Bayesian Quantile Regression using the Skew Exponential Power Distribution

Bernardi Mauro and Marco Bottone and Petrella Lea

Abstract Traditional Bayesian quantile regression relies on the Asymmetric Laplace distribution (ALD) due primarily to its satisfactory empirical and theoretical performances. However, the ALD displays medium tails and is not suitable for data characterized by strong deviations from the Gaussian hypothesis. In this paper, we propose an extension of the ALD Bayesian quantile regression framework to account for fat tails using the Skew Exponential Power (SEP) distribution. Linear and Additive Models (AM) with penalized spline are used to show the flexibility of the SEP in the Bayesian quantile regression context. Lasso priors are used to account for the problem of shrinking parameters when the parameters space becomes wide. We propose a new adaptive Metropolis–Hastings algorithm in the linear model, and an adaptive Metropolis within Gibbs one in the AM framework. Empirical evidence of the statistical properties of the model is provided through several examples based on both simulated and real datasets.

Abstract L'analisi Bayesiana per la regressione quantile si basa sull'uso della distribuzione Laplace asimmetrica come strumento inferenzale. Tale distribuzione pur fornendo performances soddisfacenti non ha un comportamento soddisfacente nel caso in cui il fenomeno sotto indagine presenti code con andamento diverso da quello gaussiano. In questo paper, per tener conto di code pesanti del fenomeno, proponiamo l'uso della distribuzione Skew Exponential Power (SEP) in un contesto di regressione quantile. Considereremo modelli lineari e modelli additivi attraverso l'uso di spline per effettuare l'inferenza bayesiana. Una distribuzione lasso a priori sui parametri del modello viene proposta per tener conto del problema della con-

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

Bottone Marco
Bank of Italy DG Economics, Statistics and Research, Italy e-mail: marco.bottone@bancaditalia.it

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

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), 2017 Firenze University Press

trazione del numero degli stessi laddove lo spazio parametrico diventi elevato. Per effettuare l'inferenza bayesiana viene proposto un nuovo algoritmo adattivo di tipo Monte Carlo Markov Chain e analisi di simulazioni verranno proposte per validare il modello considerato.

Key words: Bayesian quantile regression; Skew Exponential Power; Additive Model.

1 Introduction

Quantile regression has become a very popular approach to provide a more complete description of the distribution of a response variable conditionally on a set of regressors. Since the seminal work of [1], several papers have been proposed in literature considering the quantile regression analysis both from a frequentist and a Bayesian points of view. Specifically, let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_T)$ be a random sample of T observations, and $\mathbf{X}_t = (1, X_{t,1}, \ \ldots, X_{t,p-1})'$, with $t = 1, 2, \ldots, T$ equal to the associated set of p covariates. Consider the following linear quantile regression model

$$Y_t = \mathbf{X}_t' \beta_\tau + \varepsilon_t, \qquad t = 1, 2, \ldots, T,$$

where $\beta_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \ldots, \beta_{\tau,p-1})'$ is the vector of p unknown regression parameters, varying with the quantile τ level. As usual, ε_t represents the error term that, in the specific case of quantile regression, has the τ quantile equal to zero and constant variance. This assumption allows us to interpret the regression line as the τ conditional quantile of Y given the set of explanatory variables $\mathbf{X} = \mathbf{x}$, i.e. $Q_\tau(Y \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}^T \beta_\tau$. In what follows we omit the subscript τ for simplicity. The estimation procedure of the τ – th regression quantile in the frequentist approach is based on the minimization of the following loss

$$\min_\beta \sum_t \rho_\tau \left(y_t - \mathbf{x}_t^T \beta\right)$$

with $\rho_\tau(u) = u(\tau - I(u < 0))$. From a Bayesian point of view [8] introduces the ALD as likelihood function to perform the inference. For a wide and recent Bayesian literature on quantile regression and ALD see for example [7], and [3]. Although the ALD is widely used in the Bayesian framework it displays medium tails which may give misleading informations for extreme quantile in particular when the data are characterized by the presence of outlier and heavy tails. The absence for the ALD of a parameter governing the tail fatness may influence the final inference. To overcome this drawback we propose an extension of the Bayesian quantile regression using the Skew Exponential Power (SEP) distribution proposed by [2]. The SEP distribution, like the ALD, has the property of having the τ-level quantile as the natural location parameter but it also has an additional parameter governing the decay of the tails. Using the proposed distribution in quantile regression we are able to robustify the inference in particular when outliers or extreme values are observed.

When dealing with model building the choice of appropriate predictors and consequently the variable selection issue plays an important role. In this paper, we approach this problem, by considering the Bayesian version of Lasso penalization methodology introduced by [6] both for the simple linear regression quantile and for the non linear additive models (AM) with Penalized Spline (P–Spline) functions. To implement the Bayesian inference we propose a new adaptive Metropolis Hastings algorithm in the linear model, and an Adaptive Metropolis within Gibbs one in the AM framework for an efficient estimate of the penalization parameter and the P–Spline coefficients. We show the robust performance of the model with simulation studies.

2 Model and Inference

In their paper [2], the authors propose a parametrization of the SEP, that allows to consider the location parameter as the τ–level quantile. With their parametrization the SEP density function can be written as:

$$f(y,\mu,\sigma,\tau,\alpha) = \begin{cases} \frac{1}{\sigma}\kappa(\alpha)\exp\left\{-\frac{1}{\alpha}\left(\frac{\mu-y}{2\tau\sigma}\right)^{\alpha}\right\}, & \text{if} \quad y \leq \mu \\ \frac{1}{\sigma}\kappa(\alpha)\exp\left\{-\frac{1}{\alpha}\left(\frac{y-\mu}{2(1-\tau)\sigma}\right)^{\alpha}\right\}, & \text{if} \quad y > \mu, \end{cases} \tag{1}$$

where $y \in \mathbb{R}$, $\mu \in \Re$ is the location parameter, $\sigma \in \Re^+$ and $\alpha \in (0,\infty)$ are the scale and shape parameters, respectively, $\tau \in (0,1)$ is the skewness parameter while $\kappa = \left[2\alpha^{\frac{1}{\alpha}}\Gamma\left(1+\frac{1}{\alpha}\right)\right]^{-1}$ and $\Gamma(\cdot)$ is the complete gamma function. It can be showed that μ is the τ quantile and that the ALD is a particular case with $\alpha = 1$. Several model specifications can be obtained using the SEP likelihood by specifying a given function for the location parameter.
In this paper we consider both the linear quantile regression framework

$$\mu = \mu(\mathbf{x}_t) = \mathbf{x}_t^T\beta \tag{2}$$

where $\mathbf{x}_t$ is a set of exogenous covariates than the Additive Models within a robust semi-parametric regression framework:

$$\mu = \mu(\mathbf{x}_t, \mathbf{z}_t) = \mathbf{x}_t^T\beta + \sum_{j=1}^{J} f_j(z_{tj})$$

where $\mathbf{x}_t^T\beta$ is the parametric component while $\mathbf{z}_t = (z_{t,1},\cdots,z_{t,J})^T$ is an additional set of covariates and each $f_j(z_{tj})$ is a nonparametric continuous smooth function. To implement the Bayesian analysis we assume that $f_j(z_{tj})$, can be approximated using a polynomial spline of order d, with $k+1$ equally spaced knots.
Let's consider more specifically the linear case where the likelihood function can be easily computed starting from (1) by using μ as in (2).

The Bayesian inferential procedure requires the specification of the prior distribution for the unknown vector of parameters $\Xi = (\beta, \gamma, \sigma, \alpha)$. Here in order to account for sparsity within the quantile regression model, we generalize the prior proposed in Park and Casella for the β parameter, assuming the hierarchical structure given below. The prior distribution is given by:

$$\pi(\Xi) = \pi(\beta \mid \gamma)\,\pi(\gamma)\,\pi(\sigma)\,\pi(\alpha),$$

with

$$\begin{aligned}
\pi(\beta \mid \gamma) &\propto \prod_{j=1}^{p} L_1(\beta_j \mid 0, \gamma_j)\\
\pi(\gamma) &\propto \prod_{j=1}^{p} \mathscr{G}(\gamma_j \mid \psi, \varpi)\\
\pi(\sigma) &\propto \mathscr{IG}(a,b)\\
\pi(\alpha) &\propto \mathscr{B}(c,d)\,\mathbb{1}_{(0,2)}(\alpha),
\end{aligned}$$

where $\beta \in \mathbb{R}^p$. Here $(\psi, \varpi, a, b, c, d)$ are given positive hyperparameters and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$ are the parameters of the univariate Laplace distribution:

$$L_1(\beta_j \mid 0, \gamma_j) = \frac{\gamma_j}{2} \exp\{-\gamma_j |\beta_j|\}\, \mathbb{1}_{(-\infty,+\infty)}(\beta_j).$$

with zero location and γ_j scale parameter. Here $\mathscr{G}$, $\mathscr{IG}$ and $\mathscr{B}$ denote the Gamma, Inverse Gamma and Beta distributions, respectively. Given its characteristics, the Laplace distribution is the Bayesian counterpart of the Lasso penalization methodology introduced by [6] to achieve sparsity within the classical regression framework. By shrinking each regression parameter in a different way, we overcome problems that may arise in the presence of regressors with different scales of measurement. The Bayesian inference is performed by building an Adaptive Independent Metropolis Hastings MCMC algorithm using the location–scale mixture representation of the the Laplace distribution, see for example [9].

3 Simulation Studies

We have performed several simulation studies to highlight the improvements of our model specification with respect to the well known ALD model tool. In particular the first simulation experiment is built in order to show the robustness properties of the proposed methodology for quantile estimation when the joint distribution of the couple $(Y_t, \mathbf{X}_t)$, for $t = 1, 2, \dots, T$, is contaminated by the presence of outliers. The second study shows the effectiveness of the shrinkage effect, obtained by imposing the Lasso–type prior, used when the multiple quantile linear model is of key concern. The last experiment aims at highlighting the ability of the model to adapt to non–linear shapes, when data come from heterogeneous fat–tailed distributions. All of the simulation studies showed the improvement in performances of the model proposed in this paper with respect to the ALD quantile regression commonly used in literature. Here we present only the second experimental study. In particular we carry out a Monte Carlo simulation study specifically tailored to evaluate the performance of the model when the Lasso prior is considered for the regression parameters. The simulations are similar to the one proposed in [4] and [5]. In particular, we simulate $T = 200$ observations from the linear model $Y_t = \mathbf{X}_t'\beta + \varepsilon_t$, where the true values for the regressors are set as follows:

$$\begin{aligned} &\text{Simulation 1.} \quad && \beta = (3, 1.5, 0, 0, 2, 0, 0, 0)', \\ &\text{Simulation 2.} \quad && \beta = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)', \\ &\text{Simulation 3.} \quad && \beta = (5, 0, 0, 0, 0, 0, 0, 0)', \end{aligned}$$

The first simulation corresponds to a sparse regression case, the second to a dense case, and the third to a very sparse case. The covariates are independently generated from a $\mathcal{N}(0, \Sigma)$ with $\sigma_{i,j} = 0.5^{|i-j|}$. Two different distributions for the error terms generating process are considered for each simulation study. The first is a Gaussian distribution $\mathcal{N}\left(\mu, \sigma^2\right)$, with μ set so that the τ-th quantile is 0, while σ^2 is set as 9, as in [4]. The second distribution is a Generalized Student't $\mathcal{GS}\left(\mu, \sigma^2, \nu\right)$ with two degrees of freedom, i.e. $\nu = 2$, $\sigma^2 = 9$ and μ set so that the τ-th quantile is 0. For three different quantile levels, $\tau = (0.10, 0.5, 0.9)$ we run 50 simulations for each vector of parameters (β) and each distribution of the error term. Table 1 reports the median of mean absolute deviation (MMAD), i.e. median$\left(\frac{1}{200}\sum_{t=1}^{200} \mid x_t'\hat{\beta} - x_t'\beta \mid\right)$, and the median of the parameters $\hat{\beta}$, over 50 estimates. Results for the first simulation are reported, since results from the other two simulations are qualitatively similar. The proposed Bayesian quantile regression method based on the SEP likelihood performs better in terms of MMAD for both distributions of the error term. This is evidence that the presence of the shape parameter α in the likelihood better capture the behavior of the data. The estimated shape parameter is indeed greater and lower than one in the Gaussian and Generalized Student't cases, respectively; this provides a more reliable estimation of the vector β, regardless of the tail weight of the error term distribution. These results are reinforced in the second and third simulation (not reported here) in which we exaggerate the density and the sparsity of the predictors structure. Furthermore, the proposed robust method reduces the bias of estimated β for all quantile confidence levels. Regarding the shrinkage ability of the proposed estimator, when the true parameters are zero, the SEP distribution performs better than the ALD in identifying the parameters .

Error distribution	Par.	ALD			SEP		
		$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$
Gaussian	MMAD	1.0131	1.1008	1.0579	0.9096	1.0955	0.9708
	β_1	3.1323	3.2209	3.2145	3.0744	3.0036	3.2127
	β_2	1.6408	1.4786	1.6165	1.7656	1.4833	1.6800
	β_3	0.0444	0.0294	0.0267	0.0428	0.0228	0.0186
	β_4	0.0453	0.0243	0.0235	0.0248	0.0191	0.0156
	β_5	1.2731	1.2379	1.3471	1.3969	1.8405	1.4702
	β_6	0.0185	0.0161	0.0205	0.0124	0.0127	0.0128
	β_7	0.0112	0.0106	0.0120	0.0067	0.0063	0.0095
	β_8	0.0073	0.0078	0.0064	0.0038	0.0047	0.0051
Generalized Student t	MMAD	0.5163	0.1807	0.4685	0.4777	0.1789	0.4275
	β_1	3.0630	2.9884	2.9874	3.0826	2.9877	2.9934
	β_2	1.0484	1.3700	1.1366	1.0952	1.3951	1.2110
	β_3	0.0304	0.0144	0.0325	0.0252	0.0135	0.0412
	β_4	0.0258	0.0181	0.0162	0.0263	0.0163	0.0138
	β_5	1.7012	1.9036	1.7701	1.7558	1.9111	1.8052
	β_6	0.0128	0.0085	0.0137	0.0074	0.0072	0.0136
	β_7	0.0055	0.0057	0.0101	0.0052	0.0066	0.0082
	β_8	0.0067	0.0009	0.0002	0.0051	0.0011	-0.0021

Table 1 Multiple regression simulated data example 1. MMADs and estimated parameters for Simulation 1 under the SEP and ALD assumption for the quantile error term.

4 Conclusion

We show how to implement the Bayesian quantile regression when the SEP distribution is considered. Linear and Additive Models (AM) with penalized spline are used with Lasso priors to account for the problem of shrinking parameters. Empirical analysis highlights how the SEP quantile regression better capture the behaviour of the data when outliers or heavy tails are concerned.

References

1. Koenker, R., Basset, G.: Regression quantiles. *Econometrica* **46** (1978) 33-50.
2. Zhu, D., Zinde-Walsh, V.: Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* **148** (2009) 86-99.
3. Bernardi, M., Gayraud, G., Petrella, L.: Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis* **10** (2015) 553-603.
4. Li, Q., Xi, R., Lin, R.: Bayesian regularized quantile regresion. *Bayesian Analysis* (2015) **5** 533–556.
5. Alhamzawi, R., and Benoit, D.F.: Bayesian adaptive lasso quantile regression *Statistical Modelling*, **12** (2012) 279–297.
6. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*,**58**, (1996) 267–288.
7. Lum, K. and Gelfand A.: Spatial quantile multiple regression using the asymmetric laplace process . *Bayesian Analysis* , **7** (2012) 235-258.
8. Yu, K. and Moyeed, R. A. (2001). Bayesian Quantile Regression. *Statist. Probab. Lett*, **54**, 437–447.
9. Kozumi, H. and Kobayashi, G. (2001). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation nd Simulation*, **81**, 1565–1575.