Bayesian Analysis (2012) **7**, Number 2, pp. 263–270

Comment on Article by Lum and Gelfand

Nan Lin* and Chao Chang†

We congratulate Lum and Gelfand on their interesting and important contribution to Bayesian spatial quantile regression (BSQR). The article proposed a conditional quantile process model that incorporates spatial dependence through the asymmetric Laplace process (ALP). Based on a decomposition of the asymmetric Laplace distribution, a Bayesian hierarchical model is fitted with Markov Chain Monte Carlo (MCMC). In the sequel, we will call the method of Lum and Gelfand ALP-BSQR.

In the past decade, quantile regression literature has grown very rapidly, and more recently, it has started to attract a lot of attentions among Bayesians. One major motivation for using quantile regression by frequentists is its semi-parametric nature, i.e., there is no need to fully specify a parametric form for the error distribution. And as a special case, median regression has long been used as a robust alternative to mean regression. Among the recent development for Bayesian quantile regression, some semi- and non-parametric approaches have been proposed, such as the Bayesian empirical likelihood method in Yang and He (2012) and the Dirichlet process mixture model in Kottas and Krnjajic (2009). However, the computational complexity of these methods is usually very high. And most Bayesian quantile regression methods, including the ALP-BSQR proposed by Lum and Gelfand (2012), rely on the asymmetric Laplace distribution and are fully parametric. These methods actually utilize mean regression models and assume an asymmetric Laplace distribution on the error term. Then maximizing the posterior density becomes equivalent to optimizing the check loss function under the frequentist context. Hence the problem essentially becomes finding the posterior mode for a mean regression model with asymmetric Laplace errors. As a result, this type of Bayesian quantile regression loses the 'semi-parametric' feature and it raises the concern of whether the Bayesian method still works for a large class of distributions as in the frequentist case. In some discussions for linear quantile regression and related models, the fully parametric Bayesian approach has shown good performance on data generated from other error distributions (Ji et al. 2012; Li et al. 2010).

An interesting question is then: Is the ALP-BSQR method flexible enough to maintain some 'semi-parametric' flavor as in the frequentist quantile regression? We will investigate the performance of ALP-BSQR on data simulated from different spatial models. And as a benchmark for comparison, we can also fit spatial mean regression to these data and obtain quantiles from the predictive distribution. In the following, we will conduct simulation studies to compare the performance of the Bayesian quantile regression model and the standard Bayesian inference for the mean regression model. The criterion of our comparison is the accuracy of quantile prediction of the two methods.

*Department of Mathematics, Washington University in St. Louis, St. Louis, MO, nlin@wustl.edu
†Department of Mathematics, Washington University in St. Louis, St. Louis, MO, cccc1987@math.wustl.edu

 DOI:10.1214/12-BA708B

1 Simulation Setup

We will use three different spatial models for data simulation, including two Gaussian process models and a Gaussian-log-Gaussian (GLG) model (Palacios and Steel 2006). The covariates are coordinates of points $s_i, i = 1, 2, \ldots, n = 100$ uniformly sampled on the unit square $[0, 1] \times [0, 1]$.

The Gaussian process model is given by

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T \boldsymbol{\beta} + \sigma w(\mathbf{s}) + \tau \epsilon(\mathbf{s}), \tag{1}$$

where $w(\mathbf{s})$ is a Gaussian process with mean 0 and unitary variance and the correlation function being exponential, i.e. $C(\mathbf{s_i}, \mathbf{s_j}) = \exp(-\theta \|\mathbf{s_i} - \mathbf{s_j}\|)$, and $\epsilon(\mathbf{s})$ is an uncorrelated Gaussian process with mean 0 and unitary variance. Note that $\mathbf{X}(\mathbf{s})$ is a $3 \times n$ matrix with the first row being all 1's, and the next two rows containing the coordinates of $\{s_i\}$. In Model (I), we set $\sigma^2 = 0.8$, $\tau^2 = 0.2$ and $\theta = 12$. In Model (II), we increase the variation and set $\sigma^2 = 4$, $\tau^2 = 2$ and $\theta = 12$. The Gaussian process component specified in Models (I) and (II) is consistent with that in the ALP-BSQR.

Model (III) is the following GLG model,

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T \boldsymbol{\beta} + \sigma \frac{w(\mathbf{s})}{\sqrt{\lambda(\mathbf{s})}} + \tau \epsilon(\mathbf{s}) \tag{2}$$

with $w(\mathbf{s})$, $\epsilon(\mathbf{s})$, σ^2 and τ^2 the same as in Model (I). The extra mixing process $\lambda(\mathbf{s})$ is independent of $w(\mathbf{s})$ and has the distribution

$$(\ln(\lambda(\mathbf{s_1})), \ldots, \ln(\lambda(\mathbf{s_n})))^{\mathbf{T}} \sim \mathbf{N_n}(-\frac{\nu}{\mathbf{2}}\mathbf{1}, \nu \mathbf{C}), \tag{3}$$

where $\mathbf{1}$ is a column vector of 1's, $\mathbf{C}$ is set as the same correlation matrix of the Gaussian process $w(\mathbf{s})$, i.e. exponential, and the hyperparameter ν controls the variance of $\lambda(\mathbf{s_i})$. From the above lognormal distribution in (3), it is easy to see that $E[\lambda(\mathbf{s_i})] = \mathbf{1}$ and $var[\lambda(\mathbf{s_i})] = \exp(\nu) - \mathbf{1}$ for all i. Thus when ν is close to 0, $\lambda(\mathbf{s_i})$'s have mean close to 1 with very small variance. When ν increases, the $\lambda(\mathbf{s_i})$'s become more spread out with mode shifting toward 0. This GLG model provides non-Gaussian tail behaviors with heavy tail for large ν. We set $\nu = 3$.

We simulate one data set from each of the three models. In all three models, we set $\boldsymbol{\beta} = (1, 1, 1)^T$ and use the same set of locations $\{s_i\}$. Our goal is to examine the impact of the variation in the data and violation of the Gaussian assumption.

2 Results

For each of the three simulated data sets, we fit the standard Bayesian spatial mean regression model (BSMR) and the ALP-BSQR model at five different quantiles $p = 0.1, 0.3, 0.5, 0.7, 0.9$. The BSMR assumes a model as specified in (1) and can be estimated using Gibbs sampling. We follow the prior specification in Palacios and Steel (2006) for

the unknown parameters $\boldsymbol{\beta}$, σ^2, $\rho^2 = \tau^2/\sigma^2$ and θ. The prior of $\boldsymbol{\beta}$ is $N(\mathbf{0}, \mathbf{10^4 I})$, the prior of σ^2 is the inverse-Gamma distribution with both parameters being 10^{-6}, the prior of ρ^2 is a generalized inverse-Gaussian with parameters 0.87 and 3, and the prior of θ is exponential with mean $0.3/m$, where m is the median pairwise distance among all $\{s_i\}$. The Gibbs sampler iteratively samples from the full conditional distributions of $\boldsymbol{\beta}$, σ^2, ρ^2 and θ. Sampling from the full conditional distributions of $\boldsymbol{\beta}$ and σ^2 is straightforward, while sampling ρ and θ is done using the Metroplis-Hastings algorithm with a log-normal proposal distribution. For both BSMR and ALP-BSQR, we ran 100,000 MCMC iterations with 90,000 burn-in steps.

For all three data simulation models, we can obtain the theoretical values of the quantiles by simulation. Therefore, we can evaluate the prediction accuracy of the two methods using the sum of squared errors $SSE(p) = \sum_{i=1}^{100} [q(p, s_i) - \hat{q}(p, s_i)]^2$, where $q(p, s_i)$ and $\hat{q}(p, s_i)$ denote the theoretical pth quantile and the estimated pth quantile of the response $Y(s_i)$, $i = 1, \ldots, 100$, respectively. The result is summarized in the following table.

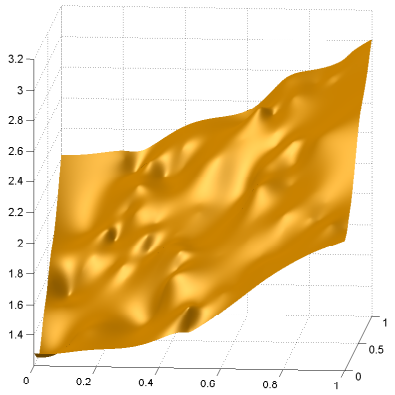
Table 1: SSE of BSMR and ALP-BSQR for three simulated data sets

Data model	Fitting method	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
I	BSMR	0.8991	0.8858	1.0923	1.3857	2.0548
I	ALP-BSQR	4.2863	4.2724	7.4521	6.1980	7.5519
II	BSMR	8.5793	7.9370	8.3825	9.9511	11.5816
II	ALP-BSQR	24.9052	14.9796	45.3208	37.0971	39.6005
III	BSMR	1944.836	347.3453	28.0514	350.9360	1945.138
III	ALP-BSQR	2378.378	497.9515	5.2371	340.1085	1972.378

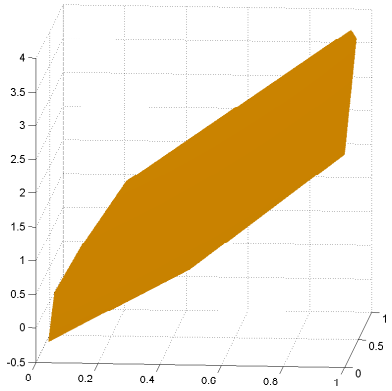
In Table 1, firstly, we notice that the BSMR outperforms the Bayesian quantile regression in most cases. For data generated from Models (I) and (II), this result is largely as expected as the BSMR method fits the correct model. But the large discrepancy in the SSEs for the two methods raises some concerns on the accuracy of the ALP-BSQR approach. While the error variance increases from Model (I) to Model (II), the SSE of both methods get larger as well, but the ALP-BSQR is shown to have a bigger increase in the prediction error. Figures 1 and 2 plot the quantile surfaces for $p = 0.5$ and 0.1 for dataset (II). They show that the restrictive linear quantile relationship in ALP-BSQR is not as flexible as the BSMR in fitting the quantile surface.

Secondly, when the data are from a non-Gaussian process, Model (III), both methods produce inaccurate predictions of the quantiles. For illustration, we plotted the quantile surfaces for $p = 0.5$ and 0.1 in Figures 3 and 4. While the large SSE of the BSMR method is as expected due to violation of the normality assumption, we would have hoped the ALP-BSQR, as a quantile regression method, would display some robustness to the non-normality. However the result shows that the ALP-BSQR is also very sensitive to non-normality, and actually more affected than BSMR at the extreme quantiles. But it is also worth noting that the ALP-BSQR method performed better at the median. This

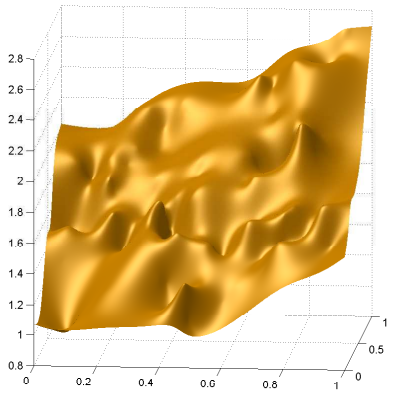
phenomenon deserves further investigation on the property of the ALP-BSQR model.



(a) The theoretical quantile surface



(b) The estimated quantile surface given by ALP-BSQR



(c) The estimated quantile surface given by the BSMR

Figure 1: The quantile surface for data set (II) at $p = 0.5$

3 Concluding Remarks

Through some simulation studies, we found that the restrictive linear quantile surface assumption in the ALP-BSQR model can be easily violated, and hence lead to inaccurate prediction of the quantiles, especially in the heavy tail situation. Results suggest that the ALP-BSQR model may not serve as a flexible model as expected for frequentist quantile regression. In addition, the introduction of the asymmetric Laplace process, though it allows one to carry out quantile regression in the Bayesian way, might be too restrictive and can not adequately capture the shape of the quantile surface of many

spatial models.

A more fundamental issue is to think about whether it is worth using Bayesian quantile regression given the fact that the Bayesian inference for a mean regression model is able to provide inferences for quantiles using the predictive distribution. In general, the mean regression models require less complicated fitting algorithms than the quantile regression models. Unlike frequentist inference, Bayesian analysis can provide the predictive distribution. So, if one has a mean regression model with a very flexible error structure, do we really need quantile regression for Bayesian inference?

References

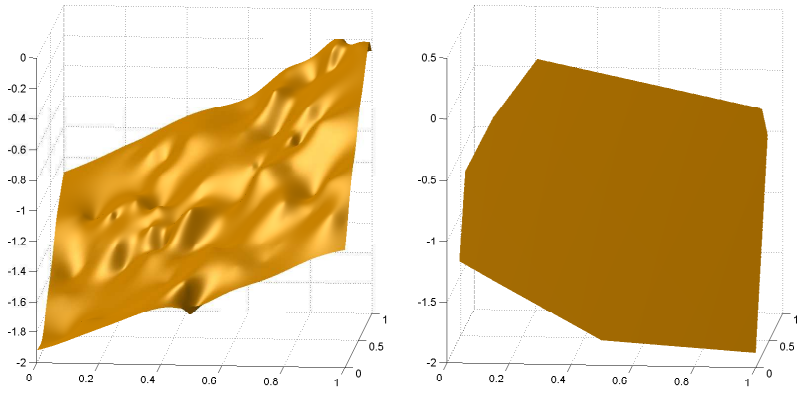
Ji, Y., Lin, N., and Zhang, B. (2012). "Model selection in binary and tobit quantile regression using the Gibbs sampler." *Computational Statistics and Data Analysis*, 56: 827–839. 263

Kottas, A. and Krnjajic, M. (2009). "Bayesian Semiparametric Modelling in Quantile Regression." *Scandinavian Journal of Statistics*, 36: 297–319. 263

Li, Q., Xi, R., and Lin, N. (2010). "Bayesian regularized quantile regression." *Bayesian Analysis*, 5: 533–556. 263

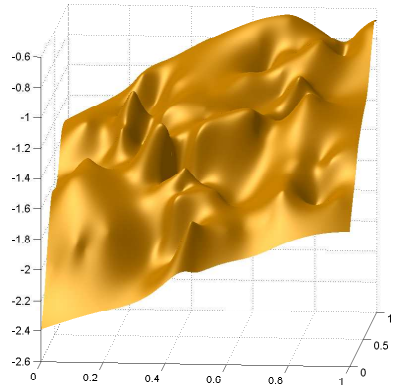
Palacios, M. B. and Steel, M. F. J. (2006). "Non-Gaussian Bayesian Geostatistical Modeling." *Journal of the American Statistical Association*, 101: 604–618. 264

Yang, Y. and He, X. (2012). "Bayesian Empirical Likelihood for Quantile Regression." *under minor revisions for Annals of Statistics*. 263



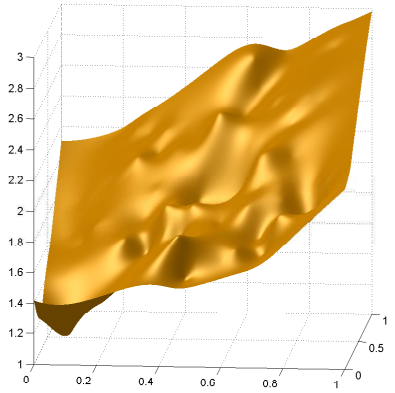
(a) The theoretical quantile surface

(b) The estimated quantile surface given by ALP-BSQR

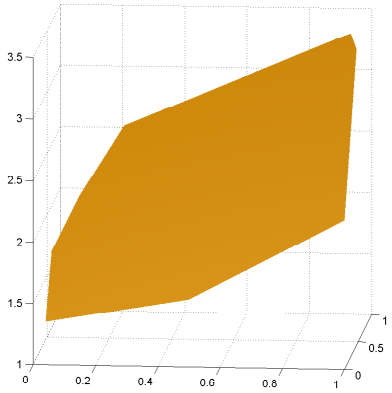


(c) The estimated quantile surface given by the BSMR

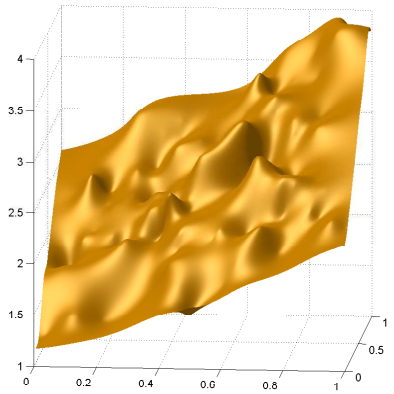
Figure 2: The quantile surface for data set (II) at $p = 0.1$



(a) The theoretical quantile surface

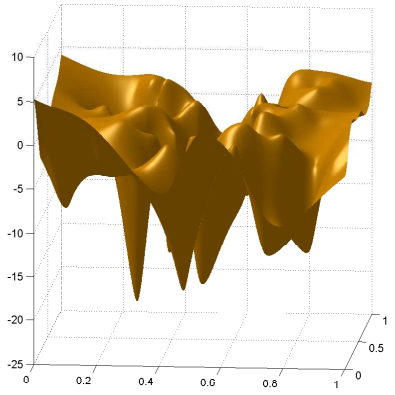


(b) The estimated quantile surface given by ALP-BSQR

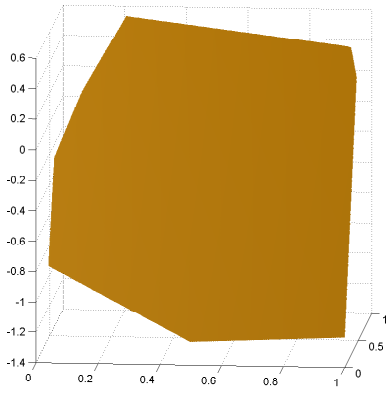


(c) The estimated quantile surface given by the BSMR

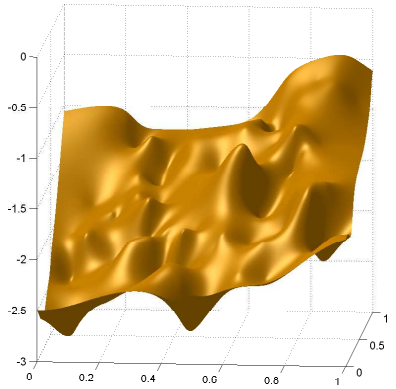
Figure 3: The quantile surface for data set (III) at $p = 0.5$



(a) The theoretical quantile surface



(b) The estimated quantile surface given by ALP-BSQR



(c) The estimated quantile surface given by the BSMR

Figure 4: The quantile surface for data set (III) at $p = 0.1$