Bayesian Analysis (2012) **7**, Number 2, pp. 235–258

Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process

Kristian Lum* and Alan E. Gelfand†

Abstract. We consider quantile multiple regression through conditional quantile models, i.e. each quantile is modeled separately. We work in the context of spatially referenced data and extend the asymmetric Laplace model for quantile regression to a spatial process, the asymmetric Laplace process (ALP) for quantile regression with spatially dependent errors. By taking advantage of a convenient conditionally Gaussian representation of the asymmetric Laplace distribution, we are able to straightforwardly incorporate spatial dependence in this process. We develop the properties of this process under several specifications, each of which induces different smoothness and covariance behavior at the extreme quantiles.

We demonstrate the advantages that may be gained by incorporating spatial dependence into this conditional quantile model by applying it to a data set of log selling prices of homes in Baton Rouge, LA, given characteristics of each house. We also introduce the asymmetric Laplace predictive process (ALPP) which accommodates large data sets, and apply it to a data set of birth weights given maternal covariates for several thousand births in North Carolina in 2000. By modeling the spatial structure in the data, we are able to show, using a check loss function, improved performance on each of the data sets for each of the quantiles at which the model was fit.

Keywords: quantile regression, conditional quantiles, spatial statistics, MCMC

1 Introduction

Quantile multiple regression supplements multiple regression for the mean by supplying information about the relationship between the response and the covariates at the tails of the response distribution. Here we focus on conditional quantile modeling since many applications areas are concerned with the changing effects of the covariates on the outcome across the quantiles of the distribution. Furthermore, we have geo-coded locations for the observations and so, in addition, we seek to account for anticipated spatial dependence.

To date, little work has been done to define a conditional quantile regression model that incorporates spatial dependence. The contribution of our work is a full development of a conditional quantile process model that incorporates spatial dependence through the spatial asymmetric Laplace process (ALP). The ALP relaxes the traditional assumption of iid errors by allowing for residual dependence in each quantile regression via a latent Gaussian process. The ALP allows closed form covariance calculation. In fact, we

*Unaffiliated, kristianlum@gmail.com
†Department of Statistical Science, Duke University, Durham NC, alan@stat.duke.edu

 DOI:10.1214/12-BA708

consider various versions of this process to allow for differing behavior as the extreme quantiles are approached and we derive the covariance structure implied by each of these models. Through the ALP, we are able to obtain *global* quantile regression coefficients. We can also provide spatial interpolation of a quantile under this process model. We fit a Bayesian hierarchical model using a straightforward Markov Chain Monte Carlo (MCMC). The estimation algorithm requires minimal tuning because of the conditional Gaussian representation of the asymmetric Laplace distribution for the errors under the model. As a result, we achieve a fully model-based approach enabling full inference regarding regression coefficients, avoiding possibly inappropriate asymptotics. Finally, we extend this process to accommodate large data sets using a reduced rank approach.

The format of this paper is as follows. In Section 2, we offer a brief literature review, discussing both the conditional and joint modeling paths for quantile regression and explaining where the ALP fits in. In Section 3, we briefly review the asymmetric Laplace distribution and the properties that make it ideal for use as the error term of a quantile regression. In Section 4, we introduce the asymmetric Laplace process along with several options for defining it. In Section 5, we formalize our quantile regression model, discuss model fitting, prior specifications, and interpolation of the quantile regression. In Section 6, we demonstrate the advantages of having a spatial component in the quantile regression for a data set of the selling price of homes in Baton Rouge, LA. In Section 7, we introduce a modification to the asymmetric Laplace process that makes it suitable for use with large data sets, which we call the asymmetric Laplace predictive process. Section 8 employs this modified process for a data set of birth weights in Durham County, North Carolina. We conclude with Section 9, which summarizes the work presented and offers some future directions.

2 Relevant Literature

As alluded to above, there are two paths for developing quantile regressions. The first approach, which essentially follows the original incarnation of this idea as presented in Koenker and Bassett (1978), offers a regression model for each of the quantiles of interest separately. Typically, inference proceeds by minimizing check loss or assuming an asymmetric Laplace error term. This approach, which we will call the "conditional quantile model", specifies a different model for each quantile of the outcome distribution.

The conditional quantile models may be further subdivided. One category specifies the conditional quantile linearly in the covariates and then introduces an asymmetric Laplace error distribution or minimizes the check loss function. Examples of conditional quantile models appear in Yu and Moyeed (2001), which specifies a Bayesian quantile regression model with iid asymmetric Laplace error terms and proves the propriety of the posterior of the regression coefficients in this framework under an improper prior. Tsionas (2003), Reed and Yu (2009), and Kozumi and Kobayashi (2011) present a Gibbs sampler for a Bayesian quantile regression model. The efficiency of the method in Reed and Yu (2009) and Kozumi and Kobayashi (2011) is due to its use of the conditionally Gaussian representation of the asymmetric Laplace distribution. Li et al. (2010) also uses this representation and considers priors that correspond to lasso, elastic net, and group lasso penalties. A non-Bayesian example is Cai and Xu (2008), which is linear in the quantile at each of several time points across which regression coefficients are smoothed. Other models relax the parametric assumption on the error term, yet retain linearity in the form of the conditional quantile. For example, Reich et al. (2010) introduces a mixture model, where the mixing components are themselves mixtures of two normal distributions which have mixing weights set to force zero to be the quantile of interest. Kottas and Krnjajić (2009) introduces two semiparametric models for the error distribution of a quantile regression. Recent work of Hallin et al. (2009) introduces conditional spatial quantile regression that is nonparametric, focusing on asymptotic behavior using assumptions associated with time series asymptotics.

The second category of conditional quantile models allows for non-linearity in the quantile yet retains the asymmetric Laplace or check loss assumption. Koenker et al. (1994), Yu (2002), and Thompson et al. (2010), for example, specify the form of the conditional quantile as a spline. Other models relax both the assumptions of linearity in the conditional quantile and the parametric form of the error term. Chaudhuri et al. (1997), Honda (2004b), and Honda (2004a) fall into this category. Accessible reviews and explanations of conditional quantile regression can be found in Buchinsky (1998), Koenker and Hallock (2001), and Yu et al. (2003).

The second approach, which we call the "joint quantile model", specifies an appropriate joint model for *all* quantiles. The roots of this approach can be traced back to density regression, as in Dunson et al. (2007) and Dunson (2007). Tokdar and Kadane (2011) introduces a model for joint quantile regression with a single covariate having finite support. Quantiles associated with covariate levels between the minimum and maximum values of its support are convex combinations of the conditional CDF at each of these boundary points. Dunson and Taylor (2005) introduces an approximate likelihood method for quantile inference while Taddy and Kottas (2010) presents a non-parametric model that jointly specifies the distribution of the covariates and the response. Quantile inference is derived from the quantiles of the conditional distribution of the response, bypassing the need for quantile regression coefficients. Recent work of Reich et al. (2011) is similar in spirit to our proposed ALP. Reich et al. (2011) develops a spatial joint quantile model that incorporates spatial dependence through spatially varying regression coefficients, which are expressed as a weighted sum of Bernstein basis polynomials where the weights are constrained spatial Gaussian processes.

In general, by modeling the whole conditional outcome distribution nonparametrically, quantile regression coefficients are not obtained. However, in many applications, researchers are specifically interested in the *varying* effect of the covariates across quantiles, so a method that can quantify the relationship between the covariates and the outcome at each of the quantiles separately would be more suitable. For instance, Miranda et al. (2009) looks at the changing relationships between the children's performance in school and covariates such as lead exposure and parental education across quantiles.

Finally, a primary advantage of the second approach is that it precludes the possibility of "crossing quantiles", i.e., it precludes $q_p(Y|\mathbf{x}) > q_{p'}(Y|\mathbf{x})$ for $p < p'$, where $q_p(Y|\mathbf{x})$ is the pth conditional quantile of Y given covariates $\mathbf{x}$. This can occur in methods that estimate and infer about quantiles separately. However, below we show that, under our conditional quantile model, if $p < p'$, then $q_{p'}(Y|\mathbf{x})$ is stochastically larger than $q_p(Y|\mathbf{x})$. Disadvantages of some of these joint methods include somewhat restrictive assumptions on covariates, such as bounded support, the infeasibility of handling more than one or two covariates, and computation which is both approximate and very intensive.

3 The asymmetric Laplace distribution

We introduce the asymmetric Laplace distribution as an error distribution for conditional quantile regression models. It has probability density function

$$\begin{aligned} f_p(\epsilon_p|\mu,\tau) &= \tau p(1-p)\exp\{-(1-p)\tau|\epsilon_p-\mu|\}1[\epsilon_p<\mu]+ \\ &\quad \tau p(1-p)\exp\{-p\tau|\epsilon_p-\mu|\}1[\epsilon_p\geq\mu] \end{aligned}$$

which can be rewritten as $f_p(\epsilon_p|\mu,\tau) = \tau p(1-p)\exp\{-\tau\rho_p(\epsilon_p-\mu)\}$, where $\rho_p(x) = -x(1-p)1[x<0] + xp1[x\geq 0]$ is the *check loss* function (Koenker and Bassett (1978)). $1[\times]$ is the indicator function, which takes the value one if the argument is true and zero otherwise. For a dataset $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, finding $argmin_\mu \sum_{i=1}^n \rho_p(x_i-\mu)$ returns the pth empirical quantile. We will denote this distribution by $AL(p,\mu,\tau)$, where we will typically set $\mu = 0$ so that $\Pr(\epsilon_p < 0) = p$.

The CDF readily enables quantiles and takes the form

$$\begin{aligned} F_p(\epsilon_p|\mu,\tau) &= p\exp\{-(1-p)\tau|\epsilon_p-\mu|\}1[\epsilon_p<\mu]+ \\ &\quad 1-(1-p)\exp\{-p\tau|\epsilon_p-\mu|\}1[\epsilon_p\geq\mu]. \end{aligned}$$

Just as minimizing L_2 loss is associated with normal errors, minimizing check loss corresponds to assuming asymmetric Laplace errors.

The asymmetric Laplace distribution as the error term achieves stochastic ordering for quantiles across p. If we let $\epsilon_p \sim AL(p,0,\tau)$ and $\epsilon_{p'} \sim AL(p',0,\tau)$ for $p < p'$, then it is straightforward to show that ϵ_p is stochastically less than $\epsilon_{p'}$, i.e $\epsilon_p \preceq \epsilon_{p'}$. Hence, for fixed Y with two quantile models, $Y = \mu_p + \epsilon_p$ and $Y = \mu_{p'} + \epsilon_{p'}$, $\epsilon_p \preceq \epsilon_{p'}$, this implies that the pth quantile prediction, μ_p, is stochastically larger than than the p'th quantile prediction, $\mu_{p'}$. Inference is still done separately for each p. However, the hard constraint of the joint quantile model is replaced by a *soft* stochastic order constraint.

3.1 An alternative representation of the asymmetric Laplace distribution

Kuzobowski and Podgorski (2000) note that the following equations,

$$\epsilon_p = \sqrt{\frac{2\xi}{\tau p(1-p)}}Z + \frac{1-2p}{p(1-p)}\xi \quad (1)$$

$$Z \sim N(0,1) \quad (2)$$

$$\xi \sim Ga(1,\tau), \quad (3)$$

provide a representation of the asymmetric Laplace distribution, where $Ga(a,b)$ is the gamma distribution with mean a/b. That is, we can express ϵ_p, which is distributed as $AL(p,0,\tau)$, as a mixture of a standard normal random variable, Z, and an exponential random variable, ξ (with mean $1/\tau$). The utility of this representation is that, conditional on $\xi \sim Ga(1,\tau)$, $\epsilon_p|\xi \sim N\left(\frac{1-2p}{p(1-p)}\xi, \frac{2\xi}{\tau p(1-p)}\right)$ is normal, leaving all of the convenient properties of the normal distribution at our disposal. The equivalence in distribution of the mixture in Equations (1) - (3) and the asymmetric Laplace distribution is derived by matching the moment generating functions. In fact, marginally, $E[\epsilon_p] = \frac{1}{\tau}\frac{1-2p}{p(1-p)}$ and $Var[\epsilon_p] = \frac{1}{\tau^2}(\frac{2}{p(1-p)} + a_p)$ where $a_p = \left(\frac{1-2p}{p(1-p)}\right)^2$.

The AL distribution, as an error specification, imposes strong assumptions on the skewness as a function of p. Yu and Moyeed (2001) presents a simple simulation to show that the posterior means obtained by assuming an asymmetric Laplace error at several values of p serve as reasonable pth quantile point estimates. In appendix C, we present encouraging results from a brief simulation investigation regarding its performance with respect to nominal coverage of credible intervals.

4 The asymmetric Laplace process (ALP)

In order to incorporate spatial structure in the model, we introduce the asymmetric Laplace process. The mixture representation of the asymmetric Laplace distribution suggests a straightforward way to create a spatial process model for quantiles. By replacing the univariate standard normal Z of Section 3.1 with a mean zero, variance one Gaussian process (GP), $Z(\mathbf{s})$, we create a process model in which each location has an asymmetric Laplace marginal distribution, provided that $\xi(\mathbf{s})$ is independent of $Z(\mathbf{s})$ and is marginally exponentially distributed. Thus, our spatial quantile process model can be written as

$$\begin{aligned}\epsilon_p(\mathbf{s}) &= \sqrt{\frac{2\xi(\mathbf{s})}{\tau p(1-p)}}Z(\mathbf{s}) + \frac{1-2p}{p(1-p)}\xi(\mathbf{s}) \\ Z(\mathbf{s}) &\sim GP(\mathbf{0}, \rho_Z(\mathbf{s},\mathbf{s}';\boldsymbol{\theta})),\end{aligned}$$

where $\rho_Z(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta})$ is a valid correlation function and we address a joint specification of $\xi(\mathbf{s})$ below.[1] Because $\xi(\mathbf{s})$ is marginally exponentially distributed with rate τ, and $Z(\mathbf{s})$ is marginally a standard normal, and because $\xi(\mathbf{s})$ is conditionally independent of $Z(\mathbf{s})$, at any given location, $\mathbf{s}_0$, $\epsilon_p(\mathbf{s}_0)|\xi(\mathbf{s}_0) \sim N\left(\frac{1-2p}{p(1-p)}\xi(\mathbf{s}_0), \frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}\right)$, i.e., an $AL(p, 0, \tau)$ marginal of $\xi(\mathbf{s}_0)$. So, the marginals of this process are distributed asymmetric Laplace. We next turn to specification for $\xi(\mathbf{s})$. Regardless of this specification, the GP model for $Z(\mathbf{s})$ ensures spatial dependence for the $\epsilon_p(\mathbf{s})$, as we illuminate in the next subsection.

4.1 Modeling $\xi(\mathrm{s})$

Here, we discuss three modeling options for the $\xi(\mathbf{s})$: (i) a common choice, (ii) an iid model, and (iii) a spatial model. We present cases (i) and (iii) in order to illustrate a full range of spatial quantile process models in terms of the behavior of the correlation of the process as the extreme quantiles are approached. Because (ii) should be adequate in practice and offers computational simplicity and good behavior when fitting, we use only iid ξ's in the ensuing sections.

(i) Common $\xi(\mathbf{s})$

First consider the simplest case in which $\xi(\mathbf{s}) \sim \xi = \text{Exp}(\text{rate} = \tau)$ is common for all locations $\mathbf{s} \in \mathcal{S}$. Through a simple iterated expectation argument, we find that, if $Cor[Z(\mathbf{s}), Z(\mathbf{s}')] = \rho_Z(\mathbf{s}, \mathbf{s}')$, then

$$Cov[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] = \frac{1}{\tau^2}\left[\frac{2\rho_Z(\mathbf{s}, \mathbf{s}')}{p(1-p)} + a_p\right].$$

From Section 2.1, we immediately obtain $Cor[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] = \frac{\rho_Z(\mathbf{s},\mathbf{s}')+a_p p(1-p)/2}{1+a_p p(1-p)/2} > \rho_Z(\mathbf{s}, \mathbf{s}')$ since $a_p p(1-p) > 0$, as expected. That is, a shared shift and scale factor, ξ, results in a higher correlation for the $\epsilon_p(\mathbf{s})$ process than for the original $Z(\mathbf{s})$ process. Equality occurs only at $p = .5$. Further, since $a_p \to \infty$ as $p \to 0$ or 1, $Cor[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] \to 1$ as $p \to 0$ or 1. Increasing correlation as we move toward the extreme quantiles may be undesirable. Finally, this common ξ produces mean square continuous surface realizations for $\epsilon_p(\mathbf{s})$ if $Z(\mathbf{s})$ is mean square continuous.

(ii) iid $\xi(\mathbf{s})$

At the other extreme, consider the case in which the $\xi(\mathbf{s})$ are independent and identically distributed exponential variables. We find that, if $Cor[Z(\mathbf{s}), Z(\mathbf{s}')] = \rho_Z(\mathbf{s}, \mathbf{s}')$, the ϵ_p process covariance is $Cov[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] = \frac{1}{\tau^2}\left(\pi\frac{\rho_Z(\mathbf{s},\mathbf{s}')}{2p(1-p)}\right)$. We see that the covariance of the quantile process model is a scaled version of that of the Gaussian process. The resulting correlation is

[1] In the above, since we specify a model for each p, we introduce a $Z_p(\mathbf{s})$ and $\xi_p(\mathbf{s})$ for each p. In the sequel, we suppress the subscript p's.

$$Cor[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] = \rho_Z(\mathbf{s}, \mathbf{s}') \frac{\frac{\pi}{2} p(1-p)}{1 - 2p + 2p^2}.$$

The panels in Figure 1 show the correlation of ϵ_p over values of ρ_Z for the iid mixture case. For the extreme quantiles, even for highly correlated Z, the correlation of ϵ_p is very weak. The heat maps of Figure 2 show the correlation and covariance of the error process on a grid of the possible values of p and ρ_Z. We see that high correlation in the quantile process model only occurs for quantiles near the median combined with high correlation in the Gaussian process. We see how the original covariance changes according to the scale factor, which is a function of p. We also see that, as $p \to 1$ or 0, $Cor[\epsilon_p(\mathbf{s}), \epsilon_p(\mathbf{s}')] \to 0$. Thus, we find that introducing iid latent variables forces discontinuities in the process. In fact, since $\xi(\mathbf{s})$ is everywhere discontinuous, the induced $\epsilon_p(\mathbf{s})$ process is as well.

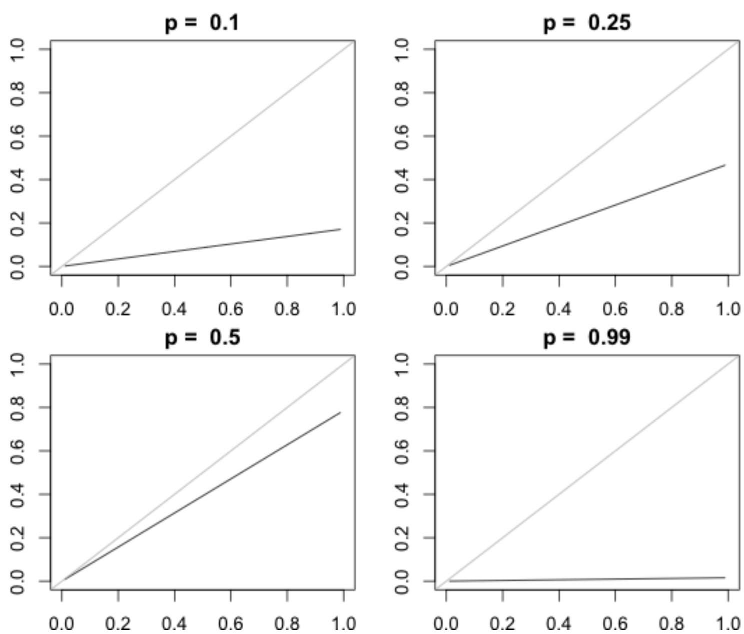

Figure 1: The correlation of $\epsilon_p(\mathbf{s})$ versus the correlation of $Z(\mathbf{s})$ for the iid ξ case, shown by the black line. The gray line, provided for reference, shows a 45 degree line.

(iii) Spatial $\xi(\mathbf{s})$

Location-specific mixing variables that are spatially structured can be introduced through a CDF (or copula) transformation. That is, let $\xi(\mathbf{s}) = -\frac{log(\Phi(V_\xi(\mathbf{s})))}{\tau} = F_\tau^{-1}(\Phi(V_\xi(\mathbf{s})))$ where $F_\tau^{-1}(\cdot)$ is the inverse CDF of the exponential distribution with rate parameter τ, Φ is the standard normal CDF, and $V_\xi(\mathbf{s})$ is a Gaussian process with

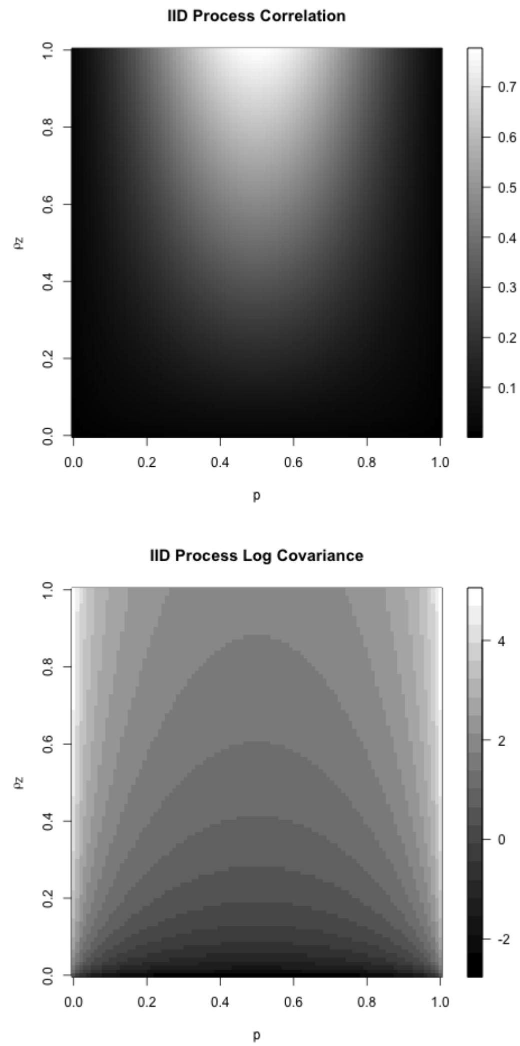

Figure 2: The correlation (top) and log covariance (bottom) of the ALP across p and ρ_Z for iid ξ.

spatial covariance function ρ_V. Monte Carlo estimates of the correlation of the resulting process at various values of ρ_ξ and ρ_Z have shown that the correlation of the $Z(\mathbf{s})$ process does not have very much effect on the resulting correlation at the extreme quantiles, while the correlation structure of $\epsilon_p(\mathbf{s})$ is more strongly dictated by ρ_Z around the median. The model with spatial ξ's is computationally demanding to fit; in the sequel we use only the case of iid ξ's.

5 The spatial quantile regression model

A general form for a spatial quantile regression is given by $Y(\mathbf{s}) - \mu_p(\mathbf{s}) = \epsilon_p(\mathbf{s})$, where $\epsilon_p(\mathbf{s})$ must satisfy the constraint that $\Pr(\epsilon_p(\mathbf{s}) \leq 0) = p$. Apart from including spatially referenced covariates in $\mu_p(\mathbf{s})$ we can incorporate spatial dependence in the $\epsilon_p(\mathbf{s})$. Hence, we arrive at our quantile regression model:

$$\begin{aligned} Y(\mathbf{s}) &= \mu_p(\mathbf{s}) + \epsilon_p(\mathbf{s}) \\ \epsilon_p(\mathbf{s}) &= \sqrt{\frac{2\xi(\mathbf{s})}{\tau p(1-p)}} Z(\mathbf{s}) + \frac{1-2p}{p(1-p)}\xi(\mathbf{s}) \end{aligned}$$

with say $\mu_p(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\beta_p$ and $\epsilon_p(\mathbf{s})$ defined as in Section 4.

Why not let $\mu_p(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_p + w_p(\mathbf{s})$ with $w_p(\mathbf{s})$ a spatial Gaussian process? If so, then $\mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_p + w_p(\mathbf{s})$ would correspond to the pth conditional quantile of $Y(\mathbf{s})$ rather than $\mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_p$, thus losing interpretability of $\boldsymbol{\beta}_p$. By embedding the spatial component within $\epsilon_p(\mathbf{s})$ in the ALP, we retain the interpretation of β_p as a global quantile regression coefficient.

5.1 Comparison to spatial mean regression

The form in which this quantile model is written invites comparison to spatial mean regression. Typically, in spatial mean regression, we write $Y(\mathbf{s}) = \mu_m(\mathbf{s}) + \epsilon_m(\mathbf{s})$ with, say, $\mu_m(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\beta_m$ and the error process as $\epsilon_m(\mathbf{s}) = w_m(\mathbf{s}) + \delta_m(\mathbf{s})$, where the m subscript denotes mean regression. In this setting, $w_m(\mathbf{s})$ is a GP realization and the $\delta_m(\mathbf{s})$ are a pure error (white noise) realization. Instead, for our quantile process we write $\epsilon_p(\mathbf{s}) = \sqrt{\frac{2\xi(\mathbf{s})}{\tau p(1-p)}} Z(\mathbf{s}) + \frac{1-2p}{p(1-p)}\xi(\mathbf{s})$ where $Z(\mathbf{s})$ is a GP realization and $\xi(\mathbf{s})$ is, say, a pure error realization. Then, the *spatial* component of $\epsilon_p(\mathbf{s})$ depends upon $\xi(\mathbf{s})$ and, in fact, is everywhere discontinuous. However, marginal spatial dependence in the $\epsilon_p(\mathbf{s})$'s is retained, according to the previous subsection. Also, in working with the ALP, we condition on and marginalize over the iid $\xi(\mathbf{s})$. For spatial mean regression, we condition on and marginalize over the $w_m(\mathbf{s})$ (see, e.g., Banerjee et al., 2004).

In the case of spatial mean regression, we may interpret the $w_m(\mathbf{s})$ as a local spatial adjustment to the mean. That is, $E[Y(\mathbf{s})|w_m(\mathbf{s})] = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_m + w_m(\mathbf{s})$ will be a better estimator in the mean squared error sense than the marginal expectation, $E[Y(\mathbf{s})] = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_m$. This argument does not extend to quantile processes or the asymmetric Laplace distribution as clarified above. However, suppose we enrich $Z(\mathbf{s})$ such that $Z(\mathbf{s}) = \sqrt{1-\alpha}w(\mathbf{s}) + \sqrt{\alpha}\delta(\mathbf{s})$ where $w(\mathbf{s})$ is a mean 0 variance 1 GP, $\delta(\mathbf{s})$ is a pure error process, and $\alpha \in [0,1]$. Evidently, the parameter α dictates what proportion of the variance of $Z(\mathbf{s})$ is due to the spatial component. Then, $Z(\mathbf{s}) \sim N(0,1)$ and we can pull apart a spatial adjustment to the marginal quantile. That is, now a spatially adjusted conditional quantile using the ALP is $q_p(Y(\mathbf{s})|w(\mathbf{s})) = \mathbf{x}^T(\mathbf{s})\beta_p + w(\mathbf{s})\sqrt{\frac{\pi}{2\tau^2 p(1-p)}}$, which is anticipated to be better in terms of check loss than the marginal quantile $q_p(Y(\mathbf{s}))$. Notice that, for each p, the $q_p(Y(\mathbf{s})|w(\mathbf{s}))$ surface will be smooth if $\mathbf{x}^T(\mathbf{s})$ and $w(\mathbf{s})$

are. In summary, to specify a quantile model that can produce spatial adjustments for more accurate local quantile regressions, we embed the spatial structure within the error process and still yield a marginal (integrating over $Z(\mathbf{s})$) quantile regression.

5.2 Model fitting

Reed and Yu (2009) develops a Gibbs sampler for non-spatial quantile regression with a model that is a special case of the ALP, the case in which $Z(\mathbf{s}) \overset{iid}{\sim} N(0,1)$. They derive full conditional distributions for each of the parameters, including the latent exponential vector of ξ_is. Similarly, we use the mixture representation of the asymmetric Laplace via the ALP, as the error term of a regression. We adopt the iid assumption for the $\xi(\mathbf{s})$ and place the following (conjugate, when possible) priors on all other parameters: $\xi(\mathbf{s}) \overset{iid}{\sim} \mathrm{Exp}(\tau)$, $\tau \sim Ga(a_\tau, b_\tau)$, and $\boldsymbol{\beta}_p \sim N_r(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$. We typically set $\boldsymbol{\mu}_\beta = \mathbf{0}$ and $\boldsymbol{\Sigma}_\beta = \sigma^2_\beta \mathbf{I}$ with σ^2_β large.

For $w(s)$, we choose the exponential correlation function with decay parameter ϕ. To achieve well behaved MCMC model fitting, we use a discrete prior for the possible values of ϕ chosen in such a way that ϕ may only take values that are within a plausible range, determined by the scale for the data locations. That is, for example, we do not allow the effective range of the process implied by a value of ϕ to go beyond the range of the data observed. The required full conditional distributions are provided in Appendix A.

5.3 The hyperparameters in the prior for τ

In practice, we have to specify the hyperparameters of the prior for τ, $Ga(a_\tau, b_\tau)$. As we have seen, $Var[\epsilon_p(\mathbf{s})] = \frac{1}{\tau^2}\left(\frac{2}{p(1-p)} + a_p\right)$, a function of both τ and p. In fact, given τ, the marginal process variance is unbounded as $p \to 0$ or 1. Hence, one may prefer to specify a prior such that $Var[\epsilon_p(\mathbf{s})]$ is the same at all quantiles and then translate this to the implied prior for τ.

Let $Var[\epsilon_p(\mathbf{s})] = \frac{1}{\kappa^2}$. Then, $\kappa = \frac{\tau}{\sqrt{\frac{2}{p(1-p)} + a_p}}$. If we specify a gamma prior on κ such that $\kappa \sim Ga(a, b_\kappa)$, by the scalability of Gamma distribution, this implies that $\tau \sim Ga\left(a, b_\tau = b_\kappa \sqrt{\frac{2}{p(1-p)} + a_p}\right)$. In this manner, we have provided a way to automatically adjust the prior on τ to imply a constant process variance across p and also to retain conjugacy for this transformed prior.

5.4 Spatial quantile interpolation

We seek to interpolate the pth conditional quantile of $Y(\mathbf{s}_0)$ at a new location $\mathbf{s}_0$. The goal here is to come up with a conditional expression for $q_p(Y(\mathbf{s}_0)|\ldots)$ such that $\Pr\left(Y(\mathbf{s}_0) \le q_p(Y(\mathbf{s}_0)|\ldots)|\mathbf{Y}\right) = p$. Focusing only on the error process model, if we were simply to condition the error at a new location, $\epsilon_p(\mathbf{s}_0)$, on $\boldsymbol{\epsilon}_p$, we would see that for

$\epsilon_p(\mathbf{s}_0) = \sqrt{\frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}} Z(\mathbf{s}_0) + \frac{1-2p}{p(1-p)}\xi(\mathbf{s}_0)$,

$$
\begin{aligned}
Z(\mathbf{s}_0)|\boldsymbol{\epsilon}_p, \xi(\mathbf{s}_0), \boldsymbol{\xi} &\sim N(m(\mathbf{s}_0), \sigma(\mathbf{s}_0)) \\
m(\mathbf{s}_0) &= \mathbf{c}^T(\mathbf{s}_0, \mathbf{s}; \theta)\mathbf{C}^{-1}(\mathbf{s}; \boldsymbol{\theta})\mathbf{Z} \\
\sigma(\mathbf{s}_0) &= 1 - \mathbf{c}^T(\mathbf{s}_0, \mathbf{s}; \theta)\mathbf{C}^{-1}(\mathbf{s}; \boldsymbol{\theta})\mathbf{c}(\mathbf{s}_0, \mathbf{s}; \boldsymbol{\theta}) \\
\mathbf{Z} &= \frac{\boldsymbol{\epsilon}_p - \frac{1-2p}{p(1-p)}\boldsymbol{\xi}}{\sqrt{\frac{2}{\tau p(1-p)}\boldsymbol{\xi}}}.
\end{aligned}
$$

Thus, $\epsilon_p(\mathbf{s}_0)|\boldsymbol{\epsilon}_p, \boldsymbol{\xi}, \xi(\mathbf{s}_0), \tau \sim N\left(\frac{1-2p}{p(1-p)}\xi(\mathbf{s}_0) + m(\mathbf{s}_0)\sqrt{\frac{2}{p(1-p)}\xi(\mathbf{s}_0)}, \frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}\sigma(\mathbf{s}_0)\right)$, which does not have zero as the pth quantile, as desired. The pth quantile of this distribution, in fact, is $m(\mathbf{s}_0)\sqrt{\frac{2}{p(1-p)}\xi(\mathbf{s}_0)}$. So, subtracting off this unwanted shift, we arrive at $\epsilon_p(\mathbf{s}_0) - m(\mathbf{s}_0)\sqrt{\frac{2}{p(1-p)}\xi(\mathbf{s}_0)}|\boldsymbol{\epsilon}_p, \boldsymbol{\xi}, \xi(\mathbf{s}_0), \tau \sim N\left(\frac{1-2p}{p(1-p)}\xi(\mathbf{s}_0), \frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}\sigma(\mathbf{s}_0)\right)$, which has a recognizable distribution. Namely, $\epsilon_p(\mathbf{s}_0) - m(\mathbf{s}_0)\sqrt{\frac{2}{p(1-p)}\xi(\mathbf{s}_0)} \sim AL(p, 0, \frac{\tau}{\sigma(\mathbf{s}_0)})$. This equation is conditional on $\xi(\mathbf{s}_0)$, which is unknown. Integrating over $\xi(\mathbf{s}_0)$, we find that $\epsilon(\mathbf{s}_0) - m(\mathbf{s}_0)\sqrt{\frac{\pi}{2p(1-p)\tau^2}} = \int \left(\epsilon(\mathbf{s}_0) - m(\mathbf{s}_0)\sqrt{\frac{2}{p(1-p)}\xi(\mathbf{s}_0)}\right) d\xi(\mathbf{s}_0) \stackrel{d}{=} AL\left(p, 0, \frac{\tau}{\sigma(\mathbf{s}_0)}\right)$.

Finally, returning to the regression setting in which $\epsilon_p(\mathbf{s}_0) = Y(\mathbf{s}_0) - \mathbf{x}^T(\mathbf{s}_0)\boldsymbol{\beta}_p$, we have that $Y(\mathbf{s}_0) - \mathbf{x}(\mathbf{s}_0)^T\boldsymbol{\beta}_p - m(\mathbf{s}_0)\sqrt{\frac{\pi}{2p(1-p)\tau^2}} \mid \tau, \boldsymbol{\beta}_p, \boldsymbol{\theta}, \mathbf{Y} \sim AL\left(p, 0, \frac{\tau}{\sigma(\mathbf{s}_0)}\right)$, which has zero as the pth quantile. So, to create the conditional spatial quantile estimate at a new location, we define $q_p(Y(\mathbf{s}_0)|\epsilon_p(\mathbf{s}_0)) = \mathbf{x}^T(\mathbf{s}_0)\boldsymbol{\beta}_p + q_p(\epsilon_p(\mathbf{s})|\boldsymbol{\epsilon}_p)$, where $q_p(\epsilon_p(\mathbf{s})|\boldsymbol{\epsilon}_p) = m(\mathbf{s}_0)\sqrt{\frac{\pi}{2p(1-p)\tau^2}}$ is a spatial quantile interpolator for the error process. By a similar argument, if we decompose $Z(\mathbf{s}_0) = w(\mathbf{s}_0) + \delta(\mathbf{s}_0)$, we can create a spatial quantile interpolator that is conditional on the spatial random effects, $\mathbf{w}$, i.e., $q_p(\epsilon_p(\mathbf{s})|\mathbf{w})) = E[w(\mathbf{s}_0) \mid \mathbf{w}]\sqrt{\frac{\pi}{2p(1-p)\tau^2}}$ replaces $q_p(\epsilon_p(\mathbf{s})|\boldsymbol{\epsilon}_p)$.

6 Example: Baton Rouge real estate data

As an illustrative example and proof of concept, we fit the ALP at several quantiles to a small dataset consisting of the log selling price of 70 homes in Baton Rouge, LA in June 1989, which can be found at www.biostat.umn.edu/ ˜brad/data2.html. As discussed in Banerjee et al. (2004), we model the quantiles of the log-selling price of 60 houses, using living area, other area, age, and number of bathrooms as explanatory variables. We leave the remaining 10 houses as a hold out set, on which we will compare the performance of the spatial to the non-spatial models. We use the pure error process specification for $\xi(\mathbf{s})$. We also partition $Z(\mathbf{s})$ into $w(\mathbf{s})$ and $\delta(\mathbf{s})$ as discussed in Section 5. More precisely, we define the covariance function as $c(\mathbf{s}_i, \mathbf{s}_j; \phi, \alpha) = (1-\alpha)\exp\{-\phi||\mathbf{s}_i - \mathbf{s}_j||\} + \alpha 1[i = j]$, i.e., an exponential covariance with parameter ϕ, a spatial variance $(1-\alpha)$, and a pure error variance α.

6.1 Results

To evaluate the performance of the spatial method, we look at the average check loss of the spatially interpolated quantile prediction from the ALP (from Section 5.4) minus the actual values of the hold out set. We compare this check loss to that produced by the Bayesian non-spatial quantile regression of Reed and Yu (2009). Losses are shown in Table 1, both in and out of sample. We see that for each of the quantiles, the average check loss of the hold out set was lower for the spatial model (the ALP) than for the non-spatial model. We also show that the check loss of the difference between the data and the spatially adjusted quantile is lower than that of the unadjusted quantile. The spatial adjustments to the marginal quantile are shown in Figure 3.

	In-Sample		Out of Sample	
p	Spatial	Non-Spatial	Spatial	Non-Spatial
0.2	0.70	0.74	0.53	0.58
0.3	0.79	0.85	0.63	0.67
0.4	0.85	0.92	0.70	0.72
0.5	0.87	0.94	0.72	0.73
0.6	0.86	0.90	0.69	0.70
0.7	0.81	0.84	0.59	0.62
0.8	0.72	0.72	0.46	0.49

Table 1: In-sample and out of sample check loss for the ALP.

Figure 4 shows the posterior mean estimates for each of the covariates across the quantiles as compared with results obtained from the quantreg R package (Koenker (2011)). In order to obtain these estimates, we used all of the 70 data points, rather than reserving some for model comparison. A surprising result is that the effect of the number of bathrooms changes sign across the quantiles. At the low quantiles, having more bathrooms has a positive effect on log selling price while at the high quantiles, more bathrooms induce lower log selling price. Although the effect of age is uniformly negative across the quantiles, its effect is not as strong at the higher quantiles– possibly accounting for an increased price for historic homes. One more point is the reversed pattern across living area and other area. It seems that at the highest quantiles, if one has a fixed amount of space to work with, it would be best to trade in other area for living area. Figure 5 shows the mean and 95% posterior credible intervals for the correlation between locations at the median inter-location distance.

7 The asymmetric Laplace predictive process

In the Bayesian setting, fitting spatial models using MCMC is computationally infeasible for large datasets because of the need to invert covariance matrices at each iteration.

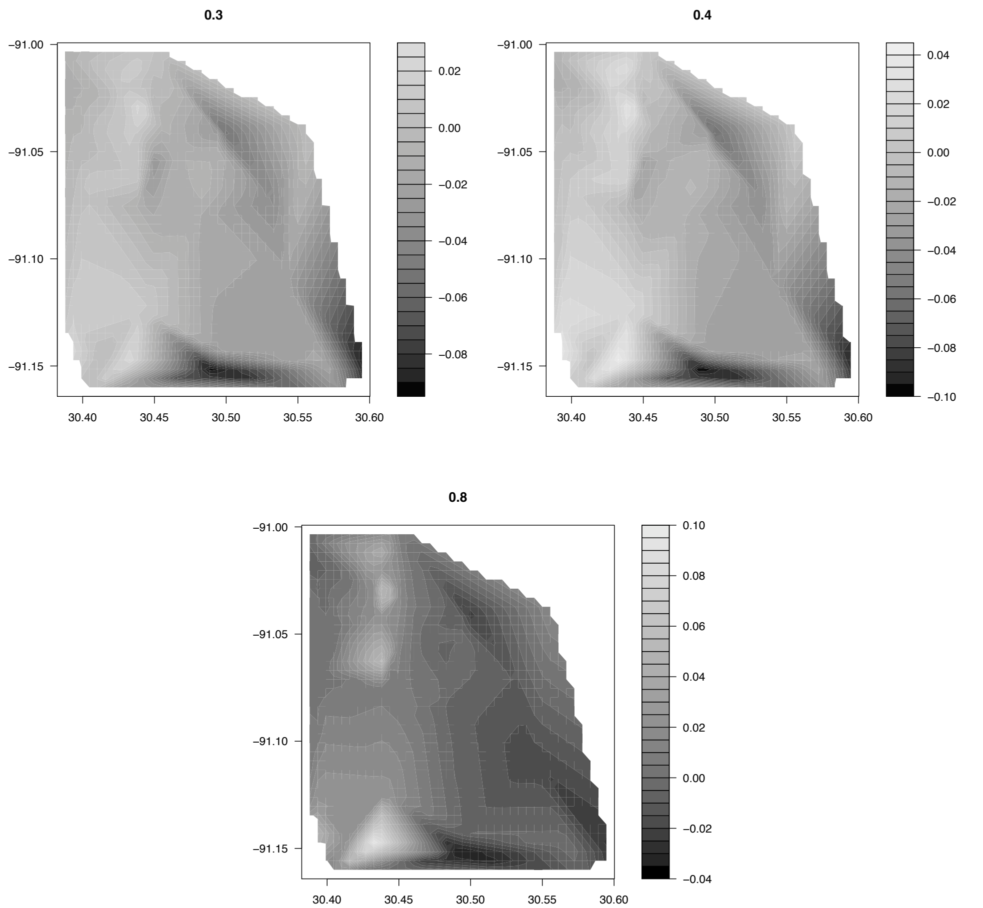

Figure 3: Contour plots of the spatial adjustment to the marginal quantile at $p = \{.3, .4, .8\}$.

7.1 The spatial predictive process

The spatial predictive process model (SPP), as introduced in Banerjee et al. (2008) and expanded upon in Finley et al. (2009), presents a reduced rank approach to accommodate the inversion of large matrices. They choose m "knots", locations $\mathbf{s}^* \in \mathcal{S}^*$, and only require the Gaussian process specification to hold for these knots, i.e. the joint distribution at these knots is the multivariate normal distribution for these knots arising from the GP. At all other locations the $w(\mathbf{s})$ are replaced by their conditional expectation given the knots.

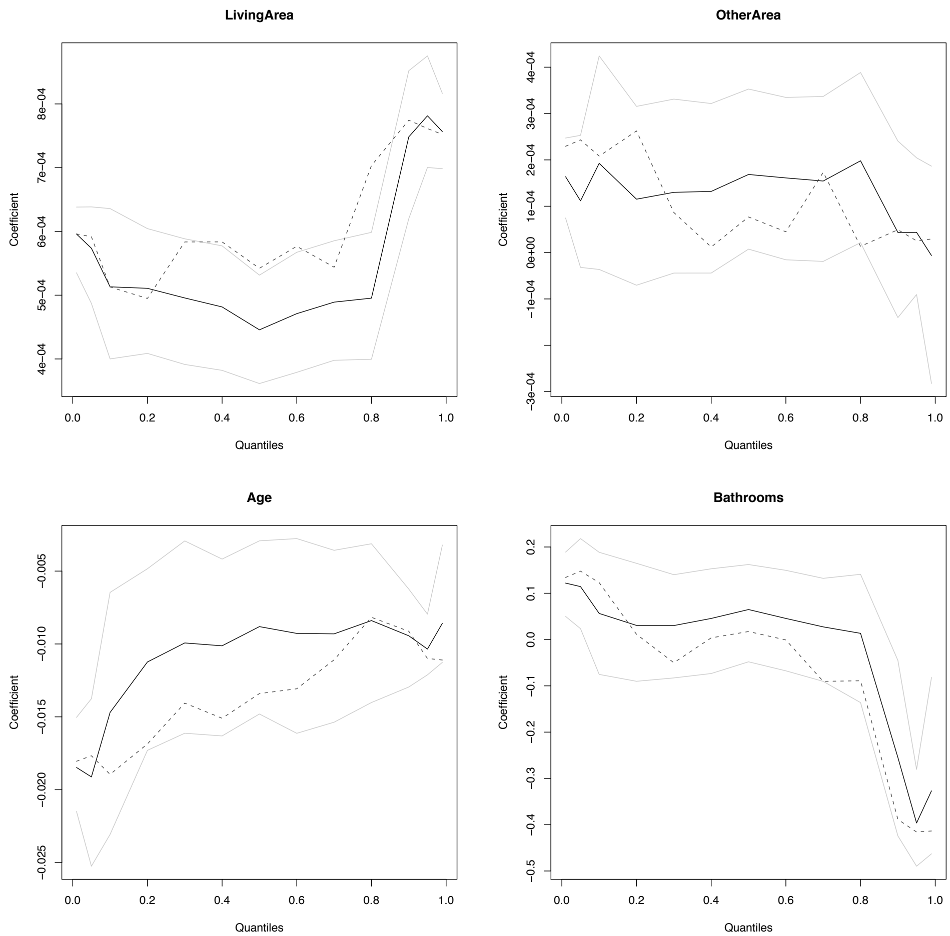

Figure 4: Posterior means for each regression coefficient across quantiles for the spatial model (in black) with 95% credible intervals (gray), and the estimate produced by the quantreg package (dashed) for comparison.

Specifically, focusing only on the error process, Banerjee et al. (2008) consider the mean regression case. Following Section 3.2, the full error term $\epsilon_m(\mathbf{s})$ is decomposed as $w_m(\mathbf{s}) + \delta_m(\mathbf{s})$, where the spatial random effects are $w_m(\mathbf{s}) \sim GP(0, \sigma_m^2 \rho(\mathbf{s}, \mathbf{s}; \boldsymbol{\theta}_\rho))$, and $\delta_m(\mathbf{s}) \overset{iid}{\sim} N(0, \tau_m^2)$ are the pure error nugget. Marginalizing over $w_m(\mathbf{s})$, $\epsilon_m(\mathbf{s}) \sim GP(0, c(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}) = \sigma_m^2 \rho(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}_\rho) + \tau_m^2 1[\mathbf{s} = \mathbf{s}'])$.

The SPP defines $\tilde{\epsilon}_m(\mathbf{s}) = \tilde{w}_m(\mathbf{s}) + \delta_m(\mathbf{s})$, with $\tilde{w}_m(\mathbf{s}) = \mathbf{c}^T(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})\mathbf{C}^{-1}\mathbf{w}^*$, where the $m \times n$ matrix $\mathbf{c}(\mathbf{s}, \mathbf{s}^*, \boldsymbol{\theta})$ contains the covariances between each $w(\mathbf{s})$ and $\mathbf{w}^*$ under the parent Gaussian process model. We see that $\tilde{w}_m(\mathbf{s})$ is simply $E[w_m(\mathbf{s})|\mathbf{w}^*]$ under

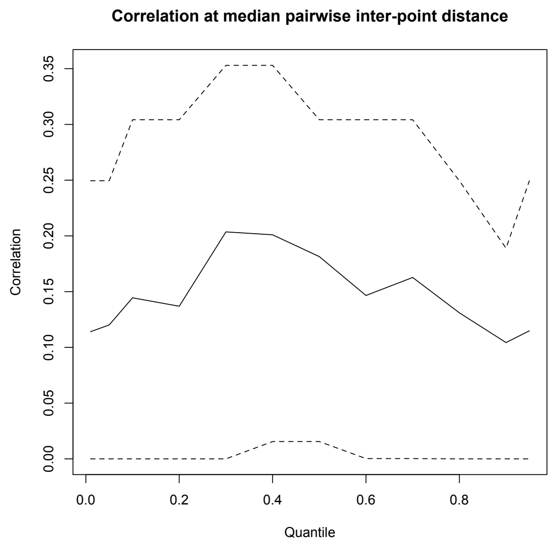

Figure 5: The mean correlation between locations at the median pairwise distance across all quantiles (solid line) and associated 95% credible intervals (dashed lines).

the multivariate normal distribution, $(\mathbf{w}^T, \mathbf{w}^{*T})^T \sim N_{n+m}(\mathbf{0}, \mathbf{C}(\{\mathbf{s}, \mathbf{s}^*\}; \boldsymbol{\theta})$. Although $\tilde{w}_m(\mathbf{s})$ is a deterministic function given $\mathbf{w}^*$, marginalizing over $\mathbf{w}^*$ returns a zero-centered normal distribution for $\epsilon_m(\mathbf{s})$ with covariance $\mathbf{c}^T(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})\mathbf{C}(\mathbf{s}^*; \boldsymbol{\theta})\mathbf{c}(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$. Also notice that $\tilde{w}_m(\mathbf{s}^*) = \mathbf{w}^*$ for any value of $\mathbf{w}^*$, so changing from the standard Gaussian process to the predictive process preserves the process at the knots.

Finley et al. (2009) present the modified predictive process, which corrects for the systematic underestimation of the variance of the $w_m(\mathbf{s})$ by the spatial predictive process. The modified predictive process fixes this problem by specifying $\ddot{w}_m(\mathbf{s}) = \tilde{w}_m(\mathbf{s}) + \tilde{\eta}_m(\mathbf{s})$, where $\tilde{\eta}_m(\mathbf{s}) \overset{iid}{\sim} N(0, \mathbf{c}(\mathbf{s}, \mathbf{s}) - \mathbf{c}^T(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})\mathbf{C}^{*-1}\mathbf{c}(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}))$. We now show how a different modification of the predictive process can enable us to extend the ALP to be able to handle large data sets.

7.2 The asymmetric Laplace predictive process

Let

$$\ddot{\epsilon}_p(\mathbf{s}) = \sqrt{\frac{2\xi(\mathbf{s})}{\tau p(1-p)}}\ddot{Z}(\mathbf{s}) + \frac{1-2p}{p(1-p)}\xi(\mathbf{s}),$$

where $\ddot{Z}(\mathbf{s}) = \ddot{w}(\mathbf{s}) + \delta(\mathbf{s})$ is the modified predictive process of above with $\ddot{w}(\mathbf{s}) = \tilde{w}(\mathbf{s}) + \tilde{\eta}(\mathbf{s})$ and $\tilde{\eta}(\mathbf{s}) \overset{iid}{\sim} N(0, \ddot{\sigma}(\mathbf{s}))$ for $\ddot{\sigma}(\mathbf{s}) = (1-\alpha)(c(\mathbf{s},\mathbf{s};\boldsymbol{\theta}) - \mathbf{c}^T(\mathbf{s},\mathbf{s}^*;\boldsymbol{\theta})\mathbf{C}^{*-1}\mathbf{c}(\mathbf{s},\mathbf{s}^*;\boldsymbol{\theta}))$ and $\delta(\mathbf{s}) \overset{iid}{\sim} N(0,\alpha)$. For the latent $\xi(\mathbf{s})$, we again assume that $\xi(\mathbf{s}) \overset{iid}{\sim} Ga(1,\tau)$. This specification retains the unit marginal variance of $\ddot{Z}(\mathbf{s})$. Marginally $\ddot{\epsilon}_p(\mathbf{s}) \sim AL(p,0,\tau)$ at every location. We refer to this modified process as the asymmetric Laplace predictive process (ALPP). Revised computation for MCMC model fitting with the ALPP is found in Appendix B. With regard to spatial quantile prediction for this predictive process, we notice that $Y(\mathbf{s}_0) - \mathbf{x}(\mathbf{s}_0)^T\boldsymbol{\beta}_p - \tilde{w}(\mathbf{s}_0)\sqrt{\frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}} \sim N\left(\frac{1-2p}{p(1-p)}\xi(\mathbf{s}_0), \frac{2\xi(\mathbf{s}_0)}{\tau p(1-p)}(\ddot{\sigma}(\mathbf{s}_0)+\alpha)\right)$, so we are back to familiar territory. We recognize the form of this distribution as the mixture that produces the asymmetric Laplace distribution with scale parameter $\frac{\tau}{\ddot{\sigma}(\mathbf{s}_0)+\alpha}$ when conditioning on $\tilde{w}(\mathbf{s}_0)$, which is a deterministic function of $w(\mathbf{s}^*)$. Integrating over $\xi(\mathbf{s}_0)$ creates asymmetric Laplace errors. Then, much like with the ALP, $Y(\mathbf{s}_0) - x(\mathbf{s}_0)^T\beta_p - \tilde{w}(\mathbf{s}_0)\sqrt{\frac{\pi}{2\tau^2 p(1-p)}} \sim AL(p,0,\ddot{\sigma}(\mathbf{s}_0)+\alpha)$, and from this we create a spatial quantile interpolator for the ALPP, i.e., $q_p(Y(\mathbf{s}_0)|\tilde{w}(\mathbf{s}_0)) = \mathbf{x}(\mathbf{s}_0)^T\boldsymbol{\beta}_p + \tilde{w}(\mathbf{s}_0)\sqrt{\frac{\pi}{2\tau^2 p(1-p)}}$.

8 Birth weight application

Quantile regression is usefully applied to birth weight because what is of interest is the effect of various risk factors on the lower quantiles of the distribution, i.e., on low birth weight. In fact, birth weight data has been a frequent source of quantile regression applications. Koenker and Hallock (2001) and Reed and Yu (2009), for example, both illustrate their quantile regression methods on birth weight data sets.

We, too, investigate the relationships between several covariates and birth weights. We look at 3,229 single births that occurred in the city of Durham, North Carolina in the year 2000, focusing on the lower quantiles of birth weight. With $n = 3229$, implementing the ALP is computationally challenging, so we turn to the ALPP. As in Banerjee et al. (2008), we randomly select 100 locations from among the 3,229 locations of the births to be the knots. Though not generally good practice, to develop our application we have omitted variables that themselves are spatially distributed, such as maternal race-group and maternal education, to encourage residual spatial structure. Variables used in this analysis include an indicator for whether the mother smoked during the pregnancy, an indicator for whether the baby was male, and one variable indicating whether this was the mother's first birth. Multiple births were excluded from this data set, as were births in which the mother used alcohol during the pregnancy. A histogram of birth weight measured in grams for the 3,229 births (not shown) reveals customary left skewness (primarily due to pre-term births).

Table 2 shows the average in-sample and out of sample check loss of the residuals averaged over the iterations of the MCMC algorithm, which are calculated using the predictive process's spatial random effects. This is in comparison to the check loss of the non-spatial model for each quantile. Estimates for the non-spatial model were obtained from the quantreg R package. We find that in-sample, we are able to achieve very

Average Check Loss				
	In-Sample		Out of Sample	
p	Spatial	QuantReg	Spatial	QuantReg
0.05	75	82	82	85
0.1	115	122	123	125
0.2	169	173	174	176
0.3	200	205	205	206
0.4	217	221	221	222
0.5	220	224	224	225
0.6	210	215	212	214
0.7	186	191	184	186
0.8	147	152	145	146
0.9	93	96	91	92
0.95	55	56	55	55

Table 2: In- and out-of-sample mean check loss for the ALPP, non-spatial quantile regression.

modest gains in terms of check loss. Out of sample check loss, which was evaluated on 300 births which were randomly chosen as a hold out set, also demonstrates very modest improvement over the non-spatial model. Perhaps, Durham is too small geographically to provide much residual spatial variation. Still, the model fitting is a useful exercise to reveal this. Moreover, our modeling provides a template for application to other geographic regions.

Figure 6 shows the posterior mean of each of the regression coefficients across quantiles. The gray lines indicate 95% credible intervals around the black line, which is the posterior mean of the coefficient at each of the quantiles. For the most part, the spatial quantile regression analysis tells the same story as the non-spatial version. We see that smoking during pregnancy has negative effects on birth weight across all quantiles. It has been observed in the past that male babies are heavier than females on average. Our analysis expands the story across quantiles. Although at all quantiles, the male babies are heavier, the shift in weight for males increases with quantile. Similarly, first born children are typically lighter than their younger siblings, which is visible in the negative quantile regression coefficients across all quantiles. However, at the highest quantiles, the advantage of being a younger sibling is much less than at the lowest quantiles.

A final bit of analysis looks at the behavior of the spatial decay parameter across quantiles. The posterior mean of ϕ is such that, at the extreme quantiles, ϕ is markedly smaller than near the median, implying a shorter range on the spatial process at the median than at the extremes. This is shown in Figure 7, which reveals the posterior mean and credible intervals of the correlation implied by ϕ for locations at the median pairwise distance in this data set.

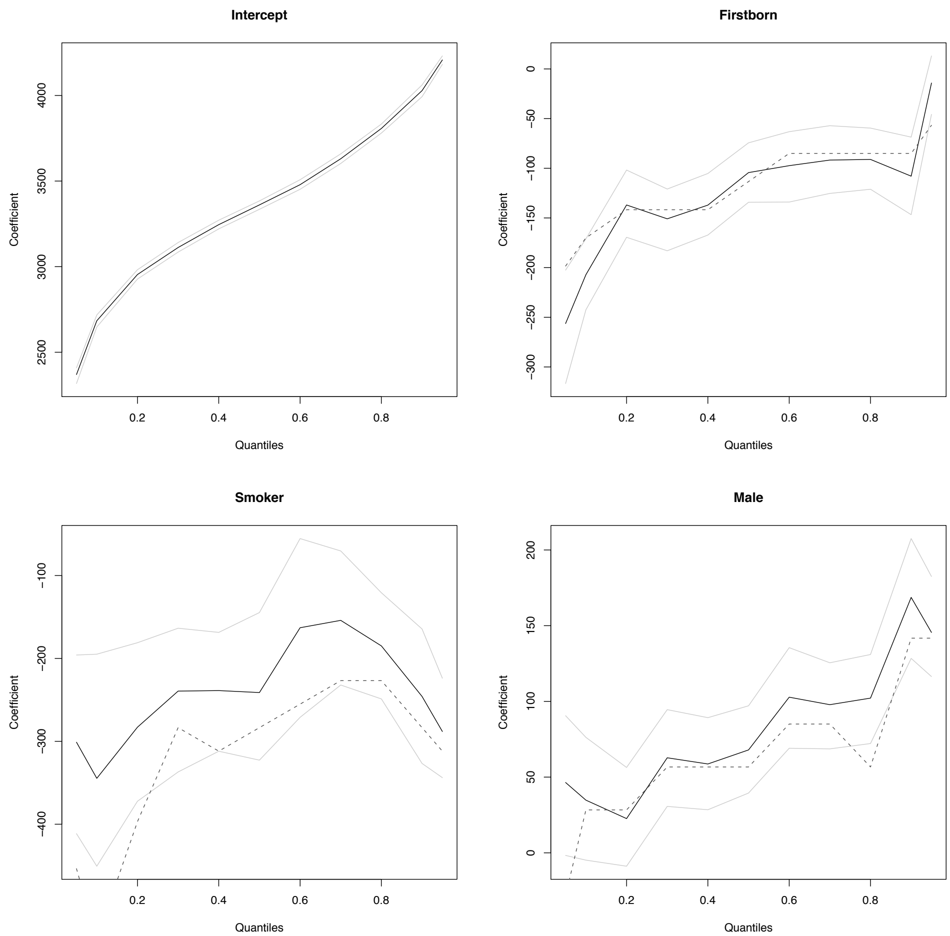

Figure 6: These plots show the mean posterior quantile regression coefficients across quantiles (black) within their 95% credible intervals (gray), as compared to the result from the quantreg package (dashed).

9 Discussion

We have introduced the asymmetric Laplace process for quantile regression with spatially dependent errors and have shown how to fit this model within a hierarchical Bayesian framework using an MCMC algorithm. We demonstrated its use on a data set of log-selling prices of homes. We have then extended the ALP to accommodate large data sets through the ALPP and illustrated use of the ALPP on a data set of birth weights given maternal covariates. Especially with the first example, we have shown the advantage, in terms of conditional quantile prediction, that can be gained by

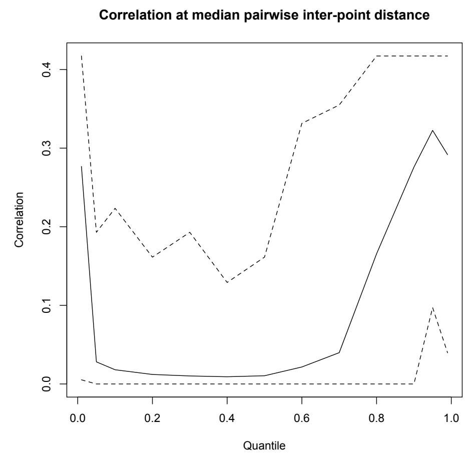

Figure 7: This plot of the correlation implied by the parameter of the exponential covariance function over the quantiles shows less correlation (smaller ϕ) at the extreme low quantiles and more correlation (larger ϕ) towards the median.

incorporating the spatial structure via the ALP or ALPP.

As for future research, following Kottas and Krnjajić (2009) one might consider introducing a different τ parameter for each side of the AL distribution in the ALP. It would also be fruitful to consider alternatives to the AL that maintain zero as the pth quantile but are more flexibly shaped and, then, ways in which these distributions could be turned into processes. A particular opportunity involves using split Gaussian or t-distributions.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. CRC Press.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets." *Journal Of The Royal Statistical Society Series B*, 70(4): 825–848.

Buchinsky, M. (1998). "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *The Journal of Human Resources*, 33(1): 88–126.

Cai, Z. and Xu, X. (2008). "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models." *Journal of the American Statistical Association*, 103(484): 1595–1607.

Chaudhuri, P., Doksum, K., and Samarov, A. (1997). "On Average Derivative Quantile Regression." *The Annals of Statistics*, 25(2): 715–744.

Dunson, D. B. (2007). "Empirical Bayes density regression." *Statistica Sinica*, 17: 481–504.

Dunson, D. B., Pillai, N., and Park, J.-H. (2007). "Bayesian density regression." *Journal Of The Royal Statistical Society Series B*, 69(2): 163–183.

Dunson, D. B. and Taylor, J. A. (2005). "Approximate Bayesian inference for quantiles." *Journal of Nonparametric Statistics*, 17(3): 3850–400.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). "Improving the performance of predictive process modeling for large datasets." *Computational Statistics & Data Analysis*, 53(8): 2873–2884.

Hallin, M., Lu, Z., and Yu, K. (2009). "Local linear spatial quantile regression." *Bernoulli*, 15(3): 659–686.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer.

Honda, T. (2004a). "Nonparametric estimation of a conditional quantile for α-mixing processes." *Annals of the Institute of Statistical Mathematics*, 121: 113–125.

— (2004b). "Quantile regression in varying coefficient models." *Journal of Statistical Planning and Inferences*, 121: 113–125.

Koenker, R. (2011). *quantreg: Quantile Regression*. R package version 4.71.
URL http://CRAN.R-project.org/package=quantreg

Koenker, R. and Bassett, G. (1978). "Regression Quantiles." *Econometrica*, 46(1): 33–50.

Koenker, R. and Hallock, K. F. (2001). "Quantile Regression." *The Journal of Economic Perspectives*, 15(4): 143–156.

Koenker, R., Ng, P., and Portnoy, S. (1994). "Quantile Smoothing Splines." *Biometrika*, 81(4): 673–680.

Kottas, A. and Krnjajić, M. (2009). "Bayesian Semiparametric Modelling in Quantile Regression." *Scandinavian Journal of Statistics*, 36: 297–319.

Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression." *Journal of Statistical Computation and Simulation*, 81(1): 1565–1578.

Kuzobowski, T. J. and Podgorski, K. (2000). "A Multivariate and Asymmetric Generalization of Laplace Distribution." *Computational Statistics*, 15(4): 531–540.

Li, Q., Xi, R., and Lin, N. (2010). "Bayesian Regularized Quantile Regression." *Bayesian Analysis*, 5(3): 533–556.

Miranda, M., Kim, D., Reiter, J., Galeano, M. O., and Maxson, P. (2009). "Environmental contributors to the achievement gap." *Neurotoxicology*, 30(6): 1019–1024.

Reed, C. and Yu, K. (2009). "An Efficient Gibbs Sampling for Bayesian quantile regression." Technical report, Department of Mathematical Sciences, Brunel University.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010). "Flexible Bayesian quantile regression for independent and clustered data." *Biostatistics*, 11(2): 337–352.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). "Bayesian Spatial Quantile Regression." *Journal of the American Statistical Association*, 106(493): 6–20.

Taddy, M. and Kottas, A. (2010). "A Bayesian Nonparametric Approach to Inference for Quantile Regression." *Journal of Business and Economic Statistics*, 28: 357–369.

Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). "Bayesian nonparametric quantile regression using splines." *Computational Statistics & Data Analysis*, 54(4): 1138 –1150.

Tokdar, S. T. and Kadane, J. B. (2011). "Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach." *Bayesian Analysis*, 6(4): 1–22.

Tsionas, E. (2003). "Bayesian Quantile Inference." *Journal of Statistical Computation and Simulation*, 73: 659–674.

Yu, K. (2002). "Quantile regression using RJMCMC algorithm." *Computational Statistics & Data Analysis*, 40: 303–315.

Yu, K., Lu, Z., and Stander, J. (2003). "Quantile Regression: Applications and Current Research Areas." *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52(3): 331–350.

Yu, K. and Moyeed, R. A. (2001). "Bayesian quantile regression." *Statistics & Probability Letters*, 54(4): 437 – 447.

Appendix A: Full conditional distributions for MCMC fitting with the ALP

We have data $\mathbf{Y}$, an $n \times 1$ vector of outcomes, and $\mathbf{X}$, an $r \times n$ matrix of covariates. Let $D_{\boldsymbol{\xi}} = \frac{2}{p(1-p)} D(\boldsymbol{\xi})$ be an appropriately scaled $n \times n$ diagonal matrix with $\frac{2}{p(1-p)}\boldsymbol{\xi}$, the $n \times 1$ vector of latent variables times the appropriate scaling factor for the quantile in question, on the diagonal. Moving to matrix notation for the points at which the data is actually observed, we write

$$\mathbf{Y} - \mathbf{X}^T \boldsymbol{\beta}_p | \boldsymbol{\xi} \sim N\left(\frac{1-2p}{p(1-p}\boldsymbol{\xi}, \tau^{-1} D_{\boldsymbol{\xi}}^{\frac{1}{2}} \mathbf{C}(\mathbf{s};\boldsymbol{\theta}) D_{\boldsymbol{\xi}}^{\frac{1}{2}}\right).$$

Then, $\boldsymbol{\beta}_p$ may be updated at each iteration of the sampler from the following multivariate normal distribution:

$$\begin{aligned}
\boldsymbol{\beta}_p|\tau,\boldsymbol{\xi},\mathbf{Y} &\sim N_r(\hat{\boldsymbol{\beta}}_p,\hat{\mathbf{S}}_\beta) \\
\hat{\mathbf{S}}_\beta &= \left(\tau\mathbf{X}^T[D_{\boldsymbol{\xi}}^{-1/2}\mathbf{C}^{-1}D_{\boldsymbol{\xi}}^{-1/2}]\mathbf{X}+\boldsymbol{\Sigma}_\beta^{-1}\right)^{-1} \\
\hat{\boldsymbol{\beta}}_p &= \hat{\mathbf{S}}_\beta\left(\tau\mathbf{X}^T[D_{\boldsymbol{\xi}}^{-1/2}\mathbf{C}^{-1}D_{\boldsymbol{\xi}}^{-1/2}]\left(\mathbf{Y}-\frac{1-2p}{p(1-p)}\boldsymbol{\xi}\right)+\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta\right).
\end{aligned}$$

For improved mixing,Reed and Yu (2009) finds a marginal distribution for τ that involves the check loss function and integrates out $\boldsymbol{\xi}$. Due to the more complicated covariance structure in this model, this same trick is not possible. However, the full conditional distribution for τ is:

$$\begin{aligned}
\tau|\boldsymbol{\xi},\boldsymbol{\beta}_p,y &\sim Ga\left(a_\tau+\frac{n}{2}+n, b_\tau+\frac{1}{2}\mathbf{u}^T[D_{\boldsymbol{\xi}}^{-1/2}\mathbf{C}^{-1}D_{\boldsymbol{\xi}}^{-1/2}]\mathbf{u}+\sum_{i=1}^{n}\xi_i\right) \\
\mathbf{u} &= \left(\mathbf{Y}-\frac{1-2p}{p(1-p)}\boldsymbol{\xi}-\mathbf{X}^T\boldsymbol{\beta}_p\right).
\end{aligned}$$

In the non-spatial case, Reed and Yu (2009) is able to derive an inverse Gaussian full conditional distribution for each element of $\boldsymbol{\xi}$. In our case, the complicated covariance matrix makes this impossible. Thus, we must sample each ξ_i using a Metropolis-Hastings step at each iteration of the algorithm from:

$$\xi_i \propto \xi_i^{1/2}\exp\{-\frac{1}{2}\mathbf{u}^T D_{\boldsymbol{\xi}}^{-\frac{1}{2}}\mathbf{C}^{-1}D_{\boldsymbol{\xi}}^{-\frac{1}{2}}\mathbf{u}+\tau\xi_i\}.$$

We have found that if we propose each ξ_i from an exponential distribution with rate τ, as is specified in the hierarchy, the pieces involved in the Metropolis-Hastings algorithm simplify nicely and the acceptance rate is generally within the range of values that is considered appropriate.

The parameters of the covariance function are updated using Metropolis-Hastings. All that is needed is that $\mathbf{Z}\sim N(\mathbf{0},\mathbf{C}(\mathbf{s};\boldsymbol{\theta}))$ and the prior, $\pi(\boldsymbol{\theta})$. $\mathbf{Z}$ is easily obtained as a function of $\boldsymbol{\beta}_p$, $\boldsymbol{\xi}$, and τ as in Section 5.4, and $\boldsymbol{\theta}$ may include α if ρ_Z has a nugget. Because α is restricted to be between zero and one, a suitable prior for it is the uniform distribution.

Appendix B: Full conditional distributions for MCMC fitting with the ALPP

Like in the ALP, by replacing $Z(\mathbf{s})$ with $\ddot{Z}(\mathbf{s})$, we lose conjugacy for $\boldsymbol{\xi}$. For a particular data set, at location $\mathbf{s}_i$, the full conditional distribution of ξ_i is

$$\begin{aligned}\xi_i|\tilde{w}_i, \tau, \boldsymbol{\beta}_p \quad &\propto \quad \xi_i^{-1/2} \exp\left\{ -\frac{1}{2} \frac{\left(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_p - \sqrt{\frac{2\xi_i}{\tau p(1-p)}} \tilde{w}_i - \frac{(1-2p)\xi_i}{p(1-p)} \right)^2}{\frac{2\xi_i}{\tau}((1-\alpha)\ddot{\sigma}_i + \alpha)} \right\} \\ &\quad \times \exp\left\{ -\xi_i \tau \right\}.\end{aligned}$$

We attain conditional independence for each of the ξ_i by conditioning on $\tilde{\mathbf{w}}$. The parameters having closed form full conditionals are the following:

$$\begin{aligned}\boldsymbol{\beta}_p &\sim N_r(A^{-1}a, A^{-1}) \\ A &= \mathbf{X}D\left(\frac{\tau p(1-p)}{2\boldsymbol{\xi}(\alpha + \ddot{\boldsymbol{\sigma}})} \right) \mathbf{X}^T + \boldsymbol{\Sigma}_\beta \\ a &= \mathbf{X}D\left(\frac{\tau p(1-p)}{2\boldsymbol{\xi}(\alpha + \ddot{\boldsymbol{\sigma}})} \right) \left(\mathbf{Y} - \mathbf{X}^T \boldsymbol{\beta} - \frac{1-2p}{p(1-p)} \boldsymbol{\xi} \right)\end{aligned}$$

and

$$\begin{aligned}\mathbf{w}^* &\sim N_m(Z^{-1}z, Z^{-1}) \\ Z &= \mathbf{C}^{*-1} \mathbf{c}(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) D^{-1}(\ddot{\boldsymbol{\sigma}} + \alpha) \mathbf{c}^T(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) \mathbf{C}^{*-1} + (1-\alpha)\mathbf{C}^{*-1} \\ z &= \mathbf{C}^{*-1} \mathbf{c}(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) D^{-1}(\ddot{\boldsymbol{\sigma}} + \alpha) \left(\mathbf{Y} - \mathbf{X}^T \beta - \frac{1-2p}{p(1-p)} \boldsymbol{\xi} \right) \sqrt{\frac{\tau p(1-p)}{2\boldsymbol{\xi}}}.\end{aligned}$$

The rest of the parameters, $\boldsymbol{\theta}$, τ, $\boldsymbol{\xi}$ are updated with a Metropolis-Hastings step. Proposing ξ_i from its prior, $\text{Exp}(\tau)$, is, again, a tuning-free solution to sampling from its distribution that has worked well in practice. We may also bypass sampling $\mathbf{w}^*$ by employing the computationally friendly Sherman-Morrison-Woodbury formula (Harville (1997))for matrix inversions to integrate over the $\mathbf{w}^*$.

Appendix C: The AL as an error distribution

We briefly present a simulation example to demonstrate the adequacy of the asymmetric Laplace distribution as an error distribution. Again, Yu and Moyeed (2001) show that the posterior means obtained by assuming an asymmetric Laplace error at several values of p serve as reasonable pth quantile point estimates. They do not, however, demonstrate that the associated credible intervals achieve nominal coverage, which may be of concern at p for which the $AL(p, 0, \tau)$ is very different from the true distribution. We simulated 50 standard normal random variables and fit an intercept only quantile regression, thus the quantile estimates are $\mu_p = \Phi^{-1}(p)$. We repeated this experiment 100 times, using the prior specified in Section 5.3 with $a = b_\kappa = 1$. Table 3 shows the results of this simulation. We find that the 95% and 90% credible intervals for μ_p contain the true value in approximately 95% and 90% of the simulations, respectively.

p	True Value	Posterior Mean	95% CI	95% CI Prop	90% CI Prop
0.1	-1.28	-1.38	[-1.74 , -1.07]	0.94	0.91
0.25	-0.67	-0.73	[-1.11 , -0.38]	0.91	0.88
0.5	0	0.16	[-0.18 , 0.46]	0.96	0.89

Table 3: The true value of μ_p, the posterior mean from an illustrative simulation, the associated 95% credible interval, and the proportion of the simulations for which the true value of μ_p fell within the 95% and 90% credible intervals, obtained from a Gibbs sampler that assumes an asymmetric Laplace error.

Acknowledgments

The authors thank Marie Lynn Miranda for valuable discussions and the use of the North Carolina birth weight dataset. They thank Merlise Clyde and Sayan Mukherjee, as well, for helpful conversations. This work was supported in part by NSF CMG 0934595 and USEPA-RD-83329301-0.