Quantile Regression for Spatially Correlated Data: an Empirical Likelihood Approach

Yunwen Yang and Xuming He

Drexel University and University of Michigan

Supplementary Material

S1 Suppenmentary Material for Computational Details

In this section, we provide the technical details of the proof about our main asymptotic theorems, and the computational details about the proposed BEL method, including the the MCMC algorithm we used to draw the posterior chain, the prior choices, the proposal distribution, and the staring values in the MCMC chain.

S1.1 Proof of Theorem 1 and Corollary 2

In this section, we provide the proof of our main results Theorem 1 and Corollary 2 in Section 3.

Proof of Theorem 1

As in Yang and He (2012), we have the following quadratic expansion on the log of the empirical likelihood:

$$\log\Big\{\prod_{l=1}^{L}\mathrm{EL}(\boldsymbol{\beta}(s_l)|D_l)\Big\} = -\sum_{l=1}^{L}\Big\{\frac{1}{2}\big(\boldsymbol{\beta}(s_l)-\hat{\boldsymbol{\beta}}(s_l)\big)^\top J^{(l)}\big(\boldsymbol{\beta}(s_l)-\hat{\boldsymbol{\beta}}(\tau,s_l)\big)+o_p(1)\Big\}, \tag{S1.1}$$

for $||\boldsymbol{\beta}(\tau,s_l)-\boldsymbol{\beta}_0(\tau,s_l)|| = O(n_l^{-1/2})$, where $\hat{\boldsymbol{\beta}}(s_l)$ is the MELE of $\boldsymbol{\beta}_0(s_l)$ for station s_l, and $J^{(l)}$ is defined in Section 3. The expression in (S1.1) is equivalent to

$$\log\Big\{\prod_{l=1}^{L}\mathrm{EL}(\boldsymbol{\beta}(\tau,s_l)|D_l)\Big\} = -\frac{1}{2}\big(\tilde{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}_{EL}\big)^\top J_{EL}\big(\tilde{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}_{EL}\big)\big)+o_p(L), \tag{S1.2}$$

$$\text{with } \tilde{\boldsymbol{\beta}}_{EL} = (\hat{\boldsymbol{\beta}}^\top(s_1), ..., \hat{\boldsymbol{\beta}}^\top(s_1), ..., \hat{\boldsymbol{\beta}}^\top(s_L), ..., \hat{\boldsymbol{\beta}}^\top(s_L))^\top,$$

$$\text{and } J_{EL} = \begin{pmatrix} J^{(1)} & 0 & \cdots & 0 \\ 0 & J^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J^{(L)} \end{pmatrix}.$$

In the prior $p_0(\tilde{\boldsymbol{\beta}})$, the prior mode is $\tilde{\boldsymbol{\beta}}_{p,0} = \boldsymbol{\beta}_{p,0} \otimes 1_{KL}$. Assuming that

$$-\frac{\alpha^2 \log p_0(\tilde{\boldsymbol{\beta}}|D)}{\alpha \tilde{\boldsymbol{\beta}}^2}\Bigg|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_{p,0}} = J_{p,0},$$

we have

$$\log p_0(\tilde{\boldsymbol{\beta}}) = \log p_0(\tilde{\boldsymbol{\beta}}_0) + \frac{1}{2}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{p,0})^\top J_{p,0}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{p,0}) + o(||\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{p,0}||^2). \quad \text{(S1.3)}$$

Combining (S1.3) and (S1.2), we have

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

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

Proof of Corollary 2

We will first show (i) in which $n||Var(\tilde{\boldsymbol{\beta}}|D) - J_n^{-1}|| = o_p(1)$. Denote $\tilde{\boldsymbol{\gamma}} = J_n^{1/2}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{post})$. Following (3.1), we have

$$p(\tilde{\boldsymbol{\gamma}}|D) \propto \exp\{-\frac{1}{2}\tilde{\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\gamma}} + R_n\} \text{ with } R_n = o_p(1).$$

Note that

$$\begin{aligned} P\big(||E(\tilde{\boldsymbol{\gamma}}|D)|| > \epsilon\big) &= P(||E(\tilde{\boldsymbol{\gamma}}|D)|| > \epsilon \,\&\, ||R_n|| < \delta\big) + P\big(||E(\tilde{\boldsymbol{\gamma}}|D)|| > \epsilon \,\&\, ||R_n|| \geq \delta\big) \\ &\leq P\big(||E(\tilde{\boldsymbol{\gamma}}|D)|| > \epsilon \;\big|\; ||R_n|| < \delta\big) + P(||R_n|| \geq \delta). \end{aligned}$$

The second term goes to 0 as $n \to \infty$. The first item can be sufficiently small for a small δ and large n. It follows that $E(\tilde{\boldsymbol{\gamma}}|D) = o_p(1)$ and similarly, $Var(\tilde{\boldsymbol{\gamma}}|D) = I + \Delta(D)$ for $\Delta(D) = o_p(1)$.

Noting that $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_{post} + J_n^{-1/2}\tilde{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\beta}}_{post}$ is a constant given D, we have

$$Var(\tilde{\boldsymbol{\beta}}|D) = J_n^{-1/2}Var(\tilde{\boldsymbol{\gamma}})J_n^{-1/2} = J_n^{-1} + J_n^{-1/2}\Delta(D)J_n^{-1/2} \text{ with } ||\Delta_D|| = o_p(1), \quad \text{(S1.4)}$$

which leads to $n||Var(\tilde{\boldsymbol{\beta}}|D) - J_n^{-1}|| = o_p(1)$.

To prove (ii), we will first show that $||\tilde{\boldsymbol{\beta}}_{BEL} - \tilde{\boldsymbol{\beta}}_{post}|| = o_p(n^{-1/2})$, and then $Avar(\tilde{\boldsymbol{\beta}}_{post}) = J_n^{-1} - J_n^{-1}J_{p,0}J_n^{-1}$.

By $\tilde{\boldsymbol{\gamma}} = J_n^{1/2}(\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{post})$, we have $\tilde{\boldsymbol{\beta}}_{BEL} = \tilde{\boldsymbol{\beta}}_{post} + J_n^{-1/2} E(\tilde{\boldsymbol{\gamma}}|D)$. Because $E(\tilde{\boldsymbol{\gamma}}|D) = o_p(1)$, we have $||\tilde{\boldsymbol{\beta}}_{BEL} - \tilde{\boldsymbol{\beta}}_{post}|| = o_p(n^{-1/2})$.

Now we note that

$$\tilde{\boldsymbol{\beta}}_{post} = J_n^{-1}(J_{p,0}\tilde{\boldsymbol{\beta}}_{p,0} + J_{EL}\tilde{\boldsymbol{\beta}}_{EL}) = J_n^{-1}J_{p,0}\tilde{\boldsymbol{\beta}}_{p,0} + (I - J_n^{-1}J_{p,0})\tilde{\boldsymbol{\beta}}_{EL},$$

we have asymptotically

$$Avar(\tilde{\boldsymbol{\beta}}_{post}) = J_n^{-1}J_{EL}J_n^{-1} = J_n^{-1} - J_n^{-1}J_{p,0}J_n^{-1}.$$

Therefore, we have (ii), that is, $Var(\tilde{\boldsymbol{\beta}}_{BEL}) = J_n^{-1} - J_n^{-1}J_{p,0}J_n^{-1} + o_p(n^{-1})$.

S1.2 The Block Metroplis-Hasting Algorithm

As in Yang and He (2012), we can use the Metroplis-Hasting algorithm to facilitate the Bayesian computation. The convergence of the Metroplis-Hasting algorithm given the data follows from the standard results in Gilks et al. (1996). Considering the large dimension of parameters involved in the joint modeling of multiple locations, we use the block Methoplis-Hasting algorithm proposed in Chib and Greenberg (1995) to update $\beta(s_l)$ seperately at each l. Let

$$\beta(s_{-l}) = \{\beta(s_{l'}), l' \neq l\}.$$

Using a normal distribution as the proposal distribution, the probability of moving from $\beta(s_1)$ to $\beta^*(s_1)$ is

$$\alpha_1 = \min\Big\{\frac{p\big(\beta^*(s_1), \beta(s_{-1})|D\big)}{p\big(\beta(s_1), \beta(s_{-1})|D\big)}, 1\Big\};$$

and for $2 \leq l \leq L$,

$$\alpha_l = \min\Big\{ \frac{EL(\beta^*(s_l)|D_l) \times g_{1,l}(\Omega_{1,l}^{-1/2}(\beta^*(\tau_1, s_l) - \beta(\tau_1, s_1))}{EL(\beta(s_l)|D_l) \times \times g_{1,l}(\Omega_{1,l}^{-1/2}(\beta(\tau_1, s_l) - \beta(\tau_1, s_1))} \times \frac{\prod_{d=2}^{K} g_{d,l}(\Omega_{d,l}^{-1/2}(\beta^*(\tau_d, s_l) - \beta^*(\tau_1, s_l))}{\prod_{d=2}^{K} g_{d,l}(\Omega_{d,l}^{-1/2}(\beta(\tau_d, s_l) - \beta(\tau_1, s_l))}, 1\Big\}.$$

In both the simulation study and the real data example, we took the same strategy as in Section 2 of Yang and He (2012) to adjust the intercept parameter estimates of the BEL.

S1.3 Computation Details in the Simulation Study

In this section, we provide the computation details for the simulation study in Section 4.

Table 1: The table gives the conditional standard deviations of the multivariate normal priors used in the simulation study for BEL. The row with rowname s_1 in the left table (Stations) provides the standard deviation (100) in the normal prior for each parameter in $\beta(\tau_1, s_1)$,respectively; The other rows provide the conditional standard deviation of each parameter in $\beta(0.9, s_i)$ given $\beta(0.9, s_1)$, respectively, for $i = 2, .., 5$. The right table (Quantiles) provides the conditional standard deviation of each parameter in $\beta(0.95, s_i)$ given $\beta(0.9, s_i)$, respectively, for $i = 1, ..., 5$.

	Stations			**Quantiles**		
	a	b_x	b_z	a	b_x	b_z
s_1	100	100	100	100	0.63	0.52
s_2	0.72	0.60	0.50	100	0.43	0.45
s_3	0.78	0.53	0.48	100	0.60	0.59
s_4	0.69	0.47	0.58	100	0.45	0.47
s_5	0.85	0.72	0.47	100	0.54	0.55

For BEL at $\tau = 0.9, 0.95$, the priors are chosen based on the posterior chains obtained for several randomly generated data sets. The MCMC chain is of length $100,000$ including a burning period $40,000$ at $\tau_1 = 0.9, \tau_2 = 0.95$ jointly, with starting values as the usual quantile regression estimates. The MCMC chain uses normal priors with mean zero and standard deviation 100 on each parameter, respectively. We run the MCMC chain on just three random data sets, to save computation time in the simulation. Then we calculate the conditional standard deviations using the covariance matrix estimated from the MCMC chain. The average of these conditional standard deviations across the three data sets for each parameter, respectively, are used as the priors in the BEL for the simulation, as listed in Table 1.

For each simulated data sets, we run a MCMC chain starting from the usual quantile regression estimates, with a total length of $120,000$ containing a burning period of length 40000. To determine the variances of the proposal distributions, we adopt the strategy of running a preliminary chain first and then a primary chain (second run) initiated from the parameter vale that maximizes the posterior in the preliminary chain. We use a common convenient value 0.3 as the proposal standard deviations for all the parameters in the preliminary run, and update the proposal standard deviations for the second chain. The latter are calculated by multiply the sample standard deviation of the draws from the preliminary chain by a "shrinkage factor", which is 1.4 multiplied by the corresponding acceptance rate for each site. We also control the acceptance rate between 0.1 and 0.3, i.e., if the acceptance rate is below 0.1, we treat it as 0.1; if the acceptance rate is above 0.3, we treat it as 0.3. In general, we use a larger proposal standard deviation in the preliminary run than in the second run. The average acceptance rates of the Metroplis-Hasting algorithm in our simulation is 52%.

In the simulation study of BEL at $\tau = 0.25, 0.5, 0.75$, we consider the following priors with the reference site (s_1) chosen as the site whose site index is the median, and $\tau_1 = 0.5$. We assume normal priors with mean zero and standard deviation 100 on

Table 2: The table presents the proposal standard deviations used in the preliminary run for the simulation study. The same proposal distributions are used at $\tau = 0.9, 0.95$.

	a	b_x	b_z
s_1	0.4	0.4	0.4
$s_i : i > 1$	0.3	0.3	0.3

$a(0.5, s_1)$, $b_x(0.5, s_1)$, and $b_z(0.5, s_1)$. Given $a(0.5, s_1)$, $b_x(0.5, s_1)$, and $b_z(0.5, s_1)$, the prior across stations assumes

$$a(0.5, s_l)|a(0.5, s_1) \sim N(0, exp(-0.01/||s_l - s_1||)),$$

$$b_x(0.5, s_l)|b_x(0.5, s_1) \sim N(0, exp(-0.05/||s_l - s_1||)),$$

$$\text{and } b_z(0.5, s_l)|b_z(0.5, s_1) \sim N(0, exp(-0.01/||s_l - s_1||)).$$

The prior across quantiles assumes that

$$a(\tau, s_l)|a(0.5, s_l) \sim N(a(0.5, s_l), 100^2),$$

$$b_x(\tau, s_l)|b_x(0.5, s_l) \sim N(b_x(0.5, s_l), 0.3^2),$$

for $\tau = 0.25, 0.75$;

$$\text{and } b_z(0.25, s_l)|b_z(0.5, s_l) \sim N(b_z(0.5, s_l), 0.3^2), b_z(0.75, s_l)|b_z(0.5, s_l) \sim N(b_z(0.5, s_l), 0.6^2).$$

In both simulation studies, the ASQR used the same priors as in Reich et al. (2011).

S2 Computational Details in the Real Data Example

To give an example of how to construct informative priors using external data information, e.g., neighboring stations, we constructed the priors using a MCMC chain on data from Joliet and Park Forest, which are in the south of the five other stations. Among the remaining five stations, we chose the Midway station as the reference station s_1 and let $\beta(\tau_d, s_i) = (a(\tau_d, s_i), b_1(\tau_d, s_i), b_2(\tau_d, s_i), b_3(\tau_d, s_i))$, for $d = 1, 2$ and $i = 1, ..., 5$.

For the MCMC chain on two stations, Park Forest (as $i = 6$) and Joliet (as $i = 7$), we use normal priors with mean zero and standard deviation 100 on each parameter. We use the posterior chain on the two stations, Park Forest and Joilet, to update the prior variances, and use them for the primary analysis of the first five stations. Table 3 contains the prior standard deviations to be used. We note that no shrinking priors are used for the intercept parameters and for the coefficients of $\tau_1 = 0.9$ at the reference station Midway (as $i = 1$), but the shrinking priors across stations and across quantile levels are being used.

The MCMC chain used in the primary analysis is of a total length $100,000$ with a burning period of 5000. The starting values are the usual quantile regression estimates for

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

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	0.161	0.082	0.111
$s_i(i = 2, .., 5)$	0.031	0.047	0.024	0.038	100	0.161	0.082	0.111

Table 4: The table gives the conditional standard deviations of the multivariate normal proposal distribution used in BEL for the real data example; the first row in the left columns (Stations) provides the standard deviations of each parameter in $\beta(\tau_1, s_1)$; the second row provides the conditional standard deviation of each parameter in $\beta(\tau_1, s_i)$ given $\beta(\tau_1, s_1)$ for $l = 2, .., 5$. The right columns (Quantiles) provides the conditional standard deviation of each parameter in $\beta(\tau_2, s_i)$ given $\beta(\tau_1, s_i)$ for $l = 1, ..., 5$.

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	0.010	0.007	0.007	0.009	0.001	0.007	0.007	0.009
s_2	0.015	0.010	0.010	0.012	0.015	0.010	0.010	0.012
s_3	0.015	0.010	0.010	0.012	0.015	0.010	0.010	0.012
s_4	0.020	0.013	0.010	0.015	0.020	0.013	0.010	0.015
s_5	0.015	0.010	0.010	0.012	0.015	0.010	0.010	0.012

each site separately. The proposal distribution is normal distribution with the conditional standard deviations given in Table 4. The ASQR uses the same prior as in Reich et al. (2011), with a joint estimation of $\tau = 0.95, 0.99$. The station index used for ASQR are $(0.21, 0.5), (0.3, 0.5), (0.61, 0.5), (0.79, 0.5), (0.9, 0.5)$.

The quantile regression coefficients obtained from the BEL, RQ and ASQR for each site at $\tau = 0.95, 0.99$ are listed in Table 5. The coefficients are calculated using the self-standarized covariates (transformed to be of mean 0 and variance 1), instead of the data in its original scale. Compared to the RQ estimates, we can see the BEL estimates of each coefficient at different quantile levels/sites are always shrinked towards some common parameters, and the magnitudes of the shrinkage are distinct for different coefficients. The corresponding standard errors of the coefficients estimates for BEL and RQ are listed in Table 6, which suggests the BEL estimates are more efficient.

Table 5: The table provides the coefficients estimates of the BEL and RQ for each site at $\tau = 0.95, 0.99$, respectively.

BEL	$\tau = 0.95$				$\tau = 0.99$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	0.999	0.331	0.037	0.393	1.626	0.352	0.115	0.495
s_2	0.945	0.319	0.045	0.390	1.554	0.359	0.135	0.563
s_3	1.003	0.336	0.032	0.377	1.601	0.419	0.087	0.466
s_4	1.026	0.325	0.024	0.374	1.579	0.370	0.060	0.483
s_5	1.004	0.359	0.031	0.389	1.532	0.383	0.116	0.514
RQ	$\tau = 0.95$				$\tau = 0.99$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	0.989	0.329	0.054	0.364	1.667	0.412	0.190	0.535
s_2	0.916	0.312	0.085	0.294	1.563	0.266	0.140	0.606
s_3	1.022	0.360	0.024	0.401	1.602	0.406	0.084	0.620
s_4	1.042	0.306	-0.011	0.433	1.585	0.376	0.056	0.489
s_5	1.014	0.397	0.021	0.386	1.536	0.414	0.151	0.492
ASQR	$\tau = 0.95$				$\tau = 0.99$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	0.989	0.344	0.084	0.298	1.276	0.366	0.115	0.355
s_2	0.978	0.338	0.091	0.304	1.263	0.358	0.125	0.362
s_3	1.040	0.361	0.067	0.343	1.327	0.383	0.095	0.406
s_4	1.057	0.353	0.025	0.358	1.347	0.373	0.047	0.421
s_5	1.033	0.385	0.055	0.358	1.323	0.411	0.082	0.424

Table 6: The table provides the standard errors of the coefficients estimates given in Tabel 5.

BEL	$\tau = 0.95$				$\tau = 0.99$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	0.023	0.024	0.020	0.035	0.064	0.043	0.054	0.050
s_2	0.035	0.029	0.024	0.048	0.091	0.094	0.060	0.079
s_3	0.025	0.036	0.025	0.038	0.101	0.078	0.070	0.087
s_4	0.029	0.031	0.026	0.041	0.061	0.061	0.051	0.060
s_5	0.029	0.032	0.025	0.041	0.057	0.040	0.062	0.086
RQ	$\tau = 0.95$				$\tau = 0.99$			
	a	b_1	b_2	b_3	a	b_1	b_2	b_3
s_1	0.038	0.041	0.031	0.054	0.080	0.057	0.121	0.149
s_2	0.026	0.028	0.032	0.045	0.120	0.116	0.118	0.201
s_3	0.038	0.059	0.034	0.050	0.084	0.082	0.120	0.145
s_4	0.049	0.065	0.061	0.078	0.073	0.064	0.117	0.140
s_5	0.042	0.045	0.051	0.074	1.536	0.414	0.151	0.492

Bibliography

Chib, S. and E. Greenberg (1995). Understanding the metroplis-hasting algorithm. *The American Statistician* **49**, 327–335.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996). *Markov chain montecarlo in practice*. London: Chapman and Hall.

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

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