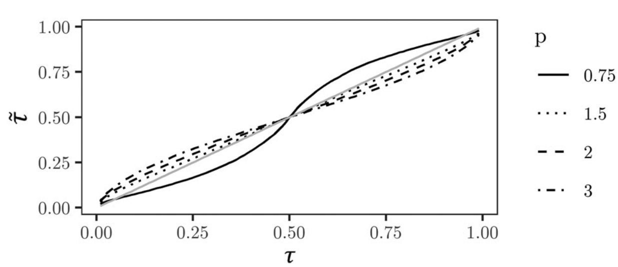

Figure 2. Relationship between L_p quantiles and L_1 quantiles, the gray line represents the case $p = 1$.

interest in Bernardi, Bottone, and Petrella (2018), it is included in Figure 2. However, as $p > 1$ is assumed for the conversion in (11), we restrict the shape parameter in our methods to $p \geq 1$ as in Daouia, Girard, and Stupfler (2019), noting the inclusion of $p = 1$ as it corresponds to quantile regression.

3. Method

This section describes the Bayesian and frequentist estimation procedures for quantile regression based on the SEPD. The complete procedures involve three steps: first, estimating all parameters of the SEPD; second, converting the quantile outlined in (11); and finally, estimating only the regression coefficients using the bias-corrected quantile.

We start by outlining the initial step, where all parameters of the SEPD are to be estimated. We consider (1) with the assumption that the τth L_p-quantile of the error distribution is 0. For estimation, we assume, as in Bernardi, Bottone, and Petrella (2018), that $\epsilon_i \sim \text{SEPD}(0, \sigma, p, \tau)$. Using the notation $q_\tau(p) = \boldsymbol{x}^T \boldsymbol{\beta}_\tau(p)$, the likelihood of $\boldsymbol{y}$ is

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}_\tau(p), \sigma, p, \tau) = \frac{1}{\sigma^n} \exp\left\{ -\frac{\Gamma(1+\frac{1}{p})^p}{\sigma^p} \left[\sum_{i\in N_1} \frac{|\boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p) - y_i|^p}{\tau^p} + \sum_{i \in N_2} \frac{|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p)|^p}{(1-\tau)^p} \right] \right\}, \tag{12}$$

where $\boldsymbol{X} = (\boldsymbol{x_1}^T, \ldots, \boldsymbol{x_n}^T)^T, N_1 = \{i : y_i < \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p)\}$, and $N_2 = \{i : y_i > \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p)\}$. The likelihood can be expressed in terms of (8) as

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}_\tau(p), \sigma, p, \tau) = \frac{1}{\sigma^n} \exp\left\{ -\frac{\Gamma(1+\frac{1}{p})^p}{\sigma^p} \sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i} - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p)) \right\}.$$

We make the following assumption for the model matrix

Assumption 1. $\boldsymbol{X} \in \mathbb{R}^{n \times k}$ *where* $k < n$

A Bayesian procedure requires the specification of the prior distribution on the unknown parameters. We consider the usual non-informative prior for the location and scale, $\pi(\boldsymbol{\beta}, \sigma) \propto \sigma^{-1}$. For the shape and assymmetry parameter we consider $\pi(p) \propto I(1 \leq p \leq 5)$ and $\pi(\tau) = I(0 < \tau < 1)$ respectively. The resulting posterior is then given by

$$\pi(\boldsymbol{\beta}_\tau(p), \sigma, p, \tau|\boldsymbol{y}) \propto \frac{1}{\sigma} p(\boldsymbol{y}|\boldsymbol{X}, \sigma, \boldsymbol{\beta}_\tau(p), p, \tau) I(1 \leq p \leq 5) I(0 < \tau < 1), \tag{13}$$

which is proper.

Theorem 1. *If* $\boldsymbol{X}$ *satisfies Assumption 1, and* $\pi(p)$ and $\pi(\tau)$ *are proper priors, the non-informative prior* $\pi(\boldsymbol{\beta}, \sigma) \propto \sigma^{-1}$ *leads to a proper posterior*

Proof. See Appendix A.1. □

We construct a MCMC sampler with (13) as target distribution. To increase the ability of the MCMC algorithm to explore the parameter space, we sample $(p, \sigma)^T$ by a collapsed Gibbs sampler (Liu 1994). The conditional distribution of $(p, \sigma)^T$ is

$$\pi(p, \sigma|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_\tau(p), \tau) \propto \frac{\pi(p)}{\sigma^{n+1}} \exp\left\{ -\frac{\Gamma(1+\frac{1}{p})^p}{\sigma^p} \sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i} - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau(p)) \right\}, \tag{14}$$

where marginalizing out σ gives

$$\begin{aligned}\pi(p|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),\tau) &\propto \int_0^\infty \pi(p,\sigma|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),\tau)d\sigma \\ &= \frac{I(1\le p\le 5)}{p}\int_0^\infty \eta^{-\frac{n}{p}-1}\exp\left\{-\frac{\Gamma(1+\frac{1}{p})^p}{\eta}\sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right\}d\eta \\ &\propto \frac{\Gamma(n/p)}{p}\left(\Gamma(1+1/p)^p\sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right)^{-\frac{n}{p}} I(1\le p\le 5),\end{aligned} \tag{15}$$

where the second step follows from the substitution $\eta=\sigma^p$, which gives an integral proportional to the Inverse-Gamma density function. As the normalizing constant of (15) is unknown, it is sampled *via* Metropolis-Hastings (MH). For $\xi_p>0$, proposals are generated as $p^{(t+1)}\sim \mathcal{N}_{(1,\infty)}(p^{(t)},\xi_p^2)$, where $\mathcal{N}_{\mathcal{A}}(\mu,\sigma)$ denotes the normal distribution with location μ and variance σ truncated to the set $\mathcal{A}$. The corresponding acceptance probability is

$$\varphi_p(p^{(t+1)},p^{(t)}) = \min\left\{1,\ \frac{\pi(p^{(t)}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),\tau)}{\pi(p^{(t+1)}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),\tau)}\frac{1-\Phi\left(\frac{1-p^{(t)}}{\xi_p}\right)}{1-\Phi\left(\frac{1-p^{(t+1)}}{\xi_p}\right)}\right\},$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (CDF). The conditional distribution of σ is

$$\pi(\sigma|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),p,\tau) \propto \frac{1}{\sigma^{n+1}}\exp\left\{-\frac{\Gamma(1+1/p)^p}{\sigma^p}\sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right\}. \tag{16}$$

The random variable η is then distributed as

$$\eta|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),p,\tau \sim \mathcal{IG}\left(\frac{n}{p},\Gamma(1+1/p)^p\sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right), \tag{17}$$

where $\mathcal{IG}(a,b)$ denotes the Inverse-Gamma distribution with shape a and scale b. Hence, σ can be sampled by first sampling η from (17), then setting $\sigma=\eta^{1/p}$. Sampling p from (15) and σ *via* (17) constitutes a blocked sample of the parameters.

The conditional distribution of $\boldsymbol{\beta}_\tau(p)$ is

$$\pi(\boldsymbol{\beta}_\tau(p)|\boldsymbol{y},\boldsymbol{X},p,\sigma,\tau) \propto \exp\left\{-\frac{\Gamma(1+1/p)^p}{\sigma^p}\sum_{i=1}^n \rho_{\tau,p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right\}, \tag{18}$$

where the normalizing constant is unknown. Due to the possibly large dimension of $\boldsymbol{\beta}_\tau(p)$, we consider the Metropolis adjusted Langevin algorithm (MALA) to sample from (18) efficiently (Girolami and Calderhead 2011). For $\xi_\beta>0$, proposals are generated as

$$\boldsymbol{\beta}_\tau^{(t+1)}(p) = \boldsymbol{\beta}_\tau^{(t)}(p) + \frac{\xi_\beta^2}{2}\boldsymbol{A}(\boldsymbol{\beta}_\tau^{(t)}(p))\nabla_{\boldsymbol{\beta}_\tau^{(t)}(p)}\mathcal{L}(\boldsymbol{\beta}^{(t)}) + \xi_\beta\sqrt{\boldsymbol{A}(\boldsymbol{\beta}_\tau^{(t)}(p))}\boldsymbol{z},$$

where $\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$, and

$$\begin{aligned}\nabla_{\boldsymbol{\beta}}\mathcal{L}(\boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}}\log\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},p,\sigma,\tau) \\ &= \frac{p\Gamma(1+\frac{1}{p})^p}{\sigma^p}\left[\sum_{i\in N_1}\frac{|\boldsymbol{x}_i^T\boldsymbol{\beta}-y_i|^{p-1}}{\tau^p}\boldsymbol{x}_i - \sum_{i\in N_2}\frac{|y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}|^{p-1}}{(1-\tau)^p}\boldsymbol{x}_i\right],\end{aligned}$$

and $\boldsymbol{A}(\boldsymbol{\beta})^{-1}$ is the Hessian,

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{\beta})^{-1} &= -\nabla^2_{\boldsymbol{\beta}} \log \pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X},p,\sigma,\tau) \\ &= \frac{p(p-1)\Gamma(1+\frac{1}{p})^p}{\sigma^p}\left[\sum_{i\in N_2}\frac{|y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}|^{p-2}}{(1-\tau)^p}\boldsymbol{x}_i\boldsymbol{x}_i^T+\sum_{i\in N_1}\frac{|\boldsymbol{x}_i^T\boldsymbol{\beta}-y_i|^{p-2}}{\tau^p}\boldsymbol{x}_i\boldsymbol{x}_i^T\right]. \end{aligned}$$

The acceptance probability is then given by

$$\varphi_\beta(\boldsymbol{\beta}_\tau^{(t+1)}(p),\boldsymbol{\beta}_\tau^{(t)}(p)) = \min\left\{1,\frac{\pi(\boldsymbol{\beta}_\tau^{(t+1)}(p))q(\boldsymbol{\beta}_\tau^{(t)}(p)|\boldsymbol{\beta}_\tau^{(t+1)}(p))}{\pi(\boldsymbol{\beta}_\tau^{(t)}(p))q(\boldsymbol{\beta}_\tau^{(t+1)}(p)|\boldsymbol{\beta}^{(t)})}\right\},$$

where $q(\boldsymbol{\beta}|\boldsymbol{\beta}')=\mathcal{N}(\boldsymbol{\beta};\boldsymbol{\beta}'+\xi_\beta^2/2\boldsymbol{A}(\boldsymbol{\beta}')\nabla_{\boldsymbol{\beta}'}\mathcal{L}(\boldsymbol{\beta}'),\xi_\beta^2\boldsymbol{A}(\boldsymbol{\beta}'))$.

Lastly, the conditional distribution of τ is given by

$$\pi(\tau|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),p,\sigma) \propto \exp\left\{-\frac{\Gamma(1+1/p)^p}{\sigma^p}\sum_{i=1}^n\rho_{\tau,p}(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau(p))\right\}I(0<\tau<1), \tag{19}$$

which is sampled *via* a MH step with proposals generated as $\tau^{(t+1)}\sim\mathcal{N}_{(0,1)}(\tau^{(t)},\xi_\tau^2)$ where $\xi_\tau>0$. Hence, the acceptance probability is

$$\varphi_\tau(\tau^{(t+1)},\tau^{(t)}) = \min\left\{1,\frac{\pi(\tau^{(t)}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),p,\sigma)}{\pi(\tau^{(t+1)}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta}_\tau(p),p,\sigma)}\frac{\Phi\left(\frac{1-\tau^{(t)}}{\xi_\tau}\right)-\Phi\left(\frac{-\tau^{(t)}}{\xi_\tau}\right)}{\Phi\left(\frac{1-\tau^{(t+1)}}{\xi_\tau}\right)-\Phi\left(\frac{-\tau^{(t+1)}}{\xi_\tau}\right)}\right\}.$$

The full MCMC sampler is outlined in Algorithm 1. One key difference compared to the method of Bernardi, Bottone, and Petrella (2018) is the sampling of the skewness parameter. Furthermore, the joint sampling of $(p,\sigma)^T$ is a modification that leads to increased efficiency in the exploration of the posterior distribution. Also, the regression coefficients are sampled differently in Algorithm 1 as the gradient and Hessian of their conditional distribution are utilized.

Algorithm 1: MCMC for sampling from the posterior

Input: $\boldsymbol{\beta}_\tau^{(1)}(p),p^{(1)},\sigma^{(1)},\tau^{(1)}\ \boldsymbol{X},\boldsymbol{y},N,\xi_p,\xi_\beta,\xi_\tau$

Output: $\left\{\boldsymbol{\beta}_\tau^{(i)}(p),p^{(i)},\sigma^{(i)},\tau^{(i)}\right\}_{i=1}^N$

for $i\leftarrow 2$ **to** N **do**

  Sample $p'\sim\mathcal{N}_{(1,\infty)}(p,\xi_p^2)$

  Accept with probability $\varphi_p(p',p^{(i-1)})$

  Sample $\eta\sim\mathcal{IG}(\frac{p^{(i)}}{n},\Gamma\left(1+\frac{1}{p^{(i)}}\right)^{p^{(i)}}\sum_{i=1}^n\rho_{\tau^{(i-1)},p}(\boldsymbol{y_i}-\boldsymbol{x}_i^T\boldsymbol{\beta}_\tau^{(i-1)}(p)))$

  Set $\sigma^{(i)}=\eta^{1/p^{(i)}}$

  Sample $\boldsymbol{\beta}'\sim\mathcal{N}(\boldsymbol{\beta}_\tau^{(i-1)}(p)+\frac{\xi_\beta^2}{2}\boldsymbol{A}(\boldsymbol{\beta}_\tau^{(i-1)}(p))\nabla_{\boldsymbol{\beta}_\tau^{(i-1)}(p)}\mathcal{L}(\boldsymbol{\beta}_\tau^{(i-1)}(p)),\xi_\beta^2\boldsymbol{A}(\boldsymbol{\beta}_\tau^{(i-1)}(p)))$

  Accept with probability $\varphi(\boldsymbol{\beta}',\boldsymbol{\beta}_\tau^{(i)}(p))$

  Sample $\tau'\sim\mathcal{N}_{(0,1)}(\tau,\xi_\tau^2)$

  Accept with probability $\varphi_\tau(\tau',\tau^{(i-1)})$

For frequentist estimation, we employ a coordinate ascent procedure to estimate the parameters $\boldsymbol{\beta}_\tau(p),\sigma,p$ and τ by maximizing (12). Let $\lambda=(\sigma,p,\tau)^T$, the method is then outlined in Algorithm 2. Notably, our procedure is similar to that presented in Yang, Gallagher, and McMahan (2019), but with some key differences. They optimize over the groups $(\boldsymbol{\beta}_\tau(p),\tau)^T$ and $(\sigma,p)^T$, where $(\boldsymbol{\beta}_\tau(p),\tau)^T$ is estimated using a coordinate ascent procedure based on initial estimates of σ and p. Estimates of σ and p are then constructed based on the estimate of $(\boldsymbol{\beta}_\tau(p), \tau)^T$. We find that Algorithm 2 is satisfactory and prefer it for its simplicity. Furthermore, separating the parameters into groups $\boldsymbol{\beta}_\tau(p)$ and λ is more natural in this context. Finally, we note that the estimates of $\boldsymbol{\beta}_\tau(p)$ and τ are sensitive to the initial values of σ and p, which we consider to be a drawback.

Algorithm 2: Frequentist estimation of the SEPD parameters

Input: $\boldsymbol{\beta}_\tau^{(0)}(p), p^{(0)}, \sigma^{(0)}, \tau^{(0)}\ \boldsymbol{X}, \boldsymbol{y}$

Output: $\widehat{\boldsymbol{\beta}_\tau(p)}, \hat{p}, \hat{\sigma}, \hat{\tau}$

1 Set $i = 1$

2 Update $\beta_\tau(p)$ by computing $\beta_\tau^{(i)}(p) = \arg\max_\beta \log p(\boldsymbol{y}|\boldsymbol{X}, \beta, \lambda^{(i-1)})$

3 Update λ by computing $\lambda^{(i)} = \arg\max_\lambda \log p(\boldsymbol{y}|\boldsymbol{X}, \beta_\tau^{(i)}(p), \lambda)$

4 Repeat steps 2 and 3 until convergence

Given estimates of all SEPD parameters using either Algorithm 1 or 2, the link to quantile regression, as given in (11), remains. We use a Monte Carlo (MC) approach, where we generate data using the estimated parameters and estimate (11) as

$$\left[\left(\frac{\frac{1}{NB}\sum_{j=1}^{B}\sum_{i=1}^{N}(y_{i,j} - \hat{q}_\tau^{(j)}(1))^{\hat{p}-1}}{\frac{1}{NB}\sum_{j=1}^{B}\sum_{i=1}^{N}(y_{i,j} - \hat{q}_\tau^{(j)}(1))^{\hat{p}-1} I(y_{i,j} \le \hat{q}_\tau^{(j)}(1))} - 1\right)^{\frac{1}{\hat{p}}} + 1\right]^{-1}, \tag{20}$$

where $y_{i,j} \sim \text{SEPD}(0, \hat{\sigma}, \hat{p}, \hat{\tau})$ and $\hat{q}_\tau^{(j)}(1)$ is the quantile estimate for the jth sample.

Finally, either Algorithm 1 or 2 is computed again with $(p, \sigma, \tau)^T = (\hat{p}, \hat{\sigma}, \hat{\tau})^T$ fixed. The estimated regression coefficients can then be interpreted as quantile regression coefficients. This process, which is the main contribution of this paper, is outlined in Algorithm 3. Note that when computing Bayesian estimates, they are always constructed as posterior mean estimates.

Algorithm 3: Quantile regression based on the SEPD

1 Estimate $\boldsymbol{\beta}_\tau(p), \sigma, p$, and τ using Alg. 1 or 2

2 Estimate $\tilde{\tau}$ by (20) using estimates $(\hat{\sigma}, \hat{p}, \hat{\tau})^T$ from Step 1

3 Estimate $\beta_{\tilde{\tau}}(p)$, fixing $p = \hat{p}, \sigma = \hat{\sigma}$, and $\tau = \tilde{\tau}$, using Alg. 1 or 2. That is, only steps 6-7 are repeated in Alg. 1 or step 2 in Alg. 2

4. Simulation study

In this section, we compare the Bayesian (BSQR) and frequentist (FSQR) estimation procedures outlined in Sec. 3 with the method of Bernardi, Bottone, and Petrella (2018), denoted by SQR.[1] We evaluate their performance as quantile estimators. To avoid confusion with quantiles, still denoted by τ, we use α to denote the skewness parameter of the SEPD. For SQR, we make no modification to the prior when the restriction $p \ge 1$ is not enforced, as we find that using vague proper priors instead yields no practical difference. Data is generated as

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i, \quad i = 1, ..., n, \tag{21}$$

[1]Julia (Bezanson et al. 2017) was used to produce the results. A package was developed for this paper and can be found at https://github.com/lukketotte/SepdQuantile.jl, with code to recreate some results

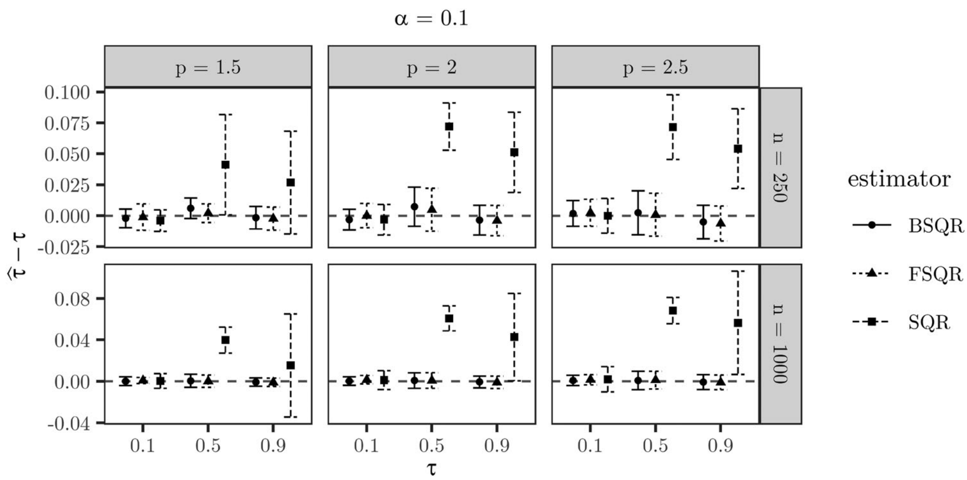

Figure 3. Comparison of differences between estimated quantiles and actual quantiles for the methods based on the SEPD with data generated by the SEPD with skewness 0.1, intervals are ± 1 standard deviation from MC replications.

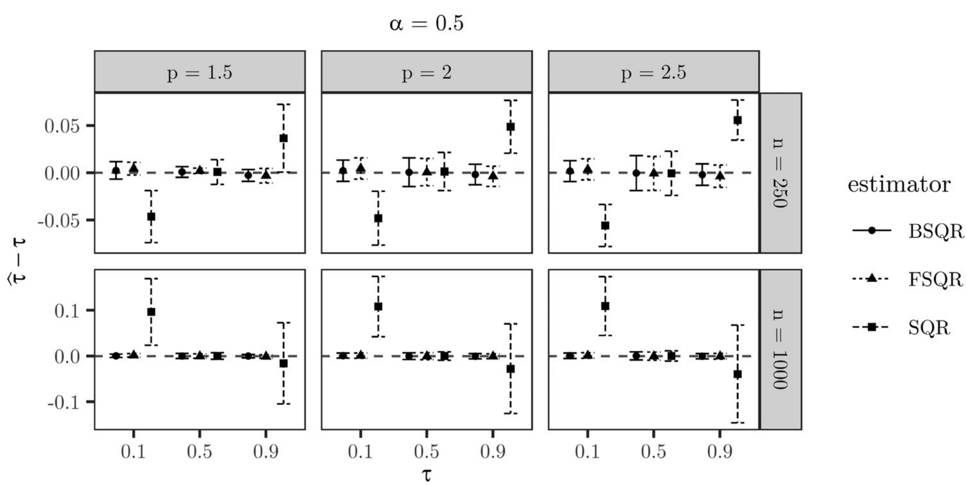

Figure 4. Comparison of differences between estimated quantiles and actual quantiles for the methods based on the SEPD with data generated by the SEPD with skewness 0.5, intervals are ±1 standard deviation from MC replications.

where x_{i1} and x_{i2} are generated from the standard normal distribution, $\beta_0 = \beta_1 = \beta_2 = 1$, and multiple distributions are considered for ε_i.

First, we compare all methods based on the SEPD, for multiple error distributions, as quantile estimators. Consider $\varepsilon_i \sim \text{SEPD}(0, 1, p, \alpha)$ for $p \in \{1.5, 2, 2.5\}, \alpha \in \{0.1, 0.5, 0.9\}$ and sample sizes $n = 250$ and 1000. In Figures 3, 4, and 5, the results are illustrated with 1000 MC replications for each combination of α, p, n and τ. Note that $\hat{\tau}$ in Figure 3, 4, and 5, given an estimator $\widehat{\boldsymbol{\beta}}$, is defined as $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} I(y_i \le \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}})$. It seems that the proposed Bayesian and frequentist methods, BSQR and FSQR, respectively, have similar performance in terms of accuracy and variance for all settings. However, it is noted that the proposed methods are sensitive to sample size, especially their variances. On the other hand, the SQR method performs poorly as a quantile estimator for most settings, except when $\alpha = \tau$, where it performs similarly to the proposed methods. It is also observed, particularly in Figure 3 and 5, that the variance of the SQR method increases with higher distance between τ and α, which is not ideal. In Figure 6, other distributions than the SEPD are considered for ε_i in (21). For the non-symmetric distributions, SQR performs poorly

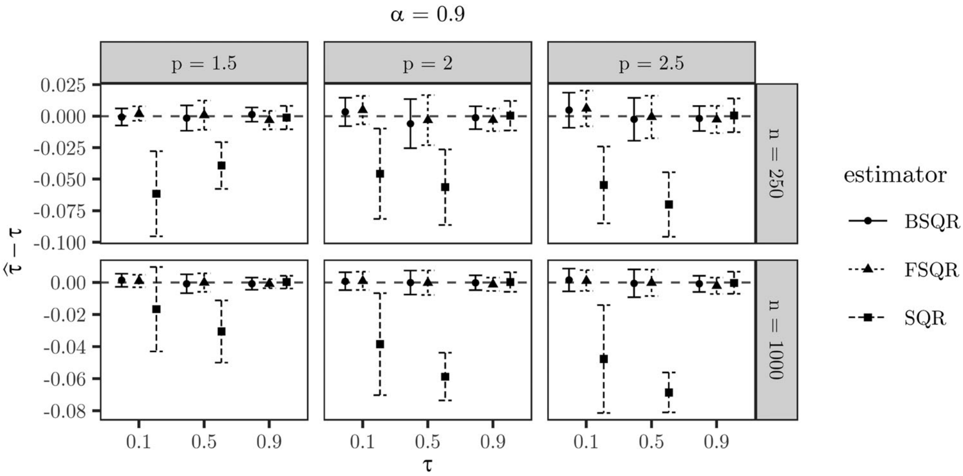

Figure 5. Comparison of differences between estimated quantiles and actual quantiles for the methods based on the SEPD with data generated by the SEPD with skewness 0.9, intervals are ±1 standard deviation from MC replications.

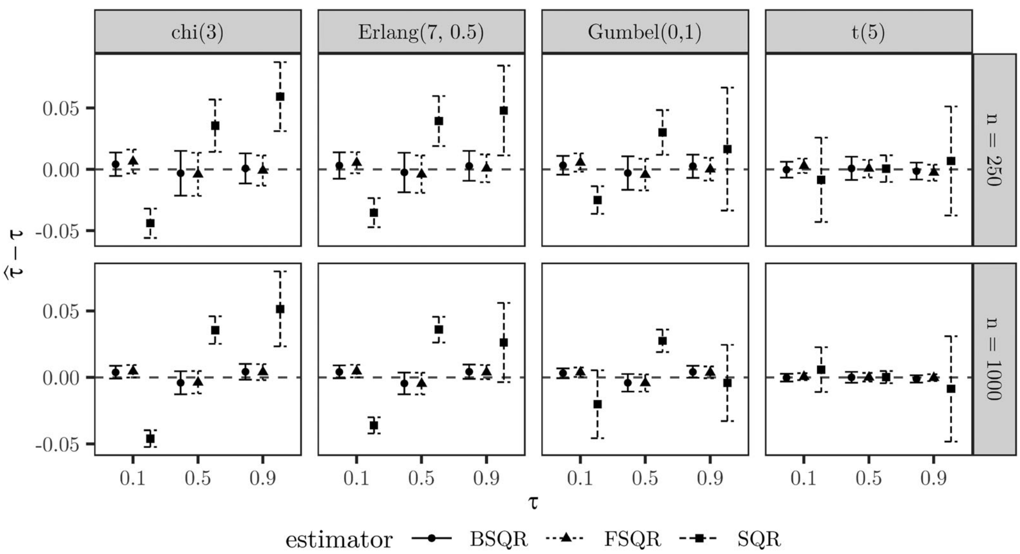

Figure 6. Comparison of differences between estimated quantiles and quantiles for the methods based on the SEPD with data generated by other distributions than the SEPD.

for all sample sizes. For the symmetric t-distribution with 5 degrees of freedom, we find that all methods perform similarly in terms of bias. For $\tau = 0.5$, this follows from the symmetry of the t -distribution. For quantile levels 0.1 and 0.9, the SQR method estimates p as close to 1. Overall, the results in Figure 6 suggest that the proposed methods perform well as quantile estimators for a variety of error distributions, and that the SQR method may perform poorly for non-symmetric distributions and for certain quantile levels.

We now turn to the comparison of the proposed methods with classical frequentist quantile regression (QR) and Bayesian quantile regression (BQR) based on the ALD likelihood.[2] The priors

[2]Frequentist estimation is done via QuantileRegressions.jl, and Bayesian estimation via the R package bayesQR.

Table 1. Comparison of QR and FSQR in terms of RMSE and quantile estimation.

			β_0		β_1		β_2	
Method	τ	$\hat{\tau}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\varepsilon \sim \mathcal{N}(0,1)$	0.5							
QR		0.500	−.500	0.095	0.002	0.099	0.000	0.091
FSQR		0.500	−.500	0.082	−.082	0.078	−.078	0.074
$\varepsilon \sim t(5)$	0.5							
QR		0.500	0.000	0.095	0.001	0.091	0.006	0.092
FSQR		0.501	0.000	0.090	0.000	0.083	0.005	0.085
$\varepsilon \sim 0.5\mathcal{N}(-1.5,1) + 0.5\mathcal{N}(-1.5,1)$	0.5							
QR		0.500	−.500	0.241	0.006	0.231	0.018	0.220
FSQR		0.501	−.501	0.120	−.120	0.105	0.000	0.105
$\varepsilon \sim (1+x_1)\mathcal{N}(0,1)$	0.5							
QR		0.500	−.500	0.104	−.104	0.112	0.005	0.047
FSQR		0.500	−.500	0.104	−.104	0.111	0.005	0.047
$\varepsilon \sim 0.12\mathcal{N}(-1,1) + 0.88\mathcal{N}(2.5,0.5)$	0.1							
QR		0.100	0.211	0.643	−.643	0.425	0.017	0.449
FSQR		0.096	0.249	0.654	0.060	0.430	−.430	0.471
$\varepsilon \sim \text{Gumbel}(4.17,5)$	0.1							
QR		0.100	0.034	0.462	0.002	0.468	0.000	0.392
FSQR		0.105	0.069	0.434	0.001	0.413	−.413	0.405
$\varepsilon \sim t(6,1.282)$	0.1							
QR		0.101	0.001	0.126	0.001	0.126	0.009	0.125
FSQR		0.105	0.009	0.114	0.001	0.119	0.008	0.118
$\varepsilon \sim \mathcal{N}(-1.282,1)$	0.9							
QR		0.900	−.900	0.121	0.006	0.116	−.116	0.120
FSQR		0.895	−.895	0.103	0.006	0.104	0.000	0.109
$\varepsilon \sim 0.88\mathcal{N}(-2.5,0.5) + 0.12\mathcal{N}(1,1)$	0.9							
QR		0.900	−.900	0.644	−.644	0.442	0.005	0.474
FSQR		0.907	−.907	0.712	−.712	0.471	0.02	0.451

Note that $t(a,b)$ denotes the non-central t-distribution with non-centrality parameter a and df b.

Table 2. Comparison of BSQR and BQR in terms of RMSE and quantile estimation.

			β_0		β_1		β_2	
Method	τ	$\hat{\tau}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\varepsilon \sim \mathcal{N}(0,1)$	0.5							
BQR		0.490	−.490	0.078	0.004	0.078	0.000	0.082
BSQR		0.500	−.500	0.081	−.081	0.078	−.078	0.073
$\varepsilon \sim t(5)$	0.5							
BQR		0.493	0.001	0.087	0.001	0.087	0.000	0.088
BSQR		0.500	0.000	0.087	0.000	0.081	0.005	0.083
$\varepsilon \sim 0.5\mathcal{N}(-1.5,1) + 0.5\mathcal{N}(-1.5,1)$	0.5							
BQR		0.499	0.011	0.238	−.238	0.209	−.209	0.221
BSQR		0.500	−.500	0.125	−.125	0.102	0.000	0.104
$\varepsilon \sim (1+x_1)\mathcal{N}(0,1)$	0.5							
BQR		0.485	−.485	0.101	−.101	0.110	0.008	0.051
BSQR		0.500	−.500	0.120	−.120	0.128	0.004	0.080
$\varepsilon \sim 0.12\mathcal{N}(-1,1) + 0.88\mathcal{N}(2.5,0.5)$	0.1							
BQR		0.096	0.167	0.520	−.520	0.362	0.003	0.338
BSQR		0.098	0.166	0.570	−.570	0.401	0.002	0.405
$\varepsilon \sim \text{Gumbel}(4.17,5)$	0.1							
BQR		0.098	0.045	0.442	0.026	0.412	0.050	0.438
BSQR		0.103	0.065	0.412	0.004	0.400	0.000	0.392
$\varepsilon \sim t(6,1.282)$	0.1							
BQR		0.89	−.89	0.124	−.124	0.111	0.003	0.109
BSQR		0.101	0.003	0.111	0.001	0.115	0.009	0.115
$\varepsilon \sim \mathcal{N}(-1.282,1)$	0.9							
BQR		0.903	0.051	0.117	0.000	0.010	−.010	0.103
BSQR		0.897	−.897	0.102	0.006	0.103	0.001	0.108
$\varepsilon \sim 0.88\mathcal{N}(-2.5,0.5) + 0.12\mathcal{N}(1,1)$	0.9							
BQR		0.901	−.901	0.515	0.018	0.371	−.371	0.371
BSQR		0.900	−.900	0.575	−.575	0.413	0.008	0.439

Note that $t(a,b)$ denotes the non-central t-distribution with non-centrality parameter a and df b.

Table 3. Coverage rate of 95% bootstrap intervals of $\beta_1 = 1$.

Distribution	Method	
	FSQR	QR
$\varepsilon \sim 0.88\mathcal{N}(-2.5, 0.5) + 0.12\mathcal{N}(1, 1)$	0.941	0.962
$\varepsilon \sim \mathcal{N}(-1.282, 1)$	0.941	0.966
$\varepsilon \sim t(6, 1.282)$	0.943	0.968
$\varepsilon \sim (1 + x_1)\mathcal{N}(0, 1)$	0.934	0.932

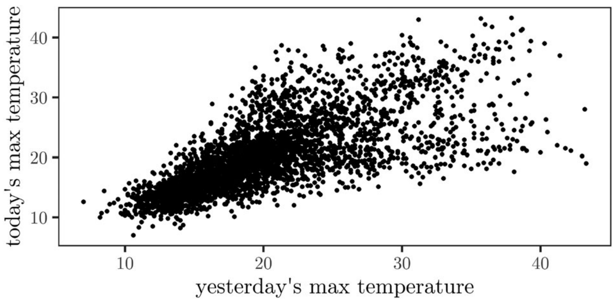

Figure 7. Daily max temperatures over 10 years for Melbourne, Australia as an AR(1) scatterplot.

used for BQR are the default settings of `BayesQR` (Benoit and Van den Poel 2017). To compare RMSE and bias for multiple quantiles, we consider distributions for ε_i such that $Q_{\varepsilon_i}(\tau) \approx 0$ for $\tau = 0.1, 0.5, 0.9$. We generate data using (21) with $n = 200$. Table 1 shows the results for the comparison of the frequentist approaches, from which we can observe that the proposed methods generally outperform classical quantile regression. However, the proposed method performs worse for the mixture distributions in the cases $\tau = 0.1$ and 0.9. This drop in performance for the mixture cases might be attributed to issues with Algorithm 2 for values of p near the boundary $p = 1$. Table 2 shows the results for the comparison of the Bayesian procedures, where the two methods perform similarly, and the proposed method generally performs slightly better except for the mixture distributions and heteroskedastic case. Nonetheless, we note that BQR performs worse than the proposed method in terms of estimating quantiles.

Finally, we evaluate the performance of the non-parametric bootstrap used to obtain standard deviations in the application to real data in Sec. 5, by comparing the coverage rates of FSQR based 95% bootstrap confidence intervals for $\beta_1 = 1$ with corresponding QR based bootstrap intervals. Results based on 250 replications with 1000 bootstrap estimates for each replication are displayed in Table 3 where we observe that the bootstrap for the two methods perform similarly. However, we find that the coverage rates for QR is higher for all cases but the heteroskedastic case. Note that only the smaller sample size of $n = 200$ is considered.

5. Application

Koenker (2000) considered the daily maximum temperature in Melbourne, Australia from January 1st, 1981, to December 30th, 1990, as shown in Figure 7. Following Koenker (2000), we consider the model

$$\log(y_t) = \beta_0 + \beta_1 \log(y_{t-1}) + \varepsilon_t, t = 2, ..., 3649. \tag{22}$$

We use Algorithm 1 and 2 to estimate the parameters in (22) and present the results in Table 4. The standard errors for frequentist estimation are based on 1000 non-parametric bootstrap replications. The estimates from both methods are nearly identical, with the largest difference being the

Table 4. Estimated parameters for the maximum temperature data, standard errors in parenthesis.

Estimator	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{p}$	$\hat{\tau}$	$\hat{\sigma}$
Bayesian (Alg. 1)	0.598	0.800	1.524	0.516	0.398
	(0.037)	(0.0136)	(0.058)	(0.012)	(0.012)
Freq. (Alg. 2)	0.603	0.798	1.527	0.513	0.398
	(0.051)	(0.012)	(0.06)	(0.013)	(0.015)

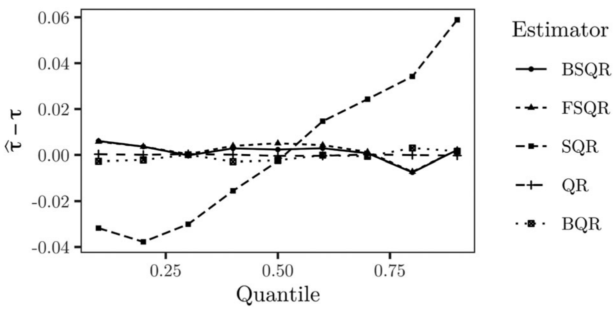

Figure 8. Comparison of SQR, proposed methods (BSQR and FSQR) and quantile regression (QR and BQR) as quantile estimators for the Melbourne temperature data.

Table 5. Comparison of $\hat{\tau}$ with 5-fold cross-validation, standard errors in parenthesis.

	τ			
Estimator	0.2	0.4	0.6	0.8
BSQR	0.211 (0.019)	0.403 (0.018)	0.597 (0.013)	0.805 (0.017)
FSQR	0.212 (0.020)	0.406 (0.018)	0.596 (0.013)	0.805 (0.018)
BQR	0.201 (0.017)	0.402 (0.017)	0.602 (0.015)	0.795 (0.019)
QR	0.201 (0.019)	0.401 (0.019)	0.599 (0.018)	0.799 (0.018)

Table 6. Comparison of the proposed estimators and quantile regression for quantile levels 0.1, 0.5 and 0.9, with standard errors in parenthesis.

	$\tau = 0.1$		$\tau = 0.5$		$\tau = 0.9$	
Estimator	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
BSQR	0.974	0.594	0.611	0.794	0.513	0.905
	(0.025)	(0.008)	(0.032)	(0.011)	(0.031)	(0.011)
FSQR	0.975	0.594	0.613	0.793	0.513	0.905
	(0.044)	(0.015)	(0.036)	(0.012)	(0.049)	(0.016)
QR	1.038	0.571	0.493	0.835	0.435	0.930
	(0.052)	(0.019)	(0.048)	(0.017)	(0.075)	(0.026)
BQR	1.025	0.574	0.491	0.835	0.435	0.931
	(0.122)	(0.043)	(0.132)	(0.046)	(0.202)	(0.069)

standard error of the intercept term. For the Bayesian approach, we obtain $N = 2 \cdot 10^4$ samples from the posterior distribution, discarding the first 1000 as burn-in.

In Figure 8, all methods are compared for quantiles $\tau \in \{0.1, 0.2, ..., 0.9\}$. It can be observed that the quantile estimators based on Algorithm 3 perform worse than standard frequentist and Bayesian quantile regression in terms of predicting the quantile levels. However, the SQR performs very poorly in terms of estimating the quantiles, and should only be considered viable for predicting the median in this application. In Table 5, we compare the proposed methods and standard methods for estimating quantiles using 5-fold cross-validation. The results are similar to those observed in Figure 8.

In Table 6, the proposed methods and standard frequentist and Bayesian quantile regression are compared for quantiles 0.1, 0.5 and 0.9. The results show that the proposed Bayesian approach has the lowest variance for all quantile levels in terms of precision. Additionally, the proposed frequentist approach produces estimates with lower variance compared to the standard procedures.

6. Conclusion

Quantile regression based on the SEPD is considered with particular focus on improving SQR, the method of Bernardi, Bottone, and Petrella (2018), by utilizing the link to L_p-quantile regression. Both frequentist and Bayesian estimation procedures are outlined, where a non-informative scale-location prior is shown to result in a proper posterior.

In a simulation study, the performance of SQR as quantile regression is found to be poor when the skewness parameter is fixed to values far from those used in the data generating process. The addition of a transitional step from L_p-quantiles to standard quantiles greatly improves the SEPD-based quantile regression in terms of estimating quantiles. However, one caveat is the assumption that the data is SEPD distributed in the quantile conversion. There is a noticeable drop in performance when the error distribution is assumed to be a mixture of normal distributions, which is more pronounced for the frequentist approach.

When comparing the proposed methods with standard frequentist and Bayesian quantile regression, we find that the standard frequentist approach performs best for estimating quantiles. However, in a simulation study, the proposed frequentist method has lower or equal RMSE for most cases considered. For the comparison of Bayesian methods, they perform close in terms of RMSE and bias for the considered error distributions. However, the proposed method generally estimates the quantiles better. In an application, the proposed and standard methods perform similarly in terms of estimating quantiles. However, the proposed methods result in estimates with lower standard deviations for all quantile levels considered.

The conversion from L_p-quantiles to quantiles is sensitive to sample size and the error distribution. However, even for the large sample size of the application, we find that the proposed method performs worse than classical quantile regression in terms of estimating quantiles.

For future research, there are several directions that could be explored to further develop the conversion from L_p-quantiles to quantiles. One possibility is to consider the asymmetric exponential power distribution, which has different shape parameters for the left and right-hand sides of the mean, instead of the SEPD. This may result in a weaker link to L_p-quantile regression, but it could allow for a better fit of the error distribution to the data and more accurate conversions of quantiles.

Another area for improvement is the Bayesian estimate based on the converted quantiles, which is not based on a posterior sample that reflects randomness in the parameters $(\mu, \sigma, p)^T$. Extensions could be considered where the sequential nature of the Algorithm 3 is avoided.

Finally, the assumption of a linear relationship between the response and covariates could be relaxed to allow for non-linear quantile regression based on the SEPD. This would be an interesting avenue for future research.

Acknowledgments

The authors are thankful to the editor, the associate editor and the reviewer for their help and suggestions for the improvement of the article.

Appendix A.

A.1. Proof of Theorem 1

To prove that the posterior is proper, we require

$$\int_{p,\tau}\int_{\mathbb{R}^k}\int_0^\infty p(\mathbf{y}|\boldsymbol{\beta},p,\sigma)\pi(\boldsymbol{\beta})\pi(\sigma)\pi(p)\pi(\tau)d\sigma d\boldsymbol{\beta}dpd\tau<\infty.$$

First, σ can be integrated out analytically as was done in (15),

$$L^1(\boldsymbol{\beta},p,\tau;y)=\int_0^\infty p(\mathbf{y}|\boldsymbol{\beta},p,\sigma)\pi(\sigma)d\sigma$$
$$\propto p^{-1}\Gamma\left(\frac{n}{p}\right)\Gamma\left(1+\frac{1}{p}\right)^{-n}\left[\sum_{i\in N_1}\frac{|\mathbf{x}_i^T\boldsymbol{\beta}-y_i|^p}{\tau^p}+\sum_{i\in N_2}\frac{|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p}{(1-\tau)^p}\right]^{-\frac{n}{p}}.$$

Next,

$$L^1(p,\tau;y)=\int_{\mathbb{R}^k}L^1(\boldsymbol{\beta},p;y)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}\propto p^{-1}\Gamma\left(\frac{n}{p}\right)\Gamma\left(1+\frac{1}{p}\right)^{-n}$$

$$\times\int_{\mathbb{R}^k}\left[\sum_{i\in N_1}\frac{|\mathbf{x}_i^T\boldsymbol{\beta}-y_i|^p}{\tau^p}+\sum_{i\in N_2}\frac{|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p}{(1-\tau)^p}\right]^{-\frac{n}{p}}d\boldsymbol{\beta}.$$

Although the integral can not be calculated analytically, it can be bounded from above and below. Firstly,

$$\sum_{i\in N_1}\frac{|\mathbf{x}_i^T\boldsymbol{\beta}-y_i|^p}{\tau^p}+\sum_{i\in N_2}\frac{|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p}{(1-\tau)^p}\le\frac{\sum_{i=1}^n|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p}{\min\{\tau^p,(1-\tau)^p\}}\le\frac{\sum_{i=1}^n(|y_i|+|\mathbf{x}_i^T\boldsymbol{\beta}|)^p}{\min\{\tau^p,(1-\tau)^p\}}\le$$
$$\le\frac{1}{\min\{\tau^p,(1-\tau)^p\}}\sum_{i=1}^n\left(\max_i|y_i|+\sum_{j=1}^k\max_i|x_{ij}||\beta_j|\right)^p\equiv h(\boldsymbol{\beta},p,\tau).$$

Next, note that $p\ge1, |\cdot|^p$ is convex. Thus, by Jensen's inequality

$$\sum_{i\in N_1}\frac{|\mathbf{x}_i^T\boldsymbol{\beta}-y_i|^p}{\tau^p}+\sum_{i\in N_2}\frac{(y_i-\mathbf{x}_i^T\boldsymbol{\beta})^p}{(1-\tau)^p}\ge\frac{1}{\max\{\tau^p,(1-\tau)^p\}}\sum_{i=1}^n|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p$$
$$=\frac{n|\bar{y}-\bar{\mathbf{x}}^T\boldsymbol{\beta}|^p}{\max\{\tau^p,(1-\tau)^p\}}\ge2^pn|\bar{y}-\bar{\mathbf{x}}^T\boldsymbol{\beta}|^p.$$

Letting $C_1=\left\{\boldsymbol{\beta}\in\mathbb{R}^k:\sum_{l=1}^k\max_i|x_{il}||\beta_l|\le|\bar{y}|\right\}, C_2=C_1^c$ where $\cdot^c$ denotes the complement, and $\widetilde{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\sum_{i=1}^n|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p$. Note that $\sum_{i=1}^n|y_i-\mathbf{x}_i^T\widetilde{\boldsymbol{\beta}}|^p>0$ with probability 1. Then, for $\boldsymbol{\beta}\in C_1$,

$$\sum_{i=1}^n|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p\ge\sum_{i=1}^n|y_i-\mathbf{x}_i^T\widetilde{\boldsymbol{\beta}}|^p\ge n|\bar{y}-\bar{\mathbf{x}}^T\widetilde{\boldsymbol{\beta}}|^p\equiv g(\widetilde{\boldsymbol{\beta}},p),$$

where the first inequality follows from the definition of $\widetilde{\boldsymbol{\beta}}$ on C_1, and the second inequality follows from Jensen's inequality. For $\boldsymbol{\beta}\in C_2$, by Jensen's inequality,

$$\sum_{i=1}^n|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p\ge n|\bar{y}-\bar{\mathbf{x}}^T\boldsymbol{\beta}|^p=g(\boldsymbol{\beta},p).$$

Therefore, we have

$$\int_{\mathbb{R}^k}h(\boldsymbol{\beta},p,\tau)^{-\frac{n}{p}}d\boldsymbol{\beta}\le\int_{\mathbb{R}^k}\left[\sum_{i\in N_1}\frac{|\mathbf{x}_i^T\boldsymbol{\beta}-y_i|^p}{\tau^p}+\sum_{i\in N_2}\frac{|y_i-\mathbf{x}_i^T\boldsymbol{\beta}|^p}{(1-\tau)^p}\right]^{-\frac{n}{p}}d\boldsymbol{\beta}\le2^{-n}\int_{\mathbb{R}^k}g(\boldsymbol{\beta},p)^{-\frac{n}{p}}d\boldsymbol{\beta}.$$

Starting with the left side, denote $\boldsymbol{\beta}_{(-j)} = (\beta_{j+1}, \dots, \beta_k)^T \in \mathbb{R}^{k-j}$. Then

$$\int_{\mathbb{R}^k} h(\boldsymbol{\beta}, p, \tau)^{-\frac{n}{p}} d\boldsymbol{\beta} = \frac{2}{\min\left\{\tau^p, (1-\tau)^p\right\}^{-\frac{n}{p}}} \times$$

$$\int_{\mathbb{R}^{k-1}} \int_0^\infty \left[n^{\frac{1}{p}} (\max|y_i| + \sum_{j=2}^k \max|x_{ij}||\beta_j| + \max|x_{i1}|\beta_1) \right]^{-n} d\beta_1 d\boldsymbol{\beta}_{(-1)}$$

$$= -2\min\left\{\tau, (1-\tau)\right\}^n \int_{\mathbb{R}^{k-1}} \frac{n^{-\frac{n}{p}} (\max|y_i| + \sum_{j=2}^k \max|x_{ij}||\beta_j| + \max|x_{i1}|\beta_1)^{-(n-1)}}{-\max|x_{i1}|(n-1)} \Bigg|_\infty^0 d\boldsymbol{\beta}_{(-1)}$$

$$= 2\min\left\{\tau, (1-\tau)\right\}^n \int_{\mathbb{R}^{k-1}} \frac{n^{-\frac{n}{p}} (\max|y_i| + \sum_{j=2}^k \max|x_{ij}||\beta_j|)^{-(n-1)}}{\max|x_{i1}|(n-1)} d\boldsymbol{\beta}_{(-1)}.$$

The integral can be computed recursively for each element of $\boldsymbol{\beta}_{(-j)}$, the result is convergent if $n > k$,

$$\int_{\mathbb{R}^k} h(\boldsymbol{\beta}, p, \tau)^{-\frac{n}{p}} d\boldsymbol{\beta} = n^{-\frac{n}{p}} m_1(y),$$

where

$$m_1(y) = \min\left\{\tau, (1-\tau)\right\}^n \frac{2^k \Gamma(n-k)(\max|y_i|)^{-(n-k)}}{\Gamma(n) \prod_{j=1}^k \max|x_{ij}|} > 0,$$

does not depend of p. The right integral is

$$2^{-n} \int_{\mathbb{R}^k} g(\boldsymbol{\beta}, p)^{-\frac{n}{p}} d\boldsymbol{\beta} = n^{-\frac{n}{p}} m_2(y),$$

where

$$m_2(y) = \frac{1}{2^n} \left(\int_{C_1} |\bar{y} - \bar{\boldsymbol{x}}^T \widetilde{\boldsymbol{\beta}}|^{-n} d\boldsymbol{\beta} + \int_{C_2} |\bar{y} - \bar{\boldsymbol{x}}^T \boldsymbol{\beta}|^{-n} d\boldsymbol{\beta} \right), \tag{23}$$

does not depend on p. Note that $m_2(y) < \infty$ as

$$\int_{C_1} |\bar{y} - \bar{\boldsymbol{x}}^T \widetilde{\boldsymbol{\beta}}|^{-n} d\boldsymbol{\beta} = |\bar{y} - \bar{\boldsymbol{x}}^T \widetilde{\boldsymbol{\beta}}|^{-n} \int_{C_1} 1 d\boldsymbol{\beta} < \infty,$$

for finite $\bar{y}$ and $\bar{x}_j, j = 1, \dots, k$ as $\sum_{i=1}^n |y_i - \boldsymbol{x}_i^T \widetilde{\boldsymbol{\beta}}|^p > 0$ with probability 1. Also,

$$\int_{C_2} |\bar{y} - \bar{\boldsymbol{x}}^T \boldsymbol{\beta}|^{-n} d\boldsymbol{\beta} < \infty,$$

as $\left|\bar{y} - \bar{\boldsymbol{x}}^T \boldsymbol{\beta}\right| > 0$ for $\boldsymbol{\beta} \in C_2$. Therefore, we have shown that

$$m_1(y) \le n^{n/p} \int_{\mathbb{R}^k} \left[\sum_{i \in N_1} \frac{|\boldsymbol{x}_i^T \boldsymbol{\beta} - y_i|^p}{\tau^p} + \sum_{i \in N_2} \frac{|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}|^p}{(1-\tau)^p} \right]^{-\frac{n}{p}} d\boldsymbol{\beta} \le m_2(y).$$

The marginal likelihood is thus finite as $\pi(p)$ is proper and the only term depending on τ is in $m_1(y)$ where $\int_0^1 \min\left\{\tau, (1-\tau)\right\}^n d\tau < \infty$.

ORCID

Lukas Arnroth http://orcid.org/0000-0002-8567-5116

References

Benoit, D. F., and D. Van den Poel. 2017. bayesQR: A Bayesian approach to quantile regression. *Journal of Statistical Software* 76 (7):1–32. doi:10.18637/jss.v076.i07.

Bernardi, M., M. Bottone, and L. Petrella. 2018. Bayesian quantile regression using the skew exponential power distribution. *Computational Statistics & Data Analysis* 126:92–111. doi:10.1016/j.csda.2018.04.008.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. Julia: A fresh approach to numerical computing. *SIAM Review* 59 (1):65–98. doi:10.1137/141000671.

Chen, Z. 1996. Conditional Lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & Probability Letters* 29 (2):107–15. doi:10.1016/0167-7152(95)00163-8.

Daouia, A., I. Gijbels, and G. Stupfler. 2019. Extremiles: a new perspective on asymmetric least squares. *Journal of the American Statistical Association* 114 (527):1366–81. doi:10.1080/01621459.2018.1498348.

Daouia, A., S. Girard, and G. Stupfler. 2019. Extreme M-quantiles as risk measures: From L1 to Lp optimization. *Bernoulli* 25 (1):264–309. doi:10.3150/17-BEJ987.

Girolami, M., and B. Calderhead. 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73 (2):123–214. doi:10.1111/j.1467-9868.2010.00765.x.

Jones, M. C. 1994. Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters* 20 (2):149–53. doi:10.1016/0167-7152(94)90031-0.

Koenker, R. 2000. Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. *Journal of Econometrics* 95 (2):347–74. doi:10.1016/S0304-4076(99)00043-3.

Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica* 46 (1):33–50. doi:10.2307/1913643.

Liu, J. S. 1994. The Collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* 89 (427):958–66. doi:10.1080/01621459.1994.10476829.

Usseglio-Carleve, A. 2018. Estimation of conditional extreme risk measures from heavytailed elliptical random vectors. *Electronic Journal of Statistics* 12 (2):4057–93. doi:10.1214/18-EJS1499.

Waltrup, L. S., F. Sobotka, T. Kneib, and G. Kauermann. 2015. Expectile and quantile regression – David and Goliath? *Statistical Modelling* 15 (5):433–56. doi:10.1177/1471082X14561155.

Xing, J., and X. Qian. 2017. Bayesian expectile regression with asymmetric normal distribution. *Communications in Statistics – Theory and Methods* 46 (9):4545–55. doi:10.1080/03610926.2015.1088030.

Yang, T., C. M. Gallagher, and C. S. McMahan. 2019. A robust regression methodology via M-estimation. *Communications in Statistics: Theory and Methods* 48 (5):1092–107. doi:10.1080/03610926.2018.1423698.

Yu, K., and R. A. Moyeed. 2001. Bayesian quantile regression. *Statistics & Probability Letters* 54 (4):437–47. doi:10.1016/S0167-7152(01)00124-9.

Zhu, D., and V. Zinde-Walsh. 2009. Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* 148 (1):86–99. doi:10.1016/j.jeconom.2008.09.038.