Statistics and Its Interface Volume 6 (2013) 261–274

Bayesian variable selection in quantile regression

Keming Yu, Cathy W.S. Chen*, Craig Reed and David B. Dunson

In many applications, interest focuses on assessing relationships between predictors and the quantiles of the distribution of a continuous response. For example, in epidemiology studies, cutoffs to define premature delivery have been based on the 10th percentile of the distribution for gestational age at delivery. Using quantile regression, one can assess how this percentile varies with predictors instead of using a pre-defined cutoff. However, there is typically uncertainty in which of the many candidate predictors should be included. In order to identify important predictors and to build accurate predictive models, Bayesian methods for variable selection and model averaging are very useful. However, such methods are currently not available for quantile regression. This article develops Bayesian methods for variable selection, with a simple and efficient stochastic search variable selection (SSVS) algorithm proposed for posterior computation. This approach can be used for moderately high-dimensional variable selection and can accommodate uncertainty in basis function selection in non-linear and additive quantile regression models. The methods are illustrated using simulated data and an application to the Boston Housing data.

Keywords and phrases: Asymmetric Laplace, Extremes, Gibbs sampling, Model averaging, Risk, Stochastic search variable selection.

1. INTRODUCTION

Quantile regression is a very widely used approach in many application areas, with uncertainty in variable selection a routinely encountered problem. Hence, it is surprising that the literature on methods for variable selection in quantile regression is so sparse. When there are p candidate predictors, the number of possible subsets is 2^p, with this number enormous even for moderate p. It has become commonplace in many applications to have data for dozens to hundreds or thousands of predictors, so automated methods for identifying promising subsets of predictors, while accounting for the substantial uncertainty that occurs in the selection process are needed. Bayesian approaches provide a convenient paradigm for accommodating uncertainty in model selection (Hoeting et al., 1999; Clyde and George, 2004).

Bayesian methods for subset selection implemented using stochastic search variable selection (SSVS) algorithms (George and McCulloch, 1997) have become widely used in linear regression, generalized linear models and other modelling frameworks. Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)' \in \Gamma$ denote a model index, where $\gamma_j = 1$ denotes that the jth of p candidate predictors is included in the model, with $\gamma_j = 0$ otherwise, for $j = 1, \ldots, p$. Bayesian variable selection proceeds by first choosing a prior over the model space, $\pi(\boldsymbol{\gamma})$, with independent Bernoulli priors providing a commonly-used choice,

$$\pi(\boldsymbol{\gamma}) = \prod_{j=1}^{p} \pi_0^{\gamma_j} (1 - \pi_0)^{1-\gamma_j}, \tag{1}$$

where π_0 is the prior probability of including a randomly-selected predictor. It is common to fix π_0 (typically at $\frac{1}{2}$). However, by allowing a hyperprior on π_0, Scott and Berger (2010) point out that if the true number of predictors in a model is fixed but the number of candidate predictors increases, the posterior distribution of π_0 concentrates near 0, so it gives similar results to what would be obtained assuming π_0 is fixed at a low value. Scott and Berger (2010) refer to this as including an adjustment for multiplicities and explain that this intuitive behaviour does not occur when fixing π_0. A mathematically convenient hyperprior is $\pi_0 \sim \text{beta}(a_0, b_0)$.

A Bayesian specification of the model uncertainty problem is completed with priors for the coefficients within each model. For normal linear regression models, let $y_i = \mathbf{x}'_{\boldsymbol{\gamma},i}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \epsilon_i, \epsilon_i \sim N(0, \phi^{-1})$, where $\boldsymbol{x}_{\boldsymbol{\gamma},\boldsymbol{i}} = \{x_{ij} : \gamma_j = 1\}$ is the vector of predictors in model $\boldsymbol{\gamma}$ for subject i, $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is the $p_{\boldsymbol{\gamma}}$ coefficients in model $\boldsymbol{\gamma}$, and ϕ is the residual precision. Due to conjugacy, it is convenient to choose a prior of the form $\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\phi, \boldsymbol{\gamma}) = N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}; \mathbf{0}, \mathbf{V}/\phi)$ with $\pi(\phi) \propto \phi^{-1}$. This leads to a closed form for the marginal likelihood

$$L(\mathbf{y}; \mathbf{X}, \boldsymbol{\gamma}) = \int \left\{ \prod_{i=1}^{n} N\left(y_i; \mathbf{x}'_{\boldsymbol{\gamma},i}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \phi^{-1}\right) \right\} \times \pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\phi, \boldsymbol{\gamma})\pi(\phi) d\boldsymbol{\beta}_{\boldsymbol{\gamma}} d\phi.$$

Two common choices of prior in the literature include ridge priors, which let $\mathbf{V} = g\mathbf{I}$ with $\mathbf{I}$ the identity matrix, and g-priors, which let $\mathbf{V} = g(\mathbf{X}'_{\boldsymbol{\gamma}}\mathbf{X}_{\boldsymbol{\gamma}})^{-1}$.

The posterior probability allocated to model $\boldsymbol{\gamma}$ is

$$\pi(\boldsymbol{\gamma} \,|\, \mathbf{y}, \mathbf{X}) = \frac{\pi(\boldsymbol{\gamma}) L(\mathbf{y}; \mathbf{X}, \boldsymbol{\gamma})}{\sum_{\boldsymbol{\gamma}^* \in \Gamma} \pi(\boldsymbol{\gamma}^*) L(\mathbf{y}; \mathbf{X}, \boldsymbol{\gamma}^*)}, \quad \boldsymbol{\gamma} \in \Gamma. \tag{2}$$

*Corresponding author.

For linear regression and conjugate priors the posterior probability can be calculated exactly using this expression. However, when the number of models in Γ is very large, it is not possible to visit every model, so the denominator cannot be calculated. To bypass the need to visit every model, stochastic search algorithms instead use Markov chain Monte Carlo (MCMC) to explore the model space, while simultaneously obtaining Monte Carlo estimates of the posterior model probabilities (PMPs) and marginal inclusion probabilities (MIPs), $\Pr(\gamma_j = 1 \,|\, \mathbf{y}, \mathbf{X})$, which provide a weight of evidence that the jth predictor should be included accounting for uncertainty in the other predictors.

SSVS proceeds by a Gibbs sampling algorithm that sequentially updates each γ_j by sampling from the full conditional posterior distribution, which is Bernoulli with the conditional probability of $\gamma_j = 1$ available in a closed form. A simple summary of the SSVS output could be to report the model with highest posterior probability, which is denoted here as $\boldsymbol{\gamma}^+$. This is recommended by Chen (1999), So and Chen (2003), and Chen et al. (2011) and corresponds to the model selected the most number of times during the MCMC run. However, it is in general difficult to obtain accurate estimates of PMPs in large model spaces. An alternative is to report MIPs and these tend to be estimated efficiently using SSVS. These MIPs can form the basis for inferences on the need to include each predictor. In addition, if the goal is to select a single model for prediction, Barbieri and Berger (2004) define the median probability model, denoted here as $\boldsymbol{\gamma}^{\mathrm{m}}$, to include those predictors having MIPs greater than 0.5. Under a loss function, which expresses the loss as the total number of predictors that are inappropriately included or excluded from the model, Barbieri and Berger (2004) show that the model that minimizes the expected posterior loss (Bayes risk) is $\boldsymbol{\gamma}^{\mathrm{m}}$. Under the usual settings for SSVS, the model with all predictors is one of the many models under consideration. This is one situation in which $\boldsymbol{\gamma}^{\mathrm{m}}$ is guaranteed to exist (Barbieri and Berger, 2004). In rare cases, it may happen that $\boldsymbol{\gamma}^{\mathrm{m}}$ does not correspond to any of the models actually visited by SSVS although we have never yet come across this situation. If such a case did arise, it would suggest that the SSVS algorithm has not fully explored the posterior distribution of models and that running the SSVS algorithm for longer would be beneficial.

Unfortunately, with the notable exceptions of a single paper by Meligkotsidou, Vrontos and Vrontos (MVV) (2009), there are no methods currently available for Bayesian variable selection in quantile regression models. One of the difficulties facing the practitioner is to specify a suitable likelihood given that the frequentist specification of quantile regression is through the result of minimizing a tilted absolute loss function defined in the next section. The MVV approach relied on the asymmetric Laplace likelihood recommended by Yu and Moyeed (2001). This likelihood has the attractive property that maximizing it corresponds directly to the frequentist procedure mentioned above. Hence, the posterior mode can be obtained using various linear programming techniques such as those in the `quantreg` package of Koenker (2009). As observed by Li et al. (2010), proper priors can be used to obtain regularized solutions, with the L1 penalized quantile regression solutions corresponding to the assumption of independent double exponential priors.

Unlike in normal linear regression models, there is unfortunately no conjugate prior available and hence calculation of the marginal likelihood $L(\mathbf{y}; \mathbf{X}, \boldsymbol{\gamma})$ requires approximation of a potentially high-dimensional integral, depending on the number of predictors in the model. MVV solve this problem using the widely-used Laplace approximation (Tierney and Kadane, 1986), which also forms the basis for the Bayesian information criterion (BIC) and for routine applications of Bayesian variable selection in generalized linear models and other model classes. In quantile regression models, one has particular concerns about the accuracy of the Laplace approximation, since even if there is a large sample size over all, there may be relatively limited data in the tails of the distribution near the quantile of interest. Even if there are substantial numbers of subjects having values near the quantile, this is unlikely to be true in all regions of the predictor space. In our experience, posterior distributions for quantile regression coefficients are commonly skewed, suggesting inaccuracy of the Laplace approximation.

The goal of this article is to develop exact SSVS methods for Bayesian variable selection in quantile regression models, with the term "exact" referring to the fact that we avoid inaccuracies introduced by analytic approximations to marginal likelihoods. Instead we rely on an innovative SSVS algorithm, which takes advantage of conditional conjugacy after re-expressing the asymmetric Laplace residual distribution as a location-scale mixture of normals. This allows us to take advantage of closed forms that are available for marginal likelihoods in normal linear mean regression models, so that the model index $\boldsymbol{\gamma}$ can be updated exactly as in the mean regression case after also updating latent variables and other unknowns that are common to the different models from simple steps.

It is for this reason that we opt for the AL distribution in the rest of this paper. Alternatives to the AL distribution have been suggested (see Reed and Yu, 2011). Unfortunately, such alternatives do not naturally generalize to handling model uncertainty.

The outline of the paper is as follows. In Section 2 we describe the quantile regression variable selection problem and review the scale mixture of normals representation. In Section 3 we describe in detail the proposed quantile regression SSVS (QR-SSVS) algorithm. Section 4 contains a simulation study in which we assess the performance of QR-SSVS on data arising from 8 different distributions. We also assess the frequentist operating characteristics relative to the direct use of the asymptotic t-test by simulating 100 replicates. Section 5 contains an application to the Boston Housing data set, and Section 6 discusses the results.

2. QUANTILE REGRESSION VARIABLE SELECTION

2.1 Augmented likelihood specification

We follow Yu and Moyeed (2001) in using the asymmetric Laplace (AL) likelihood with location parameter $\mu_i = \boldsymbol{x}_{\boldsymbol{\gamma},i}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}(\tau)$, where $\boldsymbol{\beta}_{\boldsymbol{\gamma}}(\tau)$ is a vector of parameters in model $\boldsymbol{\gamma}$ which depends on τ, $0 < \tau < 1$. Komunjer (2005) gives a theoretical justification for using this likelihood and empirical results from Reed and Yu (2011) suggest that it accurately approximates the true quantiles of many distributions having different properties. The AL likelihood is proportional to

$$\exp\left\{-\sum_{i=1}^{n}\rho_\tau\left[y_i - \boldsymbol{x}_{\boldsymbol{\gamma},i}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}(\tau)\right]\right\}, \tag{3}$$

where

$$\rho_\tau(z) = \{|z| + (2\tau - 1)z\}/2 \tag{4}$$

denotes the tilted absolute value or "check" function and z in (4) corresponds to $y_i - \boldsymbol{x}_{\boldsymbol{\gamma},i}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}(\tau)$ in (3). From this point onwards, we suppress dependence on τ to simplify notation.

In constructing an SSVS procedure for quantile regression, we can use an equivalent specification with a normal likelihood proportional to

$$\left(\prod_{i=1}^{n} w_i^{-1/2}\right)\exp\left\{-\frac{1}{4}\sum_{i=1}^{n}\frac{\{y_i - (1-2\tau)w_i + \boldsymbol{x}_{\boldsymbol{\gamma},i}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}\}^2}{w_i}\right\}. \tag{5}$$

We place independent and identical exponential priors on each w_i with rate parameter $\tau(1-\tau)$. Combining (5) with the priors on $\mathbf{w} = [w_i]_{i=1}^n$ and marginalizing over $\mathbf{w}$ recovers the asymmetric Laplace likelihood in (3) (Tsionas, 2003; Rue and Held, 2005). With this specification, we can rely directly on techniques developed in George and McCulloch (1997) for Bayesian variable selection in normal linear models. In particular, we can develop a data augmentation SSVS algorithm that relies on conjugacy after augmentation to marginalize out the regression coefficients $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ specific to model $\boldsymbol{\gamma}$. This marginalization is key in obtaining a computationally feasible algorithm. By using conditionally conjugate priors after augmentation, we avoid the need to rely on potentially-imprecise analytic approximations to the marginal likelihood, such as Laplace.

2.2 Prior specification

Following common practice, we can embed all the submodels $\boldsymbol{\gamma} \in \Gamma$ within the full model by letting $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ denote the coefficients on the p predictors in the full model, with $\beta_j = 0$ for all j such that $\gamma_j = 0$. Then, we can simultaneously induce a prior for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ by choosing a prior for $\boldsymbol{\beta}$ as follows:

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}) = \prod_{j=1}^{p}\left\{(1-\gamma_j)\delta_0 + \gamma_j N(0, \lambda_j^{-1})\right\},$$

where δ_0 denotes a degenerate distribution with all its mass at zero, $\pi_0 = \Pr(\gamma_j = 1) = \Pr(\beta_j \neq 0)$ is the prior probability of including a randomly selected predictor in the model, and letting $\lambda_j \sim \text{Gamma}(1/2, 1/2)$ induces a heavy-tailed Cauchy prior marginally for the coefficients on the predictors selected to be in the model. In particular, we have $\beta_j \sim$ Cauchy independently for j such that $\gamma_j = 1$. The Cauchy is widely-used as a robust prior. We opted for the prior specification for γ_j following Scott and Berger (2010) and letting each $\gamma_j \sim \text{Bernoulli}(\pi_0)$ independently and additionally placing a beta hyperprior on the prior inclusion probability, $\pi_0 \sim \text{beta}(a_0, b_0)$.

3. QR-SSVS ALGORITHM

In order to simultaneously search for high posterior probability models in Γ while also conducting posterior computation for the regression coefficients specific to each model, we propose a data augmentation SSVS algorithm, which proceeds by alternating between simple Gibbs sampling steps. Due to the structure of the model and the priors specified in Section 2, the Gibbs sampling steps take particularly simple forms, so sampling is straightforward. To improve mixing we marginalize out the jth regression parameter in updating the indicator γ_j for inclusion of the jth predictor. The steps proceed as follows.

1. Set initial values for π_0, $\mathbf{w}$ and $\boldsymbol{\lambda}_{\boldsymbol{\gamma}} = [\lambda_j : \gamma_j = 1]'$. The latter two could be sampled from their respective priors.
2. Update the indicator γ_j marginalizing out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$. To obtain the conditional posterior of γ_j given $\boldsymbol{\gamma}_{-j} = \{\gamma_k, k \neq j\}$ and the data, first note that the conditional posterior distribution $\pi(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\boldsymbol{\gamma}}|\mathbf{y}, \mathbf{w}, \boldsymbol{\lambda}_{\boldsymbol{\gamma}})$ can be written as

$$\left(\prod_{j:\gamma_j=1}\lambda_j^{-1/2}\right)\exp\left(-\frac{1}{2}\|\tilde{\mathbf{u}} - \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}\|^2\right),$$

where

$$\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}} = \left[\sqrt{\tfrac{1}{2}}\boldsymbol{X}_{\boldsymbol{\gamma}}'\mathbf{W}^{1/2} \quad \text{diag}\{\sqrt{\lambda_j} : \gamma_j = 1\}\right]',$$
$$\tilde{\mathbf{u}} = \left[\sqrt{\tfrac{1}{2}}(\mathbf{y} - \{1-2\tau\}\mathbf{w})'\mathbf{W}^{1/2} \quad \mathbf{0}\right]',$$
$$\mathbf{W} = \text{diag}\{w_i^{-1},\ i = 1, \ldots, n\}$$

and $\|\cdot\|$ denotes the usual Euclidean norm.
Marginalizing out $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ gives (using similar notation to George and McCulloch (1997))

$$\pi(\boldsymbol{\gamma}\,|\,\mathbf{y},\mathbf{w},\boldsymbol{\lambda}) \propto \mathrm{g}(\boldsymbol{\gamma})$$
$$\equiv \left(\prod_{j:\gamma_j=1}\lambda_j^{-\frac{1}{2}}\right)|\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}|^{-\frac{1}{2}}$$
$$\times \exp\left(-\frac{1}{2}\|\tilde{\mathbf{u}}-\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}\|^2\right),$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}})^{-1}\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\tilde{\mathbf{u}}$.
We then have

$$\gamma_j|\boldsymbol{\gamma}_{-j},\mathbf{w},\boldsymbol{\lambda},\pi_0 \sim \mathrm{Bern}(\pi_1),$$

with

$$\pi_1 = \frac{\pi_0\ \mathrm{g}(\gamma_j=1,\boldsymbol{\gamma}_{-j})}{\pi_0\ \mathrm{g}(\gamma_j=1,\boldsymbol{\gamma}_{-j})+(1-\pi_0)\ \mathrm{g}(\gamma_j=0,\boldsymbol{\gamma}_{-j})}.$$

Each of the components of $\boldsymbol{\gamma}$ can be updated either in a fixed or random order.

3. Sample the regression coefficient specific to the current model from their full conditional,

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}} \sim \mathrm{N}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}},(\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}})^{-1}).$$

4. Update the latent variables $\mathbf{w}$ from their full conditional posterior distributions, each of which can be shown to have the following form after some algebra:

$$w_i^{-1} \sim \mathrm{IG}\left(\frac{1}{|y_i-\boldsymbol{x}_{\boldsymbol{\gamma},\boldsymbol{i}}{}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}|},\frac{1}{2}\right),$$

where IG denotes the inverse Gaussian distribution.

5. Sample from the full conditional posteriors for λ_j, for $j=0,1,\ldots,p$,

$$\lambda_j \sim \mathrm{Exponential}(\tfrac{1}{2}\{\beta_j^2+1\}).$$

6. Update π_0 if a beta hyperprior was used:

$$\pi_0 \sim \mathrm{Beta}(p_{\boldsymbol{\gamma}}+a_0,p-p_{\boldsymbol{\gamma}}+b_0).$$

This algorithm can be extended to the case where predictors are a priori certain to appear in the model by expressing the location parameters of the asymmetric Laplace likelihood, μ_i, as $\mu_i = \mathbf{z}_i'\boldsymbol{\alpha}+\boldsymbol{x}_{\boldsymbol{\gamma},\boldsymbol{i}}{}'\boldsymbol{\beta}_{\boldsymbol{\gamma}}$. The most common situation is when the model is assumed to always contain an intercept term. If an improper prior is placed on $\boldsymbol{\alpha}$, then by a suitable adjustment to the definition of $\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}$, we can draw $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ in one block in step 3. This also applies for any proper normal prior on $\boldsymbol{\alpha}$.

4. SIMULATION STUDY

4.1 Example 1

We conducted a series of simulation experiments to assess the performance of the proposed QR-SSVS algorithm. In the first set of simulation examples, we focused on the case in which the data are drawn from a true model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}. \tag{6}$$

Here, $\mathbf{y}$ denotes the vector of response variables y_i, where $i=1,\ldots,120$, $\mathbf{X}$ is the 120×11 design matrix with the first column a column of 1s and the remaining columns independent $\mathrm{N}_{120}(\mathbf{0},\mathbf{I})$ random variables, $\boldsymbol{\beta}=[\beta_0\ldots\beta_{10}]'$ and $\boldsymbol{\epsilon}$ denotes the residuals. We allowed the density of the residuals $\boldsymbol{\epsilon}$ to vary across a broad variety of cases to assess robustness of the proposed approach, given that the asymmetric Laplace likelihood is used for convenience in quantile regression and not because this distributional form is thought to be an accurate reflection of the true likelihood. The conditional quantiles for model (6) are parallel, with only the intercept dependent on τ through the quantity $\beta_0+Q_\tau(\epsilon_i)$, where $Q_\tau(\epsilon_i)$ denotes the τth quantile of ϵ_i. We set $\beta=(0,1,1,0,0,0,0,0,0,1,1)'$. The following residual densities were considered:

- Gaussian (Gau): $\mathrm{N}(0,1^2)$
- Skewed (Skew): $\frac{1}{5}\mathrm{N}(-\frac{22}{25},1^2)+\frac{1}{5}\mathrm{N}(-\frac{49}{125},\frac{3}{2}{}^2)+\frac{3}{5}\mathrm{N}(\frac{49}{250},\frac{5}{9}{}^2)$
- Kurtotic (Kur): $\frac{2}{3}\mathrm{N}(0,1^2)+\frac{1}{3}\mathrm{N}(0,\frac{1}{10}{}^2)$
- Bimodal (Bim): $\frac{1}{2}\mathrm{N}(-1,\frac{2}{3}{}^2)+\frac{1}{2}\mathrm{N}(1,\frac{2}{3}{}^2)$
- Bimodal, separate modes (Sepa): $\frac{1}{2}\mathrm{N}(-\frac{3}{2},\frac{1}{2}{}^2)+\frac{1}{2}\mathrm{N}(\frac{3}{2},\frac{1}{2})^2)$
- Skewed bimodal (Skeb): $\frac{3}{4}\mathrm{N}(-\frac{43}{100},1^2)+\frac{1}{4}\mathrm{N}(\frac{107}{100},\frac{1}{3}{}^2)$
- Trimodal (Tri): $\frac{9}{20}\mathrm{N}(-\frac{6}{5},\frac{3}{5}{}^2)+\frac{9}{20}\mathrm{N}(\frac{6}{5},\frac{3}{5}{}^2)+\frac{1}{10}\mathrm{N}(0,\frac{1}{4}{}^2)$
- Cauchy (Cau): t_1

These distributions were chosen to have a median close to or equal to zero. For the first eight simulations, we drew $n=120$ observations each from model (6) with the eight error distributions specified above. The second case repeated these eight simulations with the error distributions shifted so that the 90th percentile was close to zero instead of the median. In the first eight simulations, the intercept term should be close to zero in a regression model for the median, but significant shifted from zero in a regression model for the 90th percentile, with the opposite being true for the remaining eight simulations. Although the intercept is seldom of interest in variable selection, to demonstrate how QR-SSVS can select different models at different quantiles, we include the intercept as a potential predictor, giving a total of $2^{11}=2,048$ potential models for each quantile under consideration.

Using a $\mathrm{Beta}(1,1)$, or equivalently, a uniform prior on π_0, we ran QR-SSVS for 10,000 iterations following a burn in of 1,000 iterations separately for $\tau=0.5$ and $\tau=0.9$. As a convergence diagnostic and to verify robustness of the results to starting points, we repeated the analysis with widely different starting values and obtain essentially identical results. MIPs for the intercept and each candidate predictors are

Table 1. MIPs from the 16 simulations with data generated from the 8 error distributions in the first column. The predictors are approximately independent. Predictors in the true model for each simulation are boxed. The intercept should be excluded ($\tau = 0.5$) and included ($\tau = 0.9$) when the errors have median zero. It should be included ($\tau = 0.5$) and excluded ($\tau = 0.9$) when the errors have 90th percentile equal to zero

Case 1: Errors with median = 0												
	τ	Intercept	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Gau	0.5	0.305	1.000	1.000	0.299	0.226	0.446	0.305	0.251	0.213	1.000	1.000
	0.9	1.000	0.980	0.980	0.381	0.345	0.398	0.431	0.402	0.360	0.998	0.970
Skew	0.5	0.188	1.000	1.000	0.214	0.166	0.180	0.176	0.169	0.147	1.000	1.000
	0.9	1.000	0.997	0.993	0.314	0.309	0.299	0.309	0.307	0.325	0.997	0.996
Kur	0.5	0.155	1.000	1.000	0.139	0.127	0.135	0.124	0.160	0.139	1.000	1.000
	0.9	1.000	0.994	0.977	0.450	0.407	0.398	0.469	0.407	0.429	0.999	0.985
Bim	0.5	0.371	1.000	0.994	0.267	0.322	0.326	0.326	0.287	0.333	1.000	1.000
	0.9	1.000	0.974	0.960	0.421	0.383	0.408	0.403	0.398	0.422	0.997	0.989
Sepa	0.5	0.326	1.000	0.718	0.331	0.482	0.633	0.349	0.351	0.435	1.000	1.000
	0.9	1.000	0.972	0.932	0.329	0.328	0.339	0.359	0.342	0.321	0.985	0.992
Skeb	0.5	0.539	1.000	1.000	0.253	0.242	0.277	0.316	0.248	0.348	1.000	0.987
	0.9	1.000	0.995	0.995	0.341	0.297	0.342	0.326	0.386	0.304	1.000	0.948
Tri	0.5	0.292	1.000	0.998	0.288	0.279	0.647	0.276	0.280	0.768	1.000	1.000
	0.9	1.000	0.989	0.945	0.367	0.380	0.386	0.374	0.387	0.336	0.984	0.970
Cau	0.5	0.210	1.000	1.000	0.226	0.235	0.289	0.260	0.204	0.242	1.000	0.984
	0.9	0.555	0.999	0.849	0.576	0.563	0.529	0.524	0.739	0.768	0.999	0.854
Case 2: Errors with 90th percentile = 0												
	τ	Intercept	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Gau	0.5	1.000	1.000	1.000	0.278	0.317	0.392	0.297	0.251	0.336	1.000	1.000
	0.9	0.573	0.999	0.995	0.300	0.319	0.291	0.365	0.303	0.334	0.995	1.000
Skew	0.5	1.000	1.000	1.000	0.239	0.353	0.232	0.348	0.232	0.303	1.000	1.000
	0.9	0.678	0.991	0.991	0.268	0.288	0.274	0.270	0.277	0.293	0.996	0.998
Kur	0.5	1.000	1.000	1.000	0.191	0.222	0.177	0.207	0.188	0.192	1.000	1.000
	0.9	0.798	0.993	0.998	0.349	0.322	0.325	0.330	0.359	0.344	0.939	0.997
Bim	0.5	1.000	1.000	0.997	0.322	0.389	0.328	0.433	0.294	0.438	1.000	1.000
	0.9	0.535	0.992	0.888	0.270	0.311	0.266	0.327	0.295	0.364	0.999	0.993
Sepa	0.5	1.000	0.999	0.990	0.698	0.743	0.542	0.548	0.483	0.674	0.998	1.000
	0.9	0.653	0.991	0.976	0.296	0.336	0.297	0.304	0.300	0.312	0.995	0.998
Skeb	0.5	1.000	1.000	1.000	0.285	0.280	0.278	0.282	0.310	0.303	1.000	1.000
	0.9	0.883	0.994	0.994	0.326	0.293	0.302	0.303	0.283	0.309	0.995	0.996
Tri	0.5	1.000	1.000	1.000	0.348	0.553	0.348	0.418	0.322	0.350	1.000	1.000
	0.9	0.512	0.996	0.989	0.310	0.338	0.279	0.375	0.294	0.340	0.997	0.998
Cau	0.5	1.000	1.000	0.989	0.351	0.357	0.350	0.357	0.570	0.485	1.000	1.000
	0.9	0.733	0.729	0.626	0.572	0.504	0.528	0.478	0.558	0.584	0.704	0.927

presented for each of the 16 different simulations in Table 1. For each of the 16 simulations, we expect x_1, x_2, x_9, x_{10} to be in the model with the rest of the predictors excluded. For simulations 1 to 8, we expect the intercept to be excluded when we set $\tau = 0.5$ as the argument to QR-SSVS but not when we set $\tau = 0.9$ as the argument. The reverse applies in simulations 9 to 16.

It would be useful to compare with standard linear regression model with SSVS to study the relative efficiency of the QR-SSVS model in the simulation study where the response is either Gaussian or Cauchy. Based on the idea of George and McCulloch (1993), we assume that the slope parameters can be stated as the following multivariate normal prior.

$$\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N(\mathbf{0}, \boldsymbol{D_\delta V D_\delta}) \,, \tag{7}$$

where $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_p)'$, $\boldsymbol{V}$ is the prior correlation matrix and $\boldsymbol{D_\gamma}$ is the diagonal matrix $diag\{a_0\nu_0, \ldots, a_p\nu_p\}$ with

$a_k = 1$ if $\gamma_k = 0$ and $a_k = c_k$ if $\gamma_k = 1$. In particular when we do not have any prior information about the relationship among β_k that $\boldsymbol{V} = \boldsymbol{I}$, the covariance $\boldsymbol{D}_{\boldsymbol{\gamma}}\boldsymbol{V}\boldsymbol{D}_{\boldsymbol{\gamma}}$ reduces to a diagonal matrix with elements $a_k^2\nu_k^2$, $k = 0, \ldots, p$.

The general criteria for setting c_k and ν_k are well explained by George and McCulloch (1993); we summarize: (i) when γ_k=0 and thus $\beta_k(\approx 0)$ is not included in the model, the prior can be chosen to be informative around 0, with low variance: i.e. ν_k^2 should be small and close to 0 to ensure that the posterior of β_k would be shrunk around zero when $\gamma_k = 0$. (ii) when γ_k=1 and thus $\beta_k(\neq 0)$ is included, the prior could be better chosen as flat and uninformative. Thus $c_k^2\nu_k^2$ should be large, requiring $c_k > 1$ and large enough to make $c_k^2\nu_k^2$ substantially greater than ν_k^2, and of reasonable size, so that $\beta_k|\gamma_k = 1$ has high prior variation. The SSVS method is successfully employed to autoregression (AR) and threshold AR models in Chen (1999) and So and Chen (2003). Both papers choose the ratio combinations $(\sigma_{\beta_k}/\nu_k, c_k) = (0.5, 5), (0.5, 10), (1, 5)$ and $(1, 10)$ to provide sensitivity analysis for subset selection. They conclude that the combination (0.5, 10) is optimal. Therefore, the choice of $(\sigma_{\beta_k}/\nu_k, c_k)$ is (0.5, 10) here. The results of model selection by linear regression model with SSVS is given in Table 1.

In each of the 16 simulations, the median probability model $\boldsymbol{\gamma}^{\mathrm{m}}$ included the true predictors x_1, x_2, x_9, x_{10} and the MIPs were in general quite close to one. In addition, $\boldsymbol{\gamma}^{\mathrm{m}}$ correctly included the intercept for the $\tau = 0.9$ analysis when the true median was zero and for the $\tau = 0.5$ analysis when the true 90th percentile was zero. However, using the 0.5 cutoff on the MIPs, there were some false positives, which is not surprising given the modest sample size of $n = 120$. In particular, in 10.9% of the overall cases, the MIP for a predictor that should not be included was above 0.5. In the cases involving Cauchy errors and $\tau = 0.9$, this rate was instead $11/12 = 91.7\%$. It is not surprising that for a very heavy-tailed error distribution that information available in the data regarding whether to include a predictor impacting a quantile in the tails is limited. In the absence of any information in the data about whether to include the predictor, the MIP will be expected to be right around 0.5. Nevertheless, for both values of τ, predictors x_1, x_2, x_9 and x_{10} have larger MIPs than $x_3, \ldots, x_8$ under Cauchy distributed residuals. If we instead use a more stringent cutoff for significance in which predictors are considered significant if the MIP is greater than 0.9 **(say)** we obtained an overall power of 94.5% and type I error rate of 0%.

4.2 Example 2

This simulation is identical to simulation 1, but with the columns of the design matrix no longer independent. We used a similar design matrix to George and McCulloch (1997) example 5.2.1. After a column of 1s, this involved for each column j, $j = 2, \ldots, 11$, generating $n = 120$ independent standard normal variates. Then, we independently generated 120 additional standard normal variates $\mathbf{z}$ and added $2\mathbf{z}$ to each column excluding the first. This resulted in correlation between $\mathbf{x}_j$ and $\mathbf{x}_k$, $k \neq \{1, j\}$ of around 0.8. Table 2 contains the MIPs for each predictor for this simulation.

Although the results seem similar to example 1, there is unsurprisingly an overall increase in the MIP for predictors x_3 to x_8. Compared to the 10.9% from the previous example of overall cases for which the MIPs were bigger than 0.5 for a predictor that is not in the true model, this time it was 74.0%. This again suggests that if a single model is desired, increasing the threshold for including a predictor may be recommended to maintain a low false positive rate while also having reasonably large power. For a threshold of 0.9, we had a power of 96.1% and a type I error rate of 2.6%.

In practice, this ad hoc thresholding would be unsuitable unless there was a good prior reason to choose a particular threshold value. Barbieri and Berger (2004) show that the value of 0.5 has appealing properties if the error is normally distributed and yields the optimal predictive model in certain settings. However, in our case, errors are not normally distributed and the second example has correlated predictors, the situation in which Barbieri and Berger (2004) suggest that $\boldsymbol{\gamma}^{\mathrm{m}}$ may not be optimal.

The maximum probability model $\boldsymbol{\gamma}^{+}$ in the first example matched $\boldsymbol{\gamma}^{\mathrm{m}}$ in all but one case (where $\boldsymbol{\gamma}^{+}$ had 1 additional predictor) and in the second case, it didn't ever match $\boldsymbol{\gamma}^{\mathrm{m}}$ so was not reported in the tables. As mentioned previously, using $\boldsymbol{\gamma}^{+}$ as the optimal model is not generally recommended as $\boldsymbol{\gamma}^{+}$ is not well estimated for large model spaces. As a final alternative, it may be useful to consider a model where the data chooses the thresholding for the MIPs. Cross validation is a common approach for tuning parameters and could also be implemented for this purpose.

4.3 Example 3

In this example, we simulated 100 replicates $\mathbf{y}_r$, $r = 1, \ldots, 100$ to assess the frequentist performance of QR-SSVS. The simulation and analysis was conducted as described in the previous subsection focusing on the Gaussian residuals case. We compared $\boldsymbol{\gamma}^{\mathrm{m}}$ and the frequentist model, which was based on carrying out frequentist quantile regression estimation for the full model and then including those predictors with p-values less than 0.01. We focused on $\tau = 0.5$ and $\tau = 0.9$. Table 3 presents a summary of $\boldsymbol{\gamma}^{\mathrm{m}}$ selected using QR-SSVS for each replicate, and the models selected using the frequentist approach.

We calculated the type I error rate and power for each predictor based on $\boldsymbol{\gamma}^{\mathrm{m}}$, and the model containing predictors with MIP greater than 0.9. The frequentist type I error rate and power was calculated for the models containing predictors with p value less than 0.05 and 0.01 respectively. Again, it may be useful to let the data suggest a sensible p value to use by implementing cross validation. We also compared the posterior median and frequentist estimates

Table 2. MIPs from the 16 simulations with data generated from the 8 error distributions in the first column. The predictors are highly correlated. Predictors in the true model for each simulation are boxed. The intercept should be excluded ($\tau = 0.5$) and included ($\tau = 0.9$) when the errors have median zero. It should be included ($\tau = 0.5$) and excluded ($\tau = 0.9$) when the errors have 90th percentile equal to zero

		Case 1: Errors with median = 0										
	τ	Intercept	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Gau	0.5	0.245	1.000	1.000	0.399	0.400	0.543	0.363	0.378	0.361	1.000	1.000
	0.9	1.000	0.999	0.994	0.627	0.627	0.771	0.648	0.611	0.624	0.998	0.999
Skew	0.5	0.237	1.000	1.000	0.372	0.350	0.410	0.339	0.353	0.303	1.000	1.000
	0.9	1.000	0.999	0.997	0.587	0.620	0.678	0.618	0.608	0.572	0.999	1.000
Kur	0.5	0.211	1.000	1.000	0.286	0.292	0.302	0.296	0.268	0.255	1.000	1.000
	0.9	1.000	0.999	0.999	0.628	0.633	0.656	0.644	0.661	0.612	0.999	0.996
Bim	0.5	0.370	1.000	1.000	0.567	0.566	0.531	0.534	0.926	0.480	1.000	1.000
	0.9	1.000	0.999	0.999	0.678	0.652	0.675	0.657	0.738	0.643	1.000	0.994
Sepa	0.5	0.466	1.000	1.000	0.626	0.655	0.630	0.667	0.961	0.550	1.000	1.000
	0.9	1.000	0.995	0.999	0.640	0.642	0.673	0.626	0.659	0.622	0.997	0.998
Skeb	0.5	0.354	1.000	1.000	0.670	0.522	0.516	0.465	0.502	0.412	1.000	1.000
	0.9	1.000	0.998	0.999	0.641	0.590	0.616	0.630	0.605	0.582	1.000	0.997
Tri	0.5	0.375	1.000	1.000	0.499	0.540	0.479	0.548	0.609	0.472	1.000	1.000
	0.9	1.000	0.998	0.994	0.697	0.663	0.703	0.715	0.693	0.655	0.999	0.993
Cau	0.5	0.462	0.998	0.999	0.753	0.634	0.631	0.524	0.510	0.612	1.000	1.000
	0.9	1.000	0.868	0.893	0.795	0.764	0.827	0.781	0.944	0.785	0.893	0.992
		Case 2: Errors with 90th percentile = 0										
	τ	Intercept	$\boxed{x_1}$	$\boxed{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	$\boxed{x_9}$	$\boxed{x_{10}}$
Gau	0.5	1.000	1.000	1.000	0.449	0.596	0.452	0.466	0.479	0.458	1.000	1.000
	0.9	0.409	1.000	0.997	0.467	0.493	0.475	0.470	0.491	0.459	1.000	1.000
Skew	0.5	1.000	1.000	1.000	0.467	0.524	0.519	0.528	0.507	0.499	1.000	1.000
	0.9	0.396	1.000	1.000	0.472	0.506	0.471	0.603	0.508	0.491	1.000	1.000
Kur	0.5	1.000	1.000	1.000	0.357	0.416	0.368	0.403	0.391	0.394	1.000	1.000
	0.9	0.448	1.000	0.999	0.510	0.590	0.530	0.540	0.527	0.525	1.000	1.000
Bim	0.5	1.000	1.000	1.000	0.565	0.614	0.583	0.588	0.554	0.560	1.000	1.000
	0.9	0.422	1.000	1.000	0.502	0.530	0.505	0.516	0.555	0.532	1.000	1.000
Sepa	0.5	1.000	1.000	1.000	0.811	0.707	0.696	0.697	0.721	0.697	1.000	1.000
	0.9	0.417	1.000	1.000	0.476	0.551	0.561	0.481	0.489	0.496	0.999	1.000
Skeb	0.5	1.000	1.000	1.000	0.526	0.557	0.510	0.545	0.578	0.665	1.000	1.000
	0.9	0.396	1.000	1.000	0.491	0.555	0.518	0.518	0.536	0.580	1.000	0.999
Tri	0.5	1.000	1.000	1.000	0.590	0.588	0.646	0.583	0.601	0.591	1.000	1.000
	0.9	0.449	1.000	0.999	0.519	0.582	0.545	0.545	0.620	0.546	1.000	0.999
Cau	0.5	1.000	1.000	1.000	0.639	0.931	0.681	0.675	0.732	0.826	1.000	1.000
	0.9	0.723	0.981	0.816	0.764	0.897	0.864	0.820	0.851	0.903	0.798	0.987

to the true values, while assessing coverage of the 95% credible intervals and frequentist 95% confidence intervals. To obtain the frequentist intervals, we used the default rank method in the `R` package `quantreg` (Koenker, 2009). The coverage rate of the intervals was obtained using the number of replicates for which the true value of the regression parameter was contained within the interval. Finally, given that the true τth quantile of y_i is known, we obtained mean squared errors using $\frac{1}{100}\sum_{r=1}^{100}\{\sum_{i=1}^{120}(Q_\tau(y_{ir}) - \boldsymbol{x}_{\boldsymbol{\gamma},\boldsymbol{i}}\hat{\beta})^2$, again using $Q_\tau(\cdot)$ to denote the τth quantile. The overall results averaged over the different predictors not including the intercept are presented in Table 4.

Recalling that the true model contains x_1, x_2, x_9, x_{10} in the $\tau = 0.5$ case with the intercept also included in the

Table 3. The median probability models using quantile regression stochastic search (QR-SSVS), the approach of MMV using the smallest value of the Bayesian information criterion (BIC) and the frequentist models constructed by taking those predictors with p-values less than 0.01

Models for $\tau = 0.5$	QR-SSVS	MVV	Frequentist
x_1, x_2, x_9, x_{10}	89	83	81
Intercept, x_1, x_2, x_9, x_{10}	8	9	8
Intercept, x_1, x_2, x_7, x_9, x_{10}	1	1	0
x_1, x_2, x_3, x_9, x_{10}	1	1	1
x_1, x_2, x_5, x_9, x_{10}	1	2	5
x_1, x_2, x_8, x_9, x_{10}	0	1	2
x_1, x_2, x_4, x_9, x_{10}	0	1	1
x_1, x_2, x_6, x_9, x_{10}	0	1	1
x_1, x_2, x_7, x_9, x_{10}	0	1	1
Models for $\tau = 0.9$	QR-SSVS	MVV	Frequentist
Intercept, x_1, x_2, x_9, x_{10}	89	59	66
Intercept, x_1, x_2, x_7, x_9, x_{10}	3	9	7
Intercept, x_1, x_2, x_5, x_9, x_{10}	2	7	5
Intercept, x_1, x_2, x_3, x_6, x_9, x_{10}	1	3	1
Intercept, x_1, x_2, x_3, x_9, x_{10}	1	4	4
Intercept, x_1, x_2, x_4, x_7, x_9, x_{10}	1	4	0
Intercept, x_1, x_2, x_4, x_9, x_{10}	1	3	3
Intercept, x_1, x_2, x_6, x_9, x_{10}	1	3	4
Intercept, x_1, x_2, x_8, x_9, x_{10}	1	3	2
Intercept, x_1, x_2, x_3, x_4, x_9, x_{10}	0	1	2
Intercept, x_1, x_2, x_5, x_6, x_9, x_{10}	0	2	2
Intercept, x_1, x_2, x_3, x_7, x_9, x_{10}	0	1	1
Intercept, x_1, x_2, x_4, x_5, x_7, x_8, x_9, x_{10}	0	1	1
Intercept, x_1, x_2, x_5, x_6, x_8, x_9, x_{10}	0	0	1
Intercept, x_1, x_2, x_9	0	0	1

Table 4. Summary of simulation results from 100 replicates each having $n = 120$ observations. Type I error rates are averaged across predictors $x_3 - x_8$, while power is averaged across predictors x_1, x_2, x_9, x_{10}

	$\tau = 0.5$			
Summary	Bayesian		Frequentist	
	MIP > 0.5	MIP > 0.9	p < 0.05	p < 0.01
Type I error rate (%)	0.005	0.000	7.500	1.833
Power (%)	100.000	100.000	100.000	100.000
Coverage of 95% CI	100.000		88.400	
Width of 95% CI	0.585		0.390	
Absolute bias	0.043		0.100	
MSE	0.061		0.082	
	$\tau = 0.9$			
Summary	Bayesian		Frequentist	
	MIP > 0.5	MIP > 0.9	p < 0.05	p < 0.01
Type I error rate (%)	2.217	0.000	13.833	7.333
Power (%)	100.000	97.750	100.000	99.750
Coverage of 95% CI	100.000		89.700	
Width of 95% CI	1.040		0.564	
Absolute bias	0.054		0.131	
MSE	0.118		0.186	

$\tau = 0.9$ case, it is apparent from Table 3 that the frequentist approach selects the true model in a smaller proportion of the simulation replicates. The frequentist method has a tendency to add unnecessary predictors in a substantial proportion of the simulations, which is somewhat surprising given that we used a threshold of 0.01 on the p-values for inclusion instead of the more common 0.05. In addition, the Bayesian approach included predictors having above 0.5

MIPs, so we expected more type I errors than if we had used a more stringent threshold for inclusion. Indeed, as is summarized in Table 4, the frequentist method had a higher type I error rate regardless of whether a 0.05 or 0.01 threshold was used for significance. Both methods had similarly high power, but the Bayesian approach tended to produce wider interval estimates and higher coverage rates than the frequentist approach, which did not maintain the nominal level of 95%. It is also apparent that the Bayesian approach resulted in parameter estimates which had a lower absolute bias on average than the frequentist method. This is also backed up with the fact that the average MSEs for the Bayesian approach were lower, significantly so for τ=0.9.

For reference, we also computed the average MSE for the frequentist analysis of the full model including all predictors. This turns out to be 0.194 for $\tau = 0.5$ and 0.271 for $\tau = 0.9$. These are higher than the MSEs presented in Table 4. Since the MSE decomposes as the sum of squared bias and variance of the estimator, the higher average MSE will be due to the increased variability which arises when using additional irrelevant predictors to predict the τth quantile of y_i.

5. BOSTON HOUSING DATA

To illustrate how the QR-SSVS performs in relation to frequentist methods on real data, we consider the Boston housing data concerning housing values in suburbs of Boston which is given in Harrison and Rubinfeld (1978). The corrected data consists of $n = 506$ observations and $p = 16$ potential predictors of interest. These are the tract point latitudes/longitudes in decimal degrees (LAT/LON), the per capita crime (CRIM), the proportions of residential land zoned for lots over 25,000 square feet per town (ZN), the proportions of non-retail business acres per town (INDUS), whether or not the tract borders the Charles River (CHAS), nitric oxide concentration (parts per 10 million) per town (NOX), average number of rooms per dwelling (RM), the proportions of owner occupied units built prior to 1940 (AGE), the weighted distances to 5 Boston employment centres (DIS), the index of accessibility to radial highways per town (RAD), the full value property tax rate in 10,000s of US dollars per town (TAX), pupil to teacher ratios per town (PTRATIO), $1,000(\text{proportion of blacks} - 0.63)^2$ (B) and the percentage values of lower status population (LSTAT). The response variable is CMEDV, the corrected median values of owner occupied housing in 1,000s of US dollars.

Each predictor was standardized before analysis. To follow a conventional approach and provide a basis for comparison, a linear regression model with SSVS are fitted. Based on the main idea of small and big variances of George and McCulloch (1993), we assume that the slope parameters followed the mixture normal prior in equation (7). The choice of $(\sigma_{\beta_k}/\nu_k, c_k)$ is (0.5, 10) as the same in simulation study. The results of model selection by linear regression model with SSVS is given in Table 5. The MIP for each selected predictor is at least 0.5.

We analyzed the data with $\pi_0 \sim \text{Beta}(1,1)$. We used 50,000 iterations of QR-SSVS following a 5,000 iteration burn in. We analyzed 7 different quantiles which were $\tau = \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$. Given the results in section 4, we based inference on the model containing those predictors with MIP greater than 0.9. The frequentist model was obtained from predictors having p-value less than 0.01. The MIP for each predictor and its associated posterior and frequentist estimate are presented in Table 6. We note that the default rank method was unable to produce a sensible interval for variable CHAS at the extreme values of τ. Table 5 shows the predictors appearing in each of the selected models.

To compare the two methods on predictive performance, we followed Li et al. (2010) and used 10 fold cross validation. For the QR-SSVS method, as well as assessing the predictive performance of a single model, we also used model averaging, a major advantage of the Bayesian approach to variable selection (see e.g. Hoeting et al., 1999). To obtain model averaged estimates, we used the QR-SSVS method on the training datasets to sample potential models, and then a weighted average of the regression parameters was calculated, where the weights correspond to the posterior model probabilities. In each case, the algorithm was run for 11,000 iterations with the first 1,000 discarded, and π_0 was assigned a Beta(1, 1) distribution. Since the true conditional quantiles are unknown in this case, the measure of accuracy was based on the mean weighted absolute residuals (MWARs), defined as $\frac{1}{s_k}\sum_{i=1}^{s_k}\rho_\tau(y_i - \hat{y}_i)$, where s_k denotes the size of validation set k with $k = 1, \ldots, 10$ and $\rho_\tau(\cdot)$ is defined in (4). Table 7 presents the mean and standard deviation of the MWARs for each of the 7 quantiles.

From Table 7, it can be seen that in 5 out of 7 cases, the model selected by QR-SSVS outperformed the model selected by the frequentist approach in terms of the average MWAR, with there being a noticeable difference when $\tau = 0.95$. Interestingly, when $\tau = 0.1$, we can see from Table 5 that the model selected by QR-SSVS is nested within the model selected by the frequentist approach. Given that the MIPs for those variables not appearing in the QR-SSVS model but appearing in the frequentist models were 0.87 (NOX) and 0.89 (RAD) respectively, this suggests that the MIP threshold may have been a bit too high for that case, resulting in a type II error. In contrast, the use of Bayesian model averaging produced average MWARs that were uniformly smaller than the frequentist method across all quantiles.

It is interesting to note that four predictors (CRIM, ZN, CHAS, TAX) have been selected over extreme quantile 0.95 while (CRIM, TAX) have been selected in modelling at the extreme quantile 0.05. These four predictors have been excluded when a traditional linear regression with SSVS is implemented. Crime (CRIM) inflicts many costs on a city's

Table 5. Models selected by QR-SSVS with predictors having MIP > 0.9 and by frequentist method with asymptotic p-values less than 0.01

τ	Model Size	LON	LAT	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
Reg-SSVS																
	7							•	•		•	•		•	•	•
QR-SSVS																
0.05	8	•		•					•		•		•	•	•	•
0.10	9	•		•					•	•	•		•	•	•	•
0.25	11	•		•			•		•	•	•	•	•	•	•	•
0.50	13	•		•	•		•	•	•	•	•	•	•	•	•	•
0.75	10				•		•	•	•		•	•	•	•	•	•
0.90	10				•		•	•	•		•	•	•	•	•	•
0.95	10			•	•		•	•	•		•	•	•	•		•
τ	Model Size	LON	LAT	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
Frequentist																
0.05	6	•		•					•				•		•	•
0.10	11	•		•				•	•	•	•	•	•	•	•	•
0.25	11	•		•	•				•	•	•	•	•	•	•	•
0.50	11	•		•	•				•	•	•	•	•	•	•	•
0.75	9				•		•		•		•	•	•	•	•	•
0.90	7						•		•		•	•		•	•	•
0.95	3								•			•				•

residents, including feelings of a lack of security and safety, monetary value of property loss from criminal acts, and lost earnings due to injury or death. Hence, it is reasonable to see that the impact of crime on property values is significant at the extreme quantiles 0.05 and 0.95. Land (ZN) costs money; the larger the building lot, the more potential home buyers will have to pay to acquire their "house and land" as a package. Houses built on sizable lots (over 25,000 square feet per town) will tend to have larger interior space and more likely to be more luxurious, and hence may cost more. All of them will lead to a higher housing price than a smaller house built on a smaller lot. People prefer waterfront views (CHAS), and thus houses with views tend to cost more. People are also willing to pay more for waterfront properties. Scatterplots for these four predictors are displayed in Figure 1. There is a clear tendency for housing price (CMEDV) to fall with per capita crime (CRIM), but one can also detect several other features from the scatterplots.

6. CONCLUSION AND DISCUSSION

We have introduced the QR-SSVS procedure for quantile regression using the asymmetric Laplace distribution. Expressing this distribution as a location-scale mixture of normals allows us to make use of conditional conjugacy after introducing latent variables. Consequently, all conditional distribution necessary to implement Gibbs sampling are distributions frequently encountered in the statistical literature and have efficient algorithms to sample from them. The QR-SSVS is also fast, with 11,000 samples from the joint posterior distribution taking 4.5 seconds for the simulated data and around 17 seconds for the complete Boston housing data on a Pentium 4 dual core 3.2Ghz running Ubuntu Linux. Given the sample with the burn in discarded, we can simultaneously obtain PMPs and the corresponding estimates of the associated regression parameters. Simulations have shown that the assumption of the asymmetric Laplace likelihood performs well for many different true data generating likelihoods.

Simulations have also suggested that on average, the model selected by QR-SSVS outperforms the model selected by frequentist methods. This was more noticeable at the extreme quantiles, where we observed that the frequentist method had a higher type I error rate. The 95% credible intervals for $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ obtained using QR-SSVS contained the true values in a higher proportion of simulations than the frequentist 95% confidence intervals obtained using the default method in `R`. Finally, based on 10 fold cross-validation on the Boston housing data, the predictive performance of QR-SSVS using model averaging proved superior to the single model selected by the frequentist method at all quantiles. The greatest improvement over the frequentist method was observed at $\tau = 0.95$, backing up the recent findings of Li et al. (2010) who also noticed poor performance of the frequentist method at the extreme quantiles.

The QR-SSVS model can be extended to let τ be a discrete random variable taking values $\tau_1, \tau_2, \ldots, \tau_R$ with probabilities $p_1, p_2, \ldots, p_R$ respectively. Defining $\delta_r, r = 1, \ldots, R$ as $I(\tau = \tau_r)$, where $I(\tau = \tau_r)$ is equal to 1 if $\tau = \tau_r$, 0 otherwise and defining $\mathbf{d}$ to be a vector with rth element δ_r, our

Table 6. MIPs, posterior summary and frequentist estimates of the Boston Housing data, presented for $\tau = \{0.05, 0.5, 0.95\}$

		MIP	Posterior median	95% credible interval	Frequentist estimate	95% rank interval
	LON	0.987	−0.726	(−1.313, −0.092)	-0.676	(−1.057, −0.502)
	LAT	0.710	0.088	(−0.222, 0.651)	0.110	(−0.122, 0.431)
	CRIM	0.997	-0.901	(−2.322, −0.224)	−1.314	(−9.633, −0.394)
	ZN	0.705	0.000	(−0.731, 0.803)	0.126	(−0.712, 0.581)
	INDUS	0.734	0.000	(−0.774, 0.957)	0.530	(−0.566, 0.807)
	CHAS	0.816	0.000	(−2.190, 2.052)	−0.591	(−∞, 1.583)
	NOX	0.830	−0.382	(−1.691, 0.459)	-0.507	(−1.510, 0.560)
$\tau = 0.05$	RM	1.000	1.799	(0.901, 2.814)	1.525	(1.132, 2.644)
	AGE	0.852	−0.437	(−1.381, 0.233)	−0.413	(−1.493, 0.000)
	DIS	0.960	−1.019	(−2.219, 0.000)	−0.960	(−2.085, −0.409)
	RAD	0.824	0.370	(−0.610, 1.984)	1.009	(−0.266, 2.442)
	TAX	0.998	−2.180	(−3.650, −0.741)	−2.482	(−3.446, −1.614)
	PTRATIO	0.934	-0.622	(−1.338, 0.020)	−0.705	(−1.212, −0.180)
	B	0.963	0.571	(0.000, 1.207)	0.629	(0.397, 1.093)
	LSTAT	1.000	−2.687	(−3.888, −1.554)	-2.873	(−3.491, -1.554)
	LON	0.996	−0.563	(−0.960, -0.163)	−0.495	(−0.989, −0.287)
	LAT	0.867	0.194	(−0.057, 0.505)	0.249	(−0.007, 0.504)
	CRIM	0.998	−0.953	(−1.399, −0.271)	−1.203	(−1.282, −0.222)
	ZN	0.998	0.728	(0.224, 1.198)	0.835	(0.422, 1.150)
	INDUS	0.748	0.000	(−0.566, 0.355)	0.006	(−0.405, 0.302)
	CHAS	0.983	1.036	(−0.023, 2.259)	0.942	(0.367, 2.108)
	NOX	0.978	−0.651	(−1.309, 0.000)	−0.682	(−1.213, −0.059)
$\tau = 0.5$	RM	1.000	3.534	(2.893, 4.193)	3.516	(2.619, 4.335)
	AGE	0.987	−0.617	(−1.163, −0.013)	−0.637	(−1.052, −0.173)
	DIS	1.000	−1.784	(−2.406, −1.163)	−1.909	(−2.525, −1.379)
	RAD	1.000	1.482	(0.592, 2.346)	1.733	(0.962, 2.426)
	TAX	1.000	−1.917	(−2.721, −0.999)	-2.099	(−2.630, −1.312)
	PTRATIO	1.000	−1.428	(−1.828, −1.011)	−1.485	(−1.742, −1.051)
	B	1.000	1.096	(0.746, 1.445)	1.085	(0.848, 1.496)
	LSTAT	1.000	−2.281	(−2.961, −1.607)	−2.254	(−2.904, −1.620)
	LON	0.743	0.000	(−0.773, 0.643)	0.208	(−0.383, 0.969)
	LAT	0.819	0.231	(−0.481, 1.319)	0.864	(−0.841, 1.419)
	CRIM	0.905	−0.795	(−2.745, 0.941)	−1.484	(−3.215, 27.680)
	ZN	0.927	0.782	(−0.129, 1.899)	1.120	(−0.274, 1.700)
	INDUS	0.824	0.000	(−1.349, 1.622)	1.023	(−1.991, 3.836)
	CHAS	0.988	3.143	(0.000, 9.029)	3.026	(0.387, ∞)
	NOX	0.978	−1.778	(−3.938, 0.031)	−1.739	(−6.260, 0.027)
$\tau = 0.95$	RM	1.000	3.675	(2.666, 4.775)	3.647	(2.549, 5.003)
	AGE	0.815	0.000	(−1.353, 1.234)	−0.900	(−2.432, 2.194)
	DIS	1.000	−2.933	(−4.492, −1.516)	−3.047	(−5.083, −1.666)
	RAD	1.000	5.818	(3.162, 8.239)	7.968	(2.767, 9.909)
	TAX	0.920	−1.041	(−3.481, 0.716)	−2.943	(−5.380, 1.858)
	PTRATIO	1.000	−2.546	(−4.115, −1.023)	−2.190	(−5.067, −0.852)
	B	0.891	0.615	(−0.628, 2.320)	0.934	(−2.763, 3.524)
	LSTAT	1.000	−3.739	(−4.896, −2.556)	−3.704	(−4.147, −2.936)

suggestion would be to consider the likelihood $l(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{d})$ as (8)

$$\prod_{r=1}^{R} \left\{ \tau_r^n (1-\tau_r)^n \sigma^{-n} \exp\left\{ -\sigma^{-1} \sum_{i=1}^{n} \rho_{\tau_r}(y_i - \mathbf{x_i}^T \boldsymbol{\beta}) \right\} \right\}^{\delta_r}.$$

Thus, conditional on $\delta_r = 1$, we obtain the AL likelihood with skewness parameter τ_r. To complete this model specification, we assume that jointly, $\mathbf{d}$ comes from a multinomial distribution with parameters $(1, \mathbf{p})$, where $\mathbf{p}$ is a vector with rth element p_r. The probabilities $\mathbf{p}$ could all be set equal to $\frac{1}{R}$. Alternatively, to allow more flexibility, we can use the fact that the Dirichlet distribution is conjugate to the multinomial distribution and let $\mathbf{p} \sim \text{Dirichlet}(1, 1, \ldots, 1)$. This is equivalent to a multivariate version of the uniform distribution on $\mathbf{p}$.

Table 7. Mean weighted absolute deviations assessing predictive accuracy using 10 fold cross validation. M.A. denotes model averaged

	QR-SSVS		QR-SSVS M.A.		Frequentist	
τ	Mean	SD	Mean	SD	Mean	SD
0.05	0.318	0.032	0.306	0.023	0.321	0.036
0.1	0.565	0.060	0.550	0.051	0.554	0.061
0.25	1.101	0.141	1.087	0.116	1.097	0.128
0.5	1.580	0.194	1.576	0.194	1.600	0.189
0.75	1.524	0.310	1.521	0.310	1.552	0.311
0.9	1.030	0.291	1.026	0.295	1.062	0.282
0.95	0.681	0.227	0.670	0.222	0.823	0.266

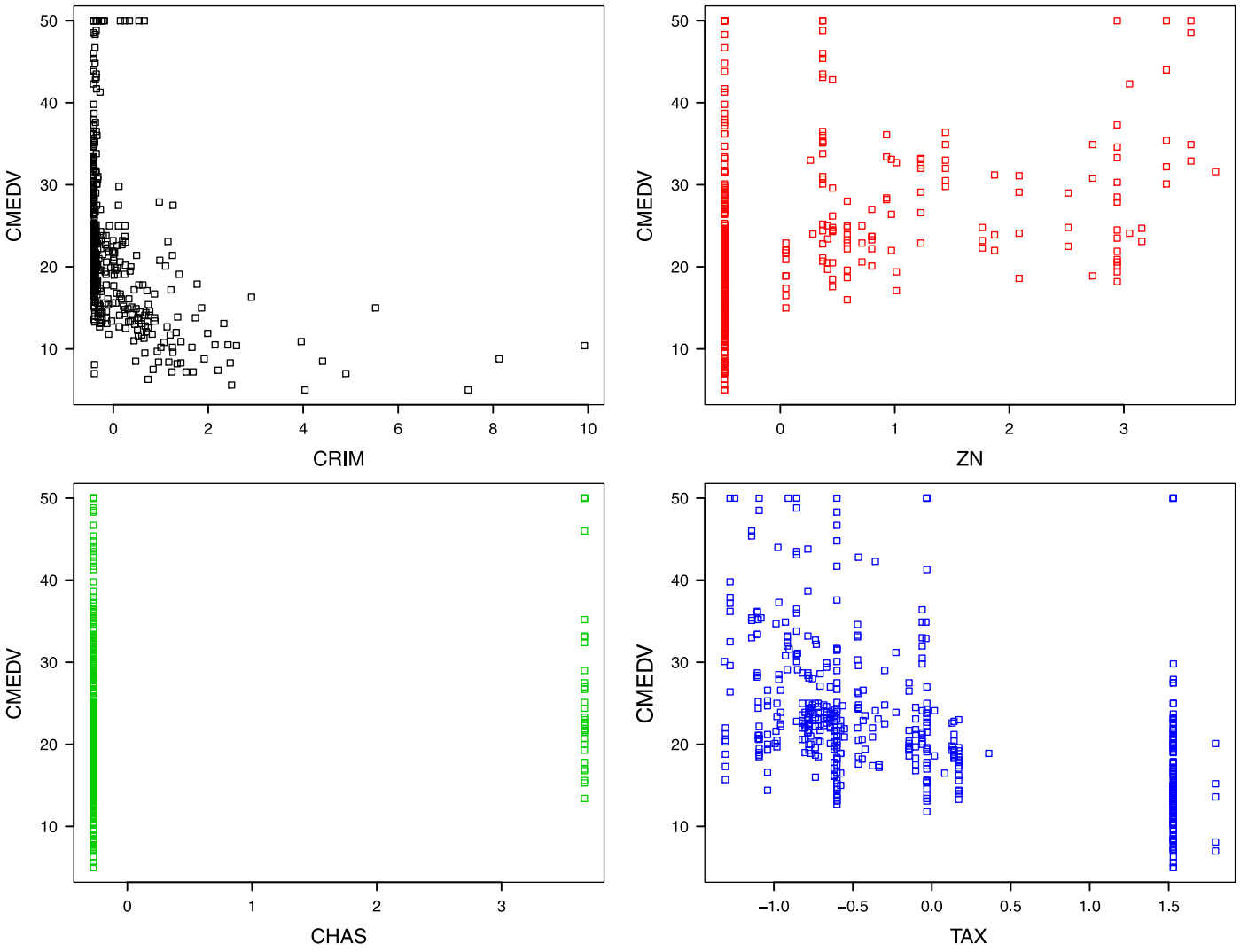

Figure 1. Scatter plots for CMEDV vs (CRIM, ZN, CHAS, TAX) where each predictor has been standardized.

A final point to note is that in applications of quantile regression, censoring is often encountered. This should not cause any real problems in general as it can be handled as missing data. This missing data could then in theory be predicted by averaging over all the models visited at all quantiles using the extensions to QR-SSVS described above.

QR-SSVS was implemented using the `R` function `SSVSquantreg`, which is to appear in the next version of the package MCMCpack (Martin et al., 2009).

ACKNOWLEDGEMENT

We thank the editor, associate editor, and two anonymous referees for their insightful and helpful comments, which improved this paper. This work was funded by EPSRC doctoral training account. Cathy Chen is supported by the grants: NSC 99-2118-M-035 -001 -MY2 and NSC 101-2118-M-035 -006 -MY2 from the National Science Council (NSC) of Taiwan.

APPENDIX

The greatest amount of computation involved in QR-SSVS is evaluating $g(\boldsymbol{\gamma})$ necessary to sample from $\gamma_j|\boldsymbol{\gamma}_{-j}, \mathbf{w}, \boldsymbol{\lambda}, \pi_0$. We can use ideas from Dongarra et al. (1979) to re-evaluate $g(\boldsymbol{\gamma})$ efficiently when 1 component of $\boldsymbol{\gamma}$ is altered. The fact that Cholesky shuffling methods of Dongarra et al. (1979) can speed up the computation of the likelihood component of the conditional posterior of each indicator variable was first suggested by Smith and Kohn (1996)to both augment a Gaussian prior with a point mass, and then marginalize out the regression coefficients analytically to give an efficient Gibbs sampler.

The idea is to find the Cholesky decomposition of the matrix

$$[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]'[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}] = \begin{bmatrix} \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}} & \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\tilde{\mathbf{u}} \\ \tilde{\mathbf{u}}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}} & \tilde{\mathbf{u}}'\tilde{\mathbf{u}} \end{bmatrix}.$$

The Cholesky matrix, $\tilde{\mathbf{S}}$, can be expressed in the following form

$$\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S} & \boldsymbol{\xi} \\ \mathbf{0}' & \psi \end{bmatrix}$$

from which it is evident that

$$\mathbf{S}'\mathbf{S} = \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}, \tag{9}$$

$$\mathbf{S}'\boldsymbol{\xi} = \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\tilde{\mathbf{u}}, \tag{10}$$

$$\boldsymbol{\xi}'\boldsymbol{\xi} + \psi^2 = \tilde{\mathbf{u}}'\tilde{\mathbf{u}}. \tag{11}$$

Equation (10) implies that $\mathbf{S}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}} = \boldsymbol{\xi}$. Substituting this into equation (11), we find that ψ^2 is equal to $\|\tilde{\mathbf{u}} - \widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}\|^2$. Also, since $\mathbf{S}$ is upper triangular, we have

$$|\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}{}'\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}| = \prod_{j=1}^{p_{\boldsymbol{\gamma}}} \mathbf{S}_{jj}^2.$$

Hence, $\tilde{\mathbf{S}}$ holds all the necessary information to calculate $g(\boldsymbol{\gamma})$. Altering 1 value of $\boldsymbol{\gamma}$ requires $\tilde{\mathbf{S}}$ to be updated when a column is added or removed from $\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}$. Deleting a column involves obtaining the Cholesky decomposition of the matrix $\mathbf{E}'[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]'[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]\mathbf{E}$, where $\mathbf{E}$ denotes a permutation matrix such that post-multiplying $[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]$ by $\mathbf{E}$ moves the column of interest to the final position. This can be achieved by first post-multiplying $\tilde{\mathbf{S}}$ by $\mathbf{E}$ and then pre-multiplying with a sequence of orthogonal transformations mainly involving Givens rotations until the final matrix satisfies the properties for it to be a Cholesky matrix. The desired Cholesky matrix is the sub matrix with the final row and column deleted. To add a column $\tilde{\mathbf{x}}_{\text{new}}$, it is necessary to solve the upper triangular system $\mathbf{S}'\boldsymbol{\xi} = [\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]'\tilde{\mathbf{x}}_{\text{new}}$, and then calculate ψ from $\boldsymbol{\xi}'\boldsymbol{\xi} + \psi^2 = \tilde{\mathbf{x}}_{\text{new}}'\tilde{\mathbf{x}}_{\text{new}}$. The vector $\boldsymbol{\xi}$ forms the final column of the new Cholesky matrix and a new row of zeros is added. The element in the bottom right position is set equal to ψ. Then, we can proceed in the same way as if we were deleting a column, except that $\mathbf{E}$ is such that post-multiplying $[\widetilde{\boldsymbol{X}}_{\boldsymbol{\gamma}}\ \tilde{\mathbf{u}}]$ by $\mathbf{E}$ moves the final column to the position that it would have appeared if it had been present.

Received 1 February 2012

REFERENCES

Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics* **32** 870–897. MR2065192

Chen, C. W. S. (1999). Subset selection of autoregressive time series models. *Journal of Forecasting* **18** 505–516.

Chen, C. W. S., Liu, F. C., and Gerlach, R. (2011). Bayesian subset selection for threshold autoregressive moving-average models. *Computational Statistics* **26** 1–30. MR2773798

Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical Science* **19** 81–94. MR2082148

Dongarra, J. J., Moler, C. B., Bunch, J. R. and Stewart, G. W. (1979). Linpack Users' Guide. *Philadelphia: Siam* MR0566367

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7** 339–373.

Harrison D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 81–102.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science* **14** 382–401. MR1765176

Koenker, R. (2009). `quantreg`: Quantile regression. R package version 4.43. http://cran.R-project.org/package=quantreg

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* **128** 137–164. MR2022929

Li, Q., Xi, R. and Lin, N. (2010). Bayesian regularized quantile regression. Technical report. MR2719666

Martin, A. D., Quinn, K. M. and Park, J. H. (2009) `MCMCpack`: Markov chain Monte Carlo package. R package version 1.0-4. http://cran.R-project.org/package=MCMCpack

Meligkotsidou, L., Vrontos, I. D. and Vrontos, S. D. (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance* **16** 264–279.

R development core team (2009). R: A language and environment for statistical computing. Vienna, Austria. http://www.R-project.org

Reed, C. and Yu, K. (2011). Efficient Gibbs sampling for Bayesian quantile regression. Technique report.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of Monographs on Statistics and Applied Probability. Chapman & Hall, London. MR2130347

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical Bayes multiplicity adjustment in the variable selection problem. *The Annals of Statistics* **38** 2587–2619. MR2722450

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75** 317–343.

So, M. K. P. and Chen, C. W. S. (2003). Subset threshold autoregression. *Journal of Forecasting* **22** 49–66.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81** 82–86. MR0830567

Tsionas, E. G. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation* **73** 659–674. MR2001612

Yu, K. M. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* **54** 437–447. MR1861390

Keming Yu
Shihezi University
China
Brunel University
UK
E-mail address: Keming.Yu@brunel.ac.uk

Cathy W.S. Chen
Department of Statistics
Feng Chia University
Taiwan
Fax: +886 4 2451 7092
E-mail address: chenws@mail.fcu.edu.tw

Craig Reed
University of Edinburgh
UK
E-mail address: craig.reed@ed.ac.uk

David B. Dunson
Duke University
USA
E-mail address: dunson@stat.duke.edu