Bayesian Variable Selection in Linear Quantile Regression Model

Mai Dao[a], Min Wang[b], Souparno Ghosh[c], Keying Ye[b]

[a]Texas Tech University, [b] The University of Texas at San Antonio, [c] University of Nebraska–Lincoln

Abstract

Asymmetric Laplace (AL) specification has become one of ideal statistical models for Bayesian quantile regression analysis. Besides fast convergence of Markov Chain Monte Carlo (MCMC), AL specification guarantees posterior consistency even under model misspecification. However, variable selection under such a specification is a daunting task because, realistically, prior specification of regression parameters should take the quantile levels into consideration. Quantile-specific Zellner's g-prior has recently been proposed for Bayesian variable selection in quantile regression, whereas it comes at a high price of the computational burden due to the intractability of the posterior distributions. This paper shows that a fast computation can be achieved with exact, but intractable posterior distributions. We devise a three-stage computational scheme starting with an expectation-maximization (EM) algorithm and then the Gibbs sampler followed by an importance re-weighting step. The performance and effectiveness of the proposed procedure are illustrated with both simulation studies and a real-data application. Numerical results suggest that the proposed procedure compares favorably with the exact MCMC algorithm.

Model

For the linear regression model

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

[5] demonstrated that for $p \in (0,1)$, β_p can be obtained by minimizing

$$Q(\beta) = \sum_{i=1}^{n} \rho_p(y_i - \mathbf{x}_i'\beta), \tag{2}$$

where $\rho_p(\cdot)$ is the quantile check function.

For the stochastic search variable selection process, we introduce the random indicator vector $\gamma = (\gamma_1, \ldots, \gamma_k)'$, with $\gamma_j = 1$ indicating the inclusion of the jth predictor. Then, for each model M_γ, we follow [7] and [1] to suggest the conjugate prior family

$$p(\beta_\gamma \mid \sigma, \boldsymbol{\nu}) \sim N\left(0, 2g\sigma\Sigma_\nu(\gamma)^{-1}\right) \quad \text{and} \quad p(\sigma) \propto \frac{1}{\sigma}, \tag{3}$$

where $\Sigma_\nu(\gamma) = \mathbf{X}_\gamma'\mathbf{V}\mathbf{X}_\gamma$, with $\mathbf{V} = \text{diag}(\nu_1, \ldots, \nu_n)$ from the location-scale mixture representation of the working asymmetric Laplace distribution likelihood, see [2].

Algorithm 1: A three-stage Gibbs−Importance sampling algorithm for $(\beta, \sigma, \boldsymbol{\nu})$

Input: The number of iterations of the proposed algorithm J and the desired number of samples I with ($I < J$).

Result: A sequence of independent posterior samples ($\beta^{(1)}, \sigma^{(1)}, \boldsymbol{\nu}^{(1)}$), $\ldots$, ($\beta^{(I)}, \sigma^{(I)}, \boldsymbol{\nu}^{(I)}$)

I. EM Algorithm Set the final iterative values of the EM algorithm as the posterior modes of β and σ, denoted by $\tilde{\beta}$ and $\tilde{\sigma}$, respectively.

II. Gibbs Sampler

for $j \leftarrow 1$ **to** J **do**

1. Generate $\boldsymbol{\nu}^{(j)}$ from the proposal generalized inverse Gaussian distribution based on $\tilde{\beta}$ and $\tilde{\sigma}$.

2. Generate $\sigma^{(j)}$ from the full conditional inverse gamma distribution conditional on $\boldsymbol{\nu}^{(j)}$.

3. Generate $\beta^{(j)}$ from the full conditional multivariate normal distribution conditional on ($\boldsymbol{\nu}^{(j)}, \sigma^{(j)}$).

4. Calculate $w^{(j)}$ based on $(\boldsymbol{\nu}^{(j)}, \sigma^{(j)}, \beta^{(j)})$ using

$$w^{(j)} = \frac{p(\beta^{(j)}, \sigma^{(j)}, \boldsymbol{\nu}^{(j)} \mid \mathbf{y})}{p(\beta^{(j)} \mid \sigma^{(j)}, \boldsymbol{\nu}^{(j)}, \mathbf{y})p(\sigma^{(j)} \mid \boldsymbol{\nu}^{(j)}, \mathbf{y})p^s(\boldsymbol{\nu}^{(j)})}$$

end

III. Importance Re-weighting Choose I samples via the importance weights without replacement.

Simulation Studies

PARAMETER ESTIMATION:

We follow [6] and simulate data from the model

$$y_i = 10 - x_i + \frac{1}{11}(11 + x_i)\epsilon_i, \quad i = 1, \ldots, 50, \tag{4}$$

with $x_i \sim$ Uniform(0, 10) and three error distributions $N(0,1)$, t_3, and χ_3^2.

The accuracy of the three methods GI, MH, and KFrQ is assessed using different criteria such as bias, root mean square error, and mean absolute error at quantile levels $p \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$. The bias is shown as below

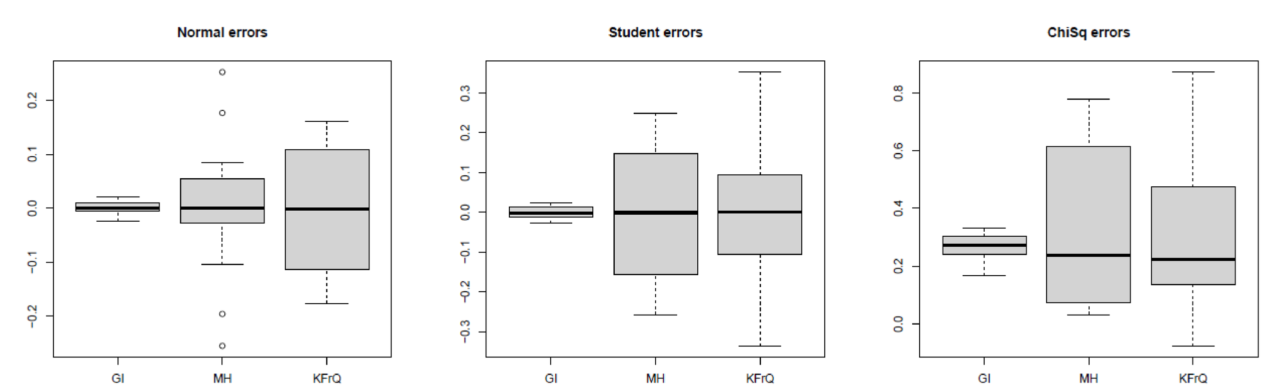

Fig. 1: Model parameter estimation performance in term of bias.

VARIABLE SELECTION:

Data are simulated from the model similar to [3]

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i, \quad i = 1, \ldots, 120, \tag{5}$$

with $\mathbf{X} \sim \mathbf{N}_{120}(0, \mathbf{I})$, $\beta = (0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1)'$, and eight different error distributions. We consider two $\mathbf{X}$ designs: (I) independent columns, (II) columns with pairwise correlations of around 0.8. For each design, consider two cases: (1) errors with median of zero, (2) errors with 90th percentile of zero.

Case I: Errors with median = 0

Model	p	Intpt	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Kur	0.05	0.9973	1.0000	1.0000	0.3846	0.3874	0.3881	0.4006	0.3985	0.4028	1.0000	1.0000
	0.50	0.2911	1.0000	1.0000	0.2892	0.2925	0.2922	0.2938	0.2912	0.2907	1.0000	1.0000
	0.95	0.9969	1.0000	1.0000	0.3821	0.3831	0.3894	0.3889	0.3965	0.4090	1.0000	1.0000
Bim	0.05	0.9991	1.0000	1.0000	0.4201	0.4245	0.4284	0.4239	0.4287	0.4358	1.0000	1.0000
	0.50	0.3394	1.0000	1.0000	0.3347	0.3317	0.3343	0.3425	0.3407	0.3403	1.0000	1.0000
	0.95	0.9998	0.9999	1.0000	0.4214	0.4175	0.4230	0.4261	0.4280	0.4408	1.0000	1.0000

Case II: Errors with 90th percentile = 0

Model	p	Intpt	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Sepa	0.05	1.0000	0.9808	0.9813	0.5098	0.4977	0.5018	0.5197	0.4924	0.5205	0.9859	0.9799
	0.50	0.9998	0.9992	0.9993	0.3454	0.3562	0.3622	0.3660	0.3637	0.3676	0.9995	0.9993
	0.95	0.3131	1.0000	1.0000	0.3402	0.3356	0.3393	0.3395	0.3404	0.3437	1.0000	1.0000
StT	0.05	1.0000	0.9886	0.9842	0.5283	0.5228	0.5226	0.5363	0.5394	0.5369	0.9889	0.9922
	0.50	1.0000	0.9990	0.9992	0.3943	0.3946	0.3922	0.4055	0.4003	0.4151	0.9995	0.9983
	0.95	0.4997	1.0000	1.0000	0.4106	0.4169	0.4132	0.4103	0.4155	0.4220	1.0000	1.0000

Table 1: Example 1: The MIPs of each predictor under the different error distributions at different quantile levels. In Case I: the intercept is included for $p = 0.05$ and $p = 0.95$ and excluded for $p = 0.50$. In Case II: the intercept is included for $p = 0.05$ and $p = 0.50$ and excluded for $p = 0.95$.

Conclusion

We have devised a three-stage computational scheme starting with an EM algorithm and then a Gibbs sampler followed by an importance re-weighting step using the ALD as a working likelihood for simultaneous parameter estimation and variable selection in quantile regression. The proposed sampling algorithm runs much faster than the exact MCMC algorithm, can generate iid samples from the target posterior distribution, and can choose accurate subset models, especially at the extreme quantiles such as $p = 0.05$ and $p = 0.95$. Our method is shown to be computationally efficient, precise, and accurate for statistical inference and variable selection procedures in both simulation studies and real-data application settings.

Boston Housing Data

We utilize the Boston housing data studied by [4] to illustrate the practicability of the proposed method for the parameter estimation and variable selection in quantile regression. The dataset consists of 506 observations on 16 variables, and the response variable is the corrected median value of owner occupied homes.

To assess the model's prediction accuracy on both parameter estimation and variable selection, we employed the 10-fold cross validation on the Boston housing data, as described in [3]. The measures of accuracy of the procedures are the mean absolute deviations (MADs) and mean weighted absolute residuals (MWARs), respectively. The formulas of these criteria are given by

$$\text{MAD} = \frac{1}{s_k}\sum_{i=1}^{s_k} y_i - \mathbf{x}_i'\hat{\beta}, \tag{6}$$

$$\text{MWAR} = \frac{1}{s_k}\sum_{i=1}^{s_k} \rho_p(y_i - \mathbf{x}_i'\hat{\beta}), \tag{7}$$

where s_k denotes the size of the validation set k, for $k = 1, \ldots, 10$.

The means and standard deviations of the MADs and MWARs over 10 cross validation folds are provided in the table below. The results of the three methods are considered at $p \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$.

		GI		MH		KFrQ	
	p	Mean	SD	Mean	SD	Mean	SD
MAD							
	0.05	5.3201	0.9911	7.1669	1.0060	5.3240	0.9531
	0.10	4.6823	0.8921	5.8458	0.9498	4.6771	0.8552
	0.25	3.5894	0.7729	4.0504	0.7913	3.5831	0.7675
	0.50	3.1337	0.5883	3.1519	0.5856	3.1198	0.5837
	0.75	3.8794	0.4324	4.1884	0.4034	3.9974	0.4293
	0.90	6.2774	0.4898	7.3172	0.5383	6.4383	0.5331
	0.95	9.0640	0.5999	10.4728	0.6136	9.2772	0.6420
MWAR							
	0.05	0.3350	0.0509	0.3757	0.0544	0.3334	0.0517
	0.10	0.5819	0.0911	0.6244	0.0871	0.5865	0.0846
	0.25	1.1580	0.1754	1.1986	0.1997	1.0962	0.1742
	0.50	1.6822	0.2861	1.7051	0.2756	1.6258	0.2889
	0.75	1.5685	0.3662	1.6050	0.2979	1.5779	0.2974
	0.90	1.0960	0.3472	1.0805	0.2603	1.2072	0.3601
	0.95	0.7733	0.2791	0.7135	0.1986	0.8147	0.2606

Table 2: Means and standard deviations of mean absolute deviations (MADs) and mean weighted absolute residuals (MWARs) under 10-fold cross validation for the GI, MH, and the KFrQ methods.

References

[1] Rahim Alhamzawi and Keming Yu. "Conjugate priors and variable selection for Bayesian quantile regression". In: *Computational Statistics & Data Analysis* 64 (2013), pp. 209–219.

[2] David F Andrews and Colin L Mallows. "Scale mixtures of normal distributions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1974), pp. 99–102. URL: https://www.jstor.org/stable/2984774?seq=1.

[3] Cathy WS Chen et al. "Bayesian variable selection in quantile regression". In: *Statistics and its Interface* 6.2 (2013), pp. 261–274. URL: https://dx.doi.org/10.4310/SII.2013.v6.n2.a9.

[4] David Harrison and Daniel L Rubinfeld. "Hedonic housing prices and the demand for clean air". In: *Journal of Environmental Economics and Management* 5.1 (1978), pp. 81–102. ISSN: 0095-0696. DOI: https://doi.org/10.1016/0095-0696(78)90006-2. URL: http://www.sciencedirect.com/science/article/pii/0095069678900062.

[5] Roger Koenker and Gilbert Bassett Jr. "Regression quantiles". In: *Journal of the Econometric Society* 46 (1978), pp. 33–50. URL: https://www.jstor.org/stable/1913643?seq=1.

[6] C. Reed and K. Yu. *A partially collapsed Gibbs sampler for Bayesian quantile regression.* Tech. rep. Department of Mathematical Sciences, Brunel University, 2009.

[7] Arnold Zellner. "On assessing prior distributions and Bayesian regression analysis with g-prior distributions". In: *Bayesian inference and decision techniques.* Vol. 6. Stud. Bayesian Econometrics Statist. Amsterdam: North-Holland, 1986, pp. 233–243.