Bayesian inference for L_p–quantile regression models

Inferenza Bayesiana per modelli di regressione L_p–quantilica

Mauro Bernardi and Valeria Bignozzi and Lea Petrella

Abstract L_p–quantiles generalise quantiles and expectiles to account for the whole distribution of the random variable of interest. In this paper, we introduce the L_p–quantile regression model, we propose a collapsed Gibbs–sampler algorithm to make Bayesian inference on the regression parameters. We also provide some theoretical results concerning the posterior distribution of the regression parameters.
Abstract *Gli L_p quantili generalizzano i quantili e gli espettili allo scopo di considerare tutta la distribuzione della variabile casuale di interesse. In questo paper, introduciamo il modello di regressione L_p quantilica, proponiamo un algoritmo Gibbs sampler collassato per l'inferenza bayesiana sui parametri di regressione. Proponiamo in aggiunta risultati teorici sulla distribuzione a posteriori dei parametri.*

Key words: Bayesian quantile regression, Skew Exponential Power distribution, MCMC.

1 Introduction

Since the seminal paper of Koenker and Bassett (1978) quantile regression has been gaining a lot of consideration in the statistic and econometric literature because of its ability to model a given quantile of a response variable as function of a set of covari-

Mauro Bernardi
University of Padua, Department of Statistical Sciences, via Cesare Battisti, 241, 25121 Padova, e-mail: mauro.bernardi@unipd.it

Valeria Bignozzi
MEMOTEF Department, Sapienza University of Rome, Via del Castro Laurenziano 9, Rome e-mail: valeria.bignozzi@uniroma1.it

Lea Petrella
MEMOTEF Department, Sapienza University of Rome, Via del Castro Laurenziano 9, Rome e-mail: valeria.bignozzi@uniroma1.it

ates with minimal assumptions on the error term, see also Koenker (2005). Despite its usefulness in handling heteroskedastic, skewed and leptokurtic data, quantile regression inherits the main drawback of quantiles being order statistics that do not keep into consideration the whole distribution of observables beyond the modelled quantile. Expectiles generalises quantiles in order to account for the whole distribution of the quantity of interest. Although expectiles have become popular as risk measures only in the very recent literature, they were already well known in the statistical literature. In fact they were introduced byAigner et al. (1976) and Newey and Powell (1987) named them expectiles, to emphasise that their loss function is obtained as a generalisation of the loss function of the *ex*pectation and of the quan*tiles*. More specifically, let Y be a continuous random variable, expectiles are defined by Newey and Powell (1987) as the unique minimiser of an asymmetric squared function

$$\mu_\tau(Y) = \arg\min_{m\in\mathbb{R}} \mathbb{E}\left[\left|\tau - \mathbb{1}_{(-\infty,0)}(Y-m)\right|(Y-m)^2\right], \tag{1}$$

where $\tau \in (0,1)$ is the expectile confidence level. Obviously, for $\tau = 1/2$, $\mu_{1/2}(Y) = \mathbb{E}(Y)$. Similarly to Yu and Moyeed (2001), De Rossi and Harvey (2009) showed that the minimisation problem in equation (1) is equivalent to maximising the following Asymmetric Gaussian Distribution (AGD)

$$f_{\text{AGD}}(y \mid \tau, \mu) = 2\left[\sqrt{\frac{\pi}{|\tau-1|}} + \sqrt{\frac{\pi}{\tau}}\right]^{-1} \exp\left(-\left|\tau - \mathbb{1}_{(-\infty,0)}\right|(y-\mu)^2\right), \tag{2}$$

for $\tau \in (0,1)$. This opens the way of performing classial, likelihood–based and Bayesian inference on expectile parametric and non–parametric models, see, e.g., Gerlach and Chen (2014).

In the present contribution we consider the class of L_p–quantiles introduced by Chen (1996). L_p–quantiles belong to the class of generalised quantiles, also investigated in Bellini et al. (2014) and share many important properties of the quantiles. For a random variable Y with cumulative distribution function F_Y the L_p–quantile at level τ is defined as

$$\rho_{\tau,p}(Y) = \arg\min_{m\in\mathbb{R}} \mathbb{E}\left[\left|\tau - \mathbb{1}_{(-\infty,0)}\right|(Y-m)^p\right], \tag{3}$$

for all $\tau \in (0,1)$, $p \in \mathbb{N} \setminus \{0\}$. As for quantiles and expectiles the minimisation problem in equation (3) is exactly equivalent to maximising the following Skew Exponential Power (SEP) distribution introduced by Zhu and Zinde-Walsh (2009) and Komunjer (2007)

$$f_{\text{SEP}}(y \mid \alpha, \mu, p, \sigma) = \begin{cases} \frac{1}{\sigma}K(p)\exp\left(-\frac{1}{p}\left|\frac{y-\mu}{2\alpha\sigma}\right|^p\right), & \text{if } y \leq \mu \\ \frac{1}{\sigma}K(p)\exp\left(-\frac{1}{p}\left|\frac{y-\mu}{2(1-\alpha)\sigma}\right|^p\right), & \text{if } y > \mu, \end{cases} \tag{4}$$

where $\alpha \in (0,1)$, $\sigma > 0$, $p > 0$ and $K(p)$ is the normalising constant which depends only on p. The SEP distribution defined in equation (4) forms the basis for the Bayesian inferential procedure for L_p–quantile regression models developed in the next sessions. The next proposition provides the link between the definition of L_p–quantiles, as minimisers of the loss function defined in equation (3), and the optimisation of the SEP probability density function.

Proposition 1. *The expected scoring function obtained from the kernel of the SEP with parameter $(\alpha, \mu, p, 1)$ has as unique minimiser the L_p–quantile at level $\tau = \frac{\alpha^p}{\alpha^p + (1-\alpha)^p}$.*

2 The L_p quantile regression model

Let $\mathbf{y} = (y_1, y_2, \ldots, y_T)$ be a random sample of T observations, $\mathbf{x}_t = (1, x_{1,t}, x_{2,t}, \ldots, x_{q,t})'$ the associated set of q covariates for $t = 1, 2, \ldots, T$, we consider the following generalised L_p quantile regression model

$$y_t = \mathbf{x}_t' \beta_{\tau,p} + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{5}$$

where $\beta_{\tau,p} = \left(\beta_{\tau,0}^p, \beta_{\tau,1}^p, \ldots, \beta_{\tau,q}^p\right)'$ is a vector of dimension $(q+1) \times 1$ unknown regression parameters which may depend on the both the p parameter and the quantile confidence level $\tau \in (0,1)$. In the basic framework, the stochastic terms ε_t, for any $t = 1, 2, \ldots, T$, are assumed to be independent identically distributed random variables with zero τ–th generalised quantile and constant variance which implies that the regression function $\mathbf{x}_t' \beta_{\tau,p}$ is the τ–th level quantile of the response variable y_t conditional to the set of exogenous covariates $\mathbf{x}_t$. The likelihood function for the model (5) based on the SEP definition (4) can be written as

$$\mathscr{L}(\mathbf{y}, \beta_{\tau,p}, \sigma \mid \mathbf{x}) \propto \frac{1}{\sigma^T} \mathrm{e}^{-\frac{1}{p} \sum_{t=1}^T \left(\frac{\mu_t - y_t}{2\tau\sigma}\right)^p} \mathbb{1}_{(y_t \le \mu_t)} \mathrm{e}^{-\frac{1}{p} \sum_{t=1}^T \left(\frac{y_t - \mu_t}{2(1-\tau)\sigma}\right)^p} \mathbb{1}_{(y_t > \mu_t)}, \tag{6}$$

where $\mu_t = \mathbf{x}_t' \beta_{\tau,p}$ and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$.

The Bayesian inferential procedure requires the specification of the prior distribution for the unknown vector of parameters $\Xi = (\beta_{\tau,p}, \sigma)$. Since in our framework, the shape parameter p is an intervention parameter that selects the generalised quantile of interest, and it is fixed and known, we can consider the following bijective reparameterisation of the SEP parameters $\Xi^* = (\beta_{\tau,p}, \tilde{\sigma}) = (\beta_{\tau,p}, \sigma^p)$. In principle non informative priors can be specified for the vector of regression parameters, i.e., $\pi(\beta_{\tau,p}) \propto 1$. The following theorems prove the existence of the posterior normalising constant of the regression parameters $\beta_{\tau,p}$ of the linear generalised quantile regression model under improper diffuse prior. Alternatively, the usual Normal–Inverse Gamma prior can be specified for regression and scale parameters, respectively, i.e., $\pi(\beta_{\tau,p}) \sim \mathscr{N}\left(\mu_\beta, \Sigma_\beta\right)$ and $\pi(\tilde{\sigma}) \sim \mathscr{IG}(\lambda_\sigma, \eta_\sigma)$.

Theorem 1. *Let $\mathbf{y} = (y_1, y_2, \ldots, y_T)$ be a sample of i.i.d. observations generated by the generalised quantile regression model with $q > 1$, under the assumption of* *improper diffuse prior for the regressor parameters, i.e.,* $\pi(\beta) \propto 1$*, then*

$$0 < \int_\beta \mathscr{L}(\mathbf{y} \mid \beta)\, \pi(\beta)\, d\beta < \infty.$$

Here, we present the theoretical results concerning Bayesian estimation of generalised quantile regression models.

Lemma 1. *Let* $(Y_1, Y_2, \dots, Y_T)$ *be a sequence of independent observations of a univariate response variable and let* $(X_1, X_2, \dots, X_T)$ *be one–dimensional non–random covariates.* $\mathscr{P}_{0,i}$ *denotes the true and unknown probability distribution of* Y_i*, with the true* τ*–level generalised quantile of order* $p \in \mathbb{N}^+$ *given by* $Q_\tau^\alpha(Y_i \mid X_i) = \delta_0 + \beta_0 X_i$*, for* $\tau \in (0,1)$*. Suppose that the misspecified model for* Y_i *is* $\mathscr{SEP}(\cdot, \mu_i^\tau, \sigma, \alpha, \tau)$ *with* $\mu_i^\tau = \delta + \beta X_i$*. Then if the condition*

$$\mathbb{E}_{\mathscr{P}_0}\left[\log\left(\frac{\mathscr{P}_{0,i}(Y_i)}{\mathscr{P}_i^{\tau,\alpha}(Y_i \mid \delta_0, \beta_0, \sigma_0 = 1)}\right)\right] < +\infty, \tag{7}$$

we have

$$\inf_{\delta,\beta} \mathbb{E}_{\mathscr{P}_0}\left[\log\left(\frac{\mathscr{P}_{0,i}(Y_i)}{\mathscr{P}_i^{\tau,\alpha}(Y_i \mid \delta, \beta, \sigma_0 = 1)}\right)\right] \geq \mathbb{E}_{\mathscr{P}_0}\left[\log\left(\frac{\mathscr{P}_{0,i}(Y_i)}{\mathscr{P}_i^{\tau,\alpha}(Y_i \mid \delta_0, \beta_0, \sigma = 1)}\right)\right], \tag{8}$$

for fixed τ *and* α*.*

The next proposition guarantees the week consistency of the posterior distribution of generalised quantile regression parameters.

Proposition 2. *The weak consistency of the posterior follows from the Schwartz (1965) theorem, since any Kullback–Leibler neighbourhood of the true density has positive probability using Lemma* (1)*.*

3 The Collapsed Gibbs sampler

Using the likelihood associated to the SEP distribution and the data–augmentation approach we can implement a partially collapsed Gibbs sampler algorithm. Assuming the same prior structure as before, we obtain the following full conditional posterior distributions. The full conditional distribution of the completion variable $\mathbf{u} = (u_1, u_2, \dots u_T)$ is proportional to $\pi(\mathbf{u} \mid \mathbf{y}, \mathbf{x}, \beta_{\tau,p}, \tilde{\sigma}) \propto \prod_{t=1}^T \pi(\mathbf{u} \mid y_t, x_t, \beta_{\tau,p}, \tilde{\sigma})$ where

$$\pi(u_t \mid y_t, \mathbf{x}_t, \beta_{\tau,p}, \tilde{\sigma}) \propto \begin{cases} \mathrm{e}^{-\frac{u_t}{p}} \mathbb{1}_{\left(u_t > \frac{(\mathbf{x}_t'\beta_{\tau,p} - y_t)^p}{(2\tau)^p \tilde{\sigma}}\right)}(y_t), & \text{if} \quad y_t \leq \mathbf{x}_t'\beta_{\tau,p} \\ \mathrm{e}^{-\frac{u_t}{p}} \mathbb{1}_{\left(u_t > \frac{(y_t - \mathbf{x}_t'\beta_{\tau,p})^p}{(2(1-\tau))^p \tilde{\sigma}}\right)}(y_t), & \text{if} \quad y_t > \mathbf{x}_t'\beta_{\tau,p}, \end{cases} \tag{9}$$

which is proportional to a truncated exponential distribution with shape parameter p. Simulation from the truncated exponential distribution can be easily performed by inverting the corresponding cumulative density function. The full condition distribution of the scale parameter σ can be obtained by marginalising out the the latent factor, and it is proportional to an Inverse Gamma distribution $\pi(\tilde{\sigma} \mid \mathbf{y}, \mathbf{x}, \beta_{\tau,p}) \propto \mathscr{IG}\left(\tilde{\lambda}_\sigma, \tilde{\eta}_\sigma\right)$, with

$$\tilde{\eta}_\sigma = \eta_\sigma + \frac{1}{p}\left[\sum_{t=1}^{T}\left(\frac{\mathbf{x}_t'\beta_{\tau,p} - y_t}{2\tau}\right)^p \mathbb{1}_{\left(y_t \leq \mathbf{x}_t'\beta_{\tau,p}\right)} + \sum_{t=1}^{T}\left(\frac{y_t - \mathbf{x}_t'\beta_{\tau,p}}{2(1-\tau)}\right)^p \mathbb{1}_{\left(y_t > \mathbf{x}_t'\beta_{\tau,p}\right)}\right]$$
$$\tilde{\lambda}_\sigma = \lambda_\sigma + \frac{T}{p}.$$

The last step of the Gibbs sampler requires the simulation of the set regressors $\beta_{\tau,p}$ form the corresponding full conditional distribution. Observe that, conditional to the latent factor $\mathbf{u}$, the full conditional of $\beta_{\tau,p}$ is proportional to

$$\pi(\beta_{\tau,p} \mid \mathbf{y}, \mathbf{x}, \mathbf{u}, \tilde{\sigma}) \propto \pi(\beta_{\tau,p}) \prod_{t=1}^{T} \mathbb{1}_{\left(\underline{\beta}_t, \overline{\beta}_t\right)}(\beta_{\tau,p}), \tag{10}$$

where $\pi(\beta_{\tau,p})$ is the prior distribution and $\underline{\beta}_t = \varepsilon_t - 2(1-\tau)\tilde{\sigma} u_t^{\frac{1}{p}}$, $\overline{\beta}_t = \varepsilon_t + 2\tau\tilde{\sigma} u_t^{\frac{1}{p}}$ with $\varepsilon_t = y_t - \beta_{\tau,p}'\mathbf{x}_t$. Under the Gaussian prior assumption for the regression parameters $\beta_{\tau,p}$, the complete set of full conditional distributions becomes

$$\beta_{\tau,p}^{(j)} \mid \mathbf{y}, \mathbf{x}, \mathbf{u}, \beta_{\tau,p}^{(-j)}, \tilde{\sigma} \sim \mathscr{N}\left(\tilde{\mu}_\beta^{(j)}, \tilde{\sigma}_\beta^{(j)}\right) \mathbb{1}_{\left(\underline{\beta}_j, \overline{\beta}_j\right)}\left(\beta_{\tau,p}^{(j)}\right), \tag{11}$$

for $j = 1, 2, \ldots, q+1$, where $\beta_{\tau,p}^{(j)}$ denotes the j–th element of the vector $\beta_{\tau,p}$, and

$$\tilde{\mu}_\beta^{(j)} = \mu_\beta^{(j)} + \Sigma_\beta^{(j,-j)} \Sigma_\beta^{(-j,-j)^{-1}}\left(\beta_{\tau,p}^{(-j)} - \mu_\beta^{(-j)}\right) \tag{12}$$
$$\tilde{\sigma}_\beta^{(j)} = \Sigma_\beta^{(j,j)} - \Sigma_\beta^{(j,-j)} \Sigma_\beta^{(-j,-j)^{-1}} \Sigma_\beta^{(-j,j)}, \tag{13}$$

where $\beta_{\tau,p}^{(-j)}$ denotes the vector of regressors with the j–th element removed, and $\Sigma_\beta^{(-m,-n)}$ denotes the sub–matrix obtained from Σ_β by removing the m–th row and the n–th column. Here $\underline{\beta}_1 = \max_{t=1,2,\ldots,T}\left\{\varepsilon_t - 2(1-\tau)(\tilde{\sigma} u_t)^{\frac{1}{p}}\right\}$, $\overline{\beta}_1 = \min_{t=1,2,\ldots,T}\left\{\varepsilon_t + 2\tau(\tilde{\sigma} u_t)^{\frac{1}{p}}\right\}$, $\underline{\beta}_j = \max_{t=1,2,\ldots,T,\ni x_{j,t}\neq 0}\left\{\frac{\varepsilon_{j,t} - 2(1-\tau)(\tilde{\sigma} u_t)^{\frac{1}{p}}}{x_{j,t}}\right\}$ and $\overline{\beta}_j = \min_{t=1,2,\ldots,T,\ni x_{j,t}\neq 0}\left\{\frac{\varepsilon_{j,t} + 2\tau(\tilde{\sigma} u_t)^{\frac{1}{p}}}{x_{j,t}}\right\}$, for $j = 2, \ldots, q+1$, with $\varepsilon_{j,t} = y_t - \beta_{\tau,p}^{(-j)'}\mathbf{x}_t^{(-j)}$ and $\mathbf{x}^{(-j)}$ denotes the vector $\mathbf{x}_t$ with the j–th element removed. Sampling from the joint full conditional distribution of the regression parameters subject to the truncation defined in equation (10), is not easy since it requires simulation from the complete set of full conditionals defined in equation (11), as suggested by Robert (1995).

The ordering of the full conditional simulation ensures that the posterior distribution is the stationary distribution of the generated Markov chain since the combination of steps above essentially ensures draws from the conditional posterior distribution $\pi(\tilde{\sigma}, \mathbf{u} \mid \mathbf{y}, \mathbf{x}, \beta_{\tau,p},)$ and the resulting partially collapsed Gibbs sampler is a blocked version of the ordinary Gibbs sampler, see, e.g., Dyk and Park (2008) and Bernardi et al. (2015).

References

Aigner, D. J., Amemiya, T., and Poirier, D. J. (1976). On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, 17(2):377–396.

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41 – 48.

Bernardi, M., Gayraud, G., Petrella, L., et al. (2015). Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis*, 10(3):553–603.

Chen, Z. (1996). Conditional lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & Probability Letters*, 29(2):107 – 115.

De Rossi, G. and Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, 152:179–185.

Dyk, D. V. and Park, T. (2008). Partially collapsed gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 110:790–796.

Gerlach, R. and Chen, C. W. S. (2014). Bayesian expected shortfall forecasting incorporating the intraday range. *Journal of Financial Econometrics*.

Koenker, B. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Koenker, B. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, 22(5):891–921.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.

Robert, C. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125.

Schwartz, L. (1965). On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 4(1):10–26.

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

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148(1):86 – 99.