

## BAYESIAN EMPIRICAL LIKELIHOOD FOR QUANTILE REGRESSION<sup>1</sup>

BY YUNWEN YANG AND XUMING HE

*Drexel University and University of Michigan*

Bayesian inference provides a flexible way of combining data with prior information. However, quantile regression is not equipped with a parametric likelihood, and therefore, Bayesian inference for quantile regression demands careful investigation. This paper considers the Bayesian empirical likelihood approach to quantile regression. Taking the empirical likelihood into a Bayesian framework, we show that the resultant *posterior* from any fixed prior is asymptotically normal; its mean shrinks toward the true parameter values, and its variance approaches that of the maximum empirical likelihood estimator. A more interesting case can be made for the Bayesian empirical likelihood when informative priors are used to explore commonality across quantiles. Regression quantiles that are computed separately at each percentile level tend to be highly variable in the data sparse areas (e.g., high or low percentile levels). Through empirical likelihood, the proposed method enables us to explore various forms of commonality across quantiles for efficiency gains. By using an MCMC algorithm in the computation, we avoid the daunting task of directly maximizing empirical likelihood. The finite sample performance of the proposed method is investigated empirically, where substantial efficiency gains are demonstrated with informative priors on common features across several percentile levels. A theoretical framework of shrinking priors is used in the paper to better understand the power of the proposed method.

**1. Introduction.** Quantile regression is a statistical methodology for the modeling and inference of conditional quantile functions. Following [Koenker and Bassett \(1978\)](#), we specify the  $\tau$ th conditional quantile function of  $Y \in \mathbb{R}$  given  $X \in \mathbb{R}^{p+1}$  as

$$(1.1) \quad Q_\tau(Y|X) = X^\top \beta(\tau),$$

where  $\tau \in (0, 1)$ , and  $\beta(\tau)$  typically includes an intercept. Quantile modeling of this type can be estimated for one or several percentile levels; we refer the details on computation and basic asymptotic theory to [Koenker \(2005\)](#). Inferential methods for quantile regression have been developed by a number of researchers, including [Gutenbrunner and Jurečková \(1992\)](#), [Horowitz \(1998\)](#), [Chen et al. \(2008\)](#)

---

Received September 2011; revised February 2012.

<sup>1</sup>Supported in part by NSF Grants DMS-07-24752 and DMS-10-07396.

*MSC2010 subject classifications.* Primary 62J05, 62F12; secondary 62G20.

*Key words and phrases.* Efficiency, empirical likelihood, high quantiles, prior, posterior.

and [Kocherginsky, He and Mu \(2005\)](#). The  $\tau$ -specific models allow for great flexibility, as  $\beta(\tau)$  for upper or lower quantiles can be distinct from central trends, but the quantile estimates are highly variable in data-sparse areas. Taking advantage of some commonality in the quantile coefficients  $\beta(\tau)$  across  $\tau$  can provide a desirable balance in the bias-variance tradeoff. In this article, we consider using prior information on  $\beta(\tau)$  across several  $\tau$  values. For example, a common slope assumption for  $\tau$  near 1 can improve the efficiency of high quantile estimation. Other forms of informative priors on  $\beta(\tau)$  may achieve a similar goal. Bayesian methods are a natural way of combining data with prior information. The main difficulty in putting the Bayesian method to work for quantile regression is that the model on  $Q_\tau(Y|X)$  for one or any small number of  $\tau$  values does not specify a parametric likelihood, which is needed in the Bayesian framework.

Several authors have attempted to use a working likelihood in the Bayesian quantile regression framework. [Kottas and Gelfand \(2001\)](#) and [Kottas and Krnjajić \(2009\)](#) used Dirichlet process mixture models. [Reich, Bondell and Wang \(2008\)](#) assumed the error distributions to be an infinite mixture of normals. [Dunson and Taylor \(2005\)](#) used an approximate method based on the Jefferey's substitution likelihood for quantiles. [Yu and Moyeed \(2001\)](#), [Geraci and Bottai \(2007\)](#) and [Yue and Rue \(2011\)](#), among others, chose (asymmetric) Laplace distributions as the working likelihood. Those approaches, mostly tailored toward a specific percentile level of  $\tau$ , use Markov chain Monte Carlo algorithms as a useful means of computation. Work of these authors provided numerical evidence that a Bayesian approach to quantile regression has merits.

In this article, we focus on estimating several quantiles together. To do so, we use the empirical likelihood (EL), introduced by [Owen \(1988\)](#), to incorporate quantile regression into a (pseudo-) Bayesian framework. Empirical likelihood makes it easy to model several quantiles at the same time, allowing informative priors on  $\beta(\tau)$  across  $\tau$  to be utilized. Statistical inference based on empirical likelihood is known to enjoy good asymptotic properties, especially if the EL is associated with moment restrictions of sufficient smoothness. [Molanes Lopez, Van Keilegom and Veraverbeke \(2009\)](#) considered the EL with nonsmooth estimating equations under a general setting. A more comprehensive review about empirical likelihood can be found in [Owen \(2001\)](#) and [Chen and Van Keilegom \(2009\)](#). Since the moment restrictions for quantiles are placed on nonsmooth functions, some researchers, including [Chen and Hall \(1993\)](#), [Whang \(2006\)](#) and [Otsu \(2008\)](#) proposed using smoothed versions of the quantile estimating equations. The smoothed EL is further extended to weakly dependent processes in [Chen and Wong \(2009\)](#) and censored data in [Ren \(2008\)](#). We choose to focus on the exact moment conditions for quantiles without the complication of choosing a smoothing parameter. Those moment conditions are also used in [Wang and Zhu \(2011\)](#) and [Kim and Yang \(2011\)](#) for clustered data. In addition, we use a standard MCMC algorithm to explore the posterior, to avoid the daunting task of directly maximizing the empirical likelihood. In fact, the EL function given any proposed

parameters is relatively easy to compute, even though the EL-maximization is notoriously difficult, even in modest dimensions.

The empirical likelihood is not a likelihood in the usual sense, so the validity of the resultant posterior does not follow automatically from the Bayes formula. Lazar (2003) discussed the validity of inference for the Bayesian empirical likelihood (BEL) approach based on earlier work of Monahan and Boos (1992). Schennach (2005) and Lancaster and Jun (2010) considered Bayesian exponentially tilted empirical likelihood (ETEL), which can be viewed as a nonparametric Bayesian procedure with noninformative priors on the space of distributions. Lancaster and Jun (2010) further considered Bayesian ETEL in quantile regression. For the inference of population means, Fang and Mukerjee (2006) investigated the asymptotic validity and accuracy of the Bayesian credible regions, and furthermore, Chang and Mukerjee (2008) showed that EL admits posterior based inference with the frequentist asymptotic validity, but many of its variants do not enjoy this property. In this article, we establish the asymptotic distributions of the posterior from the BEL approach for quantile regression, which enable us to evaluate efficiency gains from informative priors. Chernozhukov and Hong (2003) discussed the asymptotic properties of the quasi-posterior distributions defined as transformations of general statistical criterion functions. In our work, we establish the asymptotic distributions of the posterior from the BEL approach for quantile regression, and are particularly interested in the interaction of informative priors and empirical likelihood on the asymptotic distribution of the posterior, which enables us to evaluate efficiency gains from informative priors.

Ideas similar to BEL have been used by other researchers. Yin (2009) proposed the Bayesian generalized method of moments (GMM), which can be adapted to quantile estimation. Hahn (1997) considered Bayesian bootstrap in quantile regression. Note that the GMM estimators are also defined through moment restrictions, which allow them to model multiple quantiles jointly. The GMM estimators, the maximum empirical likelihood estimators (MELE) and some other EL-type estimators generally have the same asymptotic distributions, but possibly different higher order asymptotic properties; see Newey and Smith (2004) and Schennach (2007). As discussed in Newey and Smith (2004), the empirical likelihood approach has advantages over the GMM estimators. Unlike GMM, the (asymptotic) bias of the MELE does not grow with the number of moment restrictions. Furthermore, the efficiency of the GMM estimator relies on a covariance matrix estimate for the estimating equations, which could be ill-conditioned when estimating multiple quantiles.

The recent development of Bayesian (conditional) density estimation using mixture models enables nonparametric regression models on all quantiles simultaneously; see Müller, Erkanli and West (1996), Müller and Quintana (2004), Dunson, Pillai and Park (2007) and Chung and Dunson (2009), among others. Theoretical results about posterior consistency can be found in Pati, Dunson and Tokdary

(2010), Norets and Pelenis (2010) and the references therein. In contrast, our proposed BEL approach targets a small number of selected quantiles without the need to model the entire conditional distributions. A novel part of our work is its ability of employing informative priors to explore commonality across quantiles for efficiency gains.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed BEL approach for quantile regression, and discuss model assumptions, method of computation and use of informative priors. The asymptotic properties on the BEL posteriors are provided in Section 3 for both fixed and a class of shrinking priors. The theoretical framework of shrinking priors enables us to understand the efficiency gains of the BEL approach over traditional methods. Section 4 demonstrates the finite sample performance of the BEL approach through Monte Carlo simulations with a focus on frequentist properties of BEL posterior intervals, and efficiency gains from informative priors. In Section 5, we use a real data example to show that the BEL approach can be used as a useful statistical downscaling method for the projection of high quantiles of temperature from large scale climate models to a local scale. Some concluding remarks are given in Section 6. The technical details to support the theorems in Section 3 are provided in the Appendix.

**2. Bayesian empirical likelihood for quantile regression.** In this section we introduce the Bayesian empirical likelihood approach for quantile regression. We begin with notation and definitions of the underlying models and moment restrictions. Let  $D = \{(X_i, Y_i), i = 1, \dots, n\}$  be a random sample from the following quantile regression model:

$$(2.1) \quad Q_\tau(Y|X) = X^\top \beta_0(\tau),$$

where  $X \in \mathbb{R}^{p+1}$  is composed of an intercept term and  $p$  covariates. We assume that the distribution of the  $p$  covariates,  $G_X$ , has a bounded support  $\mathcal{X}$ . If the design points are nonstochastic, 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. The unknown function  $\beta_0(\tau)$ , if specified over all  $\tau \in (0, 1)$ , describes the entire conditional distribution of  $Y$  given  $X$ , which is denoted as  $F_X$  in the rest of the paper. We consider the problem of estimating  $k$  quantiles at  $\tau_1 < \tau_2 < \dots < \tau_k$ , and let  $\zeta_0 = (\beta_0(\tau_1), \dots, \beta_0(\tau_k))$  be the true parameter of interest in  $\mathbb{R}^{k(p+1)}$ . In most applications,  $k$  is a small integer. To estimate  $\zeta_0$ , we use  $k(p + 1)$  dimensional estimating functions  $m(X, Y, \zeta)$ , where  $\zeta = (\beta(\tau_1), \dots, \beta(\tau_k))$  and the components of  $m$  are

$$(2.2) \quad m_{dk+j}(X, Y, \zeta) = \psi_{\tau_{d+1}}(Y - X^\top \beta(\tau_{d+1}))X_j$$

for  $d = 0, 1, \dots, k - 1, j = 0, 1, \dots, p$ , with

$$\psi_\tau(u) = \begin{cases} 1_{\{u < 0\}} - \tau, & u \neq 0, \\ 0, & u = 0 \end{cases}$$

being the quantile score function, where  $1_{\{A\}}$  is an indicator function on the set  $A$ . We hasten to add that  $\zeta$  may contain fewer than  $k(p + 1)$  unknown parameters when some common parameters are present in  $\beta(\tau)$  at different quantile levels. In such cases, the number of moment restrictions exceeds the number of unknown parameters. As shown in [Qin and Lawless \(1994\)](#) for smooth estimating functions, the maximum empirical likelihood estimator attains the optimal asymptotic efficiency subject to those moment conditions. We expect the same for quantile estimating functions.

For any proposed  $\zeta$ , its profile empirical likelihood ratio is given by

$$(2.3) \quad \mathcal{R}(\zeta) = \max \left\{ \prod_{i=1}^n (n\omega_i) \mid \sum_{i=1}^n \omega_i m(X_i, Y_i, \zeta) = 0, \omega_i \geq 0, \sum_{i=1}^n \omega_i = 1 \right\}.$$

By a standard Lagrange multiplier argument, we have

$$\mathcal{R}(\zeta) = \prod_{i=1}^n \{n\omega_i(\zeta)\},$$

where the weights  $\omega_i(\zeta) = [n\{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)\}]^{-1}$ , and the Lagrange multiplier  $\lambda_n(\zeta)$  satisfies the following equation:

$$\sum_{i=1}^n \frac{m(X_i, Y_i, \zeta)}{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)} = 0.$$

As discussed in [Chen, Sitter and Wu \(2002\)](#) and [Qin and Lawless \(1994\)](#), the existence and uniqueness of  $\lambda_n(\zeta)$  are guaranteed when the following two conditions are satisfied:

(C1) The vector  $\underline{0} \in \mathbb{R}^{k(p+1)}$  is within the convex hull of  $\{m(X_i, Y_i, \zeta), i = 1, \dots, n\}$ .

(C2) The matrix  $\sum_{i=1}^n \{m(X_i, Y_i, \zeta)m(X_i, Y_i, \zeta)^\top\}$  is positive definite.

The first condition (C1) actually provides a feasible region of  $\zeta$  supported by the observations  $D$ , in which the proposed  $\zeta$  has a valid empirical likelihood value. If  $Y_i < X_i^\top \beta(\tau_d)$  at some  $\tau_d$  for all  $i = 1, \dots, n$ , this proposed  $\zeta$  will violate the first condition, and then we regard its empirical likelihood value as 0. The second condition (C2) requires the set of estimating functions to be linearly independent. Noting that

$$E\{m(X, Y, \zeta_0)m(X, Y, \zeta_0)^\top\} = \Psi \otimes E(XX^\top),$$

where the elements of the  $\Psi$  matrix are  $\Psi_{ij} = \tau_i \wedge \tau_j - \tau_i \tau_j$ , the second condition is generally satisfied for  $\zeta$  near  $\zeta_0$ , as long as  $E(XX^\top)$  is positive definite.

For any proposed  $\zeta$ , consider its empirical likelihood function  $\mathcal{R}(\zeta)/n^n = \prod_{i=1}^n \omega_i(\zeta)$ . With a prior specification  $p_0(\zeta)$  on the parameter  $\zeta$ , we can formally have the posterior density

$$(2.4) \quad p(\zeta|D) \propto p_0(\zeta) \times \mathcal{R}(\zeta).$$

We call  $p(\zeta|D)$  the posterior distribution from the BEL approach. This can be viewed as a misnomer, chosen for the sake of convenience, because it is not really a posterior in the strict sense. Lazar (2003) proposed a procedure to check whether the empirical likelihood is valid for posterior inference based on the criteria provided in Monahan and Boos (1992). In this paper, we focus on the asymptotic properties of the posterior distribution (2.4), and establish its frequentist validity by first-order asymptotics.

Finding the maximum empirical likelihood estimator is a daunting task computationally, because the objective function is generally multi-modal. However, the value of the empirical likelihood ratio  $\mathcal{R}(\zeta)$  is relatively easy to compute given  $\zeta$ , which makes the Metropolis–Hastings algorithm, as given in Hastings (1970), feasible for sampling from the posterior. By choosing a proper prior, the posterior in (2.4) is also proper. Therefore, by checking the detailed balance equation and Theorem 4.2 in Gilks, Richardson and Spiegelhalter (1996), the distribution of the MCMC sampler converges to the posterior in (2.4). More discussions on computation efficiency can be referred to Chernozhukov and Hong (2003). A Bayesian framework has its own merits in applications where informative priors on  $\beta(\tau)$  might be more appropriate than a strict functional relationship on some of the parameters. For example, we may believe that the slopes in  $\beta(\tau_1)$  are roughly the same as in  $\beta(\tau_2)$ . Imposing strict equalities to reduce the number of unknown parameters in  $\zeta$  might be hard to justify, but an informative prior on the difference of two neighboring  $\beta(\tau)$  can help regularize quantile estimation.

By using a standard Metropolis–Hastings algorithm for a given prior  $p_0(\zeta)$ , we may use the average of the Markov chain on  $\zeta$  as an estimate of  $\zeta$ , when the posterior looks close to normal; otherwise, we suggest using the mode of the posterior, which maximizes (2.4). In the empirical investigations in Sections 4 and 5, we use the posterior mode as the estimates.

In our empirical investigations, we have found that the posterior mode of the slope parameters behaves well, but the intercept parameter in each  $\beta(\tau)$  can be better estimated in small samples if the following strategy is followed. Suppose that  $\beta(\tau) = (\beta_I(\tau), \beta_S(\tau))$ , where  $\beta_I(\tau)$  corresponds to the intercept, and  $\beta_S(\tau)$  corresponds to the slope. Let  $\hat{\beta}_I(\tau)$  be the posterior mode/mean obtained from the MCMC chain, we use the modified estimate  $\hat{\beta}_I(\tau)$  as the  $\tau$ th sample quantile of  $Y_i - X_{S_i}^\top \hat{\beta}_S(\tau)$ , where  $X_{S_i}$  corresponds to  $X_i$  excluding the intercept term. This modification does not alter the asymptotic distributions of the  $\hat{\beta}(\tau)$ . In the rest of the paper, we always use this modification in the BEL estimate of quantile regression.

**3. Asymptotic properties of BEL.** In this section, we provide an asymptotic justification of the BEL estimator for quantile regression by deriving the limiting behavior of the posterior distribution as  $n \rightarrow \infty$ . One noticeable point about the estimating equations (2.2) is that they involve indicator functions, so the resulting empirical likelihood ratio is nonsmooth in  $\zeta$ . An asymptotic normality of

the posterior distribution in the Bayesian empirical likelihood context was derived heuristically in Lazar (2003) for smooth estimating equations. We rely on empirical process theory to establish a similar result for the BEL here.

As the first step, we shall prove the consistency of the maximum empirical likelihood estimator (MELE), which is a necessary condition for the asymptotic normality of the posterior.

3.1. *Consistency of the MELE.* We assume that the true parameter  $\zeta_0$  falls into a compact set of the parameter space, and the optimization is carried out over this compact set. For notational convenience, let

$$\hat{\zeta} = \arg \max\{\mathcal{R}(\zeta)\}$$

be the MELE, whose dependence on  $n$  and the compact set on  $\zeta$  have been suppressed in our notation. Note that the maximum empirical likelihood estimate might not be unique, but the result here applies to any maximizer of the empirical likelihood ratio, and all the maximizers converge to the same asymptotic value.

The estimating functions  $m(X, Y, \zeta)$  are not smooth in  $\zeta$ , but it is worth noting that the expectations of  $m(X, Y, \zeta)$  and the empirical likelihood function are sufficiently smooth under the following assumptions.

ASSUMPTION 3.1. There exists a neighborhood  $\mathcal{N}$  of  $\zeta_0$  such that  $P(\mathcal{R}(\zeta) > 0) \rightarrow 1$  for any  $\zeta \in \mathcal{N}$ , as  $n \rightarrow \infty$ .

ASSUMPTION 3.2. The distribution function  $G_X$  has bounded support  $\mathcal{X}$ .

ASSUMPTION 3.3. The conditional distribution  $F_X(t)$  of  $Y$  given  $X$  is twice continuously differentiable in  $t$  for all  $X \in \mathcal{X}$ .

ASSUMPTION 3.4. At any  $X \in \mathcal{X}$ , the conditional density function  $F'_X(t) = f_X(t) > 0$  for  $t$  in a neighborhood of  $F_X^{-1}(\tau_d)$  for each  $d = 1, \dots, k$ .

ASSUMPTION 3.5.  $E\{m(X, Y, \zeta_0)m(X, Y, \zeta_0)^\top\}$  is positive definite.

Assumption 3.1 is to guarantee that the interior of the convex hull of  $\{m(X_i, Y_i, \zeta) : i = 1, \dots, n\}$  for  $\zeta \in \mathcal{N}$  contains the vector of zeros with probability tending to one. By (2.1),  $F_X(X^\top \beta_0(\tau_d)) = \tau_d$  for any  $d \leq k$  and  $X \in \mathcal{X}$ . Therefore, for each  $d$ ,  $\beta_0(\tau_d)$  is a solution to  $E\{m_{dk+j}(X, Y, \zeta)\} = 0$ ,  $j = 0, \dots, p$ . Under Assumption 3.4,  $\beta_0(\tau_d)$  is indeed the unique solution. Correspondingly,  $\zeta_0$  is the unique solution for  $E\{m(X, Y, \zeta)\} = 0$ .

THEOREM 3.1. Under Assumptions 3.1–3.5, the MELE  $\hat{\zeta}$  is a consistent estimator of  $\zeta_0$ .

The proof of Theorem 3.1 is sketched in the Appendix. The basic idea is to check the conditions for consistency appearing in Theorem 5.7 of van der Vaart (1998). Because those conditions require some uniform convergence properties for collections of functions involving  $m(X, Y, \zeta)$ , we use the empirical process theory as a natural tool.

3.2. *Asymptotic normality of the posterior.* To validate the asymptotic normality of the posterior distribution (2.4), we make one more assumption.

ASSUMPTION 3.6.  $\log\{p_0(\zeta)\}$  has bounded first derivative in a neighborhood of  $\zeta_0$ .

Then we have the following theorem.

THEOREM 3.2. *Under Assumptions 3.1–3.6, the posterior density of  $\zeta$  has the following expansion on any sequence of sets  $\{\zeta : \zeta - \zeta_0 = O(n^{-1/2})\}$ :*

$$(3.1) \quad p(\zeta|D) \propto \exp\{-\frac{1}{2}(\zeta - \hat{\zeta})^\top J_n(\zeta - \hat{\zeta}) + R_n\},$$

where  $\hat{\zeta}$  is the MELE,

$$\begin{aligned} J_n &= nV_{12}^\top V_{11}^{-1}V_{12}, \\ V_{11} &= \Psi \otimes E(XX^\top), \\ V_{12} &= -\left. \frac{\partial E\{m(X, Y, \zeta)\}}{\partial \zeta} \right|_{\zeta=\zeta_0} \end{aligned}$$

and  $R_n = o_p(1)$ . When  $J_n$  is positive definite, we have  $J_n^{1/2}(\zeta - \hat{\zeta})$  converging in distribution to  $N(0, I)$ .

There are clear similarities between Theorem 3.2 here and Theorem 1 of Lazar (2003) for smooth estimating equations. We have considered fixed priors, a common scenario in the literature, where the limiting posterior distributions of  $\zeta$  are the same as the limiting sampling distribution of the MELE [cf. Qin and Lawless (1994)]. An important remark follows.

REMARK 3.1. The results in Theorem 3.2 apply to the cases where the dimension of  $\zeta$  is smaller than the dimension of the estimating functions  $m(X, Y, \zeta)$ . For  $\zeta$  with a reduced dimensionality, the definition of  $V_{12}$  is taken to be the derivative with respect to the reduced parameter vector.

Asymptotically, Theorem 3.2 justifies the use of the BEL approach for quantile regression with respect to frequentist properties. When  $f_X(X^\top \beta_0(\tau_d)) = f_{\tau_d}$  is

constant for all  $X$ , which is true for homoscedastic error models, we can simplify  $V_{12}$  to

$$V_{12} = -\text{diag}(f_{\tau_d})_{d=1,\dots,k} \otimes E(\mathbf{X}\mathbf{X}^\top),$$

if  $\zeta$  is of  $k(p + 1)$  dimensions. Because  $V_{11} = \Psi \otimes E(\mathbf{X}\mathbf{X}^\top)$ , the resultant asymptotic variance of the posterior quantity,  $J_n^{-1}$ , is equivalent to the asymptotic variance of the usual quantile regression (RQ) estimates, as proposed in [Koenker and Bassett \(1978\)](#). This property is not shared by all working likelihoods. If  $\zeta$  is of lower dimensions, the posterior variance no longer takes the same form, and improvements in the asymptotic variances over RQ become possible.

**REMARK 3.2.** An improper prior cannot guarantee a proper posterior distribution. In fact, the posterior will be improper for flat priors on  $\zeta$  in the BEL approach, and therefore we should avoid using flat priors on  $\zeta$ .

Next, we consider a more interesting scenario where the prior distribution shrinks with  $n$ . In this case, we use  $p_{0,n}(\zeta)$  as priors, and make the following assumption.

**ASSUMPTION 3.7.** The logarithm of the prior density  $p_{0,n}(\zeta)$  is twice continuously differentiable, with the prior mode  $\zeta_{0,n} = O(1)$ , and the matrix  $J_{0,n} = -\frac{\partial^2 \log\{p_{0,n}(\zeta)\}}{\partial \zeta^2} \Big|_{\zeta=\zeta_{0,n}} = O(n)$ .

By Assumption 3.7,  $\log\{p_{0,n}(\zeta)\}$  can be Taylor expanded up to the quadratic term as follows.

$$(3.2) \quad \begin{aligned} \log\{p_{0,n}(\zeta)\} &= \log\{p_{0,n}(\zeta_{0,n})\} \\ &\quad - \frac{1}{2}(\zeta - \zeta_{0,n})^\top J_{0,n}(\zeta - \zeta_{0,n}) + o(\|\zeta - \zeta_{0,n}\|^2). \end{aligned}$$

Then we have the following result.

**THEOREM 3.3.** Under Assumptions 3.1–3.5 and 3.7, the posterior density of  $\zeta$  has the following expansion on any sequence of sets  $\{\zeta : \|\zeta - \zeta_0\| = O(n^{-1/2})\}$ :

$$(3.3) \quad p(\zeta | D) \propto \exp\left\{-\frac{1}{2}(\zeta - \theta_{\text{post}})^\top J_n(\zeta - \theta_{\text{post}}) + R_n\right\},$$

where

$$\begin{aligned} J_n &= J_{0,n} + nV_{12}^\top V_{11}^{-1} V_{12}, \\ \theta_{\text{post}} &= J_n^{-1}(J_{0,n}\zeta_{0,n} + nV_{12}^\top V_{11}^{-1} V_{12}\hat{\zeta}) \end{aligned}$$

and  $R_n = o_p(1)$ .

Compared to Theorem 3.2, the additional term  $J_{0,n}$  in both  $J_n$  and  $\theta_{\text{post}}$  in Theorem 3.3 provides a balanced view of when and how an informative prior can complement the likelihood in large samples. When  $J_{0,n} = o_p(n)$ , the posterior expansion in Theorem 3.3 is the same as that of Theorem 3.2, so the empirical likelihood will dominate the prior information. Obviously, if  $J_{0,n}$  increases at a faster rate than  $n$ , the prior will dominate the empirical likelihood. For the more interesting case where  $J_{0,n}$  increases at the rate of  $n$ , the BEL produces a consistent estimate of  $\zeta_0$  if  $\|\zeta_{0,n} - \zeta_0\| = o_p(1)$ ; otherwise,  $\theta_{\text{post}}$  may not converge to  $\zeta_0$  in probability, that is, a bias may be introduced, but the variance is reduced. In the latter case, the posterior in (3.3) does not directly lead to asymptotically valid posterior inference. However, noting that  $J_n = J_{0,n} + nV_{12}^\top V_{11} V_{12}$  and  $J_{0,n}$  is known, the MCMC chain provides an estimate of the matrix  $nV_{12}^\top V_{11} V_{12}$ , which is what we need to obtain asymptotically valid confidence intervals.

Shrinking priors are relevant when the informative priors are constructed from data of a secondary source or when the hypothesis on common slope parameters are not rejected by a statistical test.

In Theorem 3.3, the prior mode  $\zeta_{0,n}$  plays a role in the posterior mean, which could be undesirable. For shrinking toward common slopes, we can use a class of priors that eliminate the bias due to a mis-specified prior mode when the common slope assumption holds. For each  $d = 1, \dots, k$ , let  $g_d$  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 a prior on  $\zeta$  as

$$(3.4) \quad \begin{aligned} \Omega^{-1/2}(\beta(\tau_1) - \beta_{p,0}) &\sim g_1 \quad \text{and} \\ \Sigma_d^{-1/2}(\beta(\tau_d) - \beta(\tau_1)) | \beta(\tau_1) &\sim g_d \quad \text{for } d = 2, \dots, k \end{aligned}$$

for any location vector  $\beta_{p,0}$  and scatter matrices  $\Omega$  and  $\Sigma_d$  of appropriate dimensions. They vary with  $n$  in our theory, but we have suppressed the dependence in notation. If we write

$$\Sigma_d = \begin{pmatrix} \Sigma_{d,I} & \mathbf{0}^\top \\ \mathbf{0} & \Sigma_{d,S} \end{pmatrix},$$

where  $\Sigma_{d,I}$  and  $\Sigma_{d,S}$  represent the components of  $\Sigma_d$  corresponding to the intercept and the slope parameters in  $\beta(\tau_d)$ , respectively, for  $d = 2, \dots, k$ , we now assume

$$(3.5) \quad \|\Omega^{-1}\| = O(\epsilon_n), \quad \|\Sigma_{d,I}^{-1}\| = O(\epsilon_n) \quad \text{and} \quad \|\Sigma_{d,S}\| = O(n^{-1})$$

for some sequence  $\epsilon_n = o(n)$ . We have the following corollary.

**COROLLARY 3.4.** *Suppose that the same conditions of Theorem 3.3 hold. If the slope parameters in  $\zeta_0$  are the same at  $\tau_1, \dots, \tau_k$ , and a (shrinking) prior satisfying (3.4) and (3.5) is used, the posterior mean of Theorem 3.3 becomes  $\theta_{\text{post}} = \zeta_0 + O_p(\epsilon_n/n + n^{-1/2})$ .*

Clearly, Corollary 3.4 indicates that the center of the posterior is asymptotically unbiased for  $\zeta_0$  with common slopes regardless of what the prior mode  $\beta_{p,0}$  is for  $\beta(\tau_d)$ . All we need is to allow the prior variances of the slope differences to be in the order of  $1/n$ , but the prior variances of the other parameters increasing with  $n$ . The idea of constructing such a class of shrinking priors applies more broadly than what we have considered here with common slopes, but in our empirical work to be reported, only independent normal and t-distributions will be used as  $g_d$ .

**4. Simulation studies.** In this section, we use Monte Carlo simulations to investigate the performance of the BEL methods (coverage probability and estimation efficiency) from the frequentist viewpoint. We use the following notation to distinguish BEL estimators with various priors on the slope parameters. The usual quantile regression estimation at each  $\tau$  will be denoted simply as RQ.

- BEL.s: BEL estimators of single quantiles using moment restrictions at each  $\tau$ .
- BEL.c: BEL estimators based on joint moment restrictions assuming a common slope parameter at several  $\tau$ 's.
- BEL.n: BEL estimators based on joint moment restrictions assuming that the differences in slope parameters across  $\tau$ 's have normal priors with zero mean and "small" variances.

4.1. *Coverage properties.* We first take a brief look at the coverage probabilities of the posterior credible intervals obtained under BEL.s. To see the impact of empirical likelihood, we also include in the comparison two other Bayesian methods, one based on the true parametric likelihood, and the other based on a working likelihood.

The data are generated from  $Y_i = \beta_I + \beta_S(X_i - 2) + e_i$  ( $i = 1, \dots, n$ ), where the true parameters are  $\beta_I = 2, \beta_S = 1, X_i$  and  $e_i$  are independently generated from the chi-square distribution with 2 degrees of freedom and  $N(0, 4)$ , respectively. We are interested in estimating the median regression coefficients  $\beta_I(0.5)$  and  $\beta_S(0.5)$ . Independent priors of  $N(0, 100^2)$ , are used on both parameters. We use the 2.5th and the 97.5th percentiles of the Markov chain from BEL.s for  $\tau = 0.5$  to form 95% interval estimates for the parameters. The simulation study uses three different sample sizes  $n = 100, 400, 1600$  to see whether the intervals have desirable coverage probabilities for modestly large  $n$ .

In addition to BEL.s, we include two other Bayesian methods:

- BTL: the Bayesian method using the true likelihood

$$\prod_{i=1}^n \sigma^{-1} \phi \left\{ \frac{y_i - \beta_I(0.5) - \beta_S(0.5)(x_i - 2)}{\sigma} \right\},$$

where  $\phi$  is the density of the standard normal distribution.

- BDL: a pseudo Bayesian method using the Laplace density as the working likelihood

$$\prod_{i=1}^n \tilde{\sigma}^{-1} \exp\left\{-\frac{|y_i - \beta_I(0.5) - \beta_S(0.5)(x_i - 2)|}{2\tilde{\sigma}}\right\},$$

where  $\tilde{\sigma}$  is estimated by the mean of the absolute residuals from the RQ estimate at  $\tau = 0.5$ .

Similar MCMC sampling algorithms are used for all the three methods. The BTL method can be viewed as a yardstick for any MCMC based method, because it uses the true parametric likelihood under the model, which is generally unknown in practice. The reason to consider BDL is that the exponential component of its working likelihood is the objective function of median regression. The BDL method has been used earlier by [Yu and Moyeed \(2001\)](#) among others, but in our empirical work, we have chosen to use a fixed value of  $\sigma$  in BDL, because we have found that the MCMC chains have better mixing properties without including  $\sigma$  as an unknown parameter. A sensible value of  $\sigma$  to use in BDL is the RQ-based scale estimate. Table 1 provides the average coverage probability and average length information for each of the three methods over 1000 samples at each choice of  $n$ .

This simple simulation study shows that as the sample size increases, the posterior intervals obtained from BEL.s and BTL approach the nominal levels 95%, although the convergence is not as fast as we might have expected. Because the underlying model has i.i.d. normal errors, the asymptotic relative efficiency of BEL.s and BDL are approximately 67% of BTL, which helps explain the differences in the interval lengths. We also note that BEL.s outperforms BDL by the frequentist measures, even after we fixed the scale parameter in BDL.

TABLE 1

*Comparison of 95% posterior intervals of the median regression parameters from three methods: (1) BEL.s, (2) BTL based on the true likelihood and (3) BDL based on a working Laplace likelihood. The coverage probability and lengths of the posterior intervals are computed over 1000 data sets of sample sizes  $n = 100, 400$  and  $1600$*

$n$		Coverage of 95% CI			Length of 95% CI		
		BEL.s	BTL	BDL	BEL.s	BTL	BDL
100	$\beta_I(0.5)$	0.97	0.94	0.98	1.06	0.80	1.11
	$\beta_S(0.5)$	0.98	0.94	0.98	0.58	0.41	0.58
400	$\beta_I(0.5)$	0.97	0.95	0.98	0.43	0.40	0.55
	$\beta_S(0.5)$	0.94	0.95	0.98	0.22	0.20	0.28
1600	$\beta_I(0.5)$	0.96	0.96	0.97	0.25	0.21	0.28
	$\beta_S(0.5)$	0.96	0.96	0.98	0.13	0.10	0.14

Similar phenomena were observed in the interval estimation for other quantiles and under several other error distributions, but we skip the details. A more extensive report on estimation efficiency is given in the next subsection.

4.2. *Efficiency of BEL under various priors.* In this section, we investigate the estimation efficiency of BEL.s, BEL.c and BEL.n for  $\zeta$  at different percentile levels, where the posterior modes are taken as the parameter estimates. The estimation efficiency is measured by the estimated mean squared error (MSE), with data generated from the following four models:

- Model 1:  $Y = X + Z + e$ , where  $X \sim \chi^2(2)$ ,  $Z/2 \sim \text{Bernoulli}(0.5)$  and  $e \sim N(0, 4)$ , with  $X$ ,  $Z$  and  $e$  being mutually independent;
- Model 2: same as Model 1 except that  $\log(e) \sim N(0, 1)$ ;
- Model 3:  $Y = X + Z + (X/2 + 1)e$ , where  $X \sim \chi^2(2)$ ,  $Z/2 \sim \text{Bernoulli}(0.5)$  and  $e \sim N(0, 4)$ , with  $X$ ,  $Z$  and  $e$  being mutually independent;
- Model 4: same as Model 3 except that  $\log(e) \sim N(0, 1)$ .

These models include two covariates, of which  $X$  is continuous, and  $Z$  is binary. Models 1 and 2 assume homoscedastic errors, and Models 3 and 4 allow the error distributions to depend on  $X$ . We use  $b_x(\tau)$ ,  $b_z(\tau)$  to denote the two slope parameters, and consider the adjusted intercept  $a(\tau)$  as the fitted value of the  $\tau$ th quantile at the sample mean of  $(X, 0)$ . The reason that we consider this adjusted intercept in the study, instead of the raw intercept, is that the fitted value at the average design point of  $X_i$  is a more meaningful value than the fitted value at the origin, which lies outside of the design space.

The three BEL methods (BEL.s, BEL.c and BEL.n) will be compared with RQ and the composite quantile regression (CQR) of Zou and Yuan (2008). The CQR assumes common slopes, and minimizes the sum of individual quantile loss functions over several  $\tau$ 's of interest. The CQR is a direct competitor of BEL.c, because they make the same assumption.

For Models 1 and 2, the common slope assumption holds, so there is no asymptotic bias for any of the methods we consider here. Table 2 shows the asymptotic efficiencies of BEL.c and CQR relative to RQ, when several quantiles are estimated jointly. It is clear that BEL.c and CQR are similar in efficiency for Model 1, but BEL.c stands out for Model 2. The asymptotic efficiency of BEL.s and that of RQ are the same; both of them are improved on by the other methods. Table 2 also includes comparisons at joint estimation of three quartiles, to indicate that the efficiency gain of BEL.c and CQR from the comparisons are not limited to high quantiles.

The asymptotic efficiencies do not depend on the choices of fixed priors. We now focus on estimation of high quantiles with  $\tau = 0.9, 0.925, 0.95$  at the sample size of  $n = 100$ , with the following priors:

- For BEL.s and BEL.c, we use the prior  $N(0, 100^2)$  for each intercept parameter, and  $N(1, 100^2)$  for each slope parameter.

TABLE 2

The table presents the ratio of the asymptotic MSE of the RQ estimators over that of the BEL.c or CQR estimator for Models 1 and 2, when jointly estimating quantiles at  $\tau = 0.25, 0.5, 0.75$  and  $\tau = 0.9, 0.925, 0.95$ , respectively

Asymptotic relative efficiencies for slope estimators						
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$	$\tau = 0.925$	$\tau = 0.95$
Model 1						
BEL.c/RQ	1.598	1.352	1.598	1.029	1.219	1.572
CQR/RQ	1.590	1.345	1.590	0.984	1.166	1.504
Model 2						
BEL.c/RQ	1.006	3.280	14.942	1.032	1.677	3.261
CQR/RQ	0.541	1.763	8.032	0.756	1.227	2.386

- For BEL.n, we use the prior  $N(0, 100^2)$  for each intercept parameter, and  $N(1, 100^2)$  for  $b_x(0.9)$  and for  $b_z(0.9)$ . The informative priors used to regulate the differences between quantiles are, conditional on  $\beta(0.90)$ ,  $b_x(0.925) \sim N(b_x(0.9), 0.16)$ ,  $b_x(0.95) \sim N(b_x(0.9), 1)$ ,  $b_z(0.925) \sim N(b_z(0.9), 0.01)$  and  $b_z(0.95) \sim N(b_z(0.9), 0.01)$ .

Additional details of the Bayesian computations can be found in the supplemental material [Yang and He (2012)]. The MSE's of various estimators of  $\beta(\tau)$  are given in Table 3 for Models 1 and 2, and in Table 4 for Models 3 and 4. We make several observations from those results:

- The performance of BEL.s is similar to or slightly better than that of RQ.
- When the common slope assumption holds, BEL.c has about the same (Model 1) or better (Model 2) efficiency when compared with CQR. The estimators that use informative priors on the slope parameters all improve on RQ. The differences among various methods are more significant at upper quantiles (say  $\tau = 0.95$ ) for heavier-tailed distributions.
- In Models 3 and 4, where the common slope assumption does not hold for  $b_x(\tau)$ , BEL.c and CQR show efficiency gains on the estimation of  $b_z(\tau)$ , but losses in the estimation of  $b_x(\tau)$ , due to bias. The BEL.n aims to reach a compromise in the bias-variance trade-off, resulting in a better MSE than RQ.

These findings are consistent with what we learned from the asymptotic comparisons shown in Table 2. The performance of BEL.n will of course depend on the choice of priors on the difference in slopes. The purpose of our study is not to demonstrate how to choose informative priors, but to show how informative priors can make a difference. Our empirical work shows that any reasonable choice of priors helps, even though an optimal choice is too much to ask for in general.

TABLE 3

The table gives the  $n \times \text{MSE}$ 's of several estimators for the adjusted intercepts and slope parameters at three quantile levels  $\tau = 0.9, 0.925, 0.95$  for Models 1 and 2, where  $n = 100$ , and the MSE is averaged over 500 samples from each model. The numbers in the brackets are the estimated standard errors

Method	Adjusted intercepts			Slopes					
	$a(0.9)$	$a(0.925)$	$a(0.95)$	$b_x(0.9)$	$b_z(0.9)$	$b_x(0.925)$	$b_z(0.925)$	$b_x(0.95)$	$b_z(0.95)$
Model 1									
BEL.s	22.0 (1.2)	26.5 (1.5)	35.6 (2.1)	3.0 (0.2)	11.7 (0.7)	3.5 (0.2)	13.8 (0.8)	4.2 (0.3)	19.8 (1.2)
BEL.n	23.1 (1.4)	25.7 (1.6)	31.5 (1.8)	3.3 (0.2)	12.3 (0.8)	3.4 (0.2)	12.3 (0.8)	3.9 (0.2)	12.4 (0.8)
BEL.c	26.6 (1.6)	27.9 (1.7)	34.1 (2)	3.4 (0.2)	13.9 (0.8)	3.4 (0.2)	13.9 (0.8)	3.4 (0.2)	13.9 (0.8)
CQR	22.8 (1.4)	25.7 (1.5)	30.0 (1.8)	3.2 (0.2)	12.7 (0.8)	3.2 (0.2)	12.7 (0.8)	3.2 (0.2)	12.7 (0.8)
RQ	22.3 (1.3)	26.9 (1.7)	36.5 (2.2)	3.3 (0.2)	12.1 (0.7)	3.7 (0.3)	14.4 (0.9)	4.4 (0.3)	19.2 (1.2)
Model 2									
BEL.s	76.4 (5.6)	126.6 (10.5)	291.3 (32)	9.5 (0.9)	42.4 (3.1)	13.5 (1.1)	71.7 (5.3)	26.2 (2.7)	159.2 (15.4)
BEL.n	78.7 (5.9)	95.0 (6.1)	150.0 (9.1)	9.4 (0.8)	43.6 (3.3)	10.3 (0.8)	43.8 (3.3)	14.5 (1.1)	43.7 (3.3)
BEL.c	86.8 (7.5)	100.5 (8.1)	158.3 (10.1)	9.1 (0.8)	46.9 (4.1)	9.1 (0.8)	46.9 (4.1)	9.1 (0.8)	46.9 (4.1)
CQR	109.3 (11.1)	125.5 (11.3)	175.9 (15.5)	12.7 (1.2)	61.7 (5.2)	12.7 (1.2)	61.7 (5.2)	12.7 (1.2)	61.7 (5.2)
RQ	76.4 (5.5)	136.4 (14.3)	280.6 (27.8)	10.0 (0.9)	41.6 (3.3)	14.9 (1.4)	73.4 (6.3)	26.5 (2.8)	144.3 (13.6)

**5. An application to temperature downscaling.** In recent decades much focus has been placed on understanding potential future climate changes. Meteorologists have developed various climate models to simulate atmospheric variables for both historical and future time periods under different greenhouse gas emission scenarios. Statistical downscaling approaches utilize those large-scale model simulations to predict small-scale regional climate changes; see [Wilby and Wigley \(1997\)](#) for a review. Quantifying nearly extreme events in climate studies is an important task, for which quantile regression is a naturally appealing tool. However, high quantiles are usually hard to estimate with RQ due to the inherently limited number of observations in the tail of the distributions. In this section, we consider the BEL methods for statistical downscaling of daily maximum temperature. We used the observed daily maximum temperature (TMAX) of Aurora, IL station from 1957–2002 as the response variable. The predictors are the simulated daily maximum temperature (RTEM) and an indicator of wet days (RAIN)

TABLE 4  
*Simulation results for Models 3 and 4; see the caption of Table 3 for more details*

Method	Adjusted intercepts			Slopes					
	$a(0.9)$	$a(0.925)$	$a(0.95)$	$b_x(0.9)$	$b_z(0.9)$	$b_x(0.925)$	$b_z(0.925)$	$b_x(0.95)$	$b_z(0.95)$
Model 3									
BEL.s	90.2 (5.2)	103.0 (5.8)	138.7 (9.0)	31.0 (1.9)	35.9 (2.4)	34.3 (2.0)	42.6 (2.6)	42.7 (2.6)	66.2 (4.7)
BEL.n	95.6 (5.6)	111.0 (6.1)	129.4 (7.5)	37.1 (2.3)	41.4 (2.8)	44.9 (2.7)	41.5 (2.9)	54.8 (3.3)	41.7 (2.8)
BEL.c	104.3 (6.9)	119.2 (7.2)	143.0 (8.1)	37.7 (2.3)	43.5 (2.9)	45.3 (2.7)	43.5 (2.9)	62.7 (3.3)	43.5 (2.9)
CQR	94.6 (5.9)	102.8 (6.3)	118.0 (7.3)	32.8 (2.0)	38.4 (2.6)	33.8 (1.9)	38.4 (2.6)	42.5 (2.4)	38.4 (2.6)
RQ	91.4 (5.3)	106.9 (6.8)	132.5 (8.3)	30.6 (1.9)	33.8 (2.2)	35.0 (2.0)	42.4 (2.8)	42.9 (2.5)	59.2 (3.9)
Model 4									
BEL.s	334.5 (25.7)	507.5 (40.0)	1085.1 (109.5)	96.5 (8.4)	134.1 (10.1)	144.6 (12.3)	213.5 (17.6)	252.4 (19.9)	547.6 (57.8)
BEL.n	277.0 (22.2)	346.8 (22.2)	518.0 (30.0)	97.3 (5.8)	124.7 (9.2)	125.6 (6.1)	124.7 (9.3)	196.9 (8.6)	125.1 (9.3)
BEL.c	391.6 (42.2)	453.4 (44.6)	659.8 (62.8)	111.9 (7.6)	160.5 (16.9)	137.2 (7.2)	160.5 (16.9)	214.7 (8.7)	160.5 (16.9)
CQR	530.1 (56.4)	520.0 (52.1)	663.9 (58.2)	142.0 (12.0)	195.3 (18.8)	140.1 (10.3)	195.3 (18.8)	175.0 (9.5)	195.3 (18.8)
RQ	340.3 (25.9)	552.0 (56.6)	1014.6 (101.1)	102.9 (8.3)	123.9 (9.6)	154.6 (11.5)	215.0 (19.0)	252.7 (17.8)	481.5 (46.0)

from the ERA-40 reanalysis model introduced in Uppala et al. (2005). A wet day is denoted by  $\text{RAIN} = 1$ , when the precipitation from ERA-40 is more than  $1.2 \text{ kg/s/m}^2$ . About 30% of the days are categorized as wet days in Aurora. We used the following linear quantile regression model:

$$(5.1) \quad Q_\tau(\text{TMAX}|\text{RTEM}, \text{RAIN}) = a(\tau) + b_x(\tau)\text{RTEM} + b_z(\tau)\text{RAIN}$$

at high quantiles  $\tau = 0.99, 0.995, 0.999$ . The quantile at  $\tau = 0.999$  is nearly extreme relative to our sample size, so the asymptotic theory developed in this paper might be questioned. We choose to consider such high quantiles partly to test the limits of our BEL methods.

We applied the following BEL methods with normal priors  $N(0, 1000^2)$  on each parameter to estimate the parameters of Model (5.1), unless otherwise specified:

- BEL.c and BEL.s as introduced in Section 4.
- BEL.z: the BEL estimator that assumes  $b_z(0.99) = b_z(0.995) = b_z(0.999)$ .
- BEL.t: the BEL estimator that assumes that given  $b_x(0.99)$  and  $b_z(0.99)$ ,  $(b_x(0.995) - b_x(0.99))/0.02$ ,  $(b_z(0.995) - b_z(0.99))/0.14$ ,  $(b_x(0.999) -$

TABLE 5  
Average effective sample sizes of the Markov chains used in the downscaling example

Method	SPLIT 1	SPLIT 2	SPLIT 3
BEL.c	976	933	364
BEL.z	542	633	747
BEL.t	672	701	515

$b_x(0.99)/0.35$  and  $(b_z(0.999) - b_z(0.99))/1.16$  are independent priors as the  $t$  distribution with degrees of freedom 3.

The scaling used in the prior distributions of BEL.t was chosen in rough proportion to the variances of those parameter estimates from RQ, and no optimality is claimed here. To assess the performances of various methods, we randomly split the data from each year into two parts, a fitting period and a testing period, with equal sizes of 7889 days in each part. We used the BEL methods and RQ for the fitting period in estimating the model parameters and then applied the fitted model to the testing period to predict the  $\tau$ th quantile of TMAX. We randomly split the data three times, and labeled them as SPLIT 1, SPLIT 2 and SPLIT 3, respectively. The average effective sample sizes of the Markov chains for the BEL methods used here are shown in Table 5, as calculated by the R function `effectiveSize()` in the R package *coda*.

Table 6 reports the normalized differences as a performance validation measure,

$$(5.2) \quad d = \frac{O - E}{\sqrt{\tau(1 - \tau)n}},$$

where  $n$  is the total number of days for prediction,  $O$  is the number of days when the observed TMAX exceeds the predicted  $\tau$ th quantile of TMAX and  $E$  indicates the expected number of days, that is,  $E = n(1 - \tau)$ . The normalized differences are shown for the whole testing period, as well as for two subsets, one subset being the lower half of RTEM, and the other subset being the wet days ( $\text{RAIN} = 1$ ). The use of these ad hoc subsets is meant to assess performances more comprehensively. The normalized differences greater than 2 in absolute values are marked as bold in Table 6, from which we have the following observations. First, over the whole testing period, the normalized differences of each BEL method are stable across random splits, but those from RQ predictions vary noticeably. For the testing periods and for the selected subsets, the BEL methods perform better than RQ, especially at  $\tau = 0.999$ . Second, among the BEL methods, BEL.c performs relatively worse, but BEL.t and BEL.z do well. When we used the ANOVA test of [Koenker and Bassett \(1982\)](#) for the null hypothesis of common slopes at  $\tau = 0.99, 0.995, 0.999$ , the hypothesis of  $b_x(0.99) = b_x(0.995) = b_x(0.999)$  was rejected at 5% level of significance. This helps explain the inferior performance of BEL.c relative to the other BEL methods, but all of them outperform RQ.

TABLE 6

The table presents the normalized differences calculated by (5.2). The row names provide the method used for model fitting. In the column names, the Whole period indicates all the data in the testing period are used; Lower RTEM indicates the testing data with RTEM below its median; Wet days indicates the testing data with RAIN equals to 1

Method	Whole period			Lower RTEM			Wet days		
	>0.99	>0.995	>0.999	>0.99	>0.995	>0.999	>0.99	>0.995	>0.999
	SPLIT 1								
BEL.c	0.012	0.089	0.040	-0.871	-0.163	1.036	0.402	0.142	-0.316
BEL.z	0.012	0.089	0.040	-0.230	-0.163	-0.476	0.804	0.142	-0.316
BEL.t	0.012	0.089	0.040	-1.351	-0.163	-1.483	-0.201	0.142	0.949
RQ	<b>-2.930</b>	<b>-2.306</b>	-1.742	<b>-2.151</b>	-1.743	-1.987	-1.005	-1.276	-1.582
	SPLIT 2								
BEL.c	0.012	0.089	0.040	0.250	0.515	1.540	<b>2.659</b>	1.591	0.329
BEL.z	0.012	0.089	0.040	1.530	-0.163	0.532	0.642	0.452	0.329
BEL.t	0.012	0.089	0.040	1.050	1.192	0.532	1.650	1.022	0.329
RQ	0.012	-0.390	<b>3.958</b>	1.370	0.063	1.540	0.843	0.737	<b>4.139</b>
	SPLIT 3								
BEL.c	0.012	0.089	0.040	-1.511	-0.163	0.532	<b>-2.188</b>	-1.543	-0.308
BEL.z	0.012	0.089	0.040	0.250	-0.163	-1.483	-1.381	-0.974	0.962
BEL.t	0.012	0.089	0.040	-0.871	-0.163	-0.979	-0.776	-0.690	1.596
RQ	-0.666	-0.869	<b>-2.454</b>	0.250	-0.388	-1.483	-1.583	-1.259	-1.577

Our empirical study shows that BEL methods can easily improve on RQ as downscaling methods for high quantiles. Informative priors will help further if the “prior makers” are well informed. In climate studies, for example, historical data are generally available from multiple stations nearby, which can lead us to helpful informative priors on slope parameters in the quantile models. In this sense, the shrinking priors considered in Theorem 3.3 are relevant.

A natural question in climate downscaling is the autocorrelation of measurements over time. In this section we have bypassed this issue on two grounds. First, the quantile regression estimation under the working assumption of independence is typically consistent under weakly dependent models; see He, Zhu and Fung (2002). Second, we verified empirically that the autocorrelation in TMAX was well represented by the autocorrelations in the predictors used in Model (5.1), and the signs of the residuals of the quantile models were nearly uncorrelated. In more general applications, however, it will be desirable to incorporate dependence in an appropriate way, and future research is clearly called for in this regard. Another interesting area of future work is to perform downscaling at a group of stations and include spatial correlation in the model. A recent paper by Reich, Fuentes and Dunson (2011) made a successful attempt at Bayesian spatial quantile regression,

and the idea of BEL with informative priors can be further explored in spatial modeling.

**6. Discussion.** In this paper, we propose using empirical likelihood as a working likelihood for quantile regression in Bayesian inference. We justify the validity of the posterior based inference by establishing its first order asymptotics. The BEL approach avoids the daunting task of directly maximizing the EL function and allows informative priors to be utilized. Although the idea of Bayesian quantile regression is not new, the work provides an important addition to the literature by providing the basic theory for incorporating possibly informative priors on multiple quantiles. The efficiency gains are demonstrated through both theoretical calculations and empirical investigations, when some common features across quantiles are explored. If common slopes are assumed, it is hard for the CQR method to find optimal weights in balancing the quantile loss function at different  $\tau$  levels, but the empirical likelihood approach does so naturally. The use of informative priors is also related in spirit to penalized optimization, but the lack of a good overall objective function for several quantile levels makes the usual regularization method difficult to formulate. The EL approach has the ability to adapt automatically across quantile levels, and the BEL approach enables flexible priors to be utilized in a simple way. Our theoretical framework of shrinking priors provides good understanding of how informative priors and likelihood can complement each other in the BEL approach.

This paper uses empirical likelihood, but some of its variants such as the ETEL, may work as well. The recent work of [Lancaster and Jun \(2010\)](#) provided an approximation to the posterior from the Bayesian ETEL of quantile regression at a given  $\tau$ . Although their approximation was not strong enough to imply posterior convergence for the Bayesian ETEL, it can be strengthened using the approach we provide for BEL. We hope that comparisons in a broader class of working likelihoods together with efficient algorithms will be further developed in the future.

#### APPENDIX: PROOFS

We begin with lemmas about the smoothness properties of functions involving the estimating functions (2.2). Note that the estimating functions (2.2) involve an indicator function, and as a result, the results obtained in [Qin and Lawless \(1994\)](#) for smooth functions do not apply. While the work of [Qin and Lawless \(1994\)](#) relies on the Taylor expansions, our proof uses the general theorem related to M-estimators in [van der Vaart \(1998\)](#) and the quadratic expansion approximating the EL function provided in [Molanes Lopez, Van Keilegom and Veraverbeke \(2009\)](#). We use  $x_j$  to indicate the  $j$ th component in the covariates vector  $X$  for  $j = 0, \dots, p$ , that is,  $X = (x_0, x_1, \dots, x_p)$  with  $x_0 = 1$ .

**A.1. Preparatory results.** We discuss the properties of functions involving the estimating function  $m(X, Y, \zeta)$ . Under Assumptions 3.2 and 3.3 about  $G_X$  and  $F_X$ ,  $E\{m(X, Y, \zeta)\}$  can be sufficiently smooth.

LEMMA A.1. *Under Assumptions 3.2 and 3.3, we have the following results:*

(L1)  $E\{m(X, Y, \zeta)\}$  and  $E\{m(X, Y, \zeta)m(X, Y, \zeta)^\top\}$  are twice continuously differentiable with respect to  $\zeta$ .

(L2) There exist  $k(p + 1)$  dimensional compact neighborhoods  $C_\xi$  and  $C_\zeta$  around 0, in which  $E\{m(X, Y, \zeta)/\{1 + \xi^\top m(X, Y, \zeta)\}\}$  is twice continuously differentiable in  $\zeta \in C_\zeta$  and  $\xi \in C_\xi$ , and  $E\{m(X, Y, \zeta)m(X, Y, \zeta)^\top/\{1 + \xi^\top m(X, Y, \zeta)\}\}$  is uniformly continuous with respect to  $\zeta \in C_\zeta$  and  $\xi \in C_\xi$ .

PROOF. To show (L1), note that for each  $d = 0, \dots, k - 1$  and  $j = 0, \dots, p$ , there is

$$\begin{aligned} E\{m_{dk+j}(X, Y, \beta(\tau))\} &= E\{(1_{\{Y \leq X^\top \beta(\tau_{d+1})\}} - \tau_{d+1})x_j\} \\ &= E_X[x_j\{E_{Y|X}(1_{\{Y \leq X^\top \beta(\tau_{d+1})\}} - \tau_{d+1})\}] \\ &= E_X[x_j\{F_X(X^\top \beta(\tau_{d+1})) - \tau_{d+1}\}]. \end{aligned}$$

Under Assumptions 3.2 and 3.3,  $E\{m(X, Y, \zeta)\}$  is twice continuously differentiable. Consider the cases  $i \leq l$  for the second moments. By the definition of regression quantiles,  $X^\top \beta(\tau_i) \leq X^\top \beta(\tau_l)$ , and therefore,

$$\begin{aligned} E\{m_{ik+j}(X, Y, \zeta)m_{lk+m}(X, Y, \zeta)\} &= E_X[x_j x_m \{E_{Y|X}(1_{\{Y \leq X^\top \beta(\tau_{i+1})\}} - \tau_{i+1})(1_{\{Y \leq X^\top \beta(\tau_{l+1})\}} - \tau_{l+1})\}] \\ &= E_X[x_j x_m \{F_X(X^\top \beta(\tau_{i+1})) - \tau_{l+1}F_X(X^\top \beta(\tau_{i+1})) \\ &\quad - \tau_{i+1}F_X(X^\top \beta(\tau_{l+1})) + \tau_{i+1}\tau_{l+1}\}], \end{aligned}$$

which is twice continuously differentiable in  $\zeta$ .

Similarly, (L2) follows from

$$\begin{aligned} E \frac{m_{dk+j}(X, Y, \zeta)}{1 + \xi^\top m(X, Y, \zeta)} &= E_X \left[ \sum_{0 \leq s \leq d} \frac{(1 - \tau_{d+1})x_j}{1 + \xi^\top m_s^*} \{F_X(X^\top \beta(\tau_{s+1})) - F_X(X^\top \beta(\tau_s))\} \right. \\ &\quad \left. - \sum_{d < s \leq k} \frac{\tau_{d+1}x_j}{1 + \xi^\top m_s^*} \{F_X(X^\top \beta(\tau_{s+1})) - F_X(X^\top \beta(\tau_s))\} \right], \end{aligned}$$

where we assume  $\tau_0 = 0$ ,  $\tau_{k+1} = 1$ ,  $m_0^* = ((1 - \tau_1)X^\top, \dots, (1 - \tau_k)X^\top)^\top$  and  $m_s^* = (-\tau_1 X^\top, \dots, -\tau_s X^\top, (1 - \tau_{s+1})X^\top, \dots, (1 - \tau_k)X^\top)^\top$  for  $s = 1, \dots, k$ .

Because  $m_s^*$  is bounded,  $1 + \xi^\top m_s^*$  could be bounded away from 0 for  $\xi$  in a sufficiently small compact neighborhood  $\mathcal{C}_\xi$ . Then  $E[m_{dk+j}(X, Y, \zeta_k)/\{1 + \xi^\top m(X, Y, \zeta)\}]$  is also twice continuously differentiable in  $\zeta$  and  $\xi$ . Similarly, we have  $E[m(X, Y, \zeta)m(X, Y, \zeta)^\top/\{1 + \xi^\top m(X, Y, \zeta)\}]$  is uniformly continuous with respect to  $\zeta \in \mathcal{C}_\zeta$  and  $\xi \in \mathcal{C}_\lambda$ .  $\square$

**A.2. Consistency of the MELE.** By Assumptions 3.2–3.4, the equation  $E\{m(X, Y, \zeta)\} = 0$  has the unique solution  $\zeta_0$ . Define

$$(A.1) \quad \Gamma_n(\zeta) = -n^{-1} \sum_{i=1}^n \log\{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)\},$$

where  $\lambda_n(\zeta)$  satisfies

$$\sum_{i=1}^n \frac{m(X_i, Y_i, \zeta)}{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)} = 0.$$

Recall that

$$\hat{\zeta} = \arg \max\{\Gamma_n(\zeta)\},$$

we define the expected value of  $\Gamma_n(\zeta)$  as

$$(A.2) \quad \Gamma(\zeta) = -E[\log\{1 + \xi(\zeta)^\top m(X, Y, \zeta)\}],$$

where  $\xi(\zeta)$  satisfies

$$E\left\{\frac{m(X, Y, \zeta)}{1 + \xi(\zeta)^\top m(X, Y, \zeta)}\right\} = 0.$$

By Lemma A.1, Assumption 3.5, and the implicit function theorem,  $\xi(\zeta)$  uniquely exists in the neighborhood  $\mathcal{C}_\lambda$  of  $\underline{0} \in \mathbb{R}^{k(p+1)}$ . To show that  $\hat{\zeta}$  is a consistent estimator of  $\zeta_0$ , it is sufficient to check the conditions of Theorem 5.7 of van der Vaart (1998). That is, we shall check

$$(A.3) \quad \sup_{\zeta} |\Gamma_n(\zeta) - \Gamma(\zeta)| \xrightarrow{P} 0$$

and

$$(A.4) \quad \sup_{\|\zeta - \zeta_0\| > \epsilon} \Gamma(\zeta) < \Gamma(\zeta_0)$$

for any  $\zeta$  within the compact neighborhood  $\mathcal{C}_\zeta$  of  $\zeta_0$  and  $\epsilon > 0$ .

LEMMA A.2. *Under Assumptions 3.1–3.5, (A.4) holds.*

PROOF. It is easy to see  $\xi(\zeta_0) = 0$  because  $E\{m(X, Y, \zeta_0)\} = 0$ , and then  $\Gamma(\zeta_0) = 0$ . By the Taylor expansion, we have

$$\Gamma(\zeta) = -\xi(\zeta)^\top E \left\{ \frac{m(X, Y, \zeta)}{1 + \xi(\zeta)^\top m(X, Y, \zeta)} \right\} - \frac{1}{2} E \left\{ \frac{(\xi(\zeta)^\top m(X, Y, \zeta))^2}{(1 + \alpha(\zeta)^\top m(X, Y, \zeta))^2} \right\}$$

for some  $\alpha(\zeta)$  on the line segment between 0 and  $\xi(\zeta)$ . On the right-hand side of the above equation, the first term equals 0, and the second term with the negative sign included is strictly negative, and thus  $\Gamma(\zeta) < 0$  for  $\zeta \neq \zeta_0$ . So within the compact neighborhood  $\mathcal{C}_\zeta$  of  $\zeta_0$ , we have

$$\sup_{\|\zeta - \zeta_0\| > \varepsilon} \Gamma(\zeta) < \Gamma(\zeta_0). \quad \square$$

To check (A.3), we first expand  $\Gamma_n(\zeta) - \Gamma(\zeta)$  as

$$(A.5) \quad \Gamma_n(\zeta) - \Gamma(\zeta) = Q_1 + Q_2,$$

where

$$\begin{aligned} Q_1 &= -n^{-1} \sum_{1 \leq i \leq n} [\log\{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)\}] \\ &\quad + E[\log\{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)\}], \\ Q_2 &= -E[\log\{1 + \lambda_n(\zeta)^\top m(X_i, Y_i, \zeta)\}] + E[\log\{1 + \xi(\zeta)^\top m(X_i, Y_i, \zeta)\}]. \end{aligned}$$

To show the uniform convergence of (A.5), we need the following lemma.

LEMMA A.3. (i) *The class of constant functions:  $\mathcal{C}_0 = \{\lambda, \lambda \in \mathcal{C}\}$  is P-Glivenko–Cantelli (P-G–C) class, where  $\mathcal{C}$  is some compact set in  $\mathbb{R}$ .* (ii) *For bounded  $X$ , the class of functions*

$$\begin{aligned} \mathcal{F}_1 &= \left\{ \frac{m(X, Y, \zeta)}{1 + \lambda_n^\top m(X, Y, \zeta)} : \zeta \in \mathcal{C}_\zeta, \lambda_n \in \mathcal{C}_\lambda \right\} \quad \text{and} \\ \mathcal{F}_2 &= \{\log(\{1 + \xi^\top m(X, Y, \zeta)\}) : \zeta \in \mathcal{C}_\zeta, \xi \in \mathcal{C}_\lambda\} \end{aligned}$$

are P-G–C, where  $\mathcal{C}_\lambda$  is a compact neighborhood around  $\underline{0} \in \mathbb{R}^{k(p+1)}$ , and  $\mathcal{C}_\zeta$  is a compact neighborhood around  $\zeta_0 \in \mathbb{R}^{k(p+1)}$ .

PROOF. (i) According to Theorem 8.14 of Kosorok (2008) and the fact that  $\mathcal{C}_0$  is a collection of bounded functions, we only need to show that  $\mathcal{C}_0$  is VC-class, as defined in Section 9.1.1 in Kosorok (2008). The P-measurability will be guaranteed by the measurability and boundedness of the constant functions in  $\mathcal{C}_0$ . The collection of all subgraphs of functions in  $\mathcal{C}_0$  is  $\mathcal{S}_0 = \{(x, y), y < \lambda\}$ . For any two points  $(x_1, y_1), (x_2, y_2)$  in  $\mathbb{R}^2$ , assume  $y_1 \leq y_2$ , it is impossible that  $\mathcal{S}_0$  would include  $(x_2, y_2)$  while excluding  $(x_1, y_1)$ . Therefore, based on the definition of VC-subgraph Class, we have  $\text{VC}(\mathcal{C}_0) = 2 < \infty$ , i.e.,  $\mathcal{C}_0$  is a VC class. (ii) From

Lemma 9.12 and Lemma 9.8 of [Kosorok \(2008\)](#), we know that the class of indicator functions  $\mathcal{G}_0 = \{1_{\{Y \leq X^\top \beta\}}, \beta \in \mathbb{R}^{p+1}\}$  is a VC-class. From (vi) and (vii) in Lemma 9.9 of [Kosorok \(2008\)](#), the sets of estimating functions

$$\mathcal{G}_d = \{(1_{\{Y \leq X^\top \beta(\tau_d)\}} - \tau_i)x_j, \beta(\tau_d) \in \mathbb{R}^{p+1}, 0 \leq j \leq p\},$$

$1 \leq d \leq k$ , are VC-class. Because  $X$  is bounded,  $\mathcal{G}_d$  is P-G-C class by Theorem 8.14 of [Kosorok \(2008\)](#). Then by Theorem 9.26 of [Kosorok \(2008\)](#), it follows that  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are P-G-C.  $\square$

We now verify (A.3). We will check the uniform convergence of  $Q_1$  and  $Q_2$  in (A.5). Because  $\mathcal{F}_2$ , in which  $\xi$  is not related to  $(X, Y)$ , is P-G-C, the uniform convergence implied by P-G-C guarantees the convergence of  $Q_1$ . For  $Q_2$ , because  $\log\{1 + \xi(\zeta)^\top m(X_i, Y_i, \zeta)\}$  is bounded, by the dominate convergence theorem, we only need to show  $\lambda_n(\zeta) \xrightarrow{P} \xi(\zeta)$  uniformly in  $\zeta$ . Because  $\lambda_n(\zeta)$  is actually a Z-estimator, the approximate zero of a data-dependent function of  $\xi(\zeta)$  as defined in Chapter 5.1 in [van der Vaart \(1998\)](#), then by using the standard arguments of Z-estimator in [van der Vaart \(1998\)](#) and by the fact that  $\mathcal{F}_1$  is P-G-C, we have  $\lambda_n(\zeta) \xrightarrow{P} \xi(\zeta)$  uniformly in  $\zeta$ .

The proof of Theorem 3.1 is now complete.

**A.3. Asymptotic normality of the posterior.** In our notation, we have

$$(A.6) \quad \log\{\mathcal{R}_n(\zeta)\} = n\Gamma_n(\zeta),$$

where  $\mathcal{R}_n(\zeta)$  is the empirical likelihood ratio of  $\zeta$ . To expand  $\Gamma_n(\zeta)$  up to the quadratic term, we use Assumption 3.5. We also use the following lemma, which is taken from the quadratic expansion provided in Lemma A.6 of [Molanes Lopez, Van Keilegom and Veraverbeke \(2009\)](#) but formulated to suit our setting.

LEMMA A.4. *Assume that the results of Lemma A.1 and Theorem 3.1 hold. Under Assumptions 3.1–3.5, and additional conditions (C1)–(C3) listed below, we have*

$$(A.7) \quad \Gamma_n(\zeta) = -\frac{1}{2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} V_{12}(\zeta - \zeta_0) + n^{-1/2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} M_n - \frac{1}{2}n^{-1}M_n^\top V_{11}^{-1} M_n + o_p(n^{-1})$$

uniformly in  $\zeta$ , for  $\zeta - \zeta_0 = O(n^{-1/2})$ , and

$$(A.8) \quad \hat{\zeta} - \zeta_0 = n^{-1/2}(V_{12}^\top V_{11}^{-1} V_{12})^{-1} V_{12}^\top V_{11}^{-1} M_n + o_p(n^{-1/2}),$$

where  $\hat{\zeta}$  is the MELE of  $\zeta_0$ ,  $M_n = n^{-1/2} \sum_{i=1}^n m(X_i, Y_i, \zeta_0)$  and  $V_{11}$  and  $V_{12}$  are the same as defined in Theorem 3.2.

(C1)  $\|\sum_{i=1}^n [m(X_i, Y_i, \zeta) - E\{m(X_i, Y_i, \zeta)\}]\| = O_p(n^{1/2})$ , uniformly in  $\zeta$  in a  $o(1)$ -neighborhood of  $\zeta_0$ .

(C2)  $\| \sum_{i=1}^n [m(X_i, Y_i, \zeta)m(X_i, Y_i, \zeta)^\top - E\{m(X_i, Y_i, \zeta)m(X_i, Y_i, \zeta)^\top\}] \| = o_p(n)$ , uniformly in  $\zeta$  in a  $o(1)$ -neighborhood of  $\zeta_0$ .

(C3)  $\| \sum_{i=1}^n [m(X_i, Y_i, \zeta) - E\{m(X_i, Y_i, \zeta)\} - m(X, Y, \zeta_0) + E\{m(X, Y, \zeta_0)\}] \| = o_p(n^{1/2})$ , uniformly in  $\zeta$  for  $\zeta - \zeta_0 = O_p(n^{-1/2})$ .

To use the expansion (A.7), we shall verify that (C1)–(C3) are satisfied.

LEMMA A.5. *Under Assumptions 3.2–3.4, Conditions (C1)–(C3) are satisfied for the estimating functions  $m(X, Y, \zeta)$  of (2.2).*

PROOF. Because the collection of estimating functions  $m(X, Y, \zeta)$  is P-Donsker class, we have (C1). By the fact that the collection of the product of the estimating functions is P-G-C, we have (C2). By applying Lemma 4.1 of He and Shao (1996) to  $m(X, Y, \zeta)$ , we obtain (C3).  $\square$

PROOF OF THEOREM 3.2. By Lemma A.5, Lemma A.4, (A.7) and (A.6), we have

$$\begin{aligned} \tilde{p}(\zeta|D) &= p_0(\zeta) \times \mathcal{R}_n(\zeta) \\ &= p_0(\zeta) \times \exp \left\{ -\frac{n}{2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} V_{12}(\zeta - \zeta_0) \right. \\ &\quad \left. + n^{1/2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} M_n - \frac{1}{2} M_n^\top V_{11}^{-1} M_n + o_p(1) \right\}. \end{aligned}$$

Because of (A.8), we have

$$\begin{aligned} \tilde{p}(\zeta|D) &= p_0(\zeta) \times \exp \left\{ -\frac{n}{2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} V_{12}(\zeta - \zeta_0) \right. \\ &\quad \left. + n(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} V_{12}(\hat{\zeta} - \zeta_0) \right. \\ &\quad \left. - \frac{1}{2} M_n^\top V_{11}^{-1} M_n + o_p(1) \right\} \\ &= p_0(\zeta) \times \exp \left\{ -\frac{n}{2}(\zeta - \zeta_0)^\top V_{12}^\top V_{11}^{-1} V_{12}(\zeta - 2\hat{\zeta} + \zeta_0) \right. \\ &\quad \left. - \frac{1}{2} M_n^\top V_{11}^{-1} M_n + o_p(1) \right\} \\ &= p_0(\zeta) \times \exp \left\{ -\frac{n}{2}(\zeta - \hat{\zeta})^\top V_{12}^\top V_{11}^{-1} V_{12}(\zeta - \hat{\zeta}) + o_p(1) \right\}. \end{aligned}$$

By Assumption 3.6, we have

$$\log\{p_0(\zeta)\} = \log\{p_0(\zeta_0)\} + O(n^{-1/2})$$

for  $\zeta - \zeta_0 = O(n^{-1/2})$ . Then we have

$$(A.9) \quad \tilde{p}(\zeta|D) = p_0(\zeta_0) \exp\{-\frac{1}{2}(\zeta - \hat{\zeta})^\top J_n(\zeta - \hat{\zeta}) + o_p(1)\},$$

where  $J_n = nV_{12}^\top V_{11}^{-1}V_{12}$ . For any  $n$ , we have  $p(\zeta|D) \propto \tilde{p}(\zeta|D)$ , and thus (3.1) holds.

Because  $J_n$  is positive definite, we have

$$(A.10) \quad p(J_n^{1/2}(\zeta - \hat{\zeta})|D) \propto \exp\{-\frac{1}{2}(J_n^{1/2}(\zeta - \hat{\zeta}))^\top (J_n^{1/2}(\zeta - \hat{\zeta})) + o_p(1)\}$$

for any  $\zeta - \zeta_0 = O(n^{-1/2})$ . Therefore, to show

$$J_n^{1/2}(\zeta - \hat{\zeta}) \xrightarrow{D} N(0, I),$$

it remains to show that

$$P(\|J_n^{1/2}(\zeta - \hat{\zeta})\| > \delta) \rightarrow 0,$$

when  $\delta \rightarrow \infty$  and  $n \rightarrow \infty$ . From (A.9), we have for any  $\zeta = \hat{\zeta} + J_n^{-1/2}t$ ,

$$\mathcal{R}_n(\zeta) \times p_0(\zeta) \xrightarrow{P} p_0(\zeta_0) \exp\{-\|t\|^2/2\}.$$

Because of  $\mathcal{R}_n(\zeta) \times p_0(\zeta) \leq p_0(\zeta)$ , by the dominate convergence theorem, we have

$$\int_{\|t\|>\delta} p_0(\hat{\zeta} + J_n^{-1/2}t)\mathcal{R}_n(\hat{\zeta} + J_n^{-1/2}t) dt \rightarrow p_0(\zeta_0) \int_{\|t\|>\delta} \exp\{-\|t\|^2/2\} dt$$

for any  $\delta \geq 0$ . Then it leads to

$$\begin{aligned} P(\|J_n^{1/2}(\zeta - \hat{\zeta})\| > \delta|D) &= \frac{\int_{\|t\|>\delta} p_0(\hat{\zeta} + J_n^{-1/2}t)\mathcal{R}_n(\hat{\zeta} + J_n^{-1/2}t) dt}{\int_{\|t\|>0} p_0(\hat{\zeta} + J_n^{-1/2}t)\mathcal{R}_n(\hat{\zeta} + J_n^{-1/2}t) dt} \\ &\rightarrow \frac{\int_{\|t\|>\delta} \exp\{-\|t\|^2/2\} dt}{\int_{\|t\|>0} \exp\{-\|t\|^2/2\} dt} \\ &= (2\pi)^{-k(p+1)/2} \int_{\|t\|>\delta} \exp\{-\|t\|^2/2\} dt \\ &< \epsilon \end{aligned}$$

for sufficiently large  $\delta$ .  $\square$

PROOF OF THEOREM 3.3. Similar to the proof of Theorem 3.2, we have

$$(A.11) \quad \tilde{p}(\zeta|D) = p_{0,n}(\zeta) \times \exp\left\{-\frac{n}{2}(\zeta - \hat{\zeta})^\top V_{12}^\top V_{11}^{-1}V_{12}(\zeta - \hat{\zeta}) + o_p(1)\right\}.$$

By Assumption 3.7, we have

$$\log\{p_{0,n}(\zeta)\} = \log\{p_{0,n}(\zeta_{0,n})\} - \frac{1}{2}(\zeta - \zeta_{0,n})^\top J_{0,n}(\zeta - \zeta_{0,n}) + o_p(1)$$

for  $\|\zeta - \zeta_0\| = O(n^{-1/2})$  and bounded  $\zeta_{0,n}$ . Combined with (A.11), we have

$$\tilde{p}(\zeta|D) = C_n \exp\left\{-\frac{1}{2}(\zeta - \theta_{\text{post}})^\top J_n(\zeta - \theta_{\text{post}}) + R_n\right\},$$

where  $J_n = J_{0,n} + nV_{12}^\top V_{11}^{-1}V_{12}$ ,  $\theta_{\text{post}} = J_n^{-1}(J_{0,n}\zeta_{0,n} + nV_{12}^\top V_{11}^{-1}V_{12}\hat{\zeta})$ ,  $R_n = o_p(1)$ , and  $C_n$  is some constant that does not depend on  $\zeta$ , and has the following expression:

$$C_n = p_{0,n}(\zeta_{0,n}) \exp\left\{-\frac{1}{2}\zeta_{0,n}^\top J_{0,n}\zeta_{0,n} - \frac{n}{2}\hat{\zeta}^\top V_{12}^\top V_{11}^{-1}V_{12}\hat{\zeta} + \frac{1}{2}\theta_{\text{post}}^\top J_n\theta_{\text{post}}\right\}.$$

Therefore, we have (3.3).  $\square$

PROOF OF COROLLARY 3.4. The prior density  $p_{0,n}(\zeta)$  can be written as

$$\begin{aligned} \log p_{0,n}(\zeta) &= C + \log\{g_1(\Omega^{-1/2}(\beta(\tau_1) - \beta_{p,0}))\} \\ &\quad + \sum_{d=2}^k \log\{g_d(\Sigma_d^{-1/2}(\beta(\tau_d) - \beta(\tau_1)))\}, \end{aligned}$$

where  $C$  is some constant not depending on  $\zeta$ . Clearly, the prior mode is  $\beta(\tau_d) = \beta_{p,0}$  for all  $d = 1, \dots, k$ . Then we have

$$\frac{\alpha^2 \log p_{0,n}(\zeta)}{\alpha\beta^2(\tau_1)} \Big|_{\zeta=1_k \otimes \beta_{p,0}} = \frac{\Omega^{-1/2}g_1''(\mathbf{0})\Omega^{-1/2}}{g_1(\mathbf{0})} + \sum_{d=2}^k \frac{\Sigma_d^{-1/2}g_d''(\mathbf{0})\Sigma_d^{-1/2}}{g_d(\mathbf{0})},$$

and for  $d = 2, \dots, k$ ,

$$\begin{aligned} \frac{\alpha^2 \log p_{0,n}(\zeta)}{\alpha\beta(\tau_1)\alpha\beta(\tau_d)} \Big|_{\zeta=1_k \otimes \beta_{p,0}} &= \frac{\alpha^2 \log\{g_d(\Sigma_d^{-1/2}(\beta(\tau_d) - \beta(\tau_1)))\}}{\alpha\beta(\tau_1)\alpha\beta(\tau_d)} \Big|_{\zeta=1_k \otimes \beta_{p,0}} \\ &= -\frac{\Sigma_d^{-1/2}g_d''(\mathbf{0})\Sigma_d^{-1/2}}{g_d(\mathbf{0})}, \\ \frac{\alpha^2 \log p_{0,n}(\zeta)}{\alpha\beta^2(\tau_d)} \Big|_{\zeta=1_k \otimes \beta_{p,0}} &= \frac{\alpha^2 \log\{g_d(\Sigma_d^{-1/2}(\beta(\tau_d) - \beta(\tau_1)))\}}{\alpha\beta^2(\tau_d)} \Big|_{\zeta=1_k \otimes \beta_{p,0}} \\ &= \frac{\Sigma_d^{-1/2}g_d''(\mathbf{0})\Sigma_d^{-1/2}}{g_d(\mathbf{0})}. \end{aligned}$$

Note that for a spherically symmetric  $g_d$  with its mode and center as zero, we have

$$\frac{g_d''(\mathbf{0})}{g_d(\mathbf{0})} = C_d I,$$

where  $I$  is the  $(p + 1) \times (p + 1)$  dimensional identity matrix, and  $C_d > 0$  are constants for  $d = 1, \dots, k$ . Then, we have

$$J_{0,n} = \begin{pmatrix} C_1 \Omega^{-1} + \sum_{d=2}^k C_d \Sigma_d^{-1} & -C_2 \Sigma_2^{-1} & \cdots & C_k \Sigma_k^{-1} \\ -C_2 \Sigma_2^{-1} & C_2 \Sigma_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -C_k \Sigma_k^{-1} & \mathbf{0}^\top & \cdots & C_k \Sigma_k^{-1} \end{pmatrix}$$

and therefore,

$$J_{0,n}(\zeta_{0,n} - \zeta_0) = \begin{pmatrix} C_1 \Omega^{-1}(\beta_{p,0} - \beta_0(\tau_1)) + \sum_{d=2}^k C_d \Sigma_{d,I}^{-1}(\beta_{0,I}(\tau_d) - \beta_{0,I}(\tau_1)) \\ C_2 \Sigma_{2,I}^{-1}(\beta_{0,I}(\tau_1) - \beta_{0,I}(\tau_2)) \\ \vdots \\ C_k \Sigma_{k,I}^{-1}(\beta_{0,I}(\tau_1) - \beta_{0,I}(\tau_k)) \end{pmatrix},$$

where  $\beta_{0,I}(\tau_d)$  is the intercept parameter in  $\beta_0(\tau_d)$ . Under the assumption in (3.4) and (3.5),  $\|J_{0,n}(\zeta_{0,n} - \zeta_0)\| = O(\epsilon_n)$  and  $\|J_{0,n}\|$  is increasing at the rate of  $n$ . Then Assumption 3.7 is satisfied, and Theorem 3.3 applies.

Note that the posterior mean  $\theta_{\text{post}}$  in Theorem 3.3 can be written as

$$\theta_{\text{post}} = \zeta_0 + n J_n^{-1} V_{12}^\top V_{11}^{-1} V_{12}(\hat{\zeta} - \zeta_0) - J_n^{-1} J_{0,n}(\zeta_{0,n} - \zeta_0).$$

By (A.8), we have  $\|\hat{\zeta} - \zeta_0\| = O_p(n^{-1/2})$ . Then we have the posterior mean

$$\theta_{\text{post}} = \zeta_0 + O_p(\epsilon_n/n + n^{-1/2}). \quad \square$$

**Acknowledgments.** The authors thank three anonymous referees and an Associate Editor for their helpful and constructive comments and suggestions that have led to several improvements in their work.

SUPPLEMENTARY MATERIAL

**Supplement to “Bayesian empirical likelihood for quantile regression”** (DOI: [10.1214/12-AOS1005SUPP](https://doi.org/10.1214/12-AOS1005SUPP); .pdf). The supplementary material contains additional details on the implementation of the Bayesian computations used in the empirical studies reported in this paper.

REFERENCES

CHANG, I. H. and MUKERJEE, R. (2008). Bayesian and frequentist confidence intervals arising from empirical-type likelihoods. *Biometrika* **95** 139–147. MR2409719

- CHEN, S. X. and HALL, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21** 1166–1181. [MR1241263](#)
- CHEN, J., SITTER, R. R. and WU, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika* **89** 230–237. [MR1888365](#)
- CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. [MR2566404](#)
- CHEN, S. X. and WONG, C. M. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statist. Sinica* **19** 71–81. [MR2487878](#)
- CHEN, K., YING, Z., ZHANG, H. and ZHAO, L. (2008). Analysis of least absolute deviation. *Biometrika* **95** 107–122. [MR2394499](#)
- CHERNOZHUKOV, V. and HONG, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115** 293–346. [MR1984779](#)
- CHUNG, Y. and DUNSON, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Amer. Statist. Assoc.* **104** 1646–1660. [MR2750582](#)
- DUNSON, D. B., PILLAI, N. and PARK, J.-H. (2007). Bayesian density regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 163–183. [MR2325270](#)
- DUNSON, D. B. and TAYLOR, J. A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Stat.* **17** 385–400. [MR2129840](#)
- FANG, K.-T. and MUKERJEE, R. (2006). Empirical-type likelihoods allowing posterior credible sets with frequentist validity: Higher-order asymptotics. *Biometrika* **93** 723–733. [MR2261453](#)
- GERACI, M. and BOTTAI, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* **8** 140–154.
- GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London. [MR1397966](#)
- GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305–330. [MR1150346](#)
- HAHN, J. (1997). Bayesian bootstrap of the quantile regression estimator: A large sample study. *Internat. Econom. Rev.* **38** 795–808. [MR1482116](#)
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
- HE, X. and SHAO, Q.-M. (1996). A general Bahadur representation of  $M$ -estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24** 2608–2630. [MR1425971](#)
- HE, X., ZHU, Z.-Y. and FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. [MR1929164](#)
- HOROWITZ, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66** 1327–1351. [MR1654307](#)
- KIM, M.-O. and YANG, Y. (2011). Semiparametric approach to a random effects quantile regression model. *J. Amer. Statist. Assoc.* **106** 1405–1417.
- KOCHERGINSKY, M., HE, X. and MU, Y. (2005). Practical confidence intervals for regression quantiles. *J. Comput. Graph. Statist.* **14** 41–55. [MR2137889](#)
- KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. [MR2268657](#)
- KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- KOENKER, R. and BASSETT, G. JR. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–61. [MR0640165](#)
- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. [MR2724368](#)
- KOTTAS, A. and GELFAND, A. E. (2001). Bayesian semiparametric median regression modeling. *J. Amer. Statist. Assoc.* **96** 1458–1468. [MR1946590](#)

- KOTTAS, A. and KRNJAJIĆ, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.* **36** 297–319. [MR2528986](#)
- LANCASTER, T. and JUN, S. J. (2010). Bayesian quantile regression methods. *J. Appl. Econometrics* **25** 287–307. [MR2758636](#)
- LAZAR, N. A. (2003). Bayesian empirical likelihood. *Biometrika* **90** 319–326. [MR1986649](#)
- MOLANES LOPEZ, E. M., VAN KEILEGOM, I. and VERAVERBEKE, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Stat.* **36** 413–432. [MR2549702](#)
- MONAHAN, J. F. and BOOS, D. D. (1992). Proper likelihoods for Bayesian analysis. *Biometrika* **79** 271–278. [MR1185129](#)
- MÜLLER, P., ERKANLI, A. and WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83** 67–79. [MR1399156](#)
- MÜLLER, P. and QUINTANA, F. A. (2004). Nonparametric Bayesian data analysis. *Statist. Sci.* **19** 95–110. [MR2082149](#)
- NEWWEY, W. K. and SMITH, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72** 219–255. [MR2031017](#)
- NORETS, A. and PELENIS, J. (2010). Posterior consistency in conditional distribution estimation by covariate dependent mixtures. Unpublished manuscript.
- OTSU, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *J. Econometrics* **142** 508–538. [MR2408747](#)
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. [MR0946049](#)
- OWEN, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton, FL.
- PATI, D., DUNSON, D. and TOKDARY, S. (2010). Posterior consistency in conditional distribution estimation. Unpublished manuscript.
- QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. [MR1272085](#)
- 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. [MR2816698](#)
- REN, J.-J. (2008). Smoothed weighted empirical likelihood ratio confidence intervals for quantiles. *Bernoulli* **14** 725–748. [MR2537809](#)
- SCHENNACH, S. M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika* **92** 31–46. [MR2158608](#)
- SCHENNACH, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.* **35** 634–672. [MR2336862](#)
- UPPALA, S. M., KÅLLBERG, P. W., SIMMONS, A. J., ANDRAE, U., DA COSTA BECHTOLD, V., 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., HÖLM, 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 Journal of the Royal Meteorological Society* **131** 2961–3012.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- WANG, H. J. and ZHU, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *J. Statist. Plann. Inference* **141** 1603–1615. [MR2747929](#)

- WHANG, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* **22** 173–205. [MR2230386](#)
- WILBY, R. L. and WIGLEY, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography* **21** 530–548.
- YANG, Y. and HE, X. (2012). Supplement to “Bayesian empirical likelihood for quantile regression.” DOI:[10.1214/12-AOS1005SUPP](#).
- YIN, G. (2009). Bayesian generalized method of moments. *Bayesian Anal.* **4** 191–207. [MR2507358](#)
- YU, K. and MOYEED, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54** 437–447. [MR1861390](#)
- YUE, Y. R. and RUE, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Statist. Data Anal.* **55** 84–96. [MR2736538](#)
- ZOU, H. and YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36** 1108–1126. [MR2418651](#)

DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS  
SCHOOL OF PUBLIC HEALTH  
DREXEL UNIVERSITY  
1505 RACE STREET  
PHILADELPHIA, PENNSYLVANIA 19102  
USA  
E-MAIL: [Yunwen.Yang@drexel.edu](mailto:Yunwen.Yang@drexel.edu)

DEPARTMENT OF STATISTICS  
UNIVERSITY OF MICHIGAN  
1085 SOUTH UNIVERSITY AVENUE  
ANN ARBOR, MICHIGAN 41809  
USA  
E-MAIL: [xmh@umich.edu](mailto:xmh@umich.edu)