A New Family of Error Distributions for Bayesian Quantile Regression

Yifei Yan, Xiaotian Zheng, and Athanasios Kottas *

April 1, 2024

Abstract

We propose a new family of error distributions for model-based quantile regression, which is constructed through a structured mixture of normal distributions. The construction enables fixing specific percentiles of the distribution while, at the same time, allowing for varying mode, skewness and tail behavior. It thus overcomes a practical limitation of the asymmetric Laplace distribution – the most commonly used error model for parametric quantile regression – for which the skewness of the error density is fully specified when a particular percentile is fixed. We develop a Bayesian formulation for the proposed quantile regression model, including conditional lasso regularized quantile regression based on a hierarchical Laplace prior for the regression coefficients, and a Tobit quantile regression model. Posterior inference is implemented via Markov Chain Monte Carlo methods. The flexibility of the new model relative to the asymmetric Laplace distribution is studied through relevant model properties, and through a simulation experiment to compare the two error distributions in regularized quantile regression. Moreover, model performance in linear quantile regression, regularized quantile regression, and Tobit quantile regression is illustrated with data examples that have been previously considered in the literature.

Keywords: Asymmetric Laplace distribution; Markov chain Monte Carlo; Regularized quantile regression; Skew normal distribution; Tobit quantile regression

*Yifei Yan is Senior Data Science Expert at Ant Group, Shanghai, China, Xiaotian Zheng is Research Fellow in the School of Mathematics and Applied Statistics, University of Wollongong, Australia, and Athanasios Kottas (thanos@soe.ucsc.edu) is Professor in the Department of Statistics, University of California, Santa Cruz.

1 Introduction

Quantile regression offers a practically important alternative to traditional mean regression, and forms an area with a growing literature in terms of both methodology and applications. Parametric quantile regression models are typically built from the asymmetric Laplace (AL) distribution the density of which is

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) \quad = \quad \frac{p(1-p)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_p\left(y-\mu\right)\right\}, \quad y \in \mathbb{R} \tag{1}$$

where $\rho_p(u) = u[p - I(u < 0)]$, with $I(\cdot)$ denoting the indicator function. Here, $\sigma > 0$ is a scale parameter, $p \in (0, 1)$, and $\mu \in \mathbb{R}$ corresponds to the pth percentile, $\int_{-\infty}^{\mu} f_p^{\mathrm{AL}}(y \mid \mu, \sigma)\mathrm{d}y = p$. Hence, a model for pth quantile regression can be developed by expressing μ as a function of available covariates $\boldsymbol{x}$, for instance, $\mu = \boldsymbol{x}^T\boldsymbol{\beta}$ yields a linear quantile regression structure. Note that maximizing the likelihood with respect to $\boldsymbol{\beta}$ under an AL response distribution corresponds to minimizing for $\boldsymbol{\beta}$ the check loss function, $\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})$, used for classical semiparametric estimation in linear quantile regression (Koenker, 2005).

The AL distribution has received attention in the Bayesian literature, originating from work on inference for linear quantile regression (Yu and Moyeed, 2001; Tsionas, 2003; Yang et al., 2016). Particularly relevant to the Bayesian framework are the different mixture representations of the distribution (Kotz et al., 2001), which have been exploited to construct posterior simulation algorithms (Kozumi and Kobayashi, 2011), as well as to explore different modeling scenarios; see, for instance, Lum and Gelfand (2012) and Waldmann et al. (2013).

The objective of this work is to address the practical limitations of the AL distribution when used as an error model for quantile regression. It has been shown that using the AL distribution for the errors in linear quantile regression can result in posterior consistency for the regression coefficients, under various conditions on the covariate space and the true underlying response distribution (Sriram et al., 2013). However, the specific asymptotic arguments do not address key inferential objectives, such as credible interval estimation or prediction for new responses. More importantly, focusing on effective finite sample inference, the AL quantile regression error distribution is arguably limited from a modeling perspective. Indeed, its limitations are reflected in different types of inference even for relatively large amounts of data, as demonstrated with the simulation study of Section 4. The most striking modeling limitation is that the skewness of the error density is fully determined when a specific percentile is chosen, that is, when p is fixed. In particular, the error density is symmetric in the case of median regression, since for $p = 0.5$, the AL reduces to the Laplace distribution. Moreover,

the mode of the error distribution is at zero, for any p, which results in rigid error density tails for extreme percentiles (an illustration is given in Figure 1 discussed in Section 2).

The literature includes Bayesian nonparametric models for the error distribution in the special case of median regression (Walker and Mallick, 1999; Kottas and Gelfand, 2001; Hanson and Johnson, 2002) and in general quantile regression (Kottas and Krnjajić, 2009; Reich et al., 2010). The Bayes nonparametrics literature has also explored inference methods for simultaneous quantile regression (Taddy and Kottas, 2010; Tokdar and Kadane, 2012; Reich and Smith, 2013; Yang and Tokdar, 2017; Das and Ghosal, 2018; Chen and Tokdar, 2021). However, work on parametric alternatives to AL quantile regression errors is limited, and the existing models do not overcome all the limitations discussed above. For instance, although the class of skew distributions studied in Wichitaksorn et al. (2014) includes the AL as a special case, it shares the same restriction with the AL as a quantile regression error model in that it has a single parameter that controls both skewness and percentiles. Zhu and Zinde-Walsh (2009) and Zhu and Galbraith (2011) explored the family of asymmetric exponential power distributions, which does not include the AL distribution. For a fixed probability p, the density function has four free parameters and allows for different decay rates in the left and the right tails. However, similar to the AL, the mode of the distribution is fixed at the quantile μ by construction.

More flexible parametric quantile regression error models are arguably useful both to expand the inferential scope of the asymmetric Laplace in the standard quantile regression setting, as well as to provide building blocks for model development under more complex data structures. The limited scope of results in this direction may be attributed to the challenge of defining sufficiently flexible distributions that are parameterized by percentiles and, at the same time, allow for practicable modeling and inference methods.

Seeking to fill this gap, we propose a new family of distributions that is parameterized in terms of percentiles, and overcomes the restrictive aspects of the AL distribution. The distribution is developed constructively building from a mixture representation of the AL distribution. In particular, we introduce a shape parameter to obtain a distribution that has more flexible skewness and tail behaviour than the AL distribution, while retaining the AL distribution as a special case of the new model (for a specific value of the extra parameter). The latter enables connections with the check loss function which are useful in studying the utility of the new model in the context of regularized quantile regression. Owing to its hierarchical mixture representation, the proposed distribution preserves the important feature of ready to implement posterior inference for Bayesian quantile regression.

In Section 2, we develop the new distribution and discuss its properties relative to the AL distribution. In Section 3, we formulate the Bayesian quantile regression model, including a prior specification for the regression coefficients that encourages shrinkage resulting in regularized quantile regression, and a Tobit quantile regression formulation. In Section 4, we present results from a simulation study to compare the performance of the AL and the proposed distribution in regularized quantile regression, under different scenarios for the underlying response distribution, for both extreme and more central percentiles, and for different sample sizes. The methodology is illustrated with three data examples in Section 5, focusing again on comparison with the AL quantile regression model. Finally, Section 6 concludes with a summary.

2 The generalized asymmetric Laplace distribution

We first construct the general, four-parameter version of the new distribution as an extension of the AL distribution (Section 2.1). Next, in Section 2.2, we develop the version of the distribution that enables its use as an error model in quantile regression.

2.1 The general construction

The construction of the new distribution is motivated by the most commonly used mixture representation of the AL density. In particular,

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma A(p) z, \sigma^2 B(p) z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{d}z \tag{2}$$

where $A(p) = (1-2p)/\{p(1-p)\}$ and $B(p) = 2/\{p(1-p)\}$ (e.g., Kotz et al., 2001). Moreover, $\mathrm{N}(m, W)$ denotes the normal distribution with mean m and variance W, and $\mathrm{Exp}(1)$ denotes the exponential distribution with mean 1. We use such notation throughout to indicate either the distribution or its density, depending on the context.

The mixture formulation in (2) enables exploration of extensions to the AL distribution. Extending the $\mathrm{Exp}(1)$ mixing distribution is not a fruitful direction in terms of evaluation of the intergal, and, more importantly, with respect to fixing percentiles of the resulting distribution. However, both goals are accomplished by replacing the normal kernel in (2) with a skew normal kernel (Azzalini, 1985). In its original parameterization, the skew normal density is given by $f^{\mathrm{SN}}(y \mid \xi, \omega, \lambda) = 2\omega^{-1}\, \phi(\omega^{-1}(y-\xi))\, \Phi(\lambda\omega^{-1}(y-\xi))$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution function, respectively, of the standard normal distribution. Here, $\xi \in \mathbb{R}$ is a location parameter, $\omega > 0$ a scale parameter, and $\lambda \in \mathbb{R}$ the skewness parameter. Key to

our construction is the fact that the skew normal density can be written as a location normal mixture with mixing distribution given by a standard normal truncated on $\mathbb{R}^+$ (Henze, 1986). More specifically, reparameterize (ξ, ω, λ) to (ξ, τ, ψ), where $\tau > 0$ and $\psi \in \mathbb{R}$, such that $\lambda = \psi/\tau$ and $\omega = (\tau^2 + \psi^2)^{1/2}$. Then, $f^{\text{SN}}(y \mid \xi, \tau, \psi) = \int_{\mathbb{R}^+} \text{N}(y \mid \xi + \psi s, \tau^2)\text{N}^+(s \mid 0, 1)\, \text{d}s$, where $\text{N}^+(0, 1)$ denotes the standard normal distribution truncated over $\mathbb{R}^+$.

The proposed model, referred to as generalized asymmetric Laplace (GAL) distribution, is built by adding a shape parameter, $\alpha \in \mathbb{R}$, to the mean of the normal kernel in (2) and mixing with respect to a $\text{N}^+(0, 1)$ variable. More specifically, the full mixture representation for the density function, $f(y \mid p, \alpha, \mu, \sigma)$, of the new distribution is as follows

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \text{N}(y \mid \mu + \sigma\alpha s + \sigma A(p)z, \sigma^2 B(p)z)\, \text{Exp}(z \mid 1)\, \text{N}^+(s \mid 0, 1)\, \text{d}z\, \text{d}s. \tag{3}$$

Note that, integrating over s in (3), the GAL density can be expressed in the form of (2) with the $\text{N}(y \mid \mu + \sigma A(p)z, \sigma^2 B(p)z)$ kernel replaced with a skew normal kernel, which, in its original parameterization, has location parameter $\mu + \sigma A(p)z$, scale parameter $\sigma\{\alpha^2 + B(p)z\}^{1/2}$, and skewness parameter $\alpha\{B(p)z\}^{-1/2}$. Evidently, when $\alpha = 0$, $f(y \mid p, 0, \mu, \sigma)$ reduces to the AL density.

To obtain the GAL density, we integrate out first z and then s in (3). The integrand of $\int_{\mathbb{R}^+} \text{N}(y \mid \mu + \sigma\alpha s + \sigma A(p)z, \sigma^2 B(p)z)\, \text{Exp}(z \mid 1)\, \text{d}z$ can be recognized to be proportional to a generalized inverse-Gaussian density. Therefore, integrating out z, $f(y \mid p, \alpha, \mu, \sigma) = \int_{\mathbb{R}^+} p(1-p)\sigma^{-1} \exp\left\{-\sigma^{-1}\left[p - I(y < \mu + \sigma\alpha s)\right]\left[y - (\mu + \sigma\alpha s)\right]\right\} \text{N}^+(s \mid 0, 1)\, \text{d}s$. This integral involves a normal density kernel, but care is needed with the limits of integration which depend on the sign of $y - \mu$ and of α. Combining the resulting expressions from all possible cases, we obtain that for $\alpha \neq 0$, the GAL density, $f(y \mid p, \alpha, \mu, \sigma)$, is given by

$$\begin{aligned} 2\,\frac{p(1-p)}{\sigma}\, \Bigg(& \left[\Phi\left(\frac{y^*}{\alpha} - p_{\alpha_-}\alpha\right) - \Phi(-p_{\alpha_-}\alpha)\right] \exp\left\{-p_{\alpha_-}y^* + \frac{1}{2}(p_{\alpha_-}\alpha)^2\right\} I\left(\frac{y^*}{\alpha} > 0\right) \\ & + \Phi\left[p_{\alpha_+}\alpha - \frac{y^*}{\alpha} I\left(\frac{y^*}{\alpha} > 0\right)\right] \exp\left\{-p_{\alpha_+}y^* + \frac{1}{2}(p_{\alpha_+}\alpha)^2\right\}\Bigg) \end{aligned} \tag{4}$$

where $y^* = (y - \mu)/\sigma$, $p_{\alpha_+} = p - I(\alpha > 0)$, $p_{\alpha_-} = p - I(\alpha < 0)$, with $p \in (0, 1)$.

2.2 Parameterization for quantile regression

The GAL distribution with density in (4) has four parameters. The introduction of the shape parameter α changes the interpretation for parameter p, in particular, for $\alpha \neq 0$, p no longer corresponds to the probability at the quantile. Here, we develop a reparameterization of the GAL distribution, which allows for a meaningful comparison between the AL and GAL

densities from a fixed quantile point of view, and thus for the use of the GAL distribution as an error model for quantile regression.

When $\alpha > 0$, the distribution function of (4) at μ is given by $\int_{-\infty}^{\mu} f(y \mid p, \alpha, \mu, \sigma)\mathrm{d}y = 2\,p\,\Phi[(p-1)\alpha]\,\exp\left\{(p-1)^2\alpha^2/2\right\}$. Hence, letting $\gamma = (1-p)\alpha$, we can write

$$\int_{-\infty}^{\mu} f(y \mid p, \gamma, \mu, \sigma)\,\mathrm{d}y = p\,g(\gamma) \qquad \text{with} \qquad g(\gamma) = 2\Phi(-|\gamma|)\exp(\gamma^2/2).$$

We use $|\gamma|$ above, since this is the general form of $g(\gamma)$ that applies also in the $\alpha < 0$ case.

Note that, for $\gamma \in \mathbb{R}^-$, $\mathrm{d}g(\gamma)/\mathrm{d}\gamma = 2h(\gamma)\exp(\gamma^2/2)$, where $h(\gamma) = \phi(\gamma) + \gamma\Phi(\gamma)$. The function $h(\gamma)$ is monotonically increasing in $\mathbb{R}^-$, since $\mathrm{d}h(\gamma)/\mathrm{d}\gamma = \Phi(\gamma) > 0$. Moreover, $h(0) = (2\pi)^{-1/2} > 0$, and $\lim_{\gamma\to-\infty} h(\gamma) = 0$. Therefore, $h(\gamma) > 0$ for $\gamma \in \mathbb{R}^-$, and thus $g(\gamma)$ is monotonically increasing in $\mathbb{R}^-$. Since $g(\gamma)$ is an even function, it also obtains that it is monotonically decreasing in $\mathbb{R}^+$.

Consider now setting $\int_{-\infty}^{\mu} f(y \mid p, \gamma, \mu, \sigma)\,\mathrm{d}y = p\,g(\gamma) = p_0$. Then, the fact that $g(\gamma)$ is decreasing in $\mathbb{R}^+$ combined with $g(\gamma) > p_0$, imply that for each $\gamma > 0$ in the domain that respects the condition of $p \in (0,1)$ and $\alpha > 0$, there is a unique solution of p that ensures $\int_{-\infty}^{\mu} f(y \mid p, \gamma, \mu, \sigma)\,\mathrm{d}y = p_0$, and subsequently a unique α based on $\gamma = (1-p)\alpha$. For $\alpha < 0$, setting $\int_{\mu}^{\infty} f(y \mid p, \gamma, \mu, \sigma)\,\mathrm{d}y = 1 - p_0$ and letting $\gamma = p\,\alpha$ leads to the same argument.

The above connection between (p_0, γ) and (p, α) suggests that by reparameterization with desired $p_0 \in (0,1)$ and $\gamma = [I(\alpha > 0) - p]|\alpha|$, we can derive a new family of distributions with the percentile for fixed p_0 given by μ, and with an additional shape parameter γ. For $\gamma \neq 0$, the density, $f_{p_0}(y \mid \gamma, \mu, \sigma)$, of such quantile-fixed GAL distribution is

$$\begin{aligned} 2\,\tfrac{p(1-p)}{\sigma}\Bigg(\left\{\Phi\left(-\tfrac{p_{\gamma_+}y^*}{|\gamma|} + \tfrac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma|\right) - \Phi\left(\tfrac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma|\right)\right\}\exp\left\{-p_{\gamma_-}y^* + \tfrac{\gamma^2}{2}\left(\tfrac{p_{\gamma_-}}{p_{\gamma_+}}\right)^2\right\} I\left(\tfrac{y^*}{\gamma} > 0\right) \\ +\ \Phi\left[-|\gamma| + \tfrac{p_{\gamma_+}y^*}{|\gamma|} I\left(\tfrac{y^*}{\gamma} > 0\right)\right]\exp\left\{-p_{\gamma_+}y^* + \tfrac{\gamma^2}{2}\right\}\Bigg) \end{aligned} \tag{5}$$

where $p \equiv p(\gamma, p_0) = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$, $p_{\gamma_+} = p - I(\gamma > 0)$, $p_{\gamma_-} = p - I(\gamma < 0)$, and $y^* = (y - \mu)/\sigma$.

In practice, we work with the mixture representation of the quantile-fixed GAL density, which emerges from (3) and the reparameterization discussed above. More specifically,

$$f_{p_0}(y \mid \gamma, \mu, \sigma) = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\,C\,|\gamma|s + \sigma A z, \sigma^2 B z)\,\mathrm{Exp}(z \mid 1)\,\mathrm{N}^+(s \mid 0, 1)\,\mathrm{d}z\,\mathrm{d}s \tag{6}$$

where $C = [I(\gamma > 0) - p]^{-1}$, and A and B are the functions of p given in (2). Since p is a function of γ and p_0, A, B and C are all functions of parameter γ.

The parameter γ has bounded support over interval (L, U), where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$. For instance, γ takes values

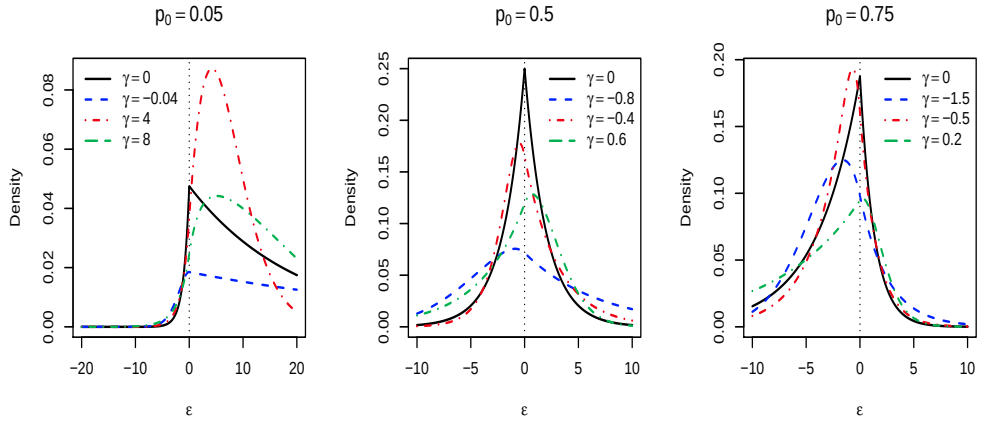


Figure 1: Density function of quantile-fixed generalized asymmetric Laplace distribution with $\mu = 0$, $\sigma = 1$ and different values of γ, for $p_0 = 0.05$, 0.5 and 0.75. In all cases, the solid line corresponds to the asymmetric Laplace density ($\gamma = 0$) with $\mu = 0$ and $\sigma = 1$.

in $(-0.07, 15.90)$, $(-1.09, 1.09)$ and $(-2.90, 0.39)$ when $p_0 = 0.05$, $p_0 = 0.5$ and $p_0 = 0.75$, respectively. When $\gamma = 0$, the density reduces to the quantile-fixed AL density, which is also a limiting case of (5). The density function is continuous for all possible γ values.

The quantile-fixed GAL distribution has three parameters, μ, σ and γ. Note that Y has density $f_{p_0}(\cdot \mid \gamma, \mu, \sigma)$ if and only if $(Y - \mu)/\sigma$ has density $f_{p_0}(\cdot \mid \gamma, 0, 1)$. Hence, similarly to the AL distribution, μ is a location parameter and σ is a scale parameter. The new parameter γ allows for more flexible distributional shapes relative to the quantile-fixed AL distribution. As demonstrated in Figure 1, γ controls skewness and tail behavior, allowing for both left and right skewness when the median is fixed, as well as for both heavier and lighter tails than the asymmetric Laplace, the difference being particularly emphatic for extreme percentiles. Moreover, as the shape parameter γ varies, the mode is no longer held fixed at μ; it is less than μ when $\gamma < 0$ and greater than μ when $\gamma > 0$. The above attributes render the proposed distribution more practical than the AL distribution in the context of quantile regression.

The complex form of the density in (5) is not an obstacle from a practical perspective, since its hierarchical mixture representation in (6) facilitates study of model properties and Markov chain Monte Carlo (MCMC) posterior simulation. The latter is developed in Section 3.1. As an example of the former, we derive the characteristic function, $\varphi_{\mathrm{GAL}}(t)$, of the GAL distribution. Based on (6),

$$\varphi_{\mathrm{GAL}}(t) = \int_{-\infty}^{\infty} e^{ity} f_{p_0}(y \mid \gamma, \mu, \sigma)\, \mathrm{d}y = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \varphi_{\mathrm{N}}(t)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

where $\varphi_{\mathrm{N}}(t)$ is the characteristic function of the $\mathrm{N}(\mu + \sigma\, C\, |\gamma| s + \sigma A z, \sigma^2 B z)$ distribution. The

integral w.r.t. z can be analytically evaluated, and the integral w.r.t. s can be recognized as the characteristic function of the $\mathrm{N}^+(0,1)$ distribution at $t\,\sigma\,C\,|\gamma|$. The resulting expression is given by

$$\varphi_{\mathrm{GAL}}(t) \;=\; \frac{2\,\exp\{it\mu - 0.5\,t^2\sigma^2C^2\gamma^2\}\,\Phi(i\,\sigma\,C\,|\gamma|\,t)}{1+0.5\,t^2\sigma^2B - i\,t\sigma A}. \tag{7}$$

Finally, we note that parameter γ satisfies likelihood identifiability. Consider the location-scale standardized density, $f_{p_0}(\cdot \mid \gamma, 0, 1)$, which is effectively the model for the errors in quantile regression. Then, assume $f_{p_0}(y \mid \gamma_1, 0, 1) = f_{p_0}(y \mid \gamma_2, 0, 1)$, for all $y \in \mathbb{R}$. Given that parameter γ controls the mode of the density, this implies that γ_1 and γ_2 must have the same sign. Working with either of the two cases (that is, $\gamma_1 > 0$ and $\gamma_2 > 0$ or $\gamma_1 < 0$ and $\gamma_2 < 0$) in expression (5), we arrive at $g(\gamma_1) = g(\gamma_2)$. Since function $g(\cdot)$ is monotonic in either $\mathbb{R}^-$ or $\mathbb{R}^+$ (increasing in $\mathbb{R}^-$, decreasing in $\mathbb{R}^+$), $g(\gamma_1) = g(\gamma_2)$ implies $\gamma_1 = \gamma_2$.

3 Bayesian quantile regression with GAL errors

3.1 Inference for linear quantile regression

Consider continuous responses y_i and the associated covariate vectors $\boldsymbol{x}_i$, for $i = 1, \ldots, n$. The linear quantile regression model is set up as $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$, where the ϵ_i arise independently from a quantile-fixed GAL distribution with $\int_{-\infty}^{0} f_{p_0}(\epsilon \mid \gamma, 0, \sigma)\mathrm{d}\epsilon = p_0$. Owing to the mixture representation of the GAL distribution, given in (6), the model for the data can be expressed hierarchically as follows

$$\begin{aligned} y_i \mid \boldsymbol{\beta}, \gamma, \sigma, z_i, s_i &\overset{ind.}{\sim} \mathrm{N}(y_i \mid \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma\,C\,|\gamma| s_i + \sigma A z_i, \sigma^2 B z_i), \quad i = 1, ..., n \\ z_i, s_i &\overset{ind.}{\sim} \mathrm{Exp}(z_i \mid 1)\,\mathrm{N}^+(s_i \mid 0, 1), \quad i = 1, ..., n \end{aligned} \tag{8}$$

where, again, A, B and C are all functions of parameter γ. The Bayesian model is completed with priors for $\boldsymbol{\beta}$, σ and γ. Here, we assume a normal prior $\mathrm{N}(\boldsymbol{m}_0, \Sigma_0)$ for $\boldsymbol{\beta}$ and an inverse-gamma prior $\mathrm{IG}(a_\sigma, b_\sigma)$ for σ, with mean $b_\sigma/(a_\sigma - 1)$ provided $a_\sigma > 1$. For any specified p_0, γ is defined over an interval (L, U) with fixed finite endpoints, and thus a natural prior for γ is given by a rescaled Beta distribution, with the uniform distribution available as a default choice.

The augmented posterior distribution, which includes the z_i and the s_i, can be explored via MCMC based on Gibbs sampling updates for all parameters other than γ. As in Kozumi and

Kobayashi (2011), we set $v_i = \sigma z_i$, $i = 1, \ldots, n$. Denote by $\text{GIG}(\nu, a, b)$ the generalized inverse-Gaussian distribution with density given by $f_{GIG}(x) \propto x^{\nu-1} \exp\{-0.5(a/x + bx)\}$. Then, the posterior simulation method is based on the following updates.

1. Sample $\boldsymbol{\beta}$ from $\text{N}(\boldsymbol{m}^*, \Sigma^*)$, with covariance matrix $\Sigma^* = [\Sigma_0^{-1} + \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T/(B\sigma v_i)]^{-1}$ and mean vector $\boldsymbol{m}^* = \Sigma^*\{\Sigma_0^{-1}\boldsymbol{m}_0 + \sum_{i=1}^n \boldsymbol{x}_i[y_i - (\sigma C|\gamma|s_i + Av_i)]/(B\sigma v_i)\}$.

2. For each $i = 1, ..., n$, sample v_i from a $\text{GIG}(0.5, a_i, b_i)$ distribution, where $a_i = [y_i - (\boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma C|\gamma|s_i)]^2/(B\sigma)$, and $b_i = 2/\sigma + A^2/(B\sigma)$.

3. For each $i = 1, ..., n$, sample s_i from a normal $\text{N}(\mu_{s_i}, \sigma_{s_i}^2)$ distribution truncated on $\mathbb{R}^+$, where $\sigma_{s_i}^2 = [(C\gamma)^2\sigma/(Bv_i) + 1]^{-1}$ and $\mu_{s_i} = \sigma_{s_i}^2 C|\gamma|[y_i - (\boldsymbol{x}_i^T\boldsymbol{\beta} + Av_i)]/(Bv_i)$.

4. Sample σ from a $\text{GIG}(\nu, c, d)$ distribution, where $\nu = -(a_\sigma + 1.5n)$, $c = 2b_\sigma + 2\sum_{i=1}^n v_i + \sum_{i=1}^n [y_i - (\boldsymbol{x}_i^T\boldsymbol{\beta} + Av_i)]^2/(Bv_i)$, and $d = \sum_{i=1}^n (C\gamma s_i)^2/(Bv_i)$.

5. We sample γ using a slice sampling update (Neal, 2003) which requires no accept/reject step and sensitive tuning parameters. In particular, we implement slice sampling based on stepping-out and shrinkage procedures (Scheme (3), Section 4.1 and Method (ii), Section 4.2 of Neal 2003).

Parameter γ can be alternatively updated with a random walk Metropolis step, implemented on the logit scale over (L, U), with a normal proposal distribution. However, in addition to requiring less tuning, we have found the slice sampling approach to provide better mixing chains.

Based on the hierarchical model structure, the posterior predictive error density can be expressed as

$$p(\epsilon \mid \text{data}) = \int \text{N}(\epsilon \mid \sigma C|\gamma|s + \sigma Az, \sigma^2 Bz)\, \text{Exp}(z \mid 1)\, \text{N}^+(s \mid 0, 1)\, \pi(\gamma, \sigma \mid \text{data})\, \text{d}s\, \text{d}z\, \text{d}\gamma\, \text{d}\sigma,$$

and thus estimated through Monte Carlo integration, using the posterior samples for (γ, σ).

3.2 Quantile regression with regularization

Since the GAL distribution is constructed through modifying the mixture representation of the AL distribution, it retains some of the useful properties of the AL distribution. In particular, working with the hierarchical representation of the GAL distribution, we are able to retrieve an extended version of the check loss function which corresponds to asymmetric Laplace errors.

Consider the collapsed posterior distribution, $\pi(\boldsymbol{\beta}, \gamma, \sigma, s_1, ..., s_n \mid \text{data})$, that arises from (8) by marginalizing over the z_i. Then, the corresponding posterior full conditional for $\boldsymbol{\beta}$ can be expressed as

$$\pi(\boldsymbol{\beta} \mid \gamma, \sigma, s_1, ..., s_n, \text{data}) \propto \pi(\boldsymbol{\beta}) \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \sigma H(\gamma)s_i)\right\}$$

where $\pi(\boldsymbol{\beta})$ is the prior density for $\boldsymbol{\beta}$, $H(\gamma) = \gamma g(\gamma)/\{g(\gamma) - |p_0 - I(\gamma < 0)|\}$, and $p = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$, with p_0 the probability associated with the specified quantile modeled through $\boldsymbol{x}_i^T\boldsymbol{\beta}$. Hence, ignoring the prior contribution, finding the mode of the posterior full conditional for $\boldsymbol{\beta}$ is equivalent to minimizing with respect to $\boldsymbol{\beta}$ the adjusted loss function $\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \sigma H(\gamma)s_i)$; note that in the special case with asymmetric Laplace errors, that is, for $\gamma = 0$, this reduces to the check loss function with $p = p_0$.

Based on the above structure, the positive-valued latent variables s_i can be viewed as response-specific weights that are adjusted by real-valued coefficient $H(\gamma)$, which is fully specified through the shape parameter γ. The result is the real-valued, response-specific terms $\sigma H(\gamma)s_i$, which reflect on the estimation of $\boldsymbol{\beta}$ the effect of outlying observations relative to the AL distribution. A promising direction to further explore this structure is in the context of variable selection. For instance, Li et al. (2010) study connections between different versions of regularized quantile regression and different priors for $\boldsymbol{\beta}$, working with asymmetric Laplace errors. The main example is lasso regularized quantile regression, which can be connected to the Bayesian asymmetric Laplace error model through a hierarchical Laplace prior for $\boldsymbol{\beta}$. We consider this prior below extending the AL error distribution to the proposed GAL distribution. This can be viewed as a more general perspective to explore regularization by adjusting the loss function, through the response distribution, in addition to the penalty term, through the prior for the regression coefficients.

Here, we denote by $\boldsymbol{\beta}$ the d-dimensional vector of regression coefficients excluding the intercept β_0. Then, the Laplace conditional prior structure for $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{k=1}^{d} \frac{\lambda}{2\sigma} \exp\left\{-\frac{\lambda}{\sigma}|\beta_k|\right\} = \prod_{k=1}^{d} \int_{\mathbb{R}^+} \frac{1}{\sqrt{2\pi\omega}} \exp\left\{-\frac{\beta_k^2}{2\omega}\right\} \frac{\eta^2}{2} \exp\left\{-\frac{\eta^2}{2}\omega\right\} \, \mathrm{d}\omega.$$

The second expression above utilizes the normal scale mixture representation for the Laplace distribution, which has been exploited for posterior simulation in the context of lasso mean regression (Park and Casella, 2008). Moreover, to facilitate Markov chain Monte Carlo sampling, we reparameterize in terms of $\eta = \lambda/\sigma$ and place a gamma prior on η^2. The lasso regularized version of model (8) is completed with a normal prior for β_0, and with the priors for the other

parameters as given in Section 3.1. The posterior simulation algorithm is similar with the one described in Section 3.1 with the exception of the updates for β_0, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$, and η^2. Using the mixture representation of the Laplace prior, each β_k can be sampled from a normal distribution, whereas η^2 has a gamma posterior full conditional distribution. Details of the posterior simulation steps are available in the Supplementary Material.

3.3 Tobit quantile regression

Tobit regression offers a modeling strategy for problems involving range constraints on the response variable (Amemiya, 1984). The standard Tobit regression model can be viewed in the context of censored regression where the responses are left censored at a threshold c; without loss of generality, we take $c = 0$. The responses can be written as $y_i = \max\{0, y_i^*\}$, where y_i are the observed values and y_i^* are latent if $y_i^* \leq 0$. In the context of quantile regression, Yu and Stander (2007) and Kozumi and Kobayashi (2011) applied the AL-based model to the latent responses y_i^*. Here, we consider the Tobit quantile regression setting with GAL errors.

Consider a data set of $n+k$ observations on covariates and associated responses $\boldsymbol{y} = (\boldsymbol{y}^{\text{o}}, \boldsymbol{0})$, where $\boldsymbol{y}^{\text{o}} = (y_1^{\text{o}}, ..., y_n^{\text{o}})$ consists of positive-valued observed responses with the remaining k responses censored from below at 0. Assuming the GAL distribution for the latent responses, we can express the likelihood as $\prod_{i=1}^{n} f_{p_0}(y_i^{\text{o}} \mid \gamma, \boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma) \prod_{j=1}^{k} \int_{-\infty}^{0} f_{p_0}(w \mid \gamma, \boldsymbol{x}_{n+j}^T\boldsymbol{\beta}, \sigma)\, \mathrm{d}w$. Using data augmentation (Chib, 1992), let $\boldsymbol{w} = (w_1, ..., w_k)$ be the unobserved (latent) responses corresponding to the k data points that are left-censored at 0. Then, using again the hierarchical representation of the GAL distribution, we can write the joint posterior distribution that includes $\boldsymbol{w}$ as

$$p(\boldsymbol{\beta}, \gamma, \sigma, \{s_i\}, \{v_i\}, \boldsymbol{w} \mid \text{data}) \propto \pi(\boldsymbol{\beta}, \gamma, \sigma) \prod_{i=1}^{n} \mathrm{N}(y_i^{\text{o}} \mid \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma C|\gamma| s_i + A v_i, \sigma B v_i)$$
$$\prod_{j=1}^{k} \mathrm{N}^-(w_j \mid \boldsymbol{x}_{n+j}^T\boldsymbol{\beta} + \sigma C|\gamma| s_{n+j} + A v_{n+j}, \sigma B v_{n+j}) \prod_{i=1}^{n+k} \mathrm{Exp}(v_i \mid \sigma^{-1})\, \mathrm{N}^+(s_i \mid 0, 1)$$

where $\pi(\boldsymbol{\beta}, \gamma, \sigma)$ denotes the prior for the model parameters, and $v_i = \sigma z_i$. Here, N^- denotes a truncated normal on $\mathbb{R}^-$, and $\mathrm{Exp}(v \,|\, \sigma^{-1})$ an exponential distribution with mean σ.

Regarding posterior inference, the posterior full conditional for each auxiliary variable w_j is given by a truncated normal distribution. And, given the augmented data $(\boldsymbol{y}^{\text{o}}, \boldsymbol{w})$, the model parameters and the latent variables $\{(v_i, s_i) : i = 1, ..., n+k\}$ can be sampled as before.

Although results are not reported here, we have tested the posterior simulation algorithm on simulated data sets based on GAL errors, with $n = 400$ observations and a censoring rate that ranged from 20% to 40%. Under this scenario, the posterior distributions successfully captured the true values of all parameters in their 95% credible intervals.

4 Simulation study

Here, we present results from a simulation study designed to compare the lasso regularized quantile regression models with AL and GAL errors, under different sample sizes. We follow a standard simulation setting from the literature regarding the linear regression component (Tibshirani, 1996; Zou and Yuan, 2008; Li et al., 2010). For the underlying data-generating error distributions, we consider four scenarios with different types of skewness and tail behavior.

For model comparison, we evaluate the accuracy in variable selection, inference for the regression coefficients (estimation accuracy, and empirical coverage of credible intervals), and posterior predictive performance, using relevant assessment criteria.

4.1 Simulation settings

We consider synthetic data generated from the linear quantile regression setting, with $p_0 = 0.05$, 0.25 and 0.5 to study model performance for both extreme and more central percentiles. The rows of the design matrix were generated independently from an 8-dimensional normal distribution with zero mean vector and covariance matrix with elements $0.5^{|i-j|}$, for $1 \leq i, j \leq 8$. We consider a relatively sparse vector of regression coefficients, in particular, $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)$.

Data were simulated under four different error distributions:

- $\text{N}(\mu, 9)$, with μ chosen such that the p_0th quantile is 0.
- $\text{Laplace}(\mu, 3)$, with μ chosen such that the p_0th quantile is 0.
- $0.1\text{N}(\mu, 1) + 0.9\text{N}(\mu + 1, 5)$, with μ chosen such that the p_0th quantile is 0.
- Log-transformed generalized $\text{Pareto}(\sigma, \xi)$, with $\xi = 3$ and σ chosen such that the p_0th quantile is 0. Based on the parameterization in Embrechts et al. (1997), the error density is given by $f(\epsilon \,|\, \sigma, \xi) = \sigma^{-1}\{1 + \xi\sigma^{-1}\exp(\epsilon)\}^{-(1+\xi^{-1})}\exp(\epsilon)$, for $\epsilon \in \mathbb{R}$.

The normal and Laplace error distributions are symmetric about zero under median regression. The parameters of the two-component normal mixture are selected such that the resulting error distribution is skewed. Finally, the log-transformed generalized Pareto distribution is included to study model performance under an error density which is both skewed and does not have exponential tails. With the exception of median regression under the second scenario, both the GAL and AL models are misspecified in terms of the error distribution.

For each setting of the simulation study, we generated 100 sets of responses and design matrices, each with n observations for training the models. Here, we report results for sample sizes $n = 100$ and $n = 1000$. In the Supplementary Material, we provide additional results based on a single data set of size $n = 10000$. One of the comparison criteria (discussed in the next section) involves out-of-sample prediction. To compute the particular criterion, for each one of the 100 replicated data sets, we generate a test set with N responses and covariates. For all sample sizes n of the training data sets, we use $N = 100$ for the size of the test data sets.

4.2 Criteria for comparison

We consider a number of criteria to assess different aspects of model performance. Details for each criterion are provided below.

Correct inclusions and exclusions for predictors. Since Bayesian lasso regression only shrinks the covariate effects, we consider a threshold on the effect size for the purpose of variable selection. Following Hoti and Sillanpää (2006), we calculate the standardized effects as $\beta_j^* = (s_{x_j}/s_y)\beta_j$, $j = 1, \ldots, d$, where s_{x_j} is the empirical standard deviation of predictor x_j and s_y is the empirical standard deviation of the response. For each posterior sample, if the standardized effect is greater than 0.1 in absolute value, we consider the predictor as included. We count the number of correct inclusions and exclusions (CIE) in the posterior sample and divide it by d to normalize it to a number between 0 and 1. By averaging over all the posterior samples, we obtain the mean standardized CIE for each simulated data set.

Root mean square error for regression coefficient estimates. To assess estimation accuracy of the regression coefficients, we consider root mean square error (RMSE). For each training data set, we calculate the RMSE for β_k using RMSE $= \sqrt{M^{-1}\sum_{m=1}^{M}(\beta_k^{(m)} - \beta_k)^2}$, where M is the posterior sample size and $\beta_k^{(m)}$ is the m-th posterior sample for β_k, for $k = 1, \ldots, d$. In Section 4.3, we aggregate results by averaging over all regression coefficient estimates to obtain the average RMSE for each training data set. In the Supplementary Material, we provide the more detailed results pertaining to each individual β_k, for $k = 1, \ldots, d$.

Coverage probability of credible intervals for the regression coefficients. Using the 100 replicated data sets, we compute the empirical coverage probability of the 95% posterior credible interval for each regression coefficient β_k, for $k = 1, \ldots, d$. The Supplementary

Material reports the results for all combinations of the percentile value ($p_0 = 0.05, 0.25, 0.5$), sample size ($n = 100, 1000$), and true error distribution.

Interval score for predictions. To assess performance in out-of-sample prediction, we apply the interval score (IS) from Gneiting and Raftery (2007). For $j = 1, \ldots, N$, denote by $\tilde{y}_j$ the j-th response in the test set, and let (l_j, u_j) be the corresponding $(1-\alpha)100\%$ prediction interval, obtained using the associated covariate vector, $\tilde{\boldsymbol{x}}_j$, from the test set. Then, the IS for predicting the j-th new response is defined as

$$\text{IS}(\tilde{y}_j, l_j, u_j) \;=\; (u_j - l_j) \;+\; \frac{2}{\alpha}(l_j - \tilde{y}_j)\mathbb{I}(\tilde{y}_j < l_j) \;+\; \frac{2}{\alpha}(\tilde{y}_j - u_j)\mathbb{I}(\tilde{y}_j > u_j).$$

Note that the particular score combines a reward for a narrow prediction interval with a penalty incurred when the realized response is outside the prediction interval. The aggregated prediction interval score for the test data set is $\text{IS}_{\text{test}} = N^{-1}\sum_{j=1}^{N}\text{IS}(\tilde{y}_j, l_j, u_j)$.

4.3 Results

We use the same hierarchical Laplace prior for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_8)^\top$ under the AL and GAL models, with a gamma prior for η^2 with prior mean 1 and variance 10. Such prior specification is relatively non-informative in the sense that it does not favor shrinkage for the regression coefficients, resulting in marginal prior densities for the β_k that place substantial probability mass away from 0. The shape parameter γ of the GAL error distribution is assigned a uniform prior. For both models, the scale parameter σ receives an inverse gamma prior IG$(2, 2)$, and the intercept β_0 a N$(0, 100)$ prior.

Results under both models and for each simulated data set are based on 5000 posterior samples, collected after appropriate burn-in and thinning. As an example of computing times, for a data set with $n = 1000$, it took about 5/3.2 minutes for the MCMC sampler of the GAL/AL lasso regularized quantile regression model to complete 55000 iterations, implemented in the R environment on a computer with a 2-GHz Intel Core i5 processor and 32-GB RAM.

Within each simulation scenario, we summarize results from the 100 data sets using the median and standard deviation (SD) of the values for the performance assessment criteria discussed in Section 4.2. Results are reported in Table 1 through Table 3, where we use boldface to indicate the model supported by the particular criterion under each setting.

Results from the first comparison criterion (CIE) are reported in Table 1. When $n = 100$, the GAL-based model correctly includes/excludes regression coefficient values more often than

		Error distribution			
p_0	Model	Normal	Laplace	Normal mixture	Log-transformed generalized Pareto
			$n = 100$		
0.05	GAL	**0.842** (0.057)	**0.774** (0.062)	**0.911** (0.047)	**0.884** (0.046)
	AL	0.721 (0.099)	0.610 (0.100)	0.820 (0.079)	0.817 (0.087)
0.25	GAL	**0.857** (0.062)	**0.840** (0.058)	**0.918** (0.047)	0.879 (0.055)
	AL	0.812 (0.070)	0.784 (0.073)	0.896 (0.059)	**0.884** (0.058)
0.50	GAL	0.844 (0.067)	0.841 (0.058)	**0.896** (0.055)	**0.889** (0.055)
	AL	**0.848** (0.070)	**0.847** (0.058)	0.882 (0.057)	0.863 (0.063)
			$n = 1000$		
0.05	GAL	**1.000** (0.003)	**0.997** (0.011)	1.000 (0.000)	1.000 (0.001)
	AL	0.995 (0.033)	0.863 (0.096)	1.000 (0.009)	1.000 (0.011)
0.25	GAL	1.000 (0.002)	**1.000** (0.002)	1.000 (0.000)	1.000 (0.000)
	AL	1.000 (0.012)	0.998 (0.013)	1.000 (0.000)	1.000 (0.000)
0.50	GAL	1.000 (0.002)	1.000 (0.001)	1.000 (0.000)	1.000 (0.000)
	AL	1.000 (0.002)	1.000 (0.001)	1.000 (0.000)	1.000 (0.001)

Table 1: Simulation study. Median (SD) across the 100 simulated data sets for the standardized number of correctly included/excluded predictors, based on the training data for sample size $n = 100$ and $n = 1000$.

the AL model for almost all combinations of lower percentiles ($p_0 = 0.05$ and $p_0 = 0.25$) and the underlying error distribution. The differences in the mean standardized CIE values are largest for $p_0 = 0.05$, they decrease when $p_0 = 0.25$, and they become fairly small in the case of median regression where the results for the two models are balanced. The CIE performance is improved for both models when the sample size increases to $n = 1000$. Except for three cases (where the GAL model is favored), both models achieve optimal performance. The results in Table 1 suggest that we should not expect to distinguish the GAL and AL models in terms of CIE for sample sizes greater than 1000. Indeed, using one data set with sample size $n = 10000$, both models result in mean standardized CIE value of 1 for all 12 combinations of p_0 and true error distribution (refer to the Supplementary Material).

Table 2 summarizes the average RMSE results. As with the CIE criterion, the GAL and AL models are comparable in median regression (under both sample sizes), although it is worth noting that the GAL model fares better when the underlying error distribution is skewed.

		Error distribution			
p_0	Model	Normal	Laplace	Normal mixture	Log-transformed generalized Pareto
			$n = 100$		
0.05	GAL	**0.48** (0.07)	**0.68** (0.11)	**0.35** (0.05)	**0.43** (0.06)
	AL	0.66 (0.14)	1.06 (0.29)	0.47 (0.10)	0.52 (0.12)
0.25	GAL	**0.46** (0.07)	**0.56** (0.08)	**0.34** (0.05)	0.44 (0.07)
	AL	0.51 (0.09)	0.63 (0.11)	0.36 (0.06)	**0.43** (0.07)
0.50	GAL	**0.48** (0.09)	0.55 (0.08)	**0.36** (0.06)	**0.44** (0.08)
	AL	0.49 (0.09)	**0.54** (0.08)	0.38 (0.06)	0.48 (0.08)
			$n = 1000$		
0.05	GAL	**0.16** (0.02)	**0.23** (0.03)	**0.11** (0.02)	**0.14** (0.02)
	AL	0.22 (0.04)	0.44 (0.11)	0.17 (0.03)	0.18 (0.04)
0.25	GAL	**0.16** (0.02)	**0.18** (0.02)	**0.11** (0.02)	0.14 (0.02)
	AL	0.17 (0.03)	0.22 (0.04)	0.12 (0.02)	**0.13** (0.02)
0.50	GAL	0.16 (0.02)	0.16 (0.02)	0.12 (0.02)	**0.14** (0.02)
	AL	0.16 (0.02)	0.16 (0.02)	0.12 (0.02)	0.16 (0.02)

Table 2: Simulation study. Median (SD) across the 100 simulated data sets for the average RMSE of the regression coefficient estimates, based on the training data for sample size $n = 100$ and $n = 1000$.

For essentially all other combinations of p_0, sample size, and true error distribution, the GAL model outperforms the AL model in terms of RMSE, with the differences more notable when $p_0 = 0.05$. As previously mentioned, the Supplementary includes the full RMSE results for each regression coefficient, which reveal the same pattern on the comparison between the two models. As also shown in the Supplementary Material, if the sample size is increased to $n = 10000$ (using a single training data set), the GAL model again produces smaller average RMSE when $p_0 = 0.05$, whereas the two models are practically indistinguishable for $p_0 = 0.25$ and $p_0 = 0.5$.

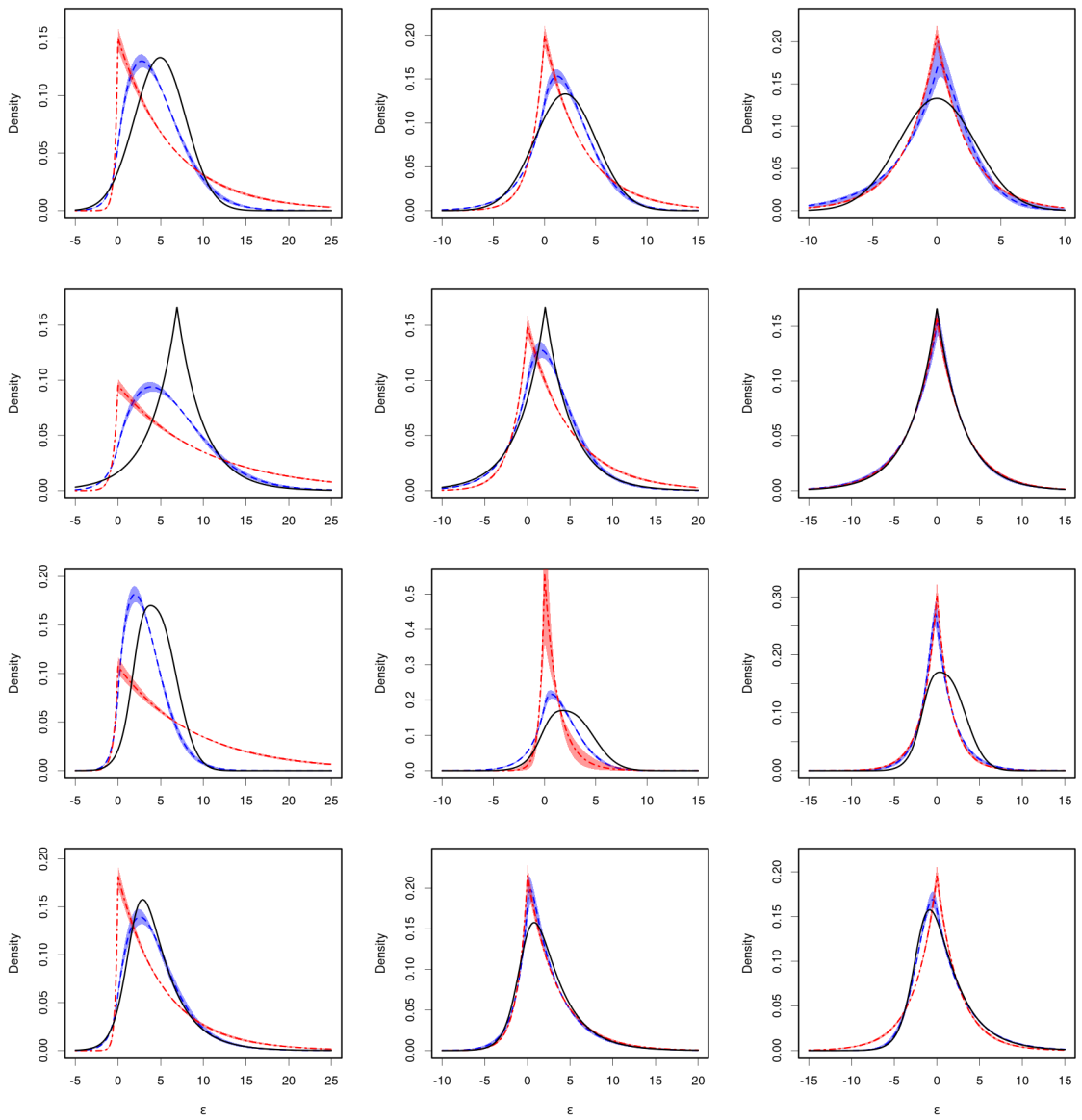
For additional assessment of inference for the regression coefficients, we obtained the empirical coverage probability of the 95% posterior credible interval for each β_k, for $k = 1, \ldots, 8$ (refer to the Supplementary Material). When $p_0 = 0.05$, the GAL model outperforms the AL model across all β_k, for all underlying error distributions, and for both sample sizes. For each of the 4 remaining quantile and sample size combinations ($p_0 = 0.25, 0.5$, and $n = 100, 1000$),

		Error distribution			
p_0	Model	Normal	Laplace	Normal mixture	Log-transformed generalized Pareto
			$n = 100$		
0.05	GAL	**15.14** (1.58)	**25.54** (4.22)	**11.07** (1.13)	**18.05** (3.15)
	AL	23.87 (2.16)	39.04 (5.12)	18.19 (1.56)	21.97 (2.40)
0.25	GAL	**15.39** (1.99)	**23.87** (4.56)	**11.22** (1.16)	18.60 (2.78)
	AL	18.68 (1.98)	28.47 (4.46)	13.29 (1.31)	**18.08** (2.12)
0.50	GAL	16.81 (2.29)	24.64 (4.62)	12.00 (1.49)	19.77 (2.90)
	AL	**15.57** (0.99)	**24.40** (4.15)	**11.32** (0.90)	**19.02 (3.37)**
			$n = 1000$		
0.05	GAL	**14.18** (1.41)	**25.20** (4.80)	**10.20** (1.13)	**17.29** (2.46)
	AL	24.24 (1.41)	41.63 (3.91)	17.45 (1.17)	21.58 (1.52)
0.25	GAL	**16.72** (2.30)	**23.79** (4.37)	**12.03** (2.07)	**16.97** (2.84)
	AL	18.04 (1.62)	28.43 (4.65)	12.77 (1.13)	17.20 (2.39)
0.50	GAL	15.16 (3.25)	23.96 (3.86)	11.12 (1.11)	23.37 (3.90)
	AL	**14.98** (0.77)	**23.61** (3.95)	**10.79** (0.61)	**19.82** (3.16)

Table 3: Simulation study. Median (SD) across the 100 simulated data sets for the prediction interval scores, based on test data sets of size $N = 100$. Results are shown for sample sizes of $n = 100$ and $n = 1000$ for the training data sets.

there are 32 coverage probabilities to compare (8 regression coefficients, 4 error distributions). Excluding the ties between the two models, the empirical coverage of the GAL model is closer to the nominal coverage than the AL model for 15/29 ($p_0 = 0.25$, $n = 100$), 28/31 ($p_0 = 0.25$, $n = 1000$), 19/25 ($p_0 = 0.5$, $n = 100$), and 11/23 ($p_0 = 0.5$, and $n = 1000$) of the credible intervals. However, the differences in the coverage probabilities are smaller/much smaller for $p_0 = 0.25/0.5$ compared to the results under $p_0 = 0.05$. With respect to this criterion, the empirical evidence in favor of GAL errors is most compelling for extreme percentiles, and it is noteworthy that such evidence is equally strong for the sample sizes of $n = 100$ and $n = 1000$.

Turning to out-of-sample predictive performance, Table 3 reports results for the interval score criterion. We observe a similar pattern with the earlier results, that is, the GAL model is favored at lower percentiles. The differences in the interval scores are smallest in the median regression case (with the AL model preferred). The relative differences are largest for $p_0 = 0.05$,



(a) $p_0 = 0.05$ (b) $p_0 = 0.25$ (c) $p_0 = 0.5$

Figure 2: Simulation study. Posterior mean and 95% interval estimates of the error density under the lasso regularized quantile regression model with AL (red) and GAL (blue) errors, for $p_0 = 0.05$ (left column), $p_0 = 0.25$ (middle column), and $p_0 = 0.5$ (right column). From top to bottom the true error density (solid line) is: normal (first row), Laplace (second row), normal mixture (third row), and log-transformed generalized Pareto (last row). Results are based on a single data set of size $n = 1000$ under each combination of p_0 value and true error distribution.

and they remain essentially the same as the training data sets size increases from $n = 100$ to $n = 1000$. The pattern in the comparative results for the GAL and AL models extends to $n = 10000$, based on analysis with a single data set (refer to the Supplementary Material).

Finally, for a graphical comparison of error density estimates, we consider results from a single simulated data set of size $n = 1000$, under each of the four scenarios for the true error density and for each of the three p_0 values. Again, both the GAL and AL model are misspecified in essentially all scenarios (the exception being median regression with Laplace errors). As shown in Figure 2, in all cases, the GAL model provides better estimates of the underlying error density. The GAL and AL error density estimates are generally difficult to distinguish when $p_0 = 0.5$, except for the log-transformed generalized Pareto scenario where the GAL model captures the tails of the true error density noticeably better than the AL model. Observing the estimates for $p_0 = 0.25$ and $p_0 = 0.05$, the superior performance of the GAL model becomes more emphatic as the percentile becomes more extreme. In the Supplementary Material, we provide additional graphical comparisons to study the effect of the sample size, based again on a single simulated data set with both smaller ($n = 100$) and larger ($n = 10000$) sample size. Regarding the comparison of the GAL and AL model error density estimates, the empirical evidence is very similar to Figure 2. Particularly important to note is that the AL model estimates do not improve as the sample size increases from $n = 1000$ to $n = 10000$.

In summary, the lasso regularized Bayesian quantile regression model performs better with GAL errors than AL errors, with the improved performance becoming more conspicuous as the percentile becomes more extreme. The simulation examples offer evidence that the GAL model is more robust than the AL model to non-standard error distributions, particularly for extreme quantiles. Overall, the two models yield comparable results in the case of median regression. Performance of both models generally improves with increasing sample size, especially under the CIE and RMSE criteria. However, we note that, for extreme percentiles, the GAL model outperforms the AL model to the same extent as the sample size increases from $n = 100$ to $n = 1000$ with respect to both coverage of credible intervals for the regression coefficients and prediction of new responses. Moreover, increasing the sample size does not overcome the rigid AL distribution tails (for lower/higher percentiles) as reflected in error density estimates.

5 Data examples

In this section, we consider three data examples to illustrate the Bayesian quantile regression models developed in Sections 3.1, 3.2, and 3.3. The main focus is on comparison of inference results between models based on the GAL distribution and those assuming an AL distribution for the errors.

We have implemented both models with priors for their parameters that result in essentially

the same prior predictive error densities. The two models were applied with the same prior distributions for $\boldsymbol{\beta}$ and σ. More specifically, for the data sets of Sections 5.1 and 5.3, we used an independent $\mathrm{N}(0, 100)$ for each regression coefficient, and an $\mathrm{IG}(2, 2)$ prior for the scale parameter σ. For the data example of Section 5.2, we used a $\mathrm{N}(0, 100)$ prior for the intercept, and the same gamma prior for η^2 and conditional Laplace prior for the remaining regression coefficients with the simulation study (see Section 4.3). Finally, a uniform prior was placed on the shape parameter γ of the GAL error distribution. For all data examples, the posterior densities for model parameters were fairly concentrated relative to the corresponding prior densities.

5.1 Immunoglobulin-G data

We illustrate the proposed model, referred to as model M_1, with a data set commonly used in additive quantile regression; see, for instance, Yu and Moyeed (2001). The analysis focuses on comparison with the simpler model based on asymmetric Laplace errors, referred to as model M_0. The data set contains the immunoglobulin-G concentration (in grams per litre) for $n = 298$ children aged between 6 months and 6 years. As in earlier applications of quantile regression for these data, we use a quadratic regression function $\beta_0 + \beta_1 x + \beta_2 x^2$ to model immunoglobulin-G concentration quantiles against covariate age (x).

We first compare the AL and GAL error models across a range of quantiles, in particular, for $p_0 = 0.05, 0.25, 0.5, 0.75, 0.85, 0.95$. The Bayesian information criterion (BIC) favors the GAL-based model at all six quantiles; see Table 4. The improvement in performance over the AL model is particularly conspicuous at the two extreme percentiles. In fact, the relative difference in BIC values is larger for $p_0 = 0.75$ and 0.95 compared to 0.25 and 0.05, respectively, suggesting that the GAL model is particularly beneficial (relative to the AL model) in modeling higher immunoglobulin-G concentration quantiles. Table 4 also reports posterior mean and interval estimates for the shape parameter γ of the GAL error distribution. For all six quantile regressions, the 95% posterior credible interval for γ does not include the value of 0, which corresponds to AL errors. Median regression is the only case where 0 is within the effective range of the posterior distribution for γ, i.e., it can be seen at the right tail of the posterior density for γ. Both the BIC values and the estimates for γ provide strong support for the GAL model relative to AL errors.

Turning to inference results, Figure 3 plots regression function and error density estimates for median regression and for 0.85-th quantile regression, under the AL and GAL models. The

Quantile	Model	Posterior mean (95% CrI) for γ	log-likelihood	BIC
$p_0 = 0.05$	M_0		-667	1357
	M_1	4.70 (3.52, 6.40)	-615	1258
$p_0 = 0.25$	M_0		-632	1287
	M_1	1.25 (0.75, 2.01)	-625	1278
$p_0 = 0.5$	M_0		-633	1289
	M_1	-0.62 (-0.82, -0.40)	-624	1276
$p_0 = 0.75$	M_0		-654	1331
	M_1	-1.54 (-1.91, -1.19)	-620	1268
$p_0 = 0.85$	M_0		-686	1395
	M_1	-2.28 (-2.79, -1.84)	-625	1278
$p_0 = 0.95$	M_0		-761	1545
	M_1	-4.55 (-5.40, -3.83)	-646	1320

Table 4: Immunoglobulin-G data. Posterior mean and 95% credible interval (CrI) for the shape parameter γ of the GAL error distribution, and log-likelihood and Bayesian information criterion (BIC) under the asymmetric Laplace and generalized asymmetric Laplace models, denoted by M_0 and M_1, respectively.

GAL model estimates a relatively small amount of skewness in the median regression error density, which can not be uncovered under the symmetric AL error density. Consistent with the BIC results and the estimates for γ (Table 4), the difference in the error density estimates is more emphatic for $p_0 = 0.85$. Under the AL model, both the shape and the skewness of the error distribution are predetermined by p_0 and the mode is forced to be at 0, resulting in a rigid heavy left tail. The GAL model, on the contrary, yields an error density that has a much thinner left tail, concentrating more of its probability mass around the mode, which is not at 0. The effect of the different error density estimates is reflected in the inference for the quantile regression function, especially at larger values of age.

5.2 Boston housing data

We apply the lasso regularized quantile regression model to the realty price data from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970 (Harrison and Rubinfeld, 1978). The data set contains 506 observations. We take the log-transformed corrected median value of owner-occupied housing in USD 1000 (LCMEDV) as the response, and consider the following predictors: point longitudes in decimal degrees (LON), point latitudes in decimal de-

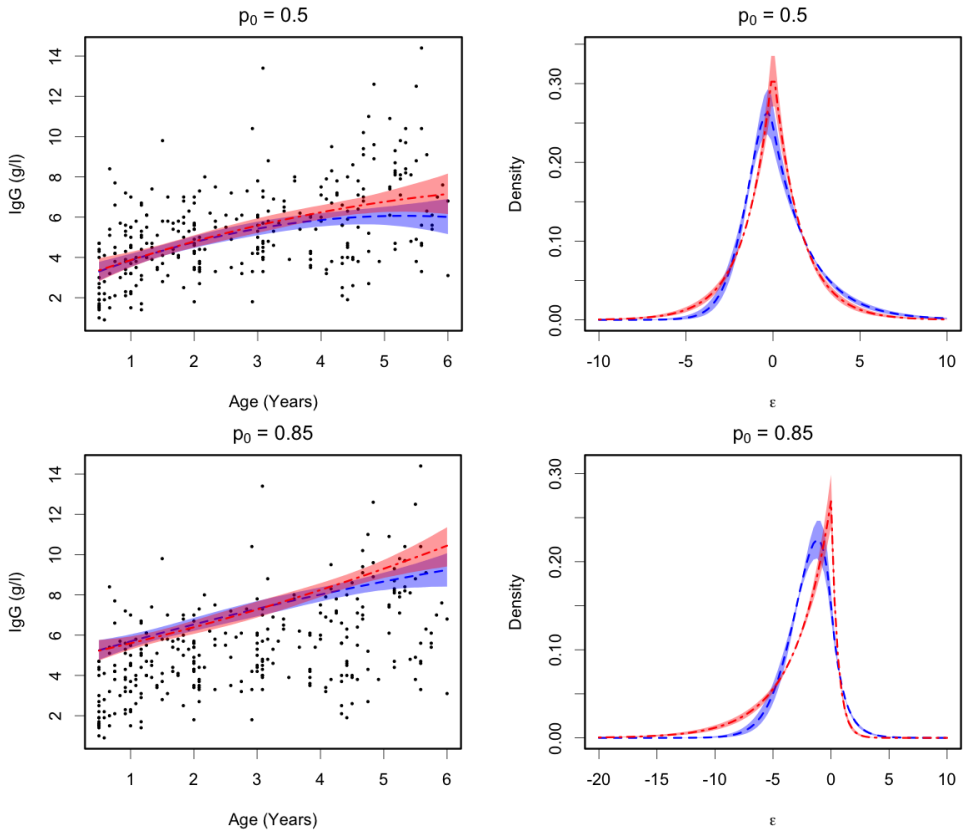


Figure 3: Immunoglobulin-G data. Posterior mean and 95% interval estimates for: the quantile regression function $\beta_0 + \beta_1 x + \beta_2 x^2$ against age (x) (left column panels); and, for the error density (right column panels). Results are reported for two quantiles, corresponding to $p_0 = 0.5$ (top row panels) and $p_0 = 0.85$ (bottom row panels), under the AL model (red) and the GAL model (blue).

grees (LAT), per capita crime (CRIM), proportions of residential land zoned for lots over 25000 square feet per town (ZN), proportions of non-retail business acres per town (INDUS), a factor indicating whether tract borders Charles River (CHAS), nitric oxides concentration (parts per 10 million) per town (NOX), average numbers of rooms per dwelling (RM), proportions of owner-occupied units built prior to 1940 (AGE), weighted distances to five Boston employment centers (DIS), index of accessibility to radial highways per town (RAD), full-value property-tax rate per USD 10,000 per town (TAX), pupil-teacher ratios per town (PTRATIO), transformed African American population proportion (B), and percentage values of lower status population (LSTAT).

We consider the quantiles that correspond to $p_0 = 0.1$ and $p_0 = 0.9$ and compare the maximum a posteriori estimates (MAP) of regression coefficients, along with 95% credible intervals, for standardized covariates under the lasso regularized quantile regression models

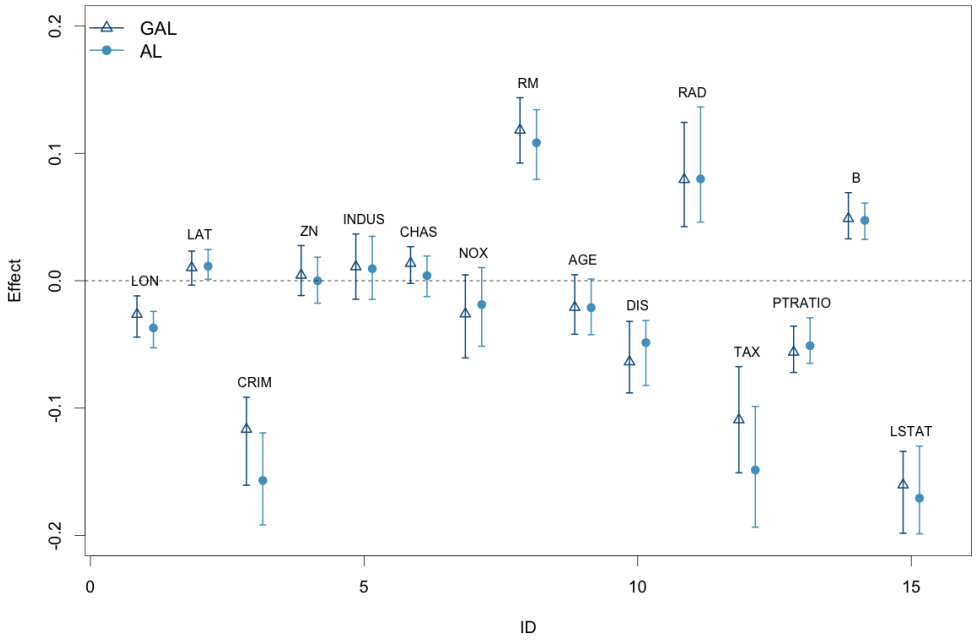


Figure 4: Boston housing data. Posterior point and 95% interval estimates for the regression coefficients of the 10th quantile lasso regularized model under AL and GAL errors.

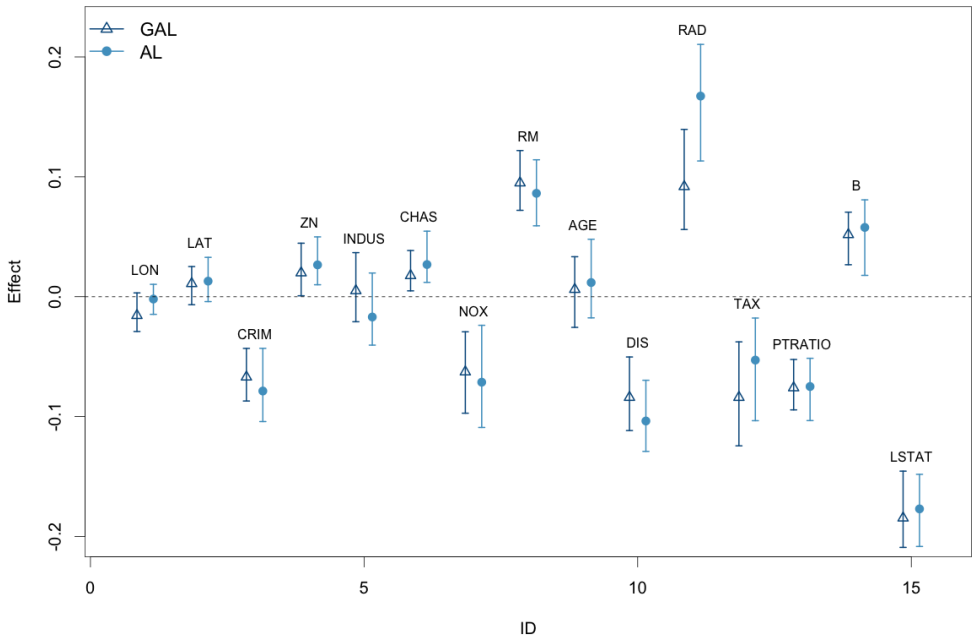


Figure 5: Boston housing data. Posterior point and 95% interval estimates for the regression coefficients of the 90th quantile lasso regularized model under AL and GAL errors.

Quantile	Model	Posterior mean (95% CrI) for γ	log-likelihood	BIC
$p_0 = 0.1$	M_0		124	-142
	M_1	1.85 (1.54, 2.20)	188	-264
$p_0 = 0.9$	M_0		-7	120
	M_1	-2.42 (-2.75, -2.11)	141	-170

Table 5: Boston housing data. Posterior mean and 95% credible interval (CrI) for the shape parameter γ of the GAL error distribution, and log-likelihood and Bayesian information criterion (BIC) under the asymmetric Laplace and generalized asymmetric Laplace models, denoted by M_0 and M_1, respectively.

with AL and GAL errors (Figure 4 and 5). For both quantiles, the widths of the 95% credible intervals for the regression coefficients are overall comparable between the two models, but some of the posterior point estimates are different. For instance, under the 10th quantile regression, the GAL model shrinks the effects of per capita crime (CRIM) and property-tax rate (TAX) to a greater extent compared to the AL model. A similar pattern can be observed for the index of accessibility to radial highways (RAD) for the 90th quantile. Moreover, the two models reach different conclusions on the effect of latitude (LAT) for the 10th quantile regression. Although the posterior point estimates suggest a higher housing price as latitude increases adjusting for all other covariates, the 95% credible interval under the GAL model includes 0, whereas the one under the AL model does not.

Focusing on inference under the GAL model, we note that, although it selects some common variables for the two quantiles, there is also some discrepancy. For instance, each of higher proportions of residential land zoned for lots over 25000 square feet per town (ZN) and having tracts bordering Charles river (CHAS) increase the price at the 90% percentile, while higher nitrogen oxide value (NOX) has a negative influence on the 90% percentile price. However, none of these covariates has a significant effect on the realty value at the 10% percentile.

Finally, note that for both the 10th and 90th quantile regression, 0 is fairly far away from the endpoints of the 95% credible interval for the GAL model shape parameter γ (Table 5), which suggests that GAL errors are more suitable than AL errors also for this data example. This is further supported by the BIC results reported in Table 5. Note in particular the considerable difference in the BIC values for the 90th quantile regression models.

Quantile	Model	Posterior mean (95% CrI) for γ	log-likelihood	BIC
$p_0 = 0.05$	M_0		-1975	4004
	M_1	5.22 (4.43, 6.24)	-1874	3809
$p_0 = 0.5$	M_0		-1867	3789
	M_1	0.58 (0.39, 0.81)	-1845	3750
$p_0 = 0.95$	M_0		-1967	3989
	M_1	-4.16 (-5.5, -3.06)	-1854	3769

Table 6: Labor supply data. Posterior mean and 95% credible interval (CrI) for the shape parameter γ of the GAL error distribution, and log-likelihood and Bayesian information criterion (BIC) values under the asymmetric Laplace and generalized asymmetric Laplace models, denoted by M_0 and M_1, respectively.

5.3 Labor supply data

We illustrate the Tobit quantile regression model with the female labor supply data from Mroz (1987), which was taken from the University of Michigan Panel Study of Income Dynamics for year 1975. The data set includes records on the work hours and other relevant information of 753 married white women aged between 30 and 60 years old. Of the 753 women, 428 worked at some time during 1975, with the corresponding fully observed responses given by the wife's work hours (in 100 hours). For the remaining 325 women, the observed zero work hours correspond to negative values for the latent "labor supply" response. We use the quantile regression function considered in Kozumi and Kobayashi (2011), where an AL-based Tobit quantile regression model was applied to the same data set. The linear predictor includes an intercept, income which is not due to the wife (`nwifeinc`), education of the wife in years (`educ`), actual labor market experience in years (`exper`) and its quadratic term (`expersq`), age of the wife (`age`), number of children less than 6 years old in household (`kidslt6`), and number of children between ages 6 and 18 in household (`kidsge6`). We compare the results from the Bayesian Tobit quantile regression model assuming AL errors (model M_0) and GAL errors (model M_1).

Table 6 summarizes the posterior distribution of γ under the GAL model, and presents results from criterion-based comparison of the two models for $p_0 = 0.05$, 0.5 and 0.95. Since there is censoring in the data, we use the revised BIC from Volinsky and Raftery (2000). In all three cases, the 95% credible interval for γ excludes 0, and the GAL-based model is associated with lower BIC values. The results support the GAL-based model more emphatically for the extreme percentiles than for median regression.

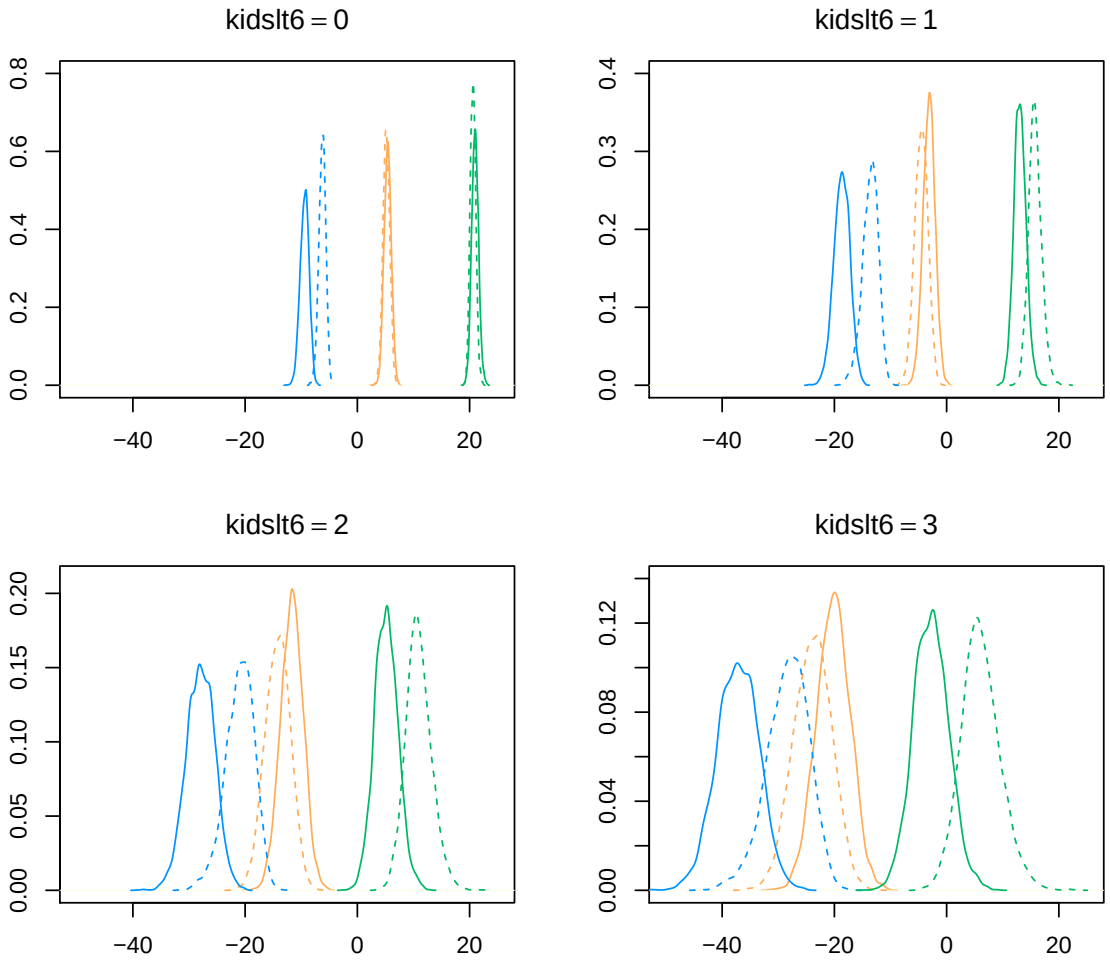


Figure 6: Labor supply data. Posterior densities for the 5th (blue), 50th (orange) and 95th quantile (green) of labor supply (in 100 hours) for women with 0, 1, 2 or 3 children less than 6 years old. The solid (dashed) lines correspond to the posterior densities under the GAL (AL) model.

Figure 6 shows the posterior distributions of labor supply quantiles corresponding to $p_0 = 0.05$, 0.5 and 0.95 for women with 0, 1, 2 and 3 children less than 6 years old. For all other predictors, we use the observed medians as input values to represent an *average* wife. As the number of young children increases, the AL model estimates the 5th quantile and the median of labor supply of an average wife to be closer to each other. Under the GAL model, the distance between the densities of the 5th quantile and median labor supply also decreases with increasing number of young children, albeit at a lower rate. When estimating the 95th quantile, the proposed model is more conservative than the AL model about the labor contribution of an average wife with an increasing number of children less than 6 years old. When there are 3 children less than 6 years old in the household, the center of the posterior distribution for the

95th quantile is below zero under the GAL model, meaning that even at the top 5th percentile of labor supply, an average wife may still produce negative labor supply as she takes care of many young family members. More specifically, the posterior probability of the 95th labor supply quantile being positive is 0.19 under the GAL model, as opposed to 0.97 under the AL model. These results demonstrate that the choice of error distribution in quantile regression can have a substantial effect on practically important conclusions for a particular application.

6 Summary

We have studied Bayesian quantile regression with a new error distribution that extends the asymmetric Laplace (AL) distribution, commonly used as a parametric error model for quantile regression. The generalized asymmetric Laplace (GAL) distribution has more flexible skewness and tail behavior than the asymmetric Laplace in the context of quantile regression. Owing to the hierarchical structure of the new distribution, posterior inference and prediction can be readily implemented via Markov chain Monte Carlo methods. Synthetic and real data examples have been used to demonstrate improved performance of the proposed error distribution, compared with asymmetric Laplace errors, in conditional lasso regularized quantile regression and in Tobit quantile regression. Improved inference and prediction are particularly noteworthy in regression modeling through extreme quantiles.

Our motivation for this work was to develop a distribution that retains the useful features of the AL distribution (parameterization in terms of percentiles, relatively straightforward posterior simulation) while extending the inferential scope of quantile regression modeling with AL errors. A further objective was to provide a building block for more general model structures that incorporate a quantile regression component. Indeed, since the first version of this paper was made available (Yan and Kottas, 2017), the utility of the GAL distribution has been explored by other authors, including: ordinal quantile regression based on latent continuous responses (Rahman and Karnawat, 2019); semiparametric additive quantile regression (Kobayashi et al., 2021); inference for dynamic quantile linear models (Barata et al., 2022); and, quantile regression methods for different types of random effects models (Nascimento and Gonçalves, 2022; Yu and Yu, 2023; Jeliazkov et al., 2023). The methodological development that prompted the GAL distribution construction involves quantile mixture regression (Yan, 2017) built from structured mixtures of generalized asymmetric Laplace densities, to synthesize information from multiple parts of the response distribution in inference for variable selection.

Acknowledgments

This research was supported in part by the National Science Foundation under award SES-1631963.

References

Amemiya, T. (1984), "Tobit models: A survey," *Journal of Econometrics*, 24, 3–61.

Azzalini, A. (1985), "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, 12, 171–178.

Barata, R., Prado, R., and Sansó, B. (2022), "Fast inference for time-varying quantiles via flexible dynamic models with application to the characterization of atmospheric rivers," *The Annals of Applied Statistics*, 16, 247–271.

Chen, X. and Tokdar, S. T. (2021), "Joint quantile regression for spatial data," *Journal of the Royal Statistical Society, Series B*, 83, 826–852.

Chib, S. (1992), "Bayes inference in the Tobit censored regression model," *Journal of Econometrics*, 51, 79–99.

Das, P. and Ghosal, S. (2018), "Bayesian non-parametric simultaneous quantile regression for complete and grid data," *Computational Statistics and Data Analysis*, 127, 172–186.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling extremal events*, volume 33, Springer Science & Business Media.

Gneiting, T. and Raftery, A. E. (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American statistical Association*, 102, 359–378.

Hanson, T. and Johnson, W. O. (2002), "Modeling regression error with a mixture of Pólya trees." *Journal of the American Statistical Association*, 97, 1020–1033.

Harrison, D. and Rubinfeld, D. L. (1978), "Hedonic housing prices and the demand for clean air," *Journal of Environmental Economics and Management*, 5, 81–102.

Henze, N. (1986), "A probabilistic representation of the skew-normal distribution," *Scandinavian Journal of Statistics*, 13, 271–275.

Hoti, F. and Sillanpää, M. (2006), "Bayesian mapping of genotype× expression interactions in quantitative and qualitative traits," *Heredity*, 97, 4–18.

Jeliazkov, I., Karnawat, S., Rahman, M. A., and Vossmeyer, A. (2023), "Flexible Bayesian quantile analysis of residential rental rates," arXiv:2305.13687.

Kobayashi, G., Roh, T., Lee, J., and Choi, T. (2021), "Flexible Bayesian quantile curve fitting with shape restrictions under the Dirichlet process mixture of the generalized asymmetric Laplace distribution," *The Canadian Journal of Statistics*, 49, 698–730.

Koenker, R. (2005), *Quantile Regression*, New York: Cambridge University Press.

Kottas, A. and Gelfand, A. E. (2001), "Bayesian semiparametric median regression modeling," *Journal of the American Statistical Association*, 96, 1458–1468.

Kottas, A. and Krnjajić, M. (2009), "Bayesian semiparametric modelling in quantile regression," *Scandinavian Journal of Statistics*, 36, 297–319.

Kotz, S., Kozubowski, T., and Podgorski, K. (2001), *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Exonomics, Engineering, and Finance*, Boston: Birkhäuser.

Kozumi, H. and Kobayashi, G. (2011), "Gibbs sampling methods for Bayesian quantile regression," *Journal of Statistical Computation and Simulation*, 81, 1565–1578.

Li, Q., Xi, R., and Lin, N. (2010), "Bayesian regularized quantile regression," *Bayesian Analysis*, 5, 533–556.

Lum, K. and Gelfand, A. E. (2012), "Spatial quantile multiple regression using the asymmetric Laplace process (with discussion)," *Bayesian Analysis*, 7, 235–276.

Mroz, T. A. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions," *Econometrica*, 55, 765–799.

Nascimento, M. G. L. and Gonçalves, K. C. M. (2022), "Bayesian variable selection in quantile regression with random effects: an application to Municipal Human Development Index," *Journal of Applied Statistics*, 49, 3436–3450.

Neal, R. M. (2003), "Slice sampling," *The Annals of Statistics*, 31, 705–767.

Park, T. and Casella, G. (2008), "The Bayesian lasso," *Journal of the American Statistical Association*, 103, 681–686.

Rahman, M. A. and Karnawat, S. (2019), "Flexible Bayesian quantile regression in ordinal models," in Jeliazkov, I. and Tobias, J. L. (editors), *Advances in Econometrics*, volume 40B, 211–251.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010), "Flexible Bayesian quantile regression for independent and clustered data," *Biostatistics*, 11, 337–352.

Reich, B. J. and Smith, L. B. (2013), "Bayesian quantile regression for censored data," *Biometrics*, 69, 651–660.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013), "Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density," *Bayesian Analysis*, 8, 479–504.

Taddy, M. and Kottas, A. (2010), "A Bayesian nonparametric approach to inference for quantile regression," *Journal of Business and Economic Statistics*, 28, 357–369.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Tokdar, S. T. and Kadane, J. B. (2012), "Simultaneous linear quantile regression: A semiparametric Bayesian approach," *Bayesian Analysis*, 7, 51–72.

Tsionas, E. G. (2003), "Bayesian quantile inference," *Journal of Statistical Computation and Simulation*, 73, 659–674.

Volinsky, C. T. and Raftery, A. E. (2000), "Bayesian information criterion for censored survival models," *Biometrics*, 56, 256–262.

Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., and Flexeder, C. (2013), "Bayesian semiparametric additive quantile regression," *Statistical Modelling*, 13, 223–252.

Walker, S. G. and Mallick, B. K. (1999), "A Bayesian semiparametric accelerated failure time model," *Biometrics*, 55, 477–483.

Wichitaksorn, N., Choy, B. S. T., and Gerlach, R. (2014), "A generalized class of skew distributions and associated robust quantile regression models," *The Canadian Journal of Statistics*, 42, 579–596.

Yan, Y. (2017), *Bayesian Modeling and Inference for Quantile Mixture Regression*, Ph.D. thesis, University of California, Santa Cruz.

Yan, Y. and Kottas, A. (2017), "A new family of error distributions for Bayesian quantile regression," arXiv:1701.05666.

Yang, Y. and Tokdar, S. T. (2017), "Joint Estimation of Quantile Planes over Arbitrary Predictor Spaces," *Journal of the American Statistical Association*, 112, 1107–1120.

Yang, Y., Wang, H. J., and He, X. (2016), "Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood," *International Statistical Review*, 84, 327–344.

Yu, H. and Yu, L. (2023), "Flexible Bayesian quantile regression for nonlinear mixed effects models based on the generalized asymmetric Laplace distribution," *Journal of Statistical Computation and Simulation*, 93, 2725–2750.

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

Yu, K. and Stander, J. (2007), "Bayesian analysis of a Tobit quantile regression model," *Journal of Econometrics*, 137, 260–276.

Zhu, D. and Galbraith, J. W. (2011), "Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions," *Journal of Empirical Finance*, 18, 765–778.

Zhu, D. and Zinde-Walsh, V. (2009), "Properties and estimation of asymmetric exponential power distribution," *Journal of Econometrics*, 148, 86–99.

Zou, H. and Yuan, M. (2008), "Composite quantile regression and the oracle model selection theory," *The Annals of Statistics*, 36, 1108–1126.

Supplementary Material for "A New Family of Error Distributions for Bayesian Quantile Regression"

A MCMC algorithms

A.1 GAL-based quantile regression with lasso regularization

Recall that in the lasso regularized model, the intercept, β_0, is separated from the rest of the regression coefficients, $\boldsymbol{\beta}$. Based on the prior structure of Section 3.2, the hierarchical model for the data can be written as

$$\begin{aligned}
y_i \mid \boldsymbol{\beta}, \gamma, \sigma, z_i, s_i &\overset{ind.}{\sim} \mathrm{N}(y_i \mid \beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma\, C\, |\gamma| s_i + \sigma A z_i, \sigma^2 B z_i), \quad i = 1, ..., n \\
z_i, s_i &\overset{ind.}{\sim} \mathrm{Exp}(z_i \mid 1)\, \mathrm{N}^+(s_i \mid 0, 1), \quad i = 1, ..., n, \\
\boldsymbol{\beta} \mid \omega_1, \ldots, \omega_d &\sim \prod_{k=1}^{d} N(\beta_k \mid 0, \omega_k), \;\; \omega_1, \ldots, \omega_d \mid \eta^2 \sim \prod_{k=1}^{d} \mathrm{Exp}(\omega_d \mid \eta^2/2), \\
\gamma, \beta_0, \sigma, \eta^2 &\sim \pi(\gamma)\, \mathrm{N}(\beta_0 \mid \mu_{\beta_0}, \sigma^2_{\beta_0})\, \mathrm{IG}(\sigma \mid a_\sigma, b_\sigma)\, \mathrm{Ga}(\eta^2 \mid a_{\eta^2}, b_{\eta^2}),
\end{aligned}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$, $\pi(\gamma)$ is the rescaled Beta distribution supported over the interval (L, U), and $\mathrm{Ga}(\cdot \mid a, b)$ denotes the gamma distribution with mean a/b.

As in Section 3.1, we set $v_i = \sigma z_i$, $i = 1, \ldots, n$, and denote by $\mathrm{GIG}(\nu, a, b)$ the generalized inverse-Gaussian distribution with density $f_{GIG}(x) \propto x^{\nu-1} \exp\{-0.5(a/x + bx)\}$. Then, posterior samples for the model parameters can be obtained through the following steps.

1. Sample β_0 from $\mathrm{N}(\mu^*_{\beta_0}, \sigma^{2*}_{\beta_0})$, with variance $\sigma^{2*}_{\beta_0} = (\sigma^{-2}_{\beta_0} + \sum_{i=1}^n (B\sigma v_i)^{-1})^{-1}$ and mean $\mu^*_{\beta_0} = \sigma^{2*}_{\beta_0}\{\sum_{i=1}^n [y_i - (\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sigma C|\gamma| s_i + A v_i)]/(B\sigma v_i)\}$.

2. Sample $\boldsymbol{\beta}$ from $\mathrm{N}(\boldsymbol{m}^*, \Sigma^*)$, with covariance matrix $\Sigma^* = [\Omega^{-1} + \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T/(B\sigma v_i)]^{-1}$ and mean vector $\boldsymbol{m}^* = \Sigma^*\{\sum_{i=1}^n \boldsymbol{x}_i [y_i - (\beta_0 + \sigma C|\gamma| s_i + A v_i)]/(B\sigma v_i)\}$, where Ω is a $d \times d$ diagonal matrix and the (k,k)-th diagonal entry is ω_k, $k = 1, \ldots, d$.

3. For each $i = 1, ..., n$, sample v_i from a $\mathrm{GIG}(0.5, a_i, b_i)$ distribution, where $a_i = [y_i - (\beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma C|\gamma| s_i)]^2/(B\sigma)$, and $b_i = 2/\sigma + A^2/(B\sigma)$.

4. For each $i = 1, ..., n$, sample s_i from a normal $\mathrm{N}(\mu_{s_i}, \sigma^2_{s_i})$ distribution truncated on $\mathbb{R}^+$, where $\sigma^2_{s_i} = [(C\gamma)^2\sigma/(Bv_i) + 1]^{-1}$ and $\mu_{s_i} = \sigma^2_{s_i} C|\gamma|[y_i - (\beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} + A v_i)]/(Bv_i)$.

5. Sample σ from a GIG(ν, c, d) distribution, where $\nu = -(a_\sigma + 1.5n)$, $c = 2b_\sigma + 2\sum_{i=1}^n v_i + \sum_{i=1}^n [y_i - (\beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} + Av_i)]^2/(Bv_i)$, and $d = \sum_{i=1}^n (C\gamma s_i)^2/(Bv_i)$.

6. Update γ using slice sampling based on stepping-out and shrinkage procedures (Scheme (3), Section 4.1 and Method (ii), Section 4.2 of Neal 2003).

7. Sample ω_k from a GIG$(0.5, \beta_k^2, \eta^2)$ distribution, for $k = 1, \ldots, d$.

8. Sample η^2 from a Ga$(\eta^2 \,|\, a_{\eta^2} + d, b_{\eta^2} + 0.5\sum_{k=1}^d \omega_k)$ distribution.

A.2 AL-based quantile regression

The hierarchical representation of the AL-based quantile regression is

$$\begin{aligned} y_i \mid \boldsymbol{\beta}, \sigma, z_i &\overset{ind.}{\sim} \mathrm{N}(y_i \mid \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma A z_i, \sigma^2 B z_i), \quad i = 1, ..., n \\ z_i &\overset{ind.}{\sim} \mathrm{Exp}(z_i \mid 1) \quad i = 1, ..., n, \\ \boldsymbol{\beta}, \sigma &\sim \mathrm{N}(\beta \mid \boldsymbol{m}_0, \Sigma_0)\,\mathrm{IG}(\sigma \mid a_\sigma, b_\sigma). \end{aligned}$$

We set $v_i = \sigma z_i$, $i = 1, \ldots, n$ and implement the MCMC algorithm based on Kozumi and Kobayashi (2011) as follows.

1. Sample $\boldsymbol{\beta}$ from $\mathrm{N}(\boldsymbol{m}^*, \Sigma^*)$, with covariance matrix $\Sigma^* = [\Sigma_0^{-1} + \sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T/(B\sigma v_i)]^{-1}$ and mean vector $\boldsymbol{m}^* = \Sigma^*\{\Sigma_0^{-1}\boldsymbol{m}_0 + \sum_{i=1}^n \boldsymbol{x}_i[y_i - Av_i)]/(B\sigma v_i)\}$.

2. For each $i = 1, ..., n$, sample v_i from a GIG$(0.5, a_i, b_i)$ distribution, where $a_i = (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2/(B\sigma)$ and $b_i = 2/\sigma + A^2/(B\sigma)$.

3. Sample σ from a IG(c, d) distribution, where $c = a_\sigma + 1.5n$ and $d = b_\sigma + \sum_{i=1}^n v_i + 0.5\sum_{i=1}^n (y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta} - Av_i)^2/(Bv_i)$.

The MCMC algorithm for the AL-based model with lasso regularization can be obtained similarly according to Section A.1.

B Additional simulation results

We provide additional results for the simulation study (Section 4 of the paper) that compares the lasso regularized quantile regression models with AL and GAL errors. Section 4.3 of the paper provides discussion of the results presented here, as well as of the ones included in the main paper.

Tables 1-6 show the RMSE of the regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_8)^\top$, for sample sizes $n = 100$ and $n = 1000$. For each $k = 1, ..., 8$, the RMSE of the corresponding regression coefficient estimate is defined as RMSE $= \sqrt{M^{-1} \sum_{m=1}^{M} (\beta_k^{(m)} - \beta_k)^2}$, where M is the posterior sample size, and $\beta_k^{(m)}$ is the m-th posterior sample for β_k.

Tables 7-12 provide the empirical coverage probability of the 95% posterior credible interval for each regression coefficient β_k, for $k = 1, \ldots, 8$. Results are reported for all combinations of the percentile value ($p_0 = 0.05, 0.25, 0.5$), sample size ($n = 100, 1000$), and true error distribution.

To further access model performance with larger samples, we simulate a single data set with sample size $n = 100, 1000, 10000$, for each p_0 and error distribution considered in the simulation study. This results in 36 data sets in total. Tables 13-15 show the results for correct inclusion and exclusion, average RMSE, and prediction interval score under both models. Moreover, using each of the 36 data sets, Figures 1-4 plot the posterior mean and 95% interval estimates of the error density under the GAL and AL models, for each of the different scenarios for p_0, sample size n, and true error distribution.

Table 1: RMSE ($p_0 = 0.05$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.50 (0.16)	0.79 (0.32)	0.37 (0.12)	0.46 (0.18)
	AL	0.61 (0.26)	1.20 (0.55)	0.46 (0.23)	0.55 (0.32)
β_2	GAL	0.56 (0.19)	0.82 (0.25)	0.41 (0.15)	0.49 (0.17)
	AL	0.68 (0.30)	1.23 (0.64)	0.53 (0.25)	0.59 (0.31)
β_3	GAL	0.50 (0.19)	0.67 (0.23)	0.35 (0.11)	0.43 (0.14)
	AL	0.64 (0.31)	1.10 (0.71)	0.46 (0.20)	0.52 (0.27)
β_4	GAL	0.45 (0.14)	0.64 (0.22)	0.35 (0.12)	0.43 (0.15)
	AL	0.62 (0.33)	1.11 (0.75)	0.45 (0.24)	0.49 (0.23)
β_5	GAL	0.58 (0.21)	0.80 (0.28)	0.41 (0.13)	0.50 (0.17)
	AL	0.75 (0.36)	1.23 (0.64)	0.54 (0.23)	0.61 (0.29)
β_6	GAL	0.49 (0.17)	0.68 (0.27)	0.35 (0.12)	0.41 (0.11)
	AL	0.73 (0.38)	1.18 (0.73)	0.46 (0.21)	0.51 (0.25)
β_7	GAL	0.45 (0.15)	0.61 (0.19)	0.34 (0.09)	0.41 (0.13)
	AL	0.62 (0.32)	0.93 (0.49)	0.42 (0.18)	0.50 (0.26)
β_8	GAL	0.42 (0.14)	0.59 (0.22)	0.31 (0.10)	0.37 (0.12)
	AL	0.56 (0.25)	0.91 (0.46)	0.42 (0.24)	0.44 (0.23)

Table 2: RMSE ($p_0 = 0.05$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.16 (0.16)	0.23 (0.23)	0.11 (0.11)	0.14 (0.14)
	AL	0.22 (0.10)	0.42 (0.22)	0.16 (0.09)	0.18 (0.08)
β_2	GAL	0.17 (0.17)	0.24 (0.24)	0.11 (0.11)	0.15 (0.15)
	AL	0.24 (0.12)	0.47 (0.26)	0.16 (0.09)	0.19 (0.08)
β_3	GAL	0.16 (0.16)	0.23 (0.23)	0.12 (0.12)	0.15 (0.15)
	AL	0.23 (0.13)	0.44 (0.24)	0.17 (0.09)	0.20 (0.10)
β_4	GAL	0.17 (0.17)	0.24 (0.24)	0.12 (0.12)	0.14 (0.14)
	AL	0.24 (0.13)	0.48 (0.26)	0.18 (0.09)	0.18 (0.08)
β_5	GAL	0.17 (0.17)	0.26 (0.26)	0.12 (0.12)	0.14 (0.14)
	AL	0.23 (0.11)	0.50 (0.29)	0.18 (0.09)	0.19 (0.10)
β_6	GAL	0.16 (0.16)	0.24 (0.24)	0.12 (0.12)	0.15 (0.15)
	AL	0.22 (0.12)	0.50 (0.31)	0.16 (0.09)	0.19 (0.09)
β_7	GAL	0.16 (0.16)	0.23 (0.23)	0.11 (0.11)	0.14 (0.14)
	AL	0.23 (0.11)	0.43 (0.23)	0.18 (0.09)	0.19 (0.10)
β_8	GAL	0.15 (0.15)	0.21 (0.21)	0.10 (0.10)	0.13 (0.13)
	AL	0.21 (0.10)	0.39 (0.22)	0.15 (0.07)	0.16 (0.09)

Table 3: RMSE ($p_0 = 0.25$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.50 (0.17)	0.61 (0.20)	0.36 (0.12)	0.45 (0.14)
	AL	0.54 (0.23)	0.68 (0.25)	0.36 (0.13)	0.43 (0.13)
β_2	GAL	0.51 (0.14)	0.64 (0.17)	0.41 (0.14)	0.52 (0.17)
	AL	0.54 (0.17)	0.71 (0.22)	0.41 (0.15)	0.52 (0.19)
β_3	GAL	0.47 (0.15)	0.55 (0.16)	0.32 (0.10)	0.45 (0.16)
	AL	0.51 (0.18)	0.64 (0.23)	0.34 (0.12)	0.45 (0.18)
β_4	GAL	0.46 (0.15)	0.54 (0.20)	0.33 (0.09)	0.40 (0.13)
	AL	0.50 (0.21)	0.58 (0.19)	0.36 (0.13)	0.39 (0.14)
β_5	GAL	0.56 (0.19)	0.67 (0.21)	0.40 (0.13)	0.52 (0.19)
	AL	0.60 (0.24)	0.72 (0.24)	0.41 (0.14)	0.50 (0.20)
β_6	GAL	0.46 (0.17)	0.54 (0.19)	0.33 (0.09)	0.41 (0.12)
	AL	0.49 (0.21)	0.61 (0.25)	0.36 (0.11)	0.41 (0.13)
β_7	GAL	0.46 (0.15)	0.52 (0.16)	0.34 (0.10)	0.42 (0.13)
	AL	0.50 (0.21)	0.58 (0.23)	0.35 (0.11)	0.41 (0.14)
β_8	GAL	0.43 (0.14)	0.47 (0.14)	0.30 (0.09)	0.41 (0.16)
	AL	0.44 (0.15)	0.53 (0.22)	0.32 (0.11)	0.40 (0.16)

Table 4: RMSE ($p_0 = 0.25$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.15 (0.15)	0.17 (0.17)	0.11 (0.11)	0.14 (0.14)
	AL	0.16 (0.07)	0.21 (0.08)	0.11 (0.04)	0.14 (0.05)
β_2	GAL	0.17 (0.17)	0.21 (0.21)	0.12 (0.12)	0.15 (0.15)
	AL	0.18 (0.07)	0.26 (0.09)	0.12 (0.05)	0.15 (0.05)
β_3	GAL	0.16 (0.16)	0.18 (0.18)	0.11 (0.11)	0.14 (0.14)
	AL	0.18 (0.08)	0.23 (0.09)	0.12 (0.05)	0.14 (0.05)
β_4	GAL	0.17 (0.17)	0.18 (0.18)	0.12 (0.12)	0.14 (0.14)
	AL	0.19 (0.08)	0.23 (0.10)	0.13 (0.05)	0.14 (0.05)
β_5	GAL	0.17 (0.17)	0.19 (0.19)	0.12 (0.12)	0.14 (0.14)
	AL	0.19 (0.08)	0.24 (0.09)	0.13 (0.05)	0.14 (0.05)
β_6	GAL	0.16 (0.16)	0.18 (0.18)	0.11 (0.11)	0.14 (0.14)
	AL	0.18 (0.08)	0.22 (0.08)	0.12 (0.04)	0.14 (0.05)
β_7	GAL	0.17 (0.17)	0.18 (0.18)	0.11 (0.11)	0.14 (0.14)
	AL	0.18 (0.07)	0.21 (0.08)	0.12 (0.05)	0.14 (0.05)
β_8	GAL	0.14 (0.14)	0.17 (0.17)	0.11 (0.11)	0.12 (0.12)
	AL	0.16 (0.06)	0.21 (0.08)	0.11 (0.05)	0.13 (0.04)

Table 5: RMSE ($p_0 = 0.5$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.51 (0.18)	0.58 (0.19)	0.37 (0.12)	0.47 (0.19)
	AL	0.52 (0.19)	0.57 (0.18)	0.38 (0.13)	0.53 (0.22)
β_2	GAL	0.57 (0.21)	0.64 (0.19)	0.42 (0.13)	0.50 (0.17)
	AL	0.57 (0.19)	0.62 (0.20)	0.42 (0.13)	0.56 (0.20)
β_3	GAL	0.49 (0.18)	0.55 (0.20)	0.35 (0.11)	0.43 (0.14)
	AL	0.49 (0.17)	0.54 (0.19)	0.35 (0.11)	0.46 (0.17)
β_4	GAL	0.49 (0.17)	0.56 (0.17)	0.37 (0.11)	0.41 (0.14)
	AL	0.50 (0.17)	0.55 (0.17)	0.38 (0.13)	0.45 (0.16)
β_5	GAL	0.61 (0.24)	0.66 (0.24)	0.42 (0.16)	0.51 (0.17)
	AL	0.61 (0.24)	0.65 (0.24)	0.42 (0.17)	0.55 (0.18)
β_6	GAL	0.47 (0.19)	0.53 (0.18)	0.38 (0.16)	0.42 (0.14)
	AL	0.48 (0.21)	0.52 (0.19)	0.38 (0.17)	0.44 (0.15)
β_7	GAL	0.45 (0.15)	0.52 (0.20)	0.36 (0.12)	0.42 (0.14)
	AL	0.46 (0.17)	0.51 (0.19)	0.36 (0.13)	0.46 (0.15)
β_8	GAL	0.40 (0.13)	0.45 (0.13)	0.31 (0.10)	0.41 (0.16)
	AL	0.41 (0.14)	0.44 (0.12)	0.32 (0.11)	0.43 (0.16)

Table 6: RMSE ($p_0 = 0.5$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.16 (0.16)	0.16 (0.16)	0.13 (0.13)	0.14 (0.14)
	AL	0.16 (0.07)	0.16 (0.05)	0.13 (0.05)	0.16 (0.06)
β_2	GAL	0.18 (0.18)	0.18 (0.18)	0.13 (0.13)	0.15 (0.15)
	AL	0.18 (0.07)	0.18 (0.06)	0.14 (0.05)	0.18 (0.06)
β_3	GAL	0.16 (0.16)	0.17 (0.17)	0.12 (0.12)	0.14 (0.14)
	AL	0.16 (0.05)	0.17 (0.05)	0.12 (0.04)	0.16 (0.06)
β_4	GAL	0.17 (0.17)	0.16 (0.16)	0.12 (0.12)	0.13 (0.13)
	AL	0.17 (0.06)	0.16 (0.05)	0.13 (0.05)	0.16 (0.05)
β_5	GAL	0.18 (0.18)	0.18 (0.18)	0.13 (0.13)	0.14 (0.14)
	AL	0.18 (0.06)	0.18 (0.06)	0.13 (0.05)	0.17 (0.06)
β_6	GAL	0.17 (0.17)	0.16 (0.16)	0.12 (0.12)	0.14 (0.14)
	AL	0.17 (0.06)	0.16 (0.04)	0.12 (0.05)	0.16 (0.05)
β_7	GAL	0.17 (0.17)	0.16 (0.16)	0.11 (0.11)	0.15 (0.15)
	AL	0.17 (0.06)	0.16 (0.05)	0.12 (0.04)	0.16 (0.05)
β_8	GAL	0.15 (0.15)	0.15 (0.15)	0.10 (0.10)	0.13 (0.13)
	AL	0.16 (0.05)	0.15 (0.05)	0.11 (0.04)	0.15 (0.05)

Table 7: Empirical coverage of the 95% credible interval ($p_0 = 0.05$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.91	0.88	0.94	0.95
	AL	0.79	0.62	0.74	0.80
β_2	GAL	0.90	0.91	0.90	0.93
	AL	0.73	0.67	0.76	0.77
β_3	GAL	0.92	0.97	0.97	0.95
	AL	0.74	0.75	0.79	0.82
β_4	GAL	0.98	0.97	0.97	0.97
	AL	0.76	0.79	0.81	0.86
β_5	GAL	0.90	0.92	0.93	0.94
	AL	0.67	0.67	0.75	0.78
β_6	GAL	0.97	0.93	0.96	0.99
	AL	0.69	0.70	0.81	0.82
β_7	GAL	0.97	0.95	0.98	0.96
	AL	0.81	0.83	0.87	0.83
β_8	GAL	0.98	0.97	0.97	0.96
	AL	0.83	0.77	0.82	0.85

Table 8: Empirical coverage of the 95% credible interval ($p_0 = 0.05$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.94	0.93	0.94	0.88
	AL	0.66	0.59	0.69	0.69
β_2	GAL	0.94	0.93	0.96	0.91
	AL	0.68	0.61	0.68	0.76
β_3	GAL	0.96	0.92	0.94	0.94
	AL	0.72	0.57	0.68	0.64
β_4	GAL	0.89	0.90	0.92	0.97
	AL	0.68	0.59	0.64	0.76
β_5	GAL	0.91	0.92	0.94	0.97
	AL	0.70	0.60	0.68	0.74
β_6	GAL	0.94	0.95	0.98	0.95
	AL	0.77	0.58	0.77	0.70
β_7	GAL	0.92	0.93	0.94	0.97
	AL	0.69	0.64	0.64	0.73
β_8	GAL	0.94	0.93	0.92	0.95
	AL	0.69	0.63	0.73	0.73

Table 9: Empirical coverage of the 95% credible interval ($p_0 = 0.25$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.93	0.97	0.94	0.98
	AL	0.83	0.88	0.91	0.97
β_2	GAL	0.98	0.99	0.90	0.91
	AL	0.94	0.93	0.91	0.89
β_3	GAL	0.97	0.99	0.97	0.94
	AL	0.91	0.94	0.96	0.91
β_4	GAL	0.96	0.97	0.97	0.98
	AL	0.88	0.99	0.93	0.97
β_5	GAL	0.91	0.93	0.92	0.94
	AL	0.86	0.87	0.87	0.92
β_6	GAL	0.97	0.99	0.99	0.99
	AL	0.93	0.94	0.94	0.99
β_7	GAL	0.99	1.00	0.97	0.98
	AL	0.90	0.96	0.98	0.96
β_8	GAL	0.97	0.99	0.98	0.94
	AL	0.95	0.97	0.95	0.93

Table 10: Empirical coverage of the 95% credible interval ($p_0 = 0.25$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.95	0.96	0.93	0.85
	AL	0.88	0.89	0.88	0.86
β_2	GAL	0.96	0.91	0.92	0.91
	AL	0.93	0.85	0.86	0.91
β_3	GAL	0.93	0.96	0.98	0.94
	AL	0.88	0.87	0.91	0.89
β_4	GAL	0.93	0.96	0.90	0.90
	AL	0.85	0.89	0.88	0.92
β_5	GAL	0.95	0.96	0.94	0.93
	AL	0.83	0.89	0.83	0.90
β_6	GAL	0.95	1.00	0.95	0.94
	AL	0.89	0.91	0.89	0.92
β_7	GAL	0.94	0.98	0.95	0.93
	AL	0.87	0.90	0.88	0.92
β_8	GAL	0.93	0.94	0.94	0.95
	AL	0.89	0.87	0.86	0.93

Table 11: Empirical coverage of the 95% credible interval ($p_0 = 0.5$, $n = 100$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.95	0.95	0.90	0.96
	AL	0.91	0.94	0.91	0.91
β_2	GAL	0.94	0.95	0.93	0.95
	AL	0.92	0.96	0.93	0.92
β_3	GAL	0.95	0.95	0.99	0.97
	AL	0.94	0.96	0.99	0.97
β_4	GAL	0.96	0.98	0.95	0.96
	AL	0.97	0.98	0.92	0.95
β_5	GAL	0.88	0.92	0.87	0.95
	AL	0.85	0.92	0.85	0.96
β_6	GAL	0.96	0.98	0.90	0.95
	AL	0.96	0.96	0.88	0.96
β_7	GAL	0.98	0.96	0.95	0.96
	AL	0.97	0.97	0.94	0.97
β_8	GAL	0.97	0.98	0.96	0.93
	AL	0.98	0.97	0.94	0.96

Table 12: Empirical coverage of the 95% credible interval ($p_0 = 0.5$, $n = 1000$)

	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
β_1	GAL	0.88	0.92	0.79	0.93
	AL	0.89	0.92	0.81	0.90
β_2	GAL	0.87	0.92	0.88	0.92
	AL	0.87	0.91	0.84	0.89
β_3	GAL	0.92	0.95	0.92	0.96
	AL	0.93	0.95	0.95	0.95
β_4	GAL	0.89	0.94	0.93	0.97
	AL	0.91	0.96	0.92	0.94
β_5	GAL	0.89	0.96	0.91	0.96
	AL	0.90	0.95	0.90	0.93
β_6	GAL	0.91	0.98	0.94	0.95
	AL	0.90	0.99	0.89	0.91
β_7	GAL	0.86	0.97	0.95	0.93
	AL	0.89	0.97	0.95	0.93
β_8	GAL	0.90	0.97	0.96	0.96
	AL	0.91	0.97	0.96	0.95

Table 13: Correct inclusion and exclusion for different n

	n	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
0.05	10^2	GAL	0.71	0.75	0.95	0.81
		AL	0.67	0.68	0.83	0.69
	10^3	GAL	1.00	0.99	1.00	1.00
		AL	0.94	0.87	1.00	1.00
	10^4	GAL	1.00	1.00	1.00	1.00
		AL	1.00	1.00	1.00	1.00
0.25	10^2	GAL	0.94	0.90	0.93	0.89
		AL	0.93	0.81	0.91	0.90
	10^3	GAL	1.00	1.00	1.00	1.00
		AL	1.00	0.99	1.00	1.00
	10^4	GAL	1.00	1.00	1.00	1.00
		AL	1.00	1.00	1.00	1.00
0.50	10^2	GAL	0.78	0.83	0.89	0.88
		AL	0.76	0.83	0.90	0.87
	10^3	GAL	1.00	1.00	1.00	1.00
		AL	1.00	1.00	1.00	1.00
	10^4	GAL	1.00	1.00	1.00	1.00
		AL	1.00	1.00	1.00	1.00

Table 14: Average RMSE for different n

	n	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
0.05	10^2	GAL	0.62	0.63	0.32	0.45
		AL	0.74	0.84	0.38	0.53
	10^3	GAL	0.16	0.24	0.13	0.14
		AL	0.31	0.39	0.11	0.13
	10^4	GAL	0.05	0.09	0.03	0.04
		AL	0.06	0.14	0.08	0.05
0.25	10^2	GAL	0.37	0.54	0.31	0.44
		AL	0.38	0.70	0.33	0.43
	10^3	GAL	0.16	0.18	0.10	0.13
		AL	0.17	0.22	0.10	0.13
	10^4	GAL	0.04	0.05	0.03	0.05
		AL	0.05	0.06	0.03	0.05
0.50	10^2	GAL	0.53	0.63	0.33	0.48
		AL	0.54	0.63	0.33	0.54
	10^3	GAL	0.15	0.20	0.10	0.17
		AL	0.14	0.19	0.12	0.19
	10^4	GAL	0.05	0.05	0.04	0.05
		AL	0.05	0.05	0.04	0.05

Table 15: Prediction interval score for different n

	n	Model	Normal	Laplace	Normal mixture	Log generalized Pareto
0.05	10^2	GAL	15.26	21.92	14.40	24.33
		AL	23.34	33.05	19.33	25.48
	10^3	GAL	12.30	18.92	10.65	16.52
		AL	23.84	36.92	17.88	21.76
	10^4	GAL	16.93	22.52	10.43	23.36
		AL	27.71	39.75	17.05	23.97
0.25	10^2	GAL	15.84	23.46	10.17	27.59
		AL	21.64	29.36	12.25	19.29
	10^3	GAL	19.43	22.91	11.57	20.95
		AL	16.17	27.73	11.33	20.69
	10^4	GAL	15.33	27.94	11.32	18.40
		AL	18.21	31.29	12.53	20.56
0.50	10^2	GAL	17.61	27.37	12.22	21.13
		AL	16.22	27.46	11.51	24.38
	10^3	GAL	22.60	20.30	13.20	22.33
		AL	16.43	20.39	11.92	18.85
	10^4	GAL	14.28	33.78	10.90	28.85
		AL	14.15	33.97	10.60	22.57

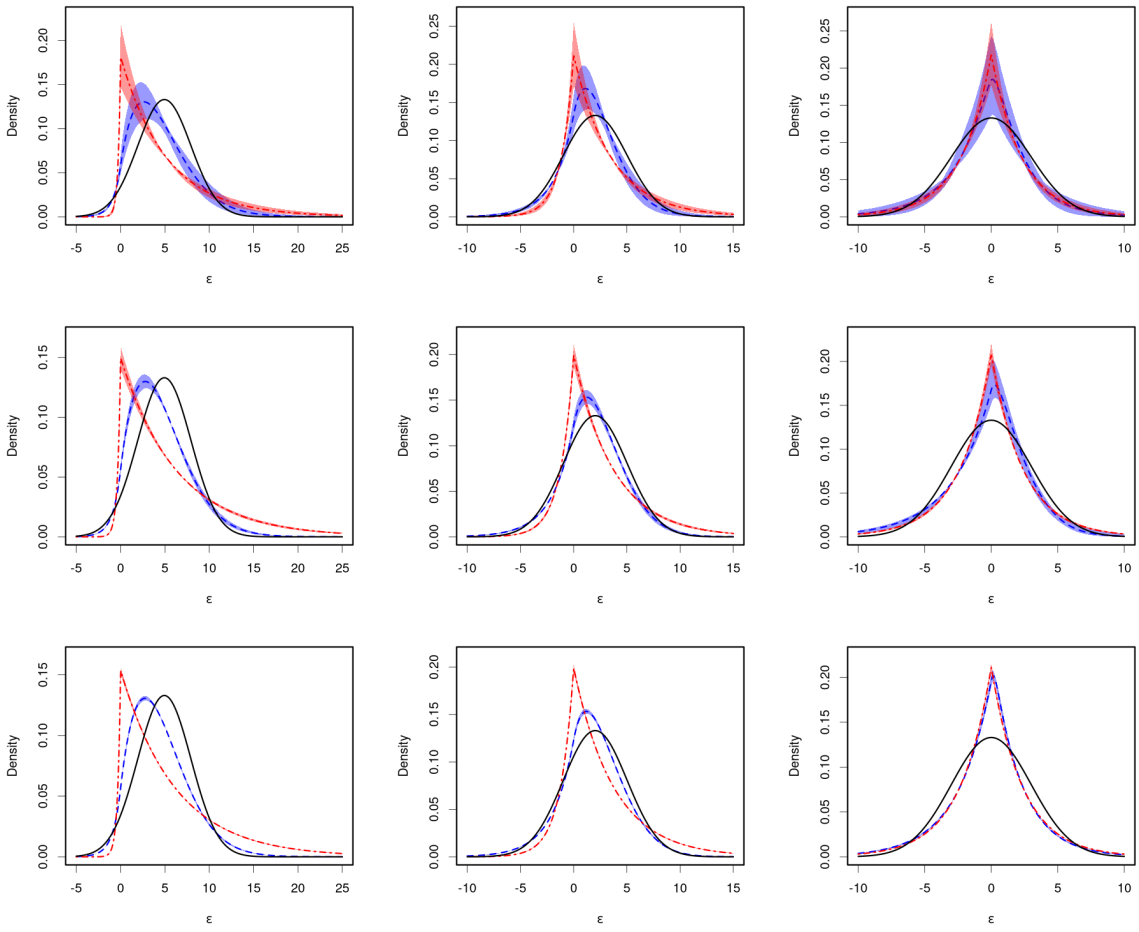


Figure 1: Simulation under the normal error density (solid black line). Posterior mean and 95% interval estimates of the error density under the asymmetric Laplace model (red) and the generalized asymmetric Laplace model (blue), for $p_0 = 0.05$ (left column), $p_0 = 0.25$ (middle column), $p_0 = 0.5$ (right column), and for sample size $n = 100$ (top row), $n = 1000$ (middle row), $n = 10000$ (bottom row).

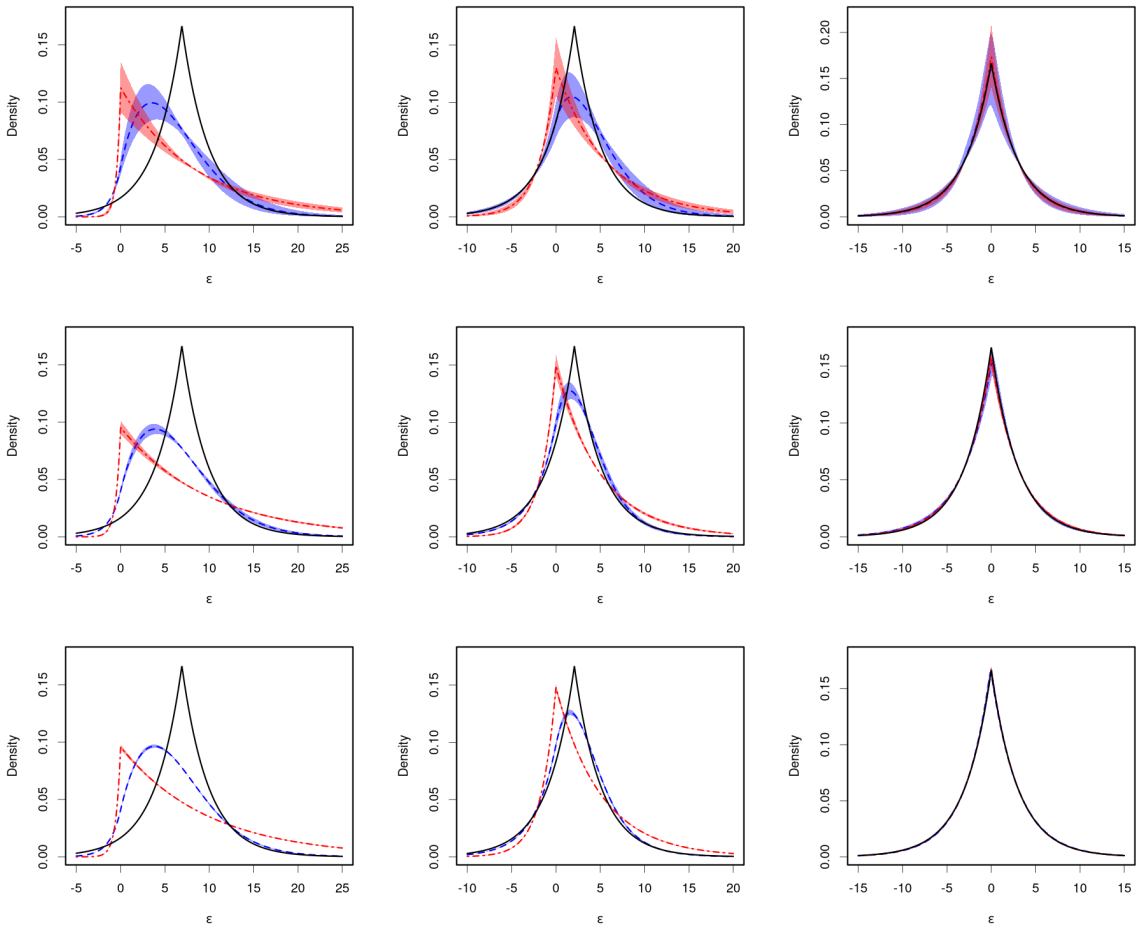


Figure 2: Simulation under the Laplace error density (solid black line). Posterior mean and 95% interval estimates of the error density under the asymmetric Laplace model (red) and the generalized asymmetric Laplace model (blue), for $p_0 = 0.05$ (left column), $p_0 = 0.25$ (middle column), $p_0 = 0.5$ (right column), and for sample size $n = 100$ (top row), $n = 1000$ (middle row), $n = 10000$ (bottom row).

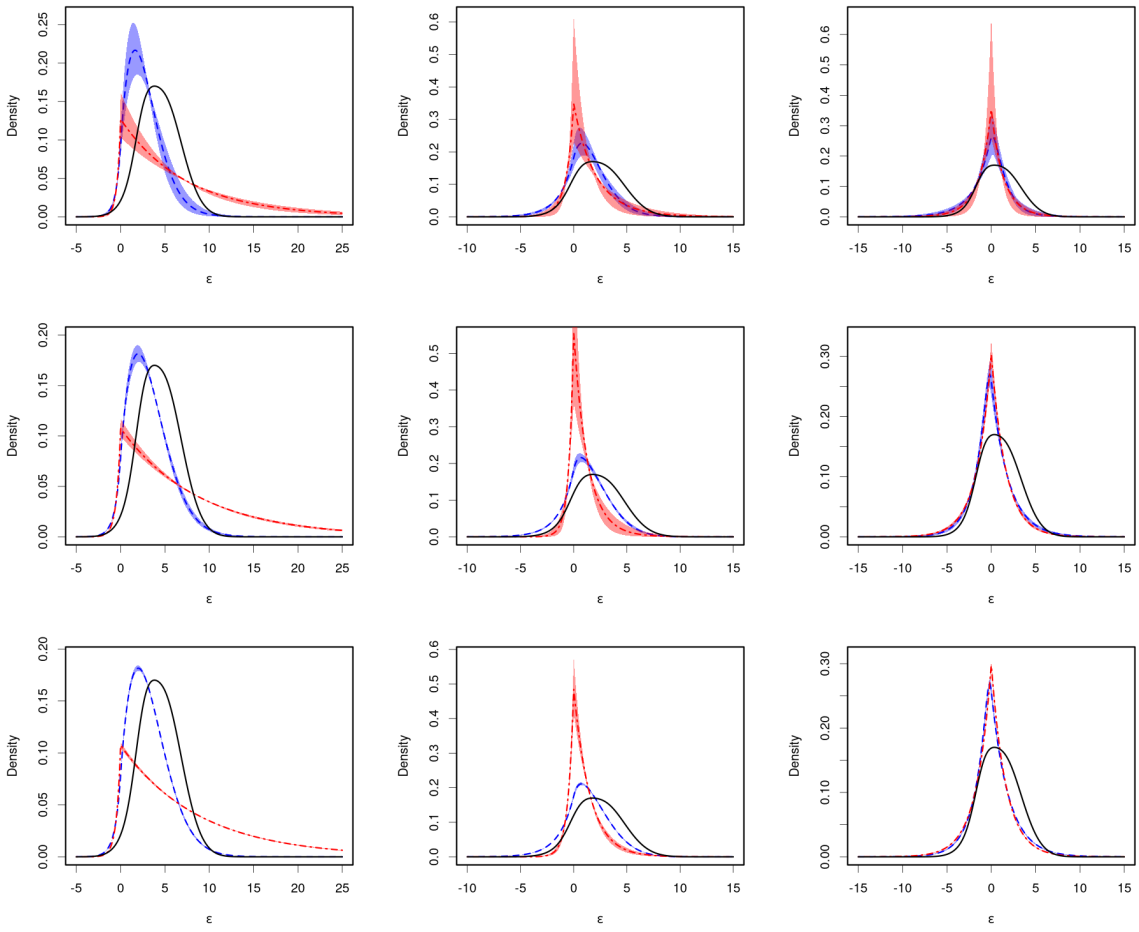


Figure 3: Simulation under the normal mixture error density (solid black line). Posterior mean and 95% interval estimates of the error density under the asymmetric Laplace model (red) and the generalized asymmetric Laplace model (blue), for $p_0 = 0.05$ (left column), $p_0 = 0.25$ (middle column), $p_0 = 0.5$ (right column), and for sample size $n = 100$ (top row), $n = 1000$ (middle row), $n = 10000$ (bottom row).

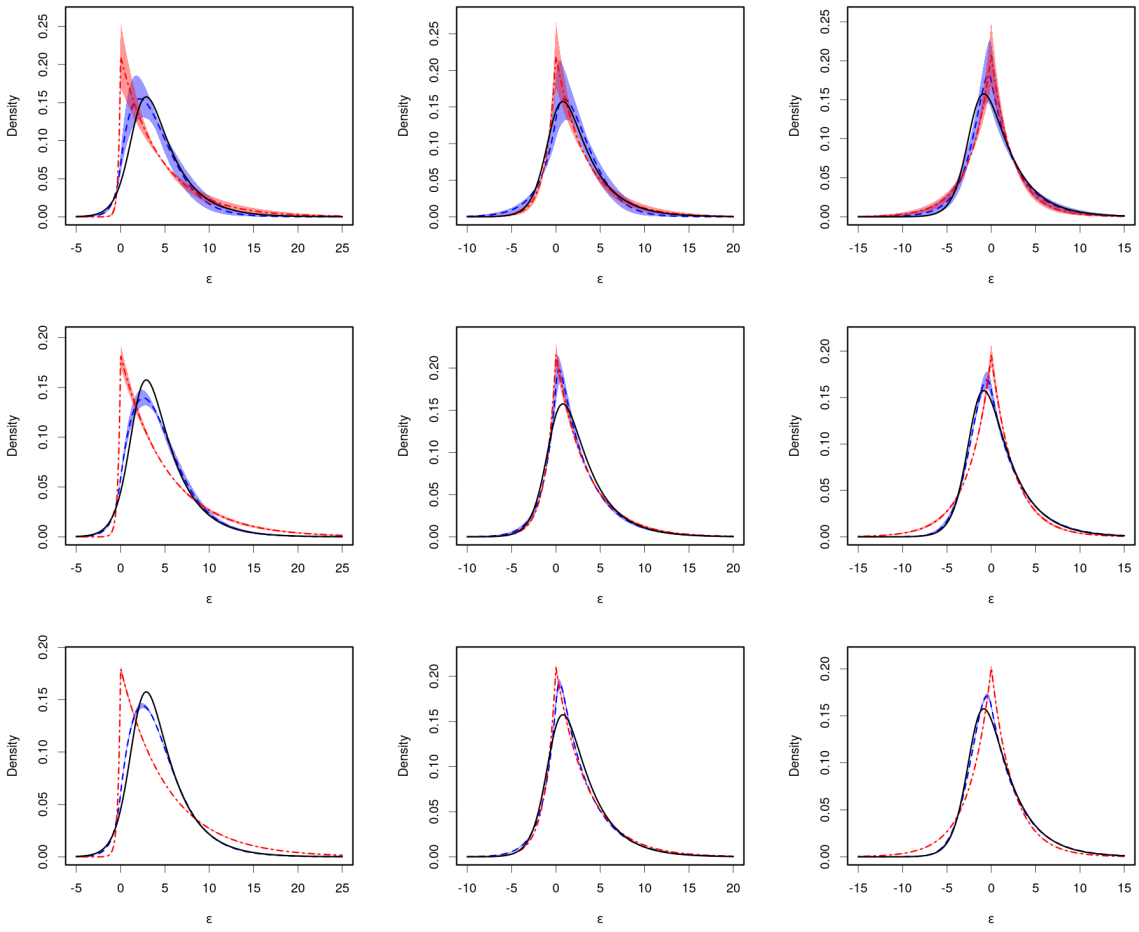


Figure 4: Simulation under the log-transformed generalized Pareto error density (solid black line). Posterior mean and 95% interval estimates of the error density under the asymmetric Laplace model (red) and the generalized asymmetric Laplace model (blue), for $p_0 = 0.05$ (left column), $p_0 = 0.25$ (middle column), $p_0 = 0.5$ (right column), and for sample size $n = 100$ (top row), $n = 1000$ (middle row), $n = 10000$ (bottom row).