Statistica Sinica **25** (2015), 261-274
doi:http://dx.doi.org/10.5705/ss.2013.065w

QUANTILE REGRESSION FOR SPATIALLY CORRELATED DATA: AN EMPIRICAL LIKELIHOOD APPROACH

Yunwen Yang and Xuming He

Drexel University and University of Michigan

Abstract: Quantile regression is a useful approach to modeling various aspects of conditional distributions. The Bayesian approach provides a natural framework for incorporating spatial correlation in a quantile regression model. This paper considers Bayesian spatial quantile regression using empirical likelihood as a working likelihood. The proposed approach inherits the merits of quantile regression in the sense that we can work with linear conditional quantile functions without having to assume a parametric form of the conditional distributions, and we allow each covariate to have differential impacts on different parts of the conditional distributions. Put into a Bayesian framework, this approach can incorporate spatial priors to smooth the conditional quantile functions across locations and across quantiles. We demonstrate both theoretically and empirically how the proposed approach can take advantage of spatial correlation to improve efficiency over the usual quantile regression estimators. An application to the statistical downscaling of daily precipitation in the Chicago area is given to illustrate the merit of our approach.

Key words and phrases: Bayesian empirical likelihood approach, informative priors, nonparametric spatial regression, spatial data.

1. Introduction

As a complement to least squares regression, quantile regression is a useful approach for the modeling and inference of conditional quantile functions. By specifying the τth conditional quantile functions of Y given $\boldsymbol{X}$ as

$$Q_\tau(Y|\boldsymbol{X}) = \boldsymbol{X}^\top \boldsymbol{\beta}(\tau), \text{ for } \tau \in (0,1),$$

this τ-specific quantile regression model allows the association between the response Y and the covariate $\boldsymbol{X}$ to vary across quantiles. There is an extensive literature on estimation and inference for various quantile regression models; see Koenker (2005) for a comprehensive review. In recent years, quantile regression has been used in a spatial context, as it allows the effects of covariates on the response to vary across quantiles as well as spatially. Koenker and Mizera (2004) considered estimation of spatial quantiles in a triogram model; Hallin, Lu, and Yu (2009) proposed a local linear spatial quantile regression approach. Refer to McMillen (2012) for more examples on spatial quantile models.

The Bayesian approach is a natural way to incorporate spatial correlation in quantile regression. Several forms of Bayesian quantile regression models have been proposed; see Kottas and Gelfand (2001), Yu and Moyeed (2001), Kottas and Krnjajić (2009), Lancaster and Jun (2010), and Yang and He (2012) for independent data, and Geraci and Bottai (2006), Reich, Bondell, and Wang (2008), and Kim and Yang (2011) for clustered data. Recently, Reich, Fuentes, and Dunson (2011) and Reich (2012) proposed model-based approaches for Bayesian spatial quantile regression by specifying a semiparamteric model for the entire quantile process. Both approaches work with linear quantile functions at all quantile levels. In this article, we extend the Bayesian empirical likelihood approach (BEL) of Yang and He (2012) to spatial data, employing an informative parametric prior to account for the spatial correlation. Statistical inference based on the empirical likelihood is known to enjoy good asymptotic properties (see Owen (2001), and Chen and Keilegom (2009) for a comprehensive review). The use of empirical likelihood makes our proposed Bayesian spatial quantile regression approach appealing in several ways. It allows joint estimation of multiple quantiles that will result in efficiency gains by borrowing strength across quantiles. It allows us to examine the impact of informative priors accounting for the spatial correlation on posterior inference in an asymptotic framework. The asymptotic framework based on shrinking priors explains the efficiency gain from the proposed method by incorporating informative priors on the spatial correlation and commonality across quantiles, as observed in empirical studies. It can model selected quantile levels of the conditional distributions without making global assumptions on the conditional distributions, such as linear quantile functions at all levels.

Variants of the empirical likelihood have been used in the Bayesian quantile regression setting by a number of authors, including Lancaster and Jun (2010), Wang and Zhu (2011), and Kim and Yang (2011), but we focus on a class of problems with spatially correlated data. We are particularly interested in using an informative prior to regularize the quantile coefficient estimates across space and across quantile levels. Because quantile regression estimates are highly variable in data-sparse areas, typically the tails, the proposed use of an informative prior in the BEL approach aims to borrow strength spatially and across quantiles, resulting in potentially substantial efficiency gains.

The empirical likelihood is not meant to represent the true likelihood, so the validity of the resultant posterior does not follow from Bayes formula, see Monahan and Boos (1992), Lazar (2003), Fang and Mukerjee (2006), and Chang and Mukerjee (2008) for more detailed discussion; no working likelihood ensures posterior validity at finite samples. In this paper, we provide the asymptotic distribution of the posterior from the BEL approach for spatial quantile regression, which enables us to study the asymptotic validity of the posterior inference.

The empirical likelihood approach is natural for handling several quantile levels together. This by itself is not a unique feature of our work, some authors have considered Bayesian nonparametric spatial regression by estimating all the quantiles simultaneously, see Wood, Jiang, and Tanner (2002), Gelfand, Kottas, and MacEachern (2005), Griffin and Steel (2006), and Dunson and Park (2008). Those methods estimate the entire conditional distributions of the response variable given the location and covariates, while the proposed BEL approach uses linear models on just a small number of quantiles of interest. With linear quantile functions, we can avoid the curse of dimensionality with high dimensional covariates and interpret the covariate effects more easily. A novel part of our work is its ability to employ informative priors to explore commonality across space and across quantile levels simultaneously for efficiency gains.

The rest of the paper is organized as follows. In Section 2 we introduce the proposed BEL approach for spatial quantile regression, and discuss the model assumptions and the formulation of informative priors. The asymptotic properties of the BEL posteriors are provided in Section 3. The theoretical framework of shrinking priors enables us to understand the efficiency of the BEL approach by joint modeling over space and quantile levels. The asymptotic validity of posterior inference is also discussed here. Section 4 demonstrates the finite sample performance of the BEL approach through a Monte Carlo simulation study with a focus on efficiency gains from informative priors. In Section 5, we apply the proposed BEL approach as a useful statistical downscaling method for the projection of high quantiles of precipitation from large scale climate model outputs to seven stations in the Chicago area. Some concluding remarks are given in Section 6. Technical details are provided in the supplementary file.

2. Bayesian Empirical Likelihood Approach

In this section, we introduce a Bayesian quantile regression approach for spatially correlated data, beginning with a description of the underlying model. Let $\{s_l : l = 1, \ldots, L\}$ denote L different spatial sites, L a fixed constant. We are mainly concerned with problems where L is small or modest, but there are a large number of observations at each site. Given location s_l, we observe $D_l = \{(Y_i^{(l)}, \boldsymbol{X}_i^{(l)}) : i = 1, \ldots, n_l\}$, where $Y_i^{(l)}$ is the response, $\boldsymbol{X}_i^{(l)} \in \mathbf{R}^{p+1}$ is composed of an intercept and p covariates, and n_l is the sample size at location s_l. We assume that at all sites the distribution of the p covariates, G_X, has a bounded support $\mathcal{X}$. If the design points are non-stochastic, the basic conclusions we obtain in this paper hold under appropriate conditions on the design sequence, but we focus on the case of random designs for simplicity. We specify the τth quantile of $Y_i^{(l)}$ given $\boldsymbol{X}_i^{(l)}$ and location s_l, as $Q_\tau(Y_i^{(l)}|\boldsymbol{X}_i^{(l)}, s_l) = \boldsymbol{X}_i^{(l)\top}\boldsymbol{\beta}_0(\tau, s_l)$, where

$\boldsymbol{\beta}_0(\tau, s_l) = (\boldsymbol{\beta}_{0,I}(\tau, s_l), \boldsymbol{\beta}_{0,S}^\top(\tau, s_l))^\top$ are the $p+1$ dimensional spatially varying coefficients including the intercept term $\boldsymbol{\beta}_{0,I}(\tau, s_l)$. The unknown function $\boldsymbol{\beta}_0(\tau, s)$, if specified over all $\tau \in (0, 1)$, describes the entire conditional distribution of Y given $\boldsymbol{X}$ at location s, which is to be denoted as $F_{s,X}$. We consider the problem of estimating K conditional quantiles at $\tau_1 < \cdots < \tau_K$. In most applications, K is a small integer. Let $\boldsymbol{\beta}_0(s_l) = (\boldsymbol{\beta}_0(\tau_1, s_l)^\top, \ldots, \boldsymbol{\beta}_0(\tau_K, s_l)^\top)^\top \in \mathbf{R}^{K(p+1)}$ be the true parameter of interest at location s_l, and $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_0(s_1)^\top, \ldots, \boldsymbol{\beta}_0(s_L)^\top)^\top$ the collection of all the parameters of interest. To infer about $\boldsymbol{\beta}_0$ in a Bayesian setting, we assume parametric spatial priors on $\boldsymbol{\beta}(s_l) = (\boldsymbol{\beta}(\tau_1, s_l)^\top, \ldots, \boldsymbol{\beta}(\tau_K, s_l)^\top)^\top$ across s_l to incorporate the spatial correlation, and employ empirical likelihood as a working likelihood. Possible parametric forms of spatial priors can be found in the literature of spatial models, see Cressie (1993).

To estimate $\boldsymbol{\beta}_0(s_l)$, we use $K(p+1)$-dimensional estimating functions $m(\boldsymbol{X}, Y, \boldsymbol{\beta}(s_l))$, where the components of m are

$$m_{d(p+1)+j}(\boldsymbol{X}, Y, \boldsymbol{\beta}(s_l)) = \psi_{\tau_{d+1}}\big(Y - \boldsymbol{X}^\top \boldsymbol{\beta}(\tau_{d+1}, s_l)\big) X_j,$$

for $d = 0, 1, \ldots, K-1$, $j = 1, \ldots, p+1$, with $\psi_\tau(u) = 1_{\{u<0\}} - \tau$ for $u \neq 0$, and $\psi_\tau(u) = 0$ for $u = 0$, being the quantile score function, where $1_{\{A\}}$ is an indicator function on the set A. For any proposed $\boldsymbol{\beta}(s_l)$, its profile empirical likelihood ratio is given by

$$\mathcal{R}(\boldsymbol{\beta}(s_l)) = \max\Big\{\prod_{i=1}^{n_l}(n_l\omega_i) \;\Big|\; \sum_{i=1}^{n_l}\omega_i m(\boldsymbol{X}_i^{(l)}, Y_i^{(l)}, \boldsymbol{\beta}(s_l)) = 0, \omega_i \geq 0, \sum_{i=1}^{n_l}\omega_i = 1\Big\}.$$

More details about the computation of $\mathcal{R}(\boldsymbol{\beta}(s_l))$ can be found in Yang and He (2012).

For any proposed $\boldsymbol{\beta}(s_l)$, consider its empirical likelihood function $\mathcal{R}(\boldsymbol{\beta}(s_l))/n_l^{n_l}$. With a prior specification $p_0(\tilde{\boldsymbol{\beta}})$ on the parameter $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}(s_1)^\top, \ldots, \boldsymbol{\beta}(s_L)^\top)^\top$, we formally have the posterior density

$$p(\tilde{\boldsymbol{\beta}}|D) \propto p_0(\tilde{\boldsymbol{\beta}}) \times \prod_{l=1}^{L} \mathcal{R}(\boldsymbol{\beta}(s_l)) \text{ with } D = \{D_1, \ldots, D_L\}. \tag{2.1}$$

Unlike the block-wise empirical likelihood used for spatial regression in Nordman (2008), we account for the spatial correlation through the priors on $\boldsymbol{\beta}(\tau_d, s_l)$ in the spirit of Reich, Fuentes, and Dunson (2011). Under our framework, the observations from different stations are conditionally independent given $\boldsymbol{\beta}(\tau_d, s_l)$ at all $\tau \in (0, 1)$ and all $l = 1, \ldots, L$.

We call $p(\tilde{\boldsymbol{\beta}}|D)$ a posterior distribution from the BEL approach for the sake of convenience, even though it is not one in the strict sense (Lazar (2003)). We focus on the asymptotic properties of the posterior distribution in (2.1) within our framework.

The prior $p_0(\tilde{\boldsymbol{\beta}})$ includes information on $\boldsymbol{\beta}(\tau_d, s_l)$ across both locations s_l and quantiles τ_d. Denote $\boldsymbol{\beta}_I(\tau_d, s_l)$ as the intercept parameter in $\boldsymbol{\beta}(\tau_d, s_l)$, and $\boldsymbol{\beta}_S(\tau_d, s_l)$ as the slope parameters. Let $g_{d,l}$ be a spherically symmetric distribution with zero as its center as well as its mode, and with a finite second order derivative at zero. We consider the following priors on $\tilde{\boldsymbol{\beta}}$.

$$\begin{aligned}
&\boldsymbol{\Omega}_{1,1}^{-1/2}(\boldsymbol{\beta}(\tau_1, s_1) - \boldsymbol{\beta}_{p,0}) \sim g_{1,1}, \\
&\boldsymbol{\Omega}_{1,l}^{-1/2}(\boldsymbol{\beta}(\tau_1, s_l) - \boldsymbol{\beta}(\tau_1, s_1))|\boldsymbol{\beta}(\tau_1, s_1) \sim g_{1,l} \text{ for } l = 2, \ldots, L, \\
&\boldsymbol{\Omega}_{d,l}^{-1/2}(\boldsymbol{\beta}(\tau_d, s_l) - \boldsymbol{\beta}(\tau_1, s_l))|\boldsymbol{\beta}(\tau_1, s_l) \sim g_{d,l} \text{ for } l = 2, \ldots, L; d = 2, \ldots, K,
\end{aligned} \tag{2.2}$$

for a location vector $\boldsymbol{\beta}_{p,0} \in \mathbf{R}^{(p+1)}$ and scatter matrices $\boldsymbol{\Omega}_{d,l}$ of appropriate dimensions. These priors imply that each component in $\tilde{\boldsymbol{\beta}}(\tau_d, s_l)$ is correlated across locations s_l and quantile levels τ_d. Let $\boldsymbol{\Omega}_{d,l} = \mathrm{diag}(\boldsymbol{\Omega}_{d,l,I}, \boldsymbol{\Omega}_{d,l,S})$, where $\boldsymbol{\Omega}_{d,l,I}$ and $\boldsymbol{\Omega}_{d,l,S}$ represent the components of $\boldsymbol{\Omega}_{d,l}$ corresponding to the intercept and slope parameters in $\boldsymbol{\beta}(\tau_d, s_l)$, respectively, for $d = 1, \ldots, K$ and $l = 1, \ldots, L$. We usually choose $\boldsymbol{\Omega}_{d,l,I}$ sufficiently large for $d \geq 2$, but with $\boldsymbol{\Omega}_{1,l}$ for $l \geq 2$, and $\boldsymbol{\Omega}_{d,l,S}$ for $d \geq 2$ relatively small. This implies that the priors in (2.2) regularize the slope parameter $\boldsymbol{\beta}_S(\tau_d, s_l)$ across locations and quantiles, but regularize the intercept parameter $\boldsymbol{\beta}_I(\tau_d, s_l)$ only across locations. The specification of priors in (2.2) relies on the choice of a reference station s_1, a reference quantile level τ_1 and a prior center $\boldsymbol{\beta}_{p,0}$. By choosing $\boldsymbol{\Omega}_{1,1}$ to grow with sample sizes and $\boldsymbol{\Omega}_{d,l}$ to be small for $d \neq 1$ or $l \neq 1$, the choices of s_1, τ_1 and $\boldsymbol{\beta}_{p,0}$ have no asymptotic effects in the posterior. We recommend choosing s_1 to be the location in the center of the region of interest or the location with the largest sample size, and τ_1 to be the quantile level closest to the median.

In general, the matrices $\boldsymbol{\Omega}_{d,l}$ for $d \neq 1$ or $l \neq 1$ represent our prior belief on the correlations across locations or quantiles. In the rest of the paper, we treat them as given without hyper-parameters. In our empirical investigations in Sections 4 and 5, we describe a specific procedure of choosing the $\boldsymbol{\Omega}_{d,l}$ matrices. We also note that if $g_{d,l}$ in the prior specification is Gaussian, then our priors are consistent with Gaussian process priors but with a particular choice of its covariance function.

We take the BEL estimate $\tilde{\boldsymbol{\beta}}_{BEL}$ of the parameters as the posterior mean from the Bayesian empirical likelihood approach. Numerically, it can be obtained from a Markov Chain Monte Carlo algorithm. For the specific Metropolis–Hastings algorithm used in our empirical studies, we refer to the supplementary file.

3. Asymptotic Properties of BEL

For convenience, we assume that $n = \sum_{l=1}^{L} n_l$ and each n_l increases at the same rate as n for $l = 1, \ldots, L$. We restrict L to be a finite constant. To obtain the asymptotic results,we make assumptions similar to those of Yang and He (2012) applied to each site s_l.

(A1) For each l, there exists a neighborhood $\mathcal{N}$ of $\boldsymbol{\beta}_0(s_l)$ such that $P(\mathcal{R}(\boldsymbol{\beta}(s_l)) > 0) \to 1$ for any $\boldsymbol{\beta}(s_l) \in \mathcal{N}$, as $n_l \to \infty$.

(A2) The distribution function G_X has bounded support $\mathcal{X}$.

(A3) The conditional distribution $F_{X,l}(t)$ of Y given $\boldsymbol{X}$ and s_l is twice continuously differentiable in t for all $\boldsymbol{X} \in \mathcal{X}$.

(A4) At any $\boldsymbol{X} \in \mathcal{X}$, the conditional density function $F'_{X,l}(t) = f_{X,l}(t) > 0$ for t in a neighborhood of $F_{X,l}^{-1}(\tau_d)$ for each $d = 1, \ldots, k$.

(A5) For each l, $E\{m(\boldsymbol{X}, Y, \boldsymbol{\beta}_0(s_l))m(\boldsymbol{X}, Y, \boldsymbol{\beta}_0(s_l))^\top\}$ is positive definite.

(A6) For the prior $p_0(\tilde{\boldsymbol{\beta}})$, the logarithm of $g_{d,j}$ is twice continuously differentiable, and $||\boldsymbol{\beta}_{p,0}|| = O(1)$.

Let $\hat{\boldsymbol{\beta}}(s_l)$ be the maximum empirical likelihood estimate of $\boldsymbol{\beta}_0(s_l)$, $\tilde{\boldsymbol{\beta}}_{EL} = (\hat{\boldsymbol{\beta}}^\top(s_1), \ldots, \hat{\boldsymbol{\beta}}^\top(s_L))^\top$, and $J_{EL} = \operatorname{diag}(J^{(1)}, \ldots, J^{(L)})$, with $J^{(l)} = n_l V_{12}^{(l)\top} V_{11}^{(l)^{-1}} V_{12}^{(l)}$, $V_{11}^{(l)} = \Psi \otimes E(\boldsymbol{X}^{(l)}\boldsymbol{X}^{(l)\top})$, $\Phi = \left(\Phi_{ij}\right) \in \mathbf{R}^{K\times K}, \Phi_{ij} = \tau_i \wedge \tau_j - \tau_i\tau_j$, and $V_{12}^{(l)} = -\frac{\partial E\left\{m(\boldsymbol{X}^{(l)}, Y^{(l)}, \boldsymbol{\beta}(\tau, s_l))\right\}}{\partial \boldsymbol{\beta}(\tau, s_l)}\Big|_{\boldsymbol{\beta}(\tau,s_l)=\boldsymbol{\beta}_0(\tau,s_l)}$. In the prior $p_0(\tilde{\boldsymbol{\beta}})$, the prior mode is $\tilde{\boldsymbol{\beta}}_{p,0} = \boldsymbol{\beta}_{p,0} \otimes 1_{KL}$ and the prior information is $-\frac{\alpha^2 \log p_0(\tilde{\boldsymbol{\beta}}|D)}{\alpha\tilde{\boldsymbol{\beta}}^2}\Big|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_{p,0}} = J_{p,0}$.

Theorem 1. *Under Assumptions* (A1)$-$(A6)*, the posterior density of* $\tilde{\boldsymbol{\beta}}$ *has the following expansion on any sequence of sets* $\left\{||\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_0|| = O(n^{-1/2})\right\}$*:*

$$p(\tilde{\boldsymbol{\beta}}|D) \propto \exp\Big\{ -\frac{1}{2}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{post})^\top J_n (\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{post}) + R_n \Big\}, \tag{3.1}$$

where $J_n = J_{p,0} + J_{EL}$*,* $\tilde{\boldsymbol{\beta}}_{post} = J_n^{-1}(J_{p,0}\tilde{\boldsymbol{\beta}}_{p,0} + J_{EL}\tilde{\boldsymbol{\beta}}_{EL})$*,* $R_n = o_p(1)$.

The results of Theorem 1 extend Theorem 3.2 of Yang and He (2012) to spatially correlated priors. Based on Theorem 1, for any fixed prior (for which $J_{p,0}$ is bounded), the estimation efficiency of the BEL estimates are asymptotically equivalent to that of the usual quantile regression estimates. By using a shrinking prior $p_0(\tilde{\boldsymbol{\beta}})$, such that the largest eigenvalue of $J_{p,0}$ grows with the sample size, the additional terms $J_{p,0}$ and $\tilde{\boldsymbol{\beta}}_{p,0}$ in both J_n and $\tilde{\boldsymbol{\beta}}_{post}$ show when and how an informative prior can complement the empirical likelihood in large samples. When $||J_{p,0}|| = o(n)$, the empirical likelihood dominates the prior asymptotically. If $J_{p,0}$ increases at a faster rate than n, the prior dominates the empirical likelihood. The framework of shrinking priors helps in the identification of efficiency gain from employing informative priors accounting for spatial correlation and commonality across quantiles in a joint modeling of multiple stations.

The empirical likelihood framework allows the joint estimation of multiple quantiles with or without a common parameter across quantiles. The assumption of a common parameter across quantiles or across sites can be viewed as an extreme case of using a dominatingly strong prior. The introduction of a common parameter reduces the number of unknown parameters to be smaller than the number of estimating functions and, correspondingly, the definition of $V_{12}^{(l)}$ is taken to be the derivative with respect to the reduced parameter space. Therefore, the posterior variance may no longer take the same form as the asymptotic variance of the usual quantile regression estimates, and improvements in the asymptotic variances become possible.

Corollary 1. *Under the assumptions of Theorem 1, there is (i)* $\mathrm{Var}\,(\tilde{\boldsymbol{\beta}}|D) = J_n^{-1} + o_p(n^{-1})$*, and (ii)* $\mathrm{Var}\,(\tilde{\boldsymbol{\beta}}_{BEL}) = J_n^{-1} - J_n^{-1} J_{p,0} J_n^{-1} + o_p(n^{-1})$*.*

Corollary 1 is useful when the posterior chain is used for inference on the posterior estimate $\tilde{\boldsymbol{\beta}}_{BEL}$. If the prior information $J_{p,0}$ is dominated by the empirical likelihood, then $\mathrm{Var}\,(\tilde{\boldsymbol{\beta}}_{BEL})$ can be approximated by J_n^{-1}. Otherwise a simple adjustment using $J_{p,0}$ is needed. In either case, it is clear that $\mathrm{Var}\,(\tilde{\boldsymbol{\beta}}_{BEL})$ is smaller than J_{EL}^{-1}, so the BEL estimate has lower variance than the maximum empirical likelihood estimator (MELE) or the usual regression quantile estimator.

However, a strong shrinking prior could lead to bias. In Theorem 1, the prior mode $\tilde{\boldsymbol{\beta}}_{p,0}$ plays a role in the posterior mean. To benefit from the shrinking priors, the amount of shrinking should depend on the evidence from preliminary studies or from external information. For example, in the proposed prior (2.2), we would choose $\Omega_{1,l}$ to be in the order of n only when we believe that $\boldsymbol{\beta}(\tau_1, s_l) - \boldsymbol{\beta}(\tau_1, s_1)$ is nearly zero (or in the order of $o(n^{-1/2})$ to be exact). To strike the right balance, it is often helpful to do a preliminary BEL estimate of the parameters and their standard errors using a nearly flat priors on all the parameters, and then use the standard error estimates to construct a shrinking prior. If two parameters are indeed identical, the standard error of their estimated difference would be shrinking towards zero at the rate of n^{-1}, providing a natural way to impose shrinking priors.

There is no easy way to choose optimal priors for any given problem. Our theory and the empirical experience show that we do not have to aim for optimality; the BEL approach is advantageous for any reasonably chosen priors. We caution that the validity of posterior inference is based on a correctly specified model; because our model accounts for spatial correlation through the prior distribution, it is difficult to perform model diagnostics in any given application. The impact of model mis-specification on posterior inference needs to be carefully examined before we recommend any formal inference to be made based on the proposed BEL method. We focus now on using the BEL method for efficient estimation of the quantile parameters.

4. Simulation Study

We used Monte Carlo simulations to investigate the estimation efficiency of the BEL methods, estimation efficiency measured by the estimated mean squared error (MSE). We compared the performance of the BEL estimates with those of the approximate spatial quantile regression estimates proposed in Reich, Fuentes, and Dunson (2011), denoted as ASQR. We adopted the approximate method of Reich, Fuentes, and Dunson (2011) rather than their full Bayesian modeling approach. The latter is computationally expensive if not infeasible for many applications. The usual quantile regression estimate at each τ, denoted simply as RQ, was also included in our comparisons.

We considered a joint modeling of quantiles $\tau = 0.9, 0.95$, or quartiles $\tau = 0.25, 0.5, 0.75$. In the joint modeling of $\tau = 0.9, 0.95$, we used the fixed site indices $s_1 = 0.5, s_2 = 0.3, s_3 = 0.7, s_4 = 0.1, s_5 = 0.9$ and equal sample size $n_l = 100$ at each site. The location index d_l could be multi-dimensional, but in our simulation model, only the relative distances between sites matter. Two covariates x and z were generated from half of the χ^2 distribution with 2 degrees of freedom, independently over space and time. To generate spatially correlated data, we took the strategy of Reich, Fuentes, and Dunson (2011). We generated U_l $(l = 1, \ldots, 5)$ from the multivariate normal with mean zero and exponential spatial covariance $\exp(-|s_l - s_{l'}|/0.5)$ for $l, l' = 1, \ldots, 5$; at each site s_l, we computed $\tau = \Phi(U_l)$, and obtained y from the quantile function (4.1) using the randomly generated τ. The data generated in this way have the specified conditional quantile functions at each site, but the spatial correlations come from how the U_l are generated. The data generating mechanism here does not match our framework of conditionally independent data given the quantile parameters; we adopted this simulation setting to show that the proposed BEL approach gains efficiency over RQ even when our framework provides only a reasonable working model.

The conditional quantile functions used in our study were

$$Q_\tau(y|x, z, s) = a(\tau, s) + b_x(\tau, s)x + b_z(\tau, s)z, \tag{4.1}$$

Table 1. Estimated MSE's of several estimators for the intercept and the slope parameters at $\tau = 0.9, 0.95$. The numbers in the parentheses are standard error estimates.

	$\mathbf{a(0.9)}$	$\mathbf{b_x(0.9)}$	$\mathbf{b_z(0.9)}$	$\mathbf{a(0.95)}$	$\mathbf{b_x(0.95)}$	$\mathbf{b_z(0.95)}$
BEL	0.339	0.156	0.168	0.503	0.190	0.217
	(0.022)	(0.010)	(0.010)	(0.034)	(0.012)	(0.014)
ASQR	0.623	0.798	0.370	0.579	0.601	0.421
	(0.063)	(0.055)	(0.022)	(0.054)	(0.044)	(0.024)
RQ	0.635	0.333	0.317	0.786	0.331	0.391
	(0.042)	(0.038)	(0.023)	(0.055)	(0.027)	(0.029)

where $a(\tau, s) = 2s + (\tau + 1)\Phi^{-1}(\tau)$, $b_x(\tau, s) = -4\sqrt{0.75 - \tau}1_{\{\tau<0.75\}}$, $b_z(\tau, s) = 2s\tau^2$, $\Phi(\cdot)$ the distribution function of the standard normal.

The covariate x influenced the outcome variable y only at lower quantiles $\tau < 0.75$, $b_x(\tau, s)$ was zero at all sites for $\tau \geq 0.75$. Consider the estimation of quantiles at $\tau = 0.9, 0.95$. Let $\boldsymbol{\beta}(\tau_d, s_l) = (a(\tau_d, s_l), b_x(\tau_d, s_l), b_z(\tau_d, s_l))$. The conditional standard deviations of the priors used in the BEL, $\boldsymbol{\Omega}_{d,l}$ in Section 2, are listed in Table 1 of the supplementary file. The priors used in the ASQR were those used in the simulation study of Reich, Fuentes, and Dunson (2011). More computational details including the prior choices can be found in the supplementary file. The mean squared errors (MSE) of the competing estimators of $\boldsymbol{\beta}(\tau, s)$ are given in Table 1. By using informative priors to regularize the quantile coefficients across quantiles and across location, the BEL method improves on RQ quite noticeably. The ASQR estimators lose efficiency relative to RQ in this study, especially in the estimation of b_x. The poor performance of ASQR could be attributed to its reliance on the asymptotic normality of the regression quantile estimates, which may not work well at the quantiles with this sample size. Similar results were obtained when the prior parameters used in Table 1 of the supplementary file were moderately varied, and the advantage of the BEL method was rather insensitive to the choice of the prior variances.

Consider the joint modeling of quartiles at $\tau = 0.25, 0.5, 0.75$. We took the setting in Reich, Fuentes, and Dunson (2011), but with five sites of equal sample size $n_l = 20$ at each site. The site indices are uniformly generated from $[0, 1]$ with the constraint that the minimum difference between the station indices be no smaller than 0.05. The quantile coefficients in (4.1) were taken to be $a(\tau, s) = 2s + (\tau + 1)\Phi^{-1}(\tau)$, $b_x(\tau, s) = 0$, and $b_x(\tau, s) = 5s\tau^2$. The covariates x and z were generated uniformly from $[0, 1]$. More computational details can be found in the supplementary file. The MSE of the competing estimators of $\boldsymbol{\beta}(\tau, s)$ are given in Table 2, from which we observed that both BEL and ASQR improve on RQ quite noticeably, and BEL performs better than ASQR.

Table 2. Estimated MSE's of several estimators for the intercept and the slope parameters at $\tau = 0.25, 0.5, 0.75$. The numbers in the parentheses are standard error estimates.

	$\mathbf{a}$	$\mathbf{b_x}$	$\mathbf{b_z}$	$\mathbf{a}$	$\mathbf{b_x}$	$\mathbf{b_z}$	$\mathbf{a}$	$\mathbf{b_x}$	$\mathbf{b_z}$
	$\tau = \mathbf{0.25}$			$\tau = \mathbf{0.5}$			$\tau = \mathbf{0.75}$		
BEL	0.458	0.728	0.989	0.642	0.902	1.143	1.034	0.915	2.509
	(0.037)	(0.056)	(0.072)	(0.073)	(0.101)	(0.124)	(0.097)	(0.092)	(0.171)
ASQR	0.960	1.287	1.878	1.085	1.747	2.314	1.450	2.531	3.366
	(0.146)	(0.180)	(0.189)	(0.182)	(0.251)	(0.219)	(0.153)	(0.278)	(0.286)
RQ	1.249	2.373	2.646	2.041	4.756	3.987	3.351	6.422	7.349
	(0.106)	(0.185)	(0.230)	(0.159)	(0.267)	(0.317)	(0.219)	(0.448)	(0.520)

5. A Data Example

In this section, we consider the BEL method for statistical downscaling of daily precipitation in the Chicago area. We used the observed daily precipitation (PRCP) of seven stations (Midway, Aurora, O'Hare, Wheaton, Elgin, Park Forest, Joliet) in Illinois from 1976 – 2002 as the response variable. We focused on the rainy days only; they account for around 30% in the Chicago area. The predictors are the simulated daily precipitation (RP), daily humidity (RH), and daily maximum temperature (RT) from the ERA-40 reanalysis model introduced in Uppala et al. (2005). We used the linear quantile regression model: $Q_\tau(PRCP|RP, RH, RT, s_i) = a(\tau, s_i)+b_1(\tau, s_i)RP+b_2(\tau, s_i)RH+b_3(\tau, s_i)RT$ at quantiles $\tau_1 = 0.95, \tau_2 = 0.99$, with station index s_i for $i = 1, \ldots, 7$. To construct informative priors using external data information, e.g., neighboring stations, we constructed the priors using a preliminary chain on data from Joliet and Park Forest, which are south of the five other stations. We chose the Midway station as the reference station s_1 and let $\boldsymbol{\beta}(\tau_d, s_i) = (a(\tau_d, s_i), b_1(\tau_d, s_i), b_2(\tau_d, s_i), b_3(\tau_d, s_i))$, for $d = 1, 2$ and $i = 1, \ldots, 5$.

For the preliminary chain on two stations, Park Forest (as $i = 6$) and Joliet (as $i = 7$) has priors on each component of $\boldsymbol{\beta}(\tau_1, s_6)$, as well as $a(\tau_1, s_7)-a(\tau_1, s_6)|a(\tau_1, s_6)$, $b_j(\tau_1, s_7) - b_j(\tau_1, s_6)|b_j(\tau_1, s_6)$, $b_j(\tau_2, s_7) - b_j(\tau_1, s_6)|b_j(\tau_1, s_6)$ for $j = 1, 2, 3$, and $a(\tau_2, s_7) - a(\tau_1, s_7)|a(\tau_1, s_6)$ as independent $N(0, 100^2)$. The prior variances are so large here that the empirical likelihood is meant to dominate.

We used the posterior chain from the preliminary run to update the prior variances, and used them for the primary analysis of the first five stations. Table 3 contains the prior standard deviations used. We note that no shrinking priors were used for the intercept parameters or for the coefficients of $\tau_1 = 0.9$ at the

Table 3. The conditional standard deviations of the multivariate normal priors used in BEL; the first row in the left columns (Stations) provides the standard deviations of each parameter in $\boldsymbol{\beta}(\tau_1, s_1)$; the second row provides the conditional standard deviation of each parameter in $\boldsymbol{\beta}(\tau_1, s_i)$ given $\boldsymbol{\beta}(\tau_1, s_1)$ for $i = 2, \ldots, 5$. The right columns (Quantiles) provides the conditional standard deviation of each parameter in $\boldsymbol{\beta}(\tau_2, s_i)$ given $\boldsymbol{\beta}(\tau_1, s_i)$ for $i = 1, \ldots, 5$.

BEL	**Stations** $(s_i \mid s_1)$				**Quantiles** $(\tau_2 \mid \tau_1)$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	100	100	100	100	100	0.125	0.079	0.103
$s_i (i = 2, \ldots, 5)$	0.030	0.044	0.025	0.041	100	0.125	0.079	0.103

reference station Midway (as $i = 1$), but shrinking priors across stations and across quantile levels were used.

The data are split into a training period (odd years) and a validation period (even years), of sizes 13,026 and 12,713. An MCMC chain, as described in the supplementary file, of length 100,000 was obtained for the BEL estimate of the model parameters for the training data. The model estimates were then applied to the validating data to predict the 0.95th and 0.99th quantiles of $PRCP$.

Table 4 reports performance validation measure $d = (O - E)/\sqrt{\tau(1 - \tau)n}$, where n is the total number of days for prediction, O is the number of days when the observed $PRCP$ exceeds the predicted τth quantile of $PRCP$, and E indicates the expected number of days, i.e., $E = n(1 - \tau)$. The normalized differences are shown for the full testing period at $\tau = 0.95$ and 0.99. The BEL method is compared to RQ, the quantile regression estimate at individual τ, and ASQR, the approximate spatial quantile regression method at the same quantile levels. Normalized differences greater than 3 in absolute values are marked as bold in Table 4. By using informative priors across stations and quantiles, we see that the BEL performs better than RQ overall, especially for the station Wheaton, where the RQ estimate at $\tau_2 = 0.99$ was unstable. The ASQR predictions are quite good at $\tau_1 = 0.95$, but inaccurate at $\tau_2 = 0.99$. The normal approximations of ASQR at high quantiles are risky even at this sample size, and it assumes linear conditional quantile functions at all quantile levels, when examination of the data suggests that lower quantiles are nonlinear. We do not include the method of Reich (2012) in this example, because it also assumes linear quantiles at all levels.

Based on Corollary 1, we can use the posterior chain obtained in the study to test the hypothesis of equal quantile coefficients at different quantile levels or different stations. For example, the standardized differences (or t-statistics) for $b_3(0.99, s_l) - b_3(0.95, s_l)$ from the BEL method are 2.75 and 2.47 at Midway and

Table 4. The table gives the normalized performance measures in the validation data when the quantiles are predicted using the parameters obtained from the training data with three different methods.

	Midway	**Ohare**	**Wheaton**	**Elgin**	**Aurora**
			$\tau = 0.95$		
BEL	-0.844	0.682	0.655	1.616	0.761
ASRQ	-2.418	-1.974	-0.456	0.277	-0.242
RQ	-0.931	-0.115	0.841	1.521	0.488
			$\tau = 0.99$		
BEL	0.665	-0.167	0.531	2.509	2.131
ASRQ	**5.645**	**4.487**	**7.247**	**6.700**	**6.324**
RQ	1.431	-0.555	**3.787**	2.509	2.530

O'Hare Stations, respectively. This indicates that the daily maximum temperature is likely to have more impact on precipitation at the $\tau = 0.99$ quantile than at the 0.95 quantile there. This difference could not be detected by hypothesis testing of RQ at each individual station, noting that the p-values from the F-test for the null hypothesis of equal slope coefficients b_3 at these two high quantiles are 0.21 and 0.10, respectively at the two stations.

6. Discussion

We propose an empirical likelihood approach for spatial quantile regression. This directly targets a small number of quantile levels of interest. It is possible that the estimated quantiles may cross, but the use of informative priors toward common slopes helps avoid this since the intercept parameters are always ordered in the MCMC chain. If prediction at several quantile levels is a goal, estimated quantiles can be re-ordered through monotonization, and estimation accuracy enhanced, see Chernozhukov, Fernández-Val, and Galichon (2010).

The proposed method is not strictly Bayesian and we are open to prior construction based on preliminary estimates. It is also possible to rely on external data or data from some of the locations to construct informative priors on other locations, as in the example in Section 5. In our experience, such informative priors are helpful in the estimation of tail quantiles and the improvement is not sensitive to the scaling of the priors within a reasonable range.

Our approach have the following limitations. The joint modeling of data and priors may not be completely consistent with a frequentist approach, and our approach is not Bayesian due to the use of a working likelihood and a shrinking prior. We view our working model as a means of generating efficient estimation of quantiles when data from several stations in a local region can be pooled.

The proposed approach does not predict the response in a location without data unless a prior is specified to include the location of interest. Our main objective is to make use of spatial correlation through the priors to enhance estimation of quantiles at each location where data is available. We have not considered modeling spatial data when a large number of sites are spread across a large area but few observations at each site. The specification of the priors could be challenging in such cases, and an asymptotic theory would be different. We hope that future research can address this general form of spatial data under the framework of Bayesian empirical likelihood.

Acknowledgement

The research is partially supported by the NSF Award DMS-1237234, NIH Grant R01GM080503-04, and NNSFC Grant 11129101. We thank the referees for constructive comments and criticism on an earlier draft of the paper.

References

Chang, I. H. and Mukerjee, R. (2008). Bayesian and frequentist confidence intervals arising from empirical-type likelihoods. *Biometrika* **95**, 139-147.

Chen, S. X. and Keilegom, I. V. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415-447.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78**, 1093-1125.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley, New York.

Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307-323.

Fang, K.-T. and Mukerjee, R. (2006). Empirical-type likelihoods allowing posterior credible sets with frequentist validity: Higher-order asymptotics. *Biometrika* **93**, 723-733.

Gelfand, A. E., Kottas, A. and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with dirichlet process mixing. *J. Amer. Statist. Assoc.* **100**, 1021-1035.

Geraci, M. and Bottai, M. (2006). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics* **8**, 140-154.

Griffin, J. E. and Steel, M. F. J. (2006). Order-based dependent dirichlet proesses. *J. Amer. Statist. Assoc.* **101**, 179-194.

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

Kim, M.-O. and Yang, Y. (2011). Semiparametric approach to a random effects quantile regression model. *J. Amer. Statist. Assoc.* **106**, 1405-1417.

Koenker, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge.

Koenker, R. and Mizera, I. (2004). Penalized triograms: total variation regularization for bivariate smoothing. *J. Roy. Statist. Soc. Ser. B* **66**, 145-164.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *J. Amer. Statist. Assoc.* **96**, 1458-1468.

Kottas, A. and M. Krnjajić (2009). Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.* **36**, 297-319.

Lancaster, T. and S. J. Jun (2010). Bayesian quantile regression methods. *J. Appl. Econ.* **25**, 287-307.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika* **90**, 319-326.

McMillen, D. (2012). *Quantile Regression for Spatial Data.* Springer, Heidelberg.

Monahan, J. F. and Boos, D. D. (1992). Proper likelihoods for bayesian analysis. *Biometrika* **79**, 271-278.

Nordman, D. J. (2008). An empirical likelihood method for spatial quantile regression. *Metrika* **68**, 351-363.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC.

Reich, B. J. (2012). Spatiotemporal quantile regression for detecing distributional changes in environmental processes. *J. Roy. Statist. Soc. Ser. C* **61**, 535-553.

Reich, B. J., Bondell, H. D. and Wang, H. J. (2008). Flexible bayesian quantile regression for independent and clustered data. *Biostatistics* **11**, 337-352.

Reich, B. J., Fuentes, M. and Dunson, D. B. (2011). Bayesian spatial quantile regression. *J. Amer. Statist. Assoc.* **106**, 6-20.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, V. D. C., Bechtold, U., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hlm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. (2005). The era-40 re-analysis. *Quarterly J. Roy. Meteorological Soc.* **131**, 2961-3012.

Wang, H. J. and Zhu, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *J. Statist. Plann. Inference* **141**, 1603-1615.

Wood, S. A., W. Jiang and M. Tanner (2002). Bayesian mixture of splines for spatially adaptive nonparemtric regression. *Biometrika* **89**, 513-528.

Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *Ann. Statist.* **40**, 1102-1131.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54**, 437-447.

Department of Epidemiology and Biostatistics, Drexel University, Philadelphia, PA 19104 USA.

E-mail: Yunwen.Yang@drexel.edu

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail: xmhe@umich.edu

(Received March 2013; accepted March 2014)