A Bayesian approach to envelope quantile regression

Minji Lee, Saptarshi Chakraborty and Zhihua Su

Department of Biostatistics, Columbia University

Department of Biostatistics, State University of New York at Buffalo

Department of Statistics, University of Florida

Abstract: The enveloping approach employs sufficient dimension reduction techniques to gain estimation efficiency, and has been used in several multivariate analysis contexts. However, its Bayesian development has been sparse and the only Bayesian envelope construction is in linear regression. In this paper we propose a Bayesian envelope approach to quantile regression, using a general framework that may potentially aid enveloping in other contexts as well. The proposed approach is also extended to accommodate censored data. Data augmentation Markov chain Monte Carlo algorithms are derived for approximate sampling from the posterior distributions. Simulations and data examples are included for illustration.

Key words and phrases: — Envelope model; Metropolis-within-Gibbs sampling; Quantile regression; Sufficient dimension reduction; Tobit quantile.

1. Introduction

The envelope methodology (Cook, 2018) induces a class of models that uses dimension reduction to increase estimation efficiency in multivariate analysis, sometimes equivalent to taking many additional observations. First proposed for multivariate linear regression (Cook et al., 2010), it has since been extended to many other contexts, including partial least squares (Cook et al., 2013; Zhu and Su, 2020), generalized linear models (Cook and Zhang, 2015), elliptical multivariate linear regression (Forzani and Su, 2021), variable selection (Su et al., 2016), matrix or tensor variate regression (Ding and Cook, 2018; Li and Zhang, 2017), spatial regression (Rekabdarkolaee et al., 2020) and quantile regression (Ding et al., 2021). These advances have primarily been made from a frequentist's perspective. In practice, there is often strong motivation for adopting a Bayesian approach. First, a Bayesian approach can incorporate existing knowledge into the model through prior specification. Second, it coherently quantifies all model uncertainties through the posterior distribution without requiring any asymptotic assumption. Consequently, using computational methods, exact inference can be made for any given sample size. However, Bayesian approach to envelope models has thus far been sparse, primarily due to the key parameter being defined on a manifold in envelope models. This makes probabilistic

modeling of the parameter both theoretically and computationally challenging. The only existing Bayesian method (Khare et al., 2017) considers a matrix-Bingham prior. However, this approach depends on the form of a multi-response linear regression model, making extensions to other contexts difficult. See the second paragraph in Supplement S2.2 for more details.

In this article we propose a Bayesian approach to envelope quantile regression. Quantile regression was introduced in the seminal work of Koenker and Bassett (1978) and has since been an active area of research (see, e.g., Koenker et al., 2005, for an overview). Unlike the standard (mean) regression, the quantile regression focuses on conditional quantiles instead of the conditional mean of the response variable given the predictors. Consequently, it can provide a full picture of the response-predictor relationship and can be robust to outliers. Furthermore, it can incorporate heteroskedasticity, thus allowing handling of a possibly richer set of data. In a frequentist setting, quantile regression is implemented by optimizing a distribution-free quantile loss function. The frequentist envelope quantile regression follows a similar path (Ding et al., 2021), which is markedly different from other likelihood-based (frequentist) envelope models.

Several Bayesian frameworks for the quantile regression have been proposed (Remark 1). The most commonly used framework utilizes the asym-

metric Laplace distribution (ALD) associated with the quantile loss as a data-generating model (Yu and Moyeed, 2001; Kozumi and Kobayashi, 2011). We use this framework here as a working model after a reparameterization that frees the envelope model from any manifold structure. The resulting Bayesian envelope quantile regression (BEQR) model features distributions that are straightforward to interpret and compute.

The major contribution of this paper lies in the formulation of a novel enveloping strategy for Bayesian quantile regression (BQR). Owing to the connection between predictor envelope models (Cook et al., 2013) and partial least squares (Wold, 1966, PLS), the proposed approach can also be treated as a rigorous Bayesian development for PLS quantile regression. Our approach embeds an envelope structure in a BQR framework; however the key enveloping strategy is foundational and we believe that it can be used in other Bayesian modeling contexts. Implementation of our model is based on a simple and computationally efficient Markov chain Monte Carlo (MCMC) sampler that is shown to be *Harris ergodic* which ensures strong theoretical guarantees for the MCMC draws. An extension of the proposed model for censored observations is also provided. Results with both simulated and real data demonstrate the efficiency gains of the proposed BEQR approach compared to the standard BQR approach.

2. A review of envelope quantile regression model

We introduce the envelope model in the context of quantile regression. A standard quantile regression model is formulated as

$$Q_\tau(Y|\boldsymbol{X}) = \mu_\tau + \boldsymbol{\beta}_\tau^T(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}}), \tag{2.1}$$

where Y is the response variable, $\boldsymbol{X} \in \mathbb{R}^p$ is the predictor vector having mean $\boldsymbol{\mu}_{\boldsymbol{X}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, and $Q_\tau(Y|\boldsymbol{X})$ denotes the τ-th conditional quantile of Y given $\boldsymbol{X}$ ($0 \leq \tau \leq 1$; $\tau = 0.5$ produces the median regression). The unknown intercept and slope are denoted by μ_τ and $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$. An envelope model seeks for a sufficient dimension reduction of $\boldsymbol{X}$ that loses no information about $Q_\tau(Y|\boldsymbol{X})$. Formally, let $(\boldsymbol{G}_{1\tau}, \boldsymbol{G}_{2\tau}) \in \mathbb{R}^{p\times p}$ be an orthogonal matrix, where $\boldsymbol{G}_{1\tau}$ has dimension $p \times d_\tau$ ($0 \leq d_\tau \leq p$). The envelope model imposes the following two conditions on $\boldsymbol{G}_{1\tau}^T\boldsymbol{X}$ and $\boldsymbol{G}_{2\tau}^T\boldsymbol{X}$: (a) $Q_\tau(Y|\boldsymbol{X}) = Q_\tau(Y|\boldsymbol{G}_{1\tau}^T\boldsymbol{X})$ and (b) $\text{cov}(\boldsymbol{G}_{1\tau}^T\boldsymbol{X}, \boldsymbol{G}_{2\tau}^T\boldsymbol{X}) = 0$. These conditions suggest that $\boldsymbol{G}_{2\tau}^T\boldsymbol{X}$ carries no information on $Q_\tau(Y|\boldsymbol{X})$ directly or indirectly through the association with $\boldsymbol{G}_{1\tau}^T\boldsymbol{X}$. Let $\text{span}(\boldsymbol{M})$ denote the column space of a matrix $\boldsymbol{M}$. Ding et al. (2021) showed that (a) and (b) are equivalent to (i) $\text{span}(\boldsymbol{\beta}) \subseteq \text{span}(\boldsymbol{G}_{1\tau})$ and (ii) $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{G}_{1\tau}}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{G}_{1\tau}} + \boldsymbol{Q}_{\boldsymbol{G}_{1\tau}}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{Q}_{\boldsymbol{G}_{1\tau}}$, where $\boldsymbol{P}$ denotes the projec-

tion matrix and $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{P}$. The intersection of all subspaces that satisfy (i) and (ii) is called the $\boldsymbol{\Sigma_X}$-envelope of $\boldsymbol{\beta}_\tau$, denoted $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. Thus the envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)$ is the smallest subspace that satisfies (i) and (ii), or equivalently (a) and (b). Let u_τ $(0 \leq u_\tau \leq p)$ denote the dimension of $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)$, and $\boldsymbol{\Gamma}_{1\tau} \in \mathbb{R}^{p\times u_\tau}$ an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. Then $\boldsymbol{\Gamma}_{1\tau}^T\boldsymbol{X}$ contains all information about $Q_\tau(Y|\boldsymbol{X})$, and is called the material part. The envelope quantile regression model (Ding et al., 2021) is formulated as

$$Q_\tau(Y|\boldsymbol{X}) = \mu_\tau + \boldsymbol{\eta}_\tau^T\boldsymbol{\Gamma}_{1\tau}^T(\boldsymbol{X} - \boldsymbol{\mu_X}), \quad \boldsymbol{\Sigma_X} = \boldsymbol{\Gamma}_{1\tau}\boldsymbol{\Omega}_{1\tau}\boldsymbol{\Gamma}_{1\tau}^T + \boldsymbol{\Gamma}_{2\tau}\boldsymbol{\Omega}_{2\tau}\boldsymbol{\Gamma}_{2\tau}^T, \quad (2.2)$$

where $\boldsymbol{\beta}_\tau = \boldsymbol{\Gamma}_{1\tau}\boldsymbol{\eta}_\tau$, and $\boldsymbol{\eta}_\tau \in \mathbb{R}^{u_\tau}$ carries the coordinates of $\boldsymbol{\beta}_\tau$ with respect to $\boldsymbol{\Gamma}_{1\tau}$. The matrix $\boldsymbol{\Gamma}_{2\tau} \in \mathbb{R}^{p\times(p-u_\tau)}$ is an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)^\perp$, the orthogonal complement of $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. The matrices $\boldsymbol{\Omega}_{1\tau} \in \mathbb{R}^{u_\tau\times u_\tau}$ and $\boldsymbol{\Omega}_{2\tau} \in \mathbb{R}^{(p-u_\tau)\times(p-u_\tau)}$ are both positive definite, and carry the coordinates of $\boldsymbol{\Sigma_X}$ with respect to $\boldsymbol{\Gamma}_{1\tau}$ and $\boldsymbol{\Gamma}_{2\tau}$. When $u_\tau = p$, the envelope quantile regression (2.2) reduces to the standard quantile regression (2.1).

In a standard linear regression model $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}^T(\boldsymbol{X} - \boldsymbol{\mu_X}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y} \in \mathbb{R}^r$ is univariate $(r = 1)$ or multivariate $(r > 1)$ and $\boldsymbol{\varepsilon}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma_{Y|X}}$, imposing the above envelope structure (2.2) to $\boldsymbol{\beta}$

and $\boldsymbol{\Sigma_X}$ produces a so-called predictor envelope model (Cook et al., 2013), a framework closely connected with that of PLS (Wold, 1966). PLS is a popular alternative to the ordinary least squares regression due to its potential of improving prediction performance, particularly in high-dimensional problems. PLS uses a sequential moment-based algorithm, such as SIMPLS (De Jong, 1993) or NIPALS (Wold, 1975), to estimate a dimension reduction subspace of $\boldsymbol{X}$. Cook et al. (2013) showed that the target subspace that PLS pursues is essentially the predictor envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma_X}}(\boldsymbol{\beta})$. This implies that PLS can be cast into, and hence investigated through, the framework of predictor envelope models. Thus, envelope quantile regression provides a model-based formulation of PLS in quantile regression (Dodge and Whittaker, 2009, partial quantile regression). By implication, BEQR aids a Bayesian formulation for the partial quantile regression model.

3. Bayesian envelope quantile regression (BEQR)

3.1 Formulation

We begin with the ALD working model for BQR (Koenker and Machado, 1999; Yu and Moyeed, 2001; Khare and Hobert, 2012): $Y = \mu_{\tau,Y} + \boldsymbol{\beta}_\tau^T(\boldsymbol{X} - \boldsymbol{\mu_X}) + \sigma\epsilon$, where $\mu_{\tau,Y}$ is an intercept, σ is the scale parameter, and ϵ follows ALD(τ) with density $g_{ALD}(\epsilon;\tau) = \tau(1-\tau)\left[e^{(1-\tau)\epsilon}I(\epsilon<0) + e^{-\tau\epsilon}I(\epsilon>0)\right]$.

Koenker and Machado (1999) showed that the frequentist quantile regression estimator is the same as the maximum likelihood estimator under the above working model. This motivates the construction of the BEQR model:

$$\begin{aligned} Y &= \mu_{\tau,Y} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}^T (\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}}) + \sigma\epsilon, \quad \epsilon \sim \mathrm{ALD}(\tau) \\ \boldsymbol{X} &\sim N_p \left(\boldsymbol{\mu}_{\boldsymbol{X}},\ \boldsymbol{\Gamma}_{1\tau}\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_{1\tau}^T + \boldsymbol{\Gamma}_{2\tau}\boldsymbol{\Omega}_2\boldsymbol{\Gamma}_{2\tau}^T \right). \end{aligned} \tag{3.3}$$

As in (2.2), we include $\boldsymbol{X}$ in the model because it aids to identify the material part $\boldsymbol{\Gamma}_{1\tau}^T \boldsymbol{X}$.

We now consider a reparameterization that *identifies* both $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)$ and $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)^\perp$ with an unconstrained matrix (Ma and Zhu, 2013; Cook et al., 2016) described as follows. For an arbitrary basis $\boldsymbol{\Gamma}_{1\tau}$, assume that the first u_τ rows form a nonsingular matrix $\boldsymbol{G}_1$. If not, permute the rows of $\boldsymbol{\Gamma}_{1\tau}$ (equivalent to reordering elements in $\boldsymbol{X}$) to achieve that. Call the matrix formed by the remaining rows $\boldsymbol{G}_2$. Then

$$\boldsymbol{\Gamma}_{1\tau} = \begin{pmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{u_\tau} \\ \boldsymbol{G}_2\boldsymbol{G}_1^{-1} \end{pmatrix} \boldsymbol{G}_1 \equiv \begin{pmatrix} \boldsymbol{I}_{u_\tau} \\ \boldsymbol{A} \end{pmatrix} \boldsymbol{G}_1 \equiv \boldsymbol{C}_{\boldsymbol{A}} \boldsymbol{G}_1. \tag{3.4}$$

Thus $\boldsymbol{C}_{\boldsymbol{A}}$ is also a basis of $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)$. This procedure shows that $\boldsymbol{A}$ and $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)$ has a one-to-one correspondence: if a different basis of $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)$ is used, by following the procedure in (3.4), we obtain the same $\boldsymbol{A}$ ma-

trix. We can obtain a unique orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)$ from $\boldsymbol{A}$ as: $\boldsymbol{\Gamma}_{1\tau} = \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A}) = \boldsymbol{C}_{\boldsymbol{A}}(\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2}$, and of $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{\beta}_\tau)^\perp$ as $\boldsymbol{\Gamma}_{2\tau} = \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A}) = \boldsymbol{D}_{\boldsymbol{A}}(\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2}$, where $\boldsymbol{D}_{\boldsymbol{A}}^T = (-\boldsymbol{A}, \boldsymbol{I}_{r-u_\tau})$ (Chen et al., 2020). Consequently, BEQR (3.3) can be written as

$$\begin{aligned} &Y = \mu_{\tau,Y} + \boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}}\right) + \sigma\epsilon, \quad \epsilon \sim \text{ALD}(\tau), \\ &\boldsymbol{X} \sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{X}},\ \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T + \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})\boldsymbol{\Omega}_2\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})^T\right). \end{aligned} \tag{3.5}$$

Remark 1. Alternative approaches to BQR using Dirichlet process priors (Kottas and Gelfand, 2001), substitution likelihood (Dunson et al., 2003) and empirical likelihood (Lancaster and Jun, 2010; Yang and He, 2012) have been proposed to circumvent the i.i.d. (and thus homoskedastic) error assumptions of the ALD model. These approaches could also potentially be used to derive a BEQR; however, we use the ALD approach as it is the simplest to implement, and provides some robustness to likelihood misspecification as evidenced in empirical (Yu and Moyeed, 2001) as well as theoretical (Sriram et al., 2013) analyses. Furthermore, the results in Sriram (2015) and Yang et al. (2016) suggest that under certain regularity conditions, posterior samples from the ALD-based BQR model can help construct valid estimates of the asymptotic covariance matrix of the frequentist quantile regression estimators, even when the likelihood is misspecified. This il-

luminates potential usefulness of the ALD-based BQR model in assessing estimation variability even in the frequentist estimation.

3.2 Prior specification and posterior distributions

We consider a joint prior density for the model parameters $\mu_{\tau,Y}, \boldsymbol{\mu_X}, \boldsymbol{\eta}$, $\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{A}$ and σ in the BEQR model (3.5) as $\pi(\mu_{\tau,Y}, \boldsymbol{\mu_X}, \boldsymbol{\eta}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{A}, \sigma) = \pi(\mu_{\tau,Y})\pi(\boldsymbol{\mu_X})\pi(\sigma)\pi(\boldsymbol{\eta} \mid \sigma, \boldsymbol{A})\pi(\boldsymbol{A})\pi(\boldsymbol{\Omega}_1)\pi(\boldsymbol{\Omega}_2)$, where

1. $\pi(\mu_{\tau,Y})$ and $\pi(\boldsymbol{\mu_X})$ are improper flat densities for $\mu_{\tau,Y}$ and $\boldsymbol{\mu_X}$ respectively (i.e., $\pi(\mu_{\tau,Y}) \propto 1$ and $\pi(\boldsymbol{\mu_X}) \propto 1$).

2. $\pi(\sigma)$ is the density of IG $(a,\ b)$, where IG denotes the inverse gamma distribution, $a > 0$ and $b > 0$.

3. $\pi(\boldsymbol{A})$ is the density of $\mathrm{MN}_{p-u_\tau,u_\tau}(\boldsymbol{A}_0,\ \boldsymbol{K},\ \boldsymbol{L})$, where MN denotes a matrix normal distribution, $\boldsymbol{K} \in \mathbb{R}^{(p-u_\tau)\times(p-u_\tau)}$ and $\boldsymbol{L} \in \mathbb{R}^{u_\tau \times u_\tau}$ are positive definite, and $\boldsymbol{A}_0 \in \mathbb{R}^{(p-u_\tau)\times u_\tau}$.

4. Conditional on σ and $\boldsymbol{A}$, $\pi(\boldsymbol{\eta} \mid \sigma, \boldsymbol{A})$ is the density of the normal distribution $\mathrm{N}_{u_\tau}\left(\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\boldsymbol{e},\ \sigma\gamma^2\boldsymbol{M}^{-1}\right)$, where $\gamma^2 = 2/[\tau(1-\tau)]$, $\boldsymbol{e} \in \mathbb{R}^{u_\tau}$, and $\boldsymbol{M} \in \mathbb{R}^{u_\tau \times u_\tau}$ is positive definite.

5. $\pi(\boldsymbol{\Omega}_1)$ and $\pi(\boldsymbol{\Omega}_2)$ are inverse Wishart densities of $\mathrm{IW}_{u_\tau}(\boldsymbol{\Psi}_1,\ \nu_1)$ and $\mathrm{IW}_{p-u_\tau}(\boldsymbol{\Psi}_2,\ \nu_2)$ respectively, with $\nu_1 > 0$ and $\nu_2 > 0$, and $\boldsymbol{\Psi}_1 \in$

$\mathbb{R}^{u_\tau \times u_\tau}$ and $\boldsymbol{\Psi}_2 \in \mathbb{R}^{(p-u_\tau)\times(p-u_\tau)}$ positive definite.

Definitions of distributions including the matrix normal, inverse Wishart and generalized inverse Gaussian distributions are given in Supplement S1.1. Comments on the choice of the prior distributions are provided in Supplement S2.1. The resulting posterior density is intractable for direct computation and i.i.d. sampling. Instead, we derive an MCMC sampler utilizing the data augmentation scheme in Kozumi and Kobayashi (2011): Let $(\boldsymbol{X}_i, Y_i)$ be independent observations from $(\boldsymbol{X}, Y)$, and $\theta = \frac{1-2\tau}{\tau(1-\tau)}$. The augmented data $Z_1, \cdots, Z_n$ are i.i.d. Exponential(σ) random variables such that (Y_i, Z_i) are independent for $i = 1, \ldots, n$, with $Y_i \mid (\boldsymbol{X}_i, Z_i) \sim N(\mu_{\tau,Y} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}}) + \theta Z_i,\ Z_i\sigma\gamma^2)$. Straightforward derivations show that $Y_i \mid \boldsymbol{X}_i$ is $\mu_{\tau,Y} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T (\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}}) + \sigma\epsilon$, where ϵ follows ALD, as desired. To simplify the notations, we define $W_i = Y_i - \theta Z_i$, then $W_i|(\boldsymbol{X}_i, Z_i) \sim N(\mu_{\tau,Y} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}}),\ Z_i\sigma\gamma^2)$. Let $\mathbb{Y}^T = (Y_1, \cdots, Y_n) \in \mathbb{R}^n$, $\mathbb{X}^T = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n) \in \mathbb{R}^{n\times p}$, $\mathbb{Z}^T = (Z_1, \cdots, Z_n) \in \mathbb{R}^n$, $\mathbb{W}^T = (W_1, \cdots, W_n) \in \mathbb{R}^n$, and $\mathbf{1}_n = (1, \cdots, 1)^T \in \mathbb{R}^n$. The data-

augmented log-likelihood is:

$$
\begin{aligned}
l\left(\mathbb{W},\mathbb{X}|\mu_{\tau,Y},\boldsymbol{\mu_X},\boldsymbol{\eta},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\boldsymbol{A},\sigma,\mathbb{Z}\right) &= -\frac{n}{2}\log\sigma - \frac{n}{2}\log|\boldsymbol{\Omega}_1| - \frac{n}{2}\log|\boldsymbol{\Omega}_2| \\
&- \frac{1}{2\sigma\gamma^2}\left\{(\mathbb{W}-\mu_{\tau,Y}\mathbf{1}_n-(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{1\tau}\boldsymbol{\eta})^T\boldsymbol{D}^{-1}(\mathbb{W}-\mu_{\tau,Y}\mathbf{1}_n-(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{1\tau}\boldsymbol{\eta})\right\} \\
&- \frac{1}{2}\,\mathrm{trace}\left\{\left(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)\left(\boldsymbol{\Gamma}_{1\tau}\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Gamma}_{1\tau}^T+\boldsymbol{\Gamma}_{2\tau}\boldsymbol{\Omega}_2^{-1}\boldsymbol{\Gamma}_{2\tau}^T\right)\left(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T\right\},
\end{aligned}
$$

where $\boldsymbol{D}=\mathrm{diag}\,(Z_1,\cdots,Z_n)$. Expressions for the resulting data-augmented unnormalized log posterior density is provided in (S3.1) of the Supplement.

Note that, the real benefit of introducing the augmented data $Z_1,\ldots,Z_n$ lies in the simplification to conditional posterior densities of the model parameters. This facilitates construction of a computationally efficient MCMC sampler, as we will see in Section 3.3. Theorem 1 shows the propriety of the target posterior density. Proofs of all theoretical results are provided in Supplement S3.

Theorem 1. *The posterior density of* $(\mu_{\tau,Y},\boldsymbol{\mu_X},\boldsymbol{\eta},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\boldsymbol{A},\sigma,\mathbb{Z})$ *as provided in* (S3.1) *of the Supplement is proper.*

3.3 Data augmentation Markov chain Monte Carlo sampler

This section proposes a data augmentation algorithm for MCMC sampling from the posterior density Algorithm 1 displays one iteration of the pro-

posed sampler. Derivations are provided in Supplement S4.

Algorithm 1. One iteration of the data augmentation Metropolis-within-Gibbs sampler for the BEQR parameters

Step 1 Generate independent $Z_1, \ldots, Z_n$ from

$$Z_i \sim \text{GIG}\left(\frac{\left\{Y_i - \mu_{\tau,Y} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}}\right)\right\}^2}{\sigma\gamma^2}, \frac{\theta^2 + 2\gamma^2}{\sigma\gamma^2}, \frac{1}{2}\right),$$

where GIG denotes a generalized inverse Gaussian distribution. Then update $W_i = Y_i - \theta Z_i$ for $i = 1, \ldots, n$.

Step 2 Generate $\mu_{\tau,Y} \sim \text{N}\left(\overline{W}_Z + \boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}_Z\right),\ \frac{1}{\sum_{i=1}^n \frac{1}{Z_i}}\sigma\gamma^2\right)$, where

$$\overline{W}_Z = \frac{1}{\sum_{i=1}^n \frac{1}{Z_i}}\sum_{i=1}^n \frac{W_i}{Z_i}, \quad \overline{\boldsymbol{X}}_Z = \frac{1}{\sum_{i=1}^n \frac{1}{Z_i}}\sum_{i=1}^n \frac{1}{Z_i}\boldsymbol{X}_i.$$

Step 3 Generate $\boldsymbol{\mu}_{\boldsymbol{X}}$ from $\text{N}\left(\Delta_{\boldsymbol{\mu}_{\boldsymbol{X}}}^{-1}\Xi_{\boldsymbol{\mu}_{\boldsymbol{X}}},\ \Delta_{\boldsymbol{\mu}_{\boldsymbol{X}}}^{-1}\right)$, where $\overline{\boldsymbol{X}} = \mathbf{1}_n^T\mathbb{X}/n$,

$$\begin{aligned}
\Xi_{\boldsymbol{\mu}_{\boldsymbol{X}}} &= \frac{1}{\sigma\gamma^2}\left(\sum_{i=1}^n \frac{1}{Z_i}\right)\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\eta}\left(\boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\overline{\boldsymbol{X}}_z + \mu_{\tau,Y} - \overline{W}_Z\right) \\
&\qquad + n\left(\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T + \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})\boldsymbol{\Omega}_2^{-1}\boldsymbol{\Gamma}_{2\tau}^T(\boldsymbol{A})\right)\overline{\boldsymbol{X}}, \\
\Delta_{\boldsymbol{\mu}_{\boldsymbol{X}}} &= \frac{1}{\sigma\gamma^2}\left(\sum_{i=1}^n \frac{1}{Z_i}\right)\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\eta}\boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T \\
&\qquad + n\left(\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T + \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})\boldsymbol{\Omega}_2^{-1}\boldsymbol{\Gamma}_{2\tau}^T(\boldsymbol{A})\right).
\end{aligned}$$

Step 4 Generate $\boldsymbol{\eta}$ from $\mathrm{N}\left(\widetilde{\boldsymbol{\eta}}_0,\ \Delta_{\boldsymbol{\eta}}^{-1}\right)$, where

$$\Delta_{\boldsymbol{\eta}} = \frac{1}{\sigma\gamma^2}\left\{\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T\boldsymbol{D}^{-1}\left(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})+\boldsymbol{M}\right\},$$
$$\widetilde{\boldsymbol{\eta}}_0 = \frac{1}{\sigma\gamma^2}\Delta_{\boldsymbol{\eta}}^{-1}\left\{\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)\boldsymbol{D}^{-1}\left(\mathbb{W}-\mu_{\tau,Y}\mathbf{1}_n\right)+\boldsymbol{M}\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\boldsymbol{e}\right\}.$$

Step 5 Generate σ from $\mathrm{IG}\left(\frac{3n}{2}+\frac{u_\tau}{2}+a,\ \widetilde{b}\right)$, where

$$\widetilde{b} = b+\sum_{i=1}^{n}Z_i+\frac{1}{2\gamma^2}\left\{\left(\mathbb{W}-\mu_{\tau,Y}\mathbf{1}_n-(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\eta}\right)^T\boldsymbol{D}^{-1}\right.$$
$$\left.\left(\mathbb{W}-\mu_{\tau,Y}\mathbf{1}_n-(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{1\tau}\boldsymbol{\eta}\right)+\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\boldsymbol{e}\right)\right\}.$$

Step 6 Generate $\boldsymbol{\Omega}_1$ from $\mathrm{IW}_{u_\tau}\left(\boldsymbol{\Psi}_1+\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)^T(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A}),\ \nu_1+n\right)$.

Step 7 Generate $\boldsymbol{\Omega}_2$ from $\mathrm{IW}_{p-u_\tau}\left(\boldsymbol{\Psi}_2+\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})^T(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)^T(\mathbb{X}-\mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T)\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A}),\ \nu_2+n\right)$.

Step 8 Generate a Markov chain realization for $\boldsymbol{A}$ with stationary density proportional to $H(\boldsymbol{A})$, which is the full conditional posterior density of $\boldsymbol{A}$ (see (S4.5) in the Supplement). Let $\boldsymbol{a}_j \in \mathbb{R}^{p-u_\tau}$ denote the j-th column of $\boldsymbol{A}$, $j = 1,\ldots,u_\tau$. Given the tuning parameter $\xi > 0$, for $j = i_1,\ldots,i_{u_\tau}$, where $\{i_1,\ldots,i_{u_\tau}\}$ denotes a random permutation of $\{1,\ldots,u_\tau\}$, perform the following:

(a) Generate $\boldsymbol{a}_j^*$ from $\mathrm{N}_{p-u_\tau}\left(\boldsymbol{a}_j,\ \xi^2\boldsymbol{I}_{p-u_\tau}\right)$. Replace the j-th column

of $\boldsymbol{A}$ by $\boldsymbol{a}_j^*$ and denote the resulting matrix as $\boldsymbol{A}^*$. Compute $\rho\left(\boldsymbol{A},\ \boldsymbol{A}^*\right) = \exp\left[H(\boldsymbol{A}^*) - H(\boldsymbol{A})\right]$.

(b) Perform a Bernoulli experiment with probability of success $\min(1, \rho\left(\boldsymbol{A},\ \boldsymbol{A}^*\right))$. If a success is obtained, update $\boldsymbol{a}_j^*$ to $\boldsymbol{a}_j$; otherwise retain $\boldsymbol{a}_j$.

(c) After updating $\boldsymbol{A}$, update $\boldsymbol{C}_{\boldsymbol{A}}$, $\boldsymbol{D}_{\boldsymbol{A}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$.

Remark 2. Algorithm 1 can account for the two degenerated cases $u_\tau = 0$ and $u_\tau = p$ as follows: when $u_\tau = 0$, $\boldsymbol{A}$ does not exist and we have $\boldsymbol{\eta} = 0$, $\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A}) = \boldsymbol{I}_p$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Omega}_2$. Thus the steps involving $\boldsymbol{\eta}$, $\boldsymbol{\Omega}_1$ and $\boldsymbol{A}$ (*Step 4*, *Step 6* and *Step 8* respectively) are not required. On the other hand, when $u_\tau = p$, the BEQR reduces to the BQR: $\boldsymbol{A}$ does not exist, $\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A}) = \boldsymbol{I}_p$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Omega}_1$, and the steps involving $\boldsymbol{A}$ and $\boldsymbol{\Omega}_2$ (*Step 7* and *Step 8*) are not needed. In each case, Algorithm 1 becomes a full Gibbs sampler.

Theorem 2 shows that a Markov chain generated using Algorithm 1 is *Harris ergodic* (see Supplement S1.2 for a technical definition). This provides the theoretical guarantee that for *all* starting points, Markov chains generated using Algorithm 1 converge to the target density. This property ensures consistent estimation of posterior expectations (Meyn and Tweedie, 2012). Without this property, an MCMC algorithm in general is only guaranteed to converge when the starting point falls outside a pathological set

of measure zero. Such pathological sets can be consequential and may arise naturally in practice, see Roberts and Rosenthal (2006). Typically, virtually no information on such *pathological* starting points is available in applications. The Harris ergodic property is thus essential for an MCMC algorithm to be useful in practice.

Theorem 2. *A Markov chain generated using the data augmentation Metropolis-within-Gibbs sampler in Algorithm 1, or its generalization to cases* $u_\tau = 0$ *and* $u_\tau = p$*, is Harris ergodic, i.e.,* ϕ*-irreducible with respect to some measure* ϕ*, aperiodic, and Harris recurrent.*

Remark 3. An algorithm for fast computation of posterior modes can be constructed using the full conditional distributions described in the above sampler. The details are provided in S5 of the Supplement.

4. Bayesian envelope quantile regression with censored data

This section discusses an extension of the BEQR that handles censored data. Censored data occurs naturally in many applications, and their statistical analyses have received increased attention over the past few decades (L.Powell, 1986; Buchinsky and Hahn, 1998; Bilias et al., 2000, among many others). The Bayesian paradigm provides a natural way of handling such data by modeling the censored points as augmented observations (Yu and

Stander, 2007; Kozumi and Kobayashi, 2011). We adapt the BEQR in a setting with Tobit censored (i.e., left censored at 0) responses for illustration, and provide an extension of Algorithm 1 for sampling from the posterior. The proposed model and algorithm could be easily adjusted for handling other types of censoring, such as interval and right censoring (Remark 4).

Let Y_i^* be an unobserved response; the corresponding observed response is $Y_i = Y_i^* \mathbb{1}_{\{Y_i^*>0\}}$, for $i = 1, \ldots, n$, where $\mathbb{1}_{\{.\}}$ denotes the indicator function. Using Y_i^* for the response in BEQR (3.5), we obtain the formulation of Bayesian envelope quantile regression with Tobit-censored response (BETQR)

$$\begin{aligned} Y_i^* &= \mu_{\tau,Y} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T \left(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}}\right) + \sigma\epsilon_i, \quad \epsilon_i \sim \text{ALD}(\tau), \\ \boldsymbol{X}_i &\sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{X}},\ \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T + \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})\boldsymbol{\Omega}_2\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})^T\right). \end{aligned}$$

We treat the unobserved response Y_i^* corresponding to the censored Y_i (i.e., the zero values) as augmented data in the model. To sample from the posterior, we iteratively generate random draws of the unobserved data given the model parameters and vice versa. Thus, one full iteration of the Markov chain sampler consists of two data augmentation steps – one for generating Y_i^*, and one for Z_i, the latter also appears in Algorithm 1. For

distinction, we call the former the *response variable imputation* step.

Conditional on Y_i^*, the model effectively reduces to (3.5). Hence the steps for generating model parameters $\mu_{\tau,Y}$, $\boldsymbol{\mu_X}$, $\boldsymbol{\eta}$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{A}$ and σ given Y_i^* are the same as in Algorithm 1, with Y_i replaced by Y_i^*. To generate random draws for Y_i^*, first note that conditional on $Y_i > 0$, $Y_i^* = Y_i$ with probability 1, while conditional on $Y_i = 0$ (and the model parameters and the augmented data Z_i's), Y_i^* has a truncated normal distribution $\mathrm{TN}_{(-\infty,0]}\left(\mu_{\tau,Y} + \boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{X}_i - \boldsymbol{\mu_X}\right) + \theta Z_i,\ Z_i\sigma\gamma^2\right)$. Therefore, in the response variable imputation step, we set $Y_i^* = Y_i$ when $Y_i > 0$, and sample Y_i^* from the above truncated normal distribution when $Y_i = 0$. A complete MCMC sampler is provided in Algorithm S6.1 in Supplement S6. The following theorem establishes Harris ergodicity of the sampler.

Theorem 3. *The Metropolis-within-Gibbs sampler in Algorithm S6.1 and its extension to the cases $u_\tau = 0$ and $u_\tau = p$ is Harris ergodic, i.e., ϕ-irreducible with respect to some measure ϕ, aperiodic, and Harris recurrent.*

Remark 4. With appropriate adjustment, the proposed approach can also handle other types of censoring, e.g., right and interval censoring. Consider $[a_0, b_0]$ interval censored responses as an example where a_0, $b_0 \in \mathbb{R}$ are known and fixed. Then $Y_i = Y_i^*\mathbb{1}_{\{a_0<Y_i^*<b_0\}} + a_0\mathbb{1}_{\{Y_i^*\le a_0\}} + b_0\mathbb{1}_{\{Y_i^*\ge b_0\}}$, and Y_i^* is generated as follows. Set $Y_i^* = Y_i$ if $a_0 < Y_i < b_0$; generate

Y_i^* from $\mathrm{TN}_{(-\infty,a_0]}\left(\mu_{\tau,Y}+\boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{\boldsymbol{X}}\right)+\theta Z_i,\ Z_i\sigma\gamma^2\right)$ if $Y_i=a_0$, and from $\mathrm{TN}_{[b_0,\infty)}\left(\mu_{\tau,Y}+\boldsymbol{\eta}^T\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T\left(\boldsymbol{X}_i-\boldsymbol{\mu}_{\boldsymbol{X}}\right)+\theta Z_i,\ Z_i\sigma\gamma^2\right)$ if $Y_i=b_0$.

5. Illustrations

This section investigates the performance of the BEQR in estimation efficiency. The censored data are also addressed. We consider simulated and real datasets, and compare estimation performance of the envelope model with the full (BQR; $u_\tau = p$) model in each dataset. For data analysis we consider vague non-informative prior on the model parameters. More specifically, we take (i) $a = b = 10^{-4}$ in the prior $\pi(\sigma)$ (Gelman et al., 2006), (ii) each element of $\boldsymbol{e}$ in $\pi(\boldsymbol{\eta} \mid \sigma, \boldsymbol{A})$ to be 10^{-3}, $\boldsymbol{M}$ in $\pi(\boldsymbol{\eta} \mid \sigma, \boldsymbol{A})$ to be 10^{-6} times a diagonal matrix with diagonal elements generated from χ^2 distribution with degrees of freedom 1, (iii) $\boldsymbol{A}_0$ in $\pi(\boldsymbol{A})$ to be the zero matrix, (iii) the covariance matrices in matrix normal prior and the scale matrix in each inverse Wishart prior to be 10^6 times the identity matrix, and (iv) the degrees of freedom in each inverse Wishart prior to be the column dimension of the corresponding the random matrix.

Selection of the envelope dimension u_τ is a critical step in envelope modeling. Here we view it as a model selection problem and use the leave-one-out information criterion (LOOIC) (Vehtari et al., 2017a,b). LOOIC

aims to approximate the expected out-of-sample log (posterior) predictive densities (ELPD), and chooses the most parsimonious model with a high ELPD. To this end we first fit all BEQR models with $u_\tau = 0, \ldots, p$ and compute the LOOIC (-2 times the ELPD) from each fit. The smallest u_τ with a LOOIC not substantially different from (within 2 standard errors of) the smallest computed LOOIC is then regarded as optimal. The `R` package `loo` is used for LOOIC computations. Since LOOIC requires fitting BEQR models with all possible envelope dimensions, i.e., $u_\tau = 0, \ldots, p$, it can be computationally expensive to perform in high dimensional problems. However, fitting BEQR models with different u_τ's can be paralleled, which can greatly reduce the total computing time in a modern multicore computer.

5.1 Simulated data

To aid assessment under possible model misspecification (heteroskedasticity), we generated data from the following model

$$\begin{aligned} Y_i &= \mu_Y + \boldsymbol{\eta}^T \boldsymbol{\Gamma}_{1\tau}^T(\boldsymbol{A}) \boldsymbol{X}_i + (5 + \boldsymbol{\alpha}^T \boldsymbol{X}_i)\epsilon_i; \\ \boldsymbol{X}_i &\sim N_p\left(\boldsymbol{\mu}_{\boldsymbol{X}},\ \boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_{1\tau}(\boldsymbol{A})^T + \boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})\boldsymbol{\Omega}_2\boldsymbol{\Gamma}_{2\tau}(\boldsymbol{A})^T\right), \quad i = 1, \ldots, n. \end{aligned} \tag{5.6}$$

We fixed $p = 10$, $u_\tau = 2$ and varied the sample size n from 50, 100, 200, 400, and 800. For each sample size, 200 replications were generated. The elements in $\boldsymbol{A}$ and $\boldsymbol{\eta}$ were generated from the uniform distributions unif$(9, 10)$ and unif$(4, 8)$ respectively. We took $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ to be diagonal matrices, with the two diagonal entries of $\boldsymbol{\Omega}_1$ generated from unif$(70, 80)$ and unif$(40, 50)$ respectively, and those in $\boldsymbol{\Omega}_2$ sampled from unif$(1, 3)$. Elements of the predictor mean $\boldsymbol{\mu_X}$ and the intercept μ_Y were generated from unif$(-10, 10)$ and unif$(20, 50)$ respectively. The first five elements in $\boldsymbol{\alpha}$ were fixed at 0 and the rest were fixed at 0.1. Finally, the error ϵ_i was a standard normal variate. The envelope dimension u_τ was selected by LOOIC. We obtained Bayesian point estimators for BEQR (with selected u_τ) and BQR ($u_\tau = p$) through their respective posterior means computed from 5000 MCMC iterations (after discarding the first 5000 burn-in iterations).

To compare the BEQR and BQR estimators, estimation variance and mean squared error (MSE) are computed from the 200 replications for each element in $\boldsymbol{\beta}_\tau$ for each estimator. For the BQR, let $\beta_i^{(k)}$ denote the ith element of the estimated $\boldsymbol{\beta}_\tau$ from the k-th replication. The estimation variance is defined as $\sum_{k=1}^{200} \left(\beta_i^{(k)} - \overline{\beta_i}\right)^2 /200$ and the MSE is defined as $\sum_{k=1}^{200} \left(\beta_i^{(k)} - \beta_{i,\text{true}}\right)^2 /200$, where $\overline{\beta_i} = \sum_{k=1}^{200} \beta_i^{(k)}/200$ and $\beta_{i,\text{true}}$ is the ith element of the true coefficients. We perform the same calculations for the

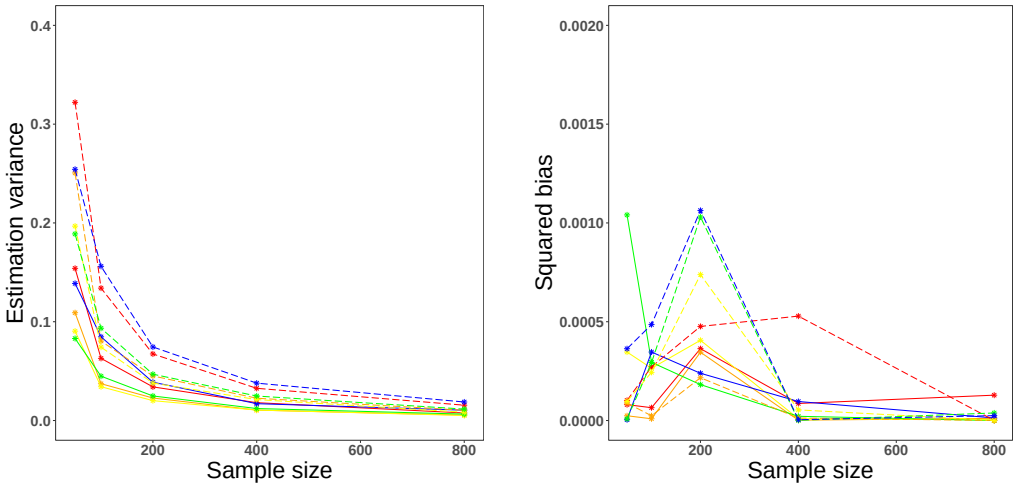

Figure 1: Estimation variances and squared biases of the second element in the BEQR estimator (solid line) and the BQR estimator (dashed line) with $\tau = 0.1$ (red line), 0.25 (orange line), 0.5 (yellow line), 0.75 (green line), and 0.9 (blue line).

BEQR. We consider quantiles 0.1, 0.25, 0.5, 0.75, and 0.9 as examples; the estimation variances and squared biases for the second element of the BEQR and BQR estimators are displayed in Figure 1. Figure 1 shows that the estimation variance is the dominant part of MSE across all quantile levels for both the BEQR and BQR estimators. The squared bias of the two estimators were comparable; however the BEQR achieved much smaller estimation variances compared to the BQR.

To better understand the efficiency gains we computed the ratio of the estimation variance and MSE of the BQR estimator versus the BEQR esti-

mator for each coordinate of $\boldsymbol{\beta}_\tau$, and then summarized these ratios (across all coordinates) by their respective medians and ranges for $\tau = 0.1$ and 0.5 in Table 1. Results for other quantile levels are similar. These ratios are strictly greater than 1 for all n demonstrating the efficiency achieved by the envelope approach.

To aid comparison with existing approaches, we also consider the R implementation `bayesQR` (Benoit and den Poel, 2017) of BQR. Note that `bayesQR` uses a similar ALD working model; however unlike BQR the predictors in `bayesQR` are assumed to be non-stochastic. Comparison of BEQR with this estimator conveys a similar message (see the Supplement S7.1).

n	$\tau = 0.1$		$\tau = 0.5$	
	Ratio_V	Ratio_M	Ratio_V	Ratio_M
50	1.58 (1.34, 2.13)	1.57 (1.34, 2.12)	1.73 (1.58, 2.63)	1.73 (1.58, 2.62)
100	1.88 (1.53, 2.52)	1.82 (1.51, 2.55)	1.80 (1.30, 2.77)	1.81 (1.30, 2.77)
200	2.01 (1.71, 2.41)	2.00 (1.64, 2.40)	1.83 (1.51, 2.43)	1.83 (1.50, 2.44)
400	1.89 (1.64, 2.92)	1.85 (1.55, 2.86)	1.88 (1.49, 2.62)	1.88 (1.49, 2.62)
800	2.02 (1.80, 3.02)	1.99 (1.71, 2.69)	1.68 (1.47, 1.99)	1.69 (1.47, 1.97)

Table 1: Medians (ranges) of the estimation variance and MSE ratios. Ratio_V: Estimation variance ratio of the BQR estimator versus the BEQR estimator. Ratio_M: Mean square error ratio of the BQR estimator versus the BEQR estimator.

We also investigated the dimension selection performance of LOOIC. Selection results of $\tau = 0.1$ and 0.5 are in Table 2 (recall that "true" u_τ is 2). Results for other quantile levels $\tau = 0.25$, 0.75 and 0.9 are similar to

those in Table 2. LOOIC selects the true u_τ most often especially for larger sample sizes. For smaller sample sizes rather than underestimating, LOOIC tends to overestimate u_τ. The overestimation issue in LOOIC, particular when p is large, is discussed in Piironen and Vehtari (2017). Overestimation curtails efficiency gains but does not produce estimation bias, while underestimation may introduce bias due to loss of material information. Thus mild overestimation of u_τ is usually less of a concern in practice (as illustrated by the efficiency gains in Table 1.) Additional simulation on the comparison of frequentist and Bayesian envelope quantile regression estimator is included in Supplement S7.2 and investigation of the effect of skewed distributions in included in Supplement S7.3. We also left censored the response in (5.6) at zero and looked into the effect of censoring on the efficiency gains, and the results are in Supplement S7.4. The effect of different censoring levels is investigated in Supplement S7.5. Additional simulation on censored data is in Supplement S7.6.

5.2 Real data analysis

5.2.1 LPGA data

The 2009 dataset for Ladies Professional Golf Association (LPGA) performance statistics (`http://users.stat.ufl.edu/~winner/data/lpga2009.`

	$\tau = 0.1$					$\tau = 0.5$				
Selected u_τ :	1	2	3	4	5	1	2	3	4	5
$n = 50$	8	101	60	30	1	5	121	47	24	3
$n = 100$	9	120	52	19	0	4	138	37	19	2
$n = 200$	2	153	35	10	0	2	157	22	17	2
$n = 400$	0	148	30	21	1	0	159	20	19	2
$n = 800$	0	160	27	12	1	0	152	18	25	5

Table 2: Number of replications (out of 200) for which a given value of u_τ is selected.

`dat`) contains winning prizes for 146 golfers with 9 measures of golfer's performances: average drive, fairways hit (%), greens reached in regulation (%), average putts per round, sand saves (%), greens in regulation putts per hole, average percentile in tournaments, rounds completed, and average strokes per round. The mean and the median of the winning prizes in USD were approximately 294K and 126K, which indicates right-skewness of winning prizes. Thus quantile regression, rather than standard linear (mean) regression is more suitable to analyze this dataset. We standardized all variables and computed the BEQR, the BQR, and `bayesQR` estimators on 11 quantile levels: $0.05, 0.10, 0.20, \ldots, 0.90$ and 0.95. We then obtained the posterior mean and 95% credible interval estimates of each regression coefficient for each model. The estimates for the coefficient of average drive are plotted in Figure 2 across all quantile levels for all three models (similar plots for other predictors are provided in Figure 5 in Supplement S7.7). As

depicted in Figure 2, the BEQR estimator has the smallest variation and `bayesQR` estimator has the largest variation, specifically for the extreme quantile levels 0.05 and 0.95.

LOOIC-based dimension selection results, and ratios of the lengths of credible intervals for the BQR versus the BEQR and the `bayesQR` versus the BEQR estimators are provided in Table 14 of the Supplement. For all quantile levels, the envelope approach is able to provide efficiency gains. For example, with $\tau = 0.1$, LOOIC chose $\widehat{u}_\tau = 2$. The ratios of the credible interval lengths of the BQR versus that BEQR varied from 1.91 to 10.4 with a median of 6.33, and the corresponding `bayesQR` to BEQR ratios ranged from 13.28 to 91.15 with a median of 59.69. The efficiency gains from the envelope estimators reflected in these ratios (especially at extreme quantile levels), permit enhanced detection of important predictor effects under the BEQR model. For example, average strokes (at levels $\tau = 0.1, 0.9$), green in regulation ($\tau = 0.1, 0.9$), average drive ($\tau = 0.9$) and average putts ($\tau = 0.1$) all have significant effects (95% credible intervals for the coefficients exclude zero) on winning prizes under the BEQR model, but identified as insignificant under the BQR model. In contrast, at $\tau = 0.5$, the significant predictors detected in the both models mostly concord, except for green in regulation, which is significant only in the BEQR model. All predictors are

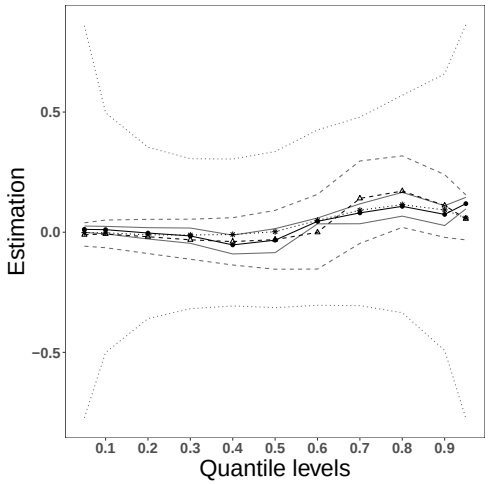

Figure 2: Point and 95% interval estimates of the coefficient of average drive. Solid lines mark BEQR estimators, dashed lines mark BQR estimators and dotted lines mark `bayesQR` estimators.

insignificant with the `bayesQR` estimator. For the BEQR and BQR, the significance of predictors at various quantile levels is detailed in Table 15 of Supplement S7.7.

5.2.2 Women's labor force data

In this section, we consider the dataset of labor force participation (Mroz, 1987) to understand the relationship between married women's working hours (in units of 100 hours) and four predictors, namely age, education (years), previous work experience (years) and other family income (in units $1000). Of the 753 women included in the dataset, 325 have zero working

hours, thus the corresponding responses are Tobit censored. Since more than 40% of the observations are left censored, we focus on the quantile levels 0.5, 0.75, and 0.9.

LOOIC selected envelope dimension u_τ as 1 and 2 for $\tau = 0.90$ and 0.75, but suggested $u_\tau = 4$ for $\tau = 0.5$ (in this case, BETQR degenerates to BTQR). We then computed the ratio of 95% credible interval lengths of the BTQR estimator versus the BETQR estimator for each coefficient. The range of the ratios are $(0.98, 1.63)$ and $(1.14, 16.2)$ with averages of 1.210 and 1.631 for $\tau = 0.75$ and $\tau = 0.9$ respectively. The improved efficiency from the envelope approach again enhances identification of important predictors in the dataset. For example, other family income (at both $\tau = 0.75$ and $\tau = 0.90$) and education (at $\tau = 0.9$) show significant effects under BETQR, while they are inferred to be insignificant in BTQR.

6. Discussion

There are several research directions in which the proposed framework could be extended. First, it will be of interest to examine if the envelope approach could be profitably combined with a non-parametric alternative to the ALD framework (Remark 1). Second, one could investigate whether the envelope modeling and posterior sampling strategy developed in this paper can aid

the development of Bayesian envelope models for other contexts, such as generalized linear and matrix/tensor variate regression. Finally, the incorporation of sparsity inducing shrinkage priors to handle large number of predictors in the proposed framework could also be studied.

References

Benoit, D. and D. V. den Poel (2017). bayesqr: A bayesian approach to quantile regression. *Journal of Statistical Software, Articles 76*(7), 1–32.

Bilias, Y., S. Chen, and Z. Ying (2000). Simple resampling methods for censored regression quantiles. *Journal of Econometrics 99*(2), 373–386.

Buchinsky, M. and J. Hahn (1998). An alternative estimator for the censored quantile regression model. *Econometrica 66*(3), 653–671.

Chen, T., Z. Su, Y. Yang, and S. Ding (2020). Efficient estimation in expectile regression using envelope models. *Electronic Journal of Statistics 14*(1), 143–173.

Cook, R., I. Helland, and Z. Su (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(5), 851–877.

Cook, R. D. (2018). *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics*. John Wiley & Sons.

Cook, R. D., L. Forzani, and Z. Su (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis 150*, 42–54.

Cook, R. D., B. Li, and F. Chiaromonte (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica 20*, 927–960.

Cook, R. D. and X. Zhang (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association 110*(510), 599–611.

De Jong, S. (1993). Simpls: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems 18*(3), 251–263.

Ding, S. and R. D. Cook (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(2), 387–408.

Ding, S., Z. Su, G. Zhu, and L. Wang (2021). Envelope quantile regression. *Statistica Sinica 31*, 79–106.

Dodge, Y. and J. Whittaker (2009). Partial quantile regression. *Metrika 70*(1), 35–57.

Dunson, D. B., M. Watson, and J. A. Taylor (2003). Bayesian latent variable models for median regression on multiple outcomes. *Biometrics 59*(2), 296–304.

Forzani, L. and Z. Su (2021). Envelope for elliptical multivariate linear regression. *Statistica Sinica 31*, 301–332.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis 1*(3), 515–534.

Khare, K. and J. P. Hobert (2012). Geometric ergodicity of the gibbs sampler for bayesian quantile regression. *Journal of Multivariate Analysis 112*, 108–116.

Khare, K., S. Pal, and Z. Su (2017). A bayesian approach for envelope models. *The Annals of Statistics 45*(1), 196–222.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koenker, R., A. Chesher, and M. Jackson (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), 1296–1310.

Kottas, A. and A. E. Gelfand (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association 96*(456), 1458–1468.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation 81*, 1565–1578.

Lancaster, T. and S. J. Jun (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics 25*(2), 287–307.

Li, L. and X. Zhang (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association 112*(519), 1131–1146.

L.Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics 32*(1), 143–155.

Ma, Y. and L. Zhu (2013). Efficiency loss and the linearity condition in dimension reduction. *Biometrika 100*(2), 371–383.

Meyn, S. P. and R. L. Tweedie (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica 55*(4), 765–799.

Piironen, J. and A. Vehtari (2017). Comparison of bayesian predictive methods for model selection. *Statistics and Computing 27*(3), 711–735.

Rekabdarkolaee, H. M., Q. Wang, Z. Naji, and M. Fuentes (2020). New parsimonious multivariate spatial model: spatial envelope. *Statistica Sinica*, to appear.

Roberts, G. O. and J. S. Rosenthal (2006). Harris recurrence of metropolis-within-gibbs and trans-dimensional markov chains. *The Annals of Applied Probability*, 2123–2139.

Sriram, K. (2015). A sandwich likelihood correction for bayesian quantile regression based on the misspecified asymmetric laplace density. *Statistics & Probability Letters 107*, 18–26.

Sriram, K., R. Ramamoorthi, and P. Ghosh (2013, 06). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Anal. 8*(2), 479–504.

Su, Z., G. Zhu, X. Chen, and Y. Yang (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika 103*(3), 579–593.

Vehtari, A., A. Gelman, and J. Gabry (2017a). Pareto smoothed importance sampling. *arXiv preprint: http://arxiv.org/abs/1507.02646/*.

Vehtari, A., A. Gelman, and J. Gabry (2017b). Practical bayesian model evaluation using leaveone-out cross-validation and waic. *Statistics and Computing 27*(5), 1413–1432.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. Krishnaiah (Ed.), *In Multivariate Analysis*, Volume 59, pp. 391–420. Academic Press, NY.

Wold, H. (1975). Path models with latent variables: The nipals approach. In *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, pp. 307–357. New York: Academic Press.

Yang, Y. and X. He (2012, 04). Bayesian empirical likelihood for quantile regression. *Ann. Statist. 40*(2), 1102–1131.

Yang, Y., H. J. Wang, and X. He (2016). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *International Statistical Review 84*(3), 327–344.

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

Yu, K. and J. Stander (2007). Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics 137*(1), 260–276.

Zhu, G. and Z. Su (2020). Envelope-based sparse partial least squares. *Ann. Statist. 48*(1), 161–182.

Department of Biostatistics, Columbia University

E-mail: ml4623@cumc.columbia.edu

Department of Biostatistics, State University of New York at Buffalo

E-mail: chakrab2@buffalo.edu

Department of Statistics, University of Florida

E-mail: zhihuasu@ufl.edu