Decoupling Shrinkage and Selection for the Bayesian Quantile Regression[*]

David Kohns [†]

Department of Economics, Heriot-Watt University

and

Tibor Szendrei

Department of Economics, Heriot-Watt University

July 20, 2021

Abstract

This paper extends the idea of decoupling shrinkage and sparsity for continuous priors to Bayesian Quantile Regression (BQR). The procedure follows two steps: In the first step, we shrink the quantile regression posterior through state of the art continuous priors and in the second step, we sparsify the posterior through an efficient variant of the adaptive lasso, the signal adaptive variable selection (SAVS) algorithm. We propose a new variant of the SAVS which automates the choice of penalisation through quantile specific loss-functions that are valid in high dimensions. We show in large scale simulations that our selection procedure decreases bias irrespective of the true underlying degree of sparsity in the data, compared to the un-sparsified regression posterior. We apply our two-step approach to a high dimensional growth-at-risk (GaR) exercise. The prediction accuracy of the un-sparsified posterior is retained while yielding interpretable quantile specific variable selection results. Our procedure can be used to communicate to policymakers which variables drive downside risk to the macro economy

Keywords: Global-Local Priors, Sparsity, Shrinkage, Machine Learning, Quantile Regression.
JEL: C110, C530, C550, C63

1 Introduction

While modern day economics, and broadly social science research, is often faced with high dimensional estimation problems in which the number of potential explanatory variables is large, often larger than the number of sample observations, the extant literature for high dimensional methods has focused developments mainly on for conditional mean models. Moving beyond the conditional mean, by estimating quantile regression on the other hand, allows to gauge potentially heterogeneous effects of variables directly across the conditional response distribution. While highly influential in the risk-management and finance literature in calculating risk measures such as VaR (i.e., the loss a portfolio's value incurs at a specific probability level), quantile regression has experienced a recent surge in popularity within the macroeconomic literature to quantify risks and vulnerabilities of output growth in response to summary measures of financial health, aptly named growth-at-risk (GaR) (Adrian et al., 2019; Figueres and Jarociński, 2020; Adams et al., 2020). As an important distinction to literature that focuses on forecasting crisis periods directly such as through Markov-switching models (Hubrich and Tetlow, 2015; Guérin and Marcellino, 2013) or probit models (McCracken et al., 2021), GaR instead gives information about the accumulation of risks facing an economy.

Since sources of risk can be numerous, high dimensional quantile problems are becoming ever more pertinent to policy makers and practitioners alike which has spurned methods that deal with variable selection and shrinkage for the quantile regression problem (Chernozhukov et al., 2010; Kohns and Szendrei, 2020; Hasenzagl et al., 2020). Likewise, recent papers such as Ferrara et al. (2020) and Carriero et al. (2020) use high dimensional Bayesian quantile regression (BQR) methods to nowcast tail risks in real time. However, most of the previous research has focused on the use of continuous shrinkage priors whose posterior is not easily amenable to interpretation due to lack of exact sparsity.

In this paper, we propose methods to simultaneously deal with high dimensional quantile estimation and variable selection from a Bayesian decision-theoretic perspective which allows to decouple shrinkage from variable selection in the spirit of Hahn and Carvalho (2015). We propose easy to implement adaptive sparsification procedures of frontier Bayesian quantile regression priors such as the horseshoe prior (Kohns and Szendrei, 2020) that adapt in a data driven manner to the given quantile. Evidence from simulations as well as a high dimensional GaR application to the US shows that our procedure either preserves fit of the unsparsified posterior or even improves both point- as well as - density fit, especially in the left tail which

characterises downside risks. Contributing further to investigating sparsity in economic data sets, we find that there is considerable quantile specific sparsity across the conditional distribution which is lost when only focusing on the conditional mean.

The Literature. The canonical way in which to do Bayesian variable selection for normal likelihood models is through Bayesian model averaging priors (Raftery et al., 1997; Clyde and George, 2004) in which the model space is discretised and posterior model probabilities allow for posterior probability statements about which is the most likely model, or to provide weights for model averaging. See Alhamzawi and Yu (2013) for an adaptation of the related g-prior approach to quantile regression. Due to combinatorial bottlenecks, much of Bayesian research in variable selection in the past decades have proposed a variety of methods (see Polson and Scott (2010) and Ishwaran et al. (2005) for excellent reviews), of which the search stochastic variable selection (SSVS) prior, dating back to George and McCulloch (1993) and George and McCulloch (1997), and adapted to the quantile regression by Korobilis (2017), has been notably influential. The SSVS prior is a discrete mixture of normals prior which stays computationally efficient as it stochastically searches among models with highest posterior probability. Although it only approximates the model space, it offers good empirical performance since low probability models are often ignored in the Markov chain.

Instead of approximating the discretised model space, continuous shrinkage priors on the other hand seek to include all variables at all times and instead shrink noise variables' coefficients toward zero so as to minimise their predictive influence. The class of global-local shrinkage priors (Polson and Scott, 2010) which include the lasso (Park and Casella, 2008), horseshoe (Carvalho et al., 2010) and Dirichlet-Laplace (DL) prior (Bhattacharya et al., 2015) increasingly supplant discrete mixture priors due to their excellent theoretical as well as empirical performance (Bhadra et al., 2019). Adapting such priors to quantile regression has been alreadyt attempted in the literature: See Li et al. (2010) for lasso type priors and Kohns and Szendrei (2020) for global-local prior adaptations to the BQR. The continuous nature of global-local priors allows additionally for fast and efficient computational methods (Bhattacharya et al., 2016; Kohns and Szendrei, 2020).

Instead of conducting estimation and variable selection in 1 step, Hahn and Carvalho (2015) propose a 2 step procedure in which the posterior is sparsified in an inferentially coherent way through integration over loss-functions which add model size penalties. In particular, they use adaptive lasso type penalisation, akin to Zou (2006). The idea is to threshold the posterior mean of the coefficients to zero, when they have a negligible effect in terms of squared-error loss to

the full model predictions. Piironen et al. (2020) extend this approach to Kullback-Leibler and Kowal (2021) to more general loss-functions. These approaches have been applied in numerous macroeconomic forecasting articles (see Huber et al. (2019); Kohns and Bhattacharjee (2020) among others) which show good forecasting and variable selection properties.

In section 2, we firstly provide a review of the BQR and shrinkage priors considered, followed by a derivation of our proposed decision theoretically motivated sparsification, and details on its implementation. In section 3, we conduct a large scale Monte Carlo experiment that tests the proposed methodology's robustness to a variety of high dimensional settings. This is followed in section 4 by a high dimensional GaR application and lastly we conclude.

2 Methodology

2.1 Bayesian Quantile Regression

Assuming a linear model such as

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, 2, \cdots, T, \tag{1}$$

where $\{y_t\}_{t=1}^T$ is a scalar response variable and $\{x_t\}_{t=1}^T$ a $K \times 1$ known covariate vector, the objective function of quantile regression can be expressed as the minimised sum of weighted residuals which are zero in expectation for the given quantile $p \in (0, 1)$:

$$\hat{\beta}_p = \min_{\beta} \sum_{t=1}^n \rho_p(y_t - x_t'\beta), \tag{2}$$

whose solution $\hat{\beta}_p$, is a $K \times 1$ quantile specific coefficient vector. Note that the expected quantile $\hat{Q}_p(Y|X) = X\hat{\beta}(p)$ is a consistent estimator of $Q_p(Y|X)$, independent of any parametric assumption about residuals $\{\epsilon\}_{t=1}^T$ (Koenker, 2005). We will maintain the assumption throughout that paper that the design X is known. The loss function $\rho_p(.)$ is often expressed as a tick loss function of the form $\rho_p(y) = [p - I(y < 0)]y$ where $I(.)$ is an indicator function taking on a value of 0 or 1 depending on whether the condition is satisfied. As noted by Koenker et al. (2017), this loss function is proportional to the negative log density of the asymmetric laplace distribution. This connection has been used to recast quantile regression as a maximum likelihood solution of model (1) with an Asymmetric-Laplace distribution, denoted as $\mathcal{ALD}(p, 0, \sigma)$, where σ denotes the scale of the $\mathcal{ALD}$. Assuming an $\mathcal{ALD}$ error distribution, the working likelihood $f(Y|X, \beta_p, \sigma)$ becomes:

$$f(Y|\beta, \sigma) = \frac{p^T(1-p)^T}{\sigma^T} \prod_{t=1}^{T} \left[e^{-\rho_p(y_t - x_t'\beta_p)/\sigma} \right]. \tag{3}$$

As posterior moments with conventional priors are not analytically available with an $\mathcal{ALD}$ working likelihood, it has become standard practice in the literature to use a mixture representation, proposed by Kozumi and Kobayashi (2011), in which the $\mathcal{ALD}$ error process can be recovered as a mixture between an exponentially distributed variable z_t, $z_t \sim exp(\sigma)$, and a standard normal variable, u_t, $u_t \sim N(0,1)$:

$$\epsilon_t = \xi z_t + \tau \sqrt{\sigma z_t} u_t$$
$$\xi = \frac{1 - 2p}{p(1-p)} \tag{4}$$
$$\tau^2 = \frac{2}{p(1-p)}$$

where ξ and τ are deterministic quantile specific parameters. The conditional likelihood stacked over all observations thus becomes:

$$f(Y|X, \beta_p, Z, \sigma) \propto det(\Sigma)^{-\frac{1}{2}} exp\left\{ -\frac{1}{2}[(y - X\beta_p - \xi Z)'\Sigma(y - X\beta_p - \xi Z)] \right\}, \tag{5}$$

where $Y = (y_1, \cdots, y_T)'$, $X = (x_1, \cdots, x_T)'$, $Z = (z_1, \cdots, z_T)'$ and $\Sigma = diag(1/\tau^2 z_1 \sigma, \cdots, 1/\tau^2 z_T \sigma)$. Hence, the mixture representation results in a normal kernel for the likelihood which enables analytical solutions for conditional posteriors as shown below.

Throughout the paper, we consider priors on β_p that take the following form:

$$\beta_p \sim N(\mathbf{0}_K, \Lambda_*), \tag{6}$$

where a prior mean of zero implies shrinkage toward sparsity and the prior variance parameters, Λ^* control the amount of shrinkage toward zero.

By applying independent priors $p(\beta_p, \sigma, Z) = p(\beta_p)p(\sigma)p(Z)$, the conditional posterior for β_p is normal:

$$p(\beta_p|\cdot) \sim N(\overline{\beta}_p, \overline{\Lambda}_*)$$
$$\overline{\beta}_p = \overline{\Lambda}_*(X'\Sigma(Y - \xi Z)) \tag{7}$$
$$\overline{\Lambda}_* = (X'\Sigma X + \Lambda_*^{-1})^{-1}.$$

The conditional posterior of the scale parameter is:

$$p(\sigma|\cdot) \sim IG(\overline{a}, \overline{b})$$
$$\overline{a} = \underline{a} + \frac{3T}{2}$$
$$\overline{b} = \underline{b} + \sum_{t=1}^{T} \frac{(y_t - x_t'\beta_p - \xi z_t)2}{2z_t + \tau^2} + \sum_{t=1}^{T} z_t,$$

(8)

where IG stands for the inverse-Gamma distribution. Finally, the conditional posterior for z_t is:

$$p(z_t|\cdot) \sim 1/iG(\overline{c}_t, \overline{d}_t)$$
$$\overline{c}_t = \frac{\sqrt{\xi^2 + 2\tau^2}}{|y_t - x_t'\beta_p|}$$
$$\overline{d}_t = \frac{\xi^2 + 2\tau^2}{\sigma\tau^2},$$

(9)

where iG stands for the inverse Gaussian density with location ($\overline{c}_t$) and rate ($\overline{d}_t$) parameter.

The conditionally conjugate posteriors (7,8,9) allow for efficient Gibbs sampling algorithms which for the independent prior setup have been shown to be geometrically ergodic by Khare and Hobert (2012), independent of any assumptions on X. Hence, X could include more variables than observations, dependent or deterministic variables.

Since taking the inverse of the posterior covariance of the regression coefficients $\overline{\Lambda}_*$ can be computationally demanding in high dimensions, we make use of the fast BQR sampler proposed by Kohns and Szendrei (2020) that reduces the computational complexity involved in obtaining a draw from $\overline{\beta}_p$ from $\mathcal{O}(K^3)$ to $\mathcal{O}(T^2 K)$. This algorithm is particularly suitable for macroeconomic data sets in which K is typically much larger than T. Details of the exact sampling steps for all priors considered are given in the appendix.

2.2 Shrinkage Priors

Shrinkage priors, both for conditional mean and quantile models, can be understood as a Bayesian generalisation to frequentist penalised regression, where the penalisation is a function of the number of active coefficients:

$$\beta = \underset{\tilde{\beta}}{\operatorname{argmin}} \sum_t h(y_t, x_t, \tilde{\beta}) + \delta Q(\tilde{\beta}).$$

(10)

h and Q are two positive functions, and δ controls the amount of penalisation. Choosing $h(\bullet)$ to be the negative log-likelihood of the data and $Q(\tilde{\beta})$ to be the ℓ_1-norm, $||\tilde{\beta}||_1$, (10) yields sparse optimal solutions for β which are maximum likelihood equivalents the popular lasso estimator of Tibshirani (1996) when the likelihood is Gaussian. Likewise, using ℓ_1-norm penalisation and setting $h(\bullet)$ to the negative log of the $\mathcal{ALD}$ likelihood, (10) recovers the maximum likelihood equivalent of the quantile lasso of Chernozhukov et al. (2010). From a Bayesian perspective, $Q(\beta)$ instead can be understood as the negative log prior distribution imposed on β, which renders (10) the negative log-posterior. The Bayesian paradigm has the added advantage of being able to set a prior also for the amount of penalisation $p(\delta)$ which gives a probabilistic way to conduct shrinkage. Different prior forms of $p(\delta)$ will result in different shrinkage properties.

Lasso Prior. Fistly introduced by Park and Casella (2008), the lasso prior generalises the frequentist lasso by placing a mixture exponential prior on δ. Adapted to the BQR by Li et al. (2010), the prior takes the following form:

$$\beta_j|\phi \sim N(0, \lambda_j),$$
$$\lambda_j|\phi \sim exp(\frac{\phi}{2}) \tag{11}$$
$$\phi \sim G(a_1, b_1)$$

where $\Lambda_* = diag(\lambda_1, \cdots, \lambda_K)$. The conditional posteriors for the hyperparameters are standard:

$$p(\lambda_j^{-1}|\bullet) \sim iG(\sqrt{\frac{\phi}{\beta_{j,p}^2}}, \phi)$$
$$p(\phi|\bullet) \sim G(K + a_1, \frac{1}{2}\sum_{j=1}^{K} \lambda_j + b_1). \tag{12}$$

If instead ϕ is allowed to vary with the same distribution as in (11) for each j, this prior will become the adaptive lasso prior of Alhamzawi et al. (2012)[1].

Horseshoe Prior. Similar to the standard result as to why the ℓ_1-norm penalty in a frequentist approach to solving (10) tends to over shrink signals in lasso regression (Zou, 2006),

---

[1]Since in simulations and the application we found there to be no significant difference in terms of performance between the Bayesian lasso and adaptive lasso BQR, we report only resuls for the Bayesian lasso. See Kohns and Szendrei (2020) for a thorough investigation of the differences in performance between these priors.

the Bayesian lasso over shrinks large coefficients as well[2]. The general class of global-local priors seek to solve this problem by adding another shrinkage scale hierarchy, which, with suitably chosen priors, approximate the ideal behaviour of shrinking noise variables aggressively toward zero, while leaving signal variables untouched. Global-local priors take the following form:

$$\beta_j | \lambda_j^2, \nu^2 \sim N(0, \lambda_j^2 \nu^2), j \in (1, \cdots, K)$$
$$\lambda_j^2 \sim \pi(\lambda_j^2) d\lambda_j^2, j \in (1, \cdots, K) \tag{13}$$
$$\nu^2 \sim \pi(\nu^2) d\nu^2.$$

The horseshoe prior of Carvalho et al. (2010) employs two independently distributed half Cauchy distributions on the positive support for both the global and local scales:

$$\lambda_j \sim C_+(0, 1) \tag{14}$$

$$\nu \sim C_+(0, 1), \tag{15}$$

where $\Lambda^* = \nu^2 diag(\lambda_1^2, \cdots, \lambda_K)$. Due to the assumption of independence on the scales (λ, ν), it is straightforward to show that their posteriors follow independent Cauchy distributions. To sample from their posteriors, we make use of the slice sampler used for the Horseshoe-BQR in Kohns and Szendrei (2020).

SSVS Prior. While the lasso and horseshoe prior are continuous shrinkage priors, the SSVS prior, dating back to Mitchell and Beauchamp (1988) and George and McCulloch (1993, 1997), discretises the model space akin to Bayesian model averaging priors (Raftery et al., 1997; Clyde and George, 2004). Since evaluating each model to compare marginal likelihoods becomes quickly infeasible in high dimensions, the SSVS saves computation time by exploring models through its Markov chain with highest posterior probability. This is achieved by modeling β by a mixture prior, where coefficients are sorted into two groups, the "spike" and the "slab". When sorted into the first group, the value of the coefficient is shrunk heavily toward zero, while in the second it is modelled through a disperse normal prior. We follow George et al.

---

[2]Carvalho et al. (2010) and Piironen et al. (2017) formalise this argument through implied shrinkage distributions for the normal means and normal regression model respectively.

(2008)'s implementation, which has been adapted to the BQR by Korobilis (2017):

$$\beta_{j,p}|\gamma_j,\lambda_j \sim (1-\gamma_j)N(0,c\lambda_j^2) + \gamma_j N(0,\lambda_j^2)\forall j \in \{1,\cdots,K\}$$

$$\lambda_j^2 \sim G(a_2,b_2)$$

$$\gamma_j|\pi_0 \sim Bern(\pi_0) \tag{16}$$

$$\pi_0 \sim B(a_3,b_3),$$

where $Bern(\bullet)$ stands for the Bernoulli distribution, $B(\bullet)$ for the Beta distribution and $c = 10^{-5}$, which effectively shrinks the spike group of coefficients to 0. π_0 controls the probability of inclusion into the slab group. Λ^* for the SSVS priors becomes $diag(\lambda_1^2,\cdots,\lambda_K^2)$ for all j if $\gamma_j = 1$ and $diag(c\lambda_1^2,\cdots,c\lambda_K^2)$ if $\gamma_j = 0$. Under prior (16), the conditional posteriors are:

$$\lambda_j^2|\bullet \sim G(a_2 + \frac{1}{2}, \frac{\beta_{j,p}^2}{2} + b_2)$$

$$\gamma_j|\bullet \sim Bern(\frac{\pi_0 N(\beta_{j,p},\lambda_j^2)}{\pi_0 N(\beta_{j,p},\lambda_j^2 + (1-\pi_0)N(\beta_{j,p},c\lambda_j^2)}) \tag{17}$$

$$\pi_0|\bullet \sim B(1+a_3, k-1+b_3),$$

where k denotes the size of the slab group. Due to the very strong shrinkage implied by the scalar c, we treat $\mathbb{1}'\gamma_j^s|\bullet$ as a posterior estimate of the model size on an iteration basis. Further, throughout all simulation and the empirical application, we set $a_3 = b_3 = 1$, which embeds the assumption that a-priori all model sizes are equally likely, and thus allows for dense as well as sparse models as recommended by Giannone et al. (2017).

2.3   Decoupling Shrinkage and Sparsity

While continuous shrinkage priors such as global-local priors in (13) yield good forecasting performance (Cross et al., 2020; Huber et al., 2019; Kohns and Bhattacharjee, 2020), interpretation of forecasts based on the posterior are impeded by the fact that the marginal posterior $p(\beta_p|Y)$ is continuous on $\mathbb{R}^K$. To aid interpretability, the assumption of sparsity is often employed, which forces small coefficients to zero when they have small enough effects on the target, Y. The Bayesian approach to enforcing sparsity can be seen as an optimal action that minimises an expected loss function that embeds the preference of sparsity as:

$$\mathcal{L}(\tilde{Y},\alpha) = \phi||\alpha||_0 + T^{-1}||X\alpha - \tilde{Y}||_2^2, \tag{18}$$

which was proposed by Hahn and Carvalho (2015), and $|| \bullet ||_0$ refers to the ℓ_0-norm. $\tilde{Y}$ refers to a realisation of the predictive distribution with density:

$$p(\tilde{Y}|Y) = \int p(Y|\tilde{X}, \theta)p(\theta|Y, \tilde{X})d\theta. \tag{19}$$

In (19), we have collected all unknown parameters into θ for the sake of brevity. For simplicity, we focus here on in-sample predictions which renders $\tilde{X}$ known and equal to X.[3] Note that in the following discussion, we omit the condition on $\tilde{X}$ for readability. Since $\tilde{Y}$ is a latent quantity, the expectation of (18) requires integration both with the respect to $\tilde{Y}$ and θ:

$$\mathcal{L}(\alpha) = \int \int \phi||\alpha||_0 + T^{-1}||X\alpha - \tilde{Y}||_2^2 \, p(\tilde{Y}|\theta)d\tilde{Y} \, p(\theta|\tilde{Y})d\theta. \tag{20}$$

Theorem 1: Assume the observation likelihood is the ALD, given by (3), and that the posterior for θ, $p(\theta|Y)$, given by (7), (8), and (9), has been obtained through posterior sampling, then the expected loss (20) is given by:

$$\mathcal{L}(\alpha) \propto \phi||\alpha||_0 + T^{-1}||X\alpha - X\overline{\beta}_p||_2^2 + tr(X'X\overline{\Sigma}_{\beta_p}) - T^{-1}\alpha'X'\xi\overline{Z}_p, \tag{21}$$

where $\overline{\beta}$ and $\overline{\Sigma}_{\beta_p}$ refer to the posterior mean and covariance of $p(\beta_p|Y)$, and $\overline{Z}_p$ is defined as $\overline{Z}_p = \left(\frac{|y_1 - x_1'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2}, \cdots, \frac{|y_T - x_T'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2} \right)'$. The proof is provided in the appendix.

Remark on Theorem 1: the integrated loss function (21) is remarkably similar to the normal likelihood case derived in Hahn and Carvalho (2015), with the difference being a term involving the posterior means of Z, $\overline{Z}_p$, which appear due to the mixture representation of the ALD. Since we are interested in sparsifying the vector $\beta_p|\bullet$, so as to minimise the Euclidean distance to the expected quantile $\hat{Q}_p = X\beta_{p|\bullet}$, instead of the expected location of the ALD directly, we proceed from this analysis by neglecting the term involving $\overline{Z}_p$. While this may seem a strong simplification, notice that 1) the contributions to this term become very small with increasing T as the entries of $\overline{Z}_p$ are multiplied one covariate at a time, which will become clear from the coordinate descent algorithm employed below; and 2) for $p = 0.5$, the term vanishes completely due to $\xi = 0$. Thus central quantiles are virtually unaffected. Lastly, ignoring the terms involving $\overline{Z}_p$ can also be understood from the perspective that the ALD in the BQR is often only treated as a working likelihood rather than the true data generating process of Y in

---

[3]Note that in general, this need not be the case and X might contain an entirely different subset to X. See Hahn and Carvalho (2015) and Piironen et al. (2020) for more discussion on that.

order to retrieve posterior estimates of the quantile regression coefficients (Yang et al., 2016). To convince on the negligible effect of dropping the last term in (21), we provide simulation evidence below.

Proceeding by dropping all constant terms, the objective function becomes:

$$\alpha_p^* = \underset{\alpha}{\text{argmin}} \ \ \phi||\alpha||_0 + T^{-1}||X\alpha - X\overline{\beta}_p||_2^2. \tag{22}$$

At this point, it is interesting to note the difference in the objective function to frequentist penalised quantile regression such as the quantile lasso of Chernozhukov et al. (2010). While Chernozhukov et al. (2010) compute the expected quantiles from observations Y, (22) finds a sparsified vector α_p^* by minimising the squared differences to the expected quantile directly. Hence, due to the availability of the posterior for $\beta_p|\bullet$, $\overline{\beta}_p$, the quantile can be treated as observed. To verify that indeed squared error loss applied in (22) results in recovering the desired quantile coefficients, we provide Lemma 1 below:

Lemma 1: Let $\beta_p^* \in \mathbb{R}^K$ be the true quantile regression parameter vector, and X be of full column rank. Then, the following minimisation problem recovers the true quantile coefficients:

$$\hat{\beta}_p = \underset{\tilde{\beta}}{\text{argmin}} \ \ ||X\beta_p^* - X\tilde{\beta}||_2^2/T. \tag{23}$$

Proof: Expanding (23) and solving the first order conditions yields: $\hat{\beta}_p = (X'X)^{-1}X'X\beta_p^* = \beta_p^*$.

2.4 Signal Adaptive Variable Selection for the BQR

Following Hahn and Carvalho (2015) and Ray and Bhattacharya (2018), we make three modifications to the objective function (22). Firstly, we make use of the ℓ_1-norm instead of ℓ_0-norm penalisation to obtain a convex objective function whose solution is computable with standard techniques such as the coordinate descent algorithm of Friedman et al. (2010). Secondly, it is well known that the ℓ_1-norm penalisation as an approximation to the ℓ_0 norm overshrinks signals which might hurt predictive performance. To tackle this we make use of adaptive penalisation akin to Zou (2006):

$$\alpha_p^* = \underset{\alpha}{\text{argmin}} \ \ \frac{1}{2}||X\overline{\beta}_p - X\alpha||_2^2 + \sum_{j=1}^K \phi_j|\alpha|_j. \tag{24}$$

Lastly, Ray and Bhattacharya (2018) have observed that starting the coordinate descent algorithm with the posterior means of $\beta|\bullet$ results in convergence after the first iteration by setting $\phi_j = \frac{1}{|\overline{\beta}_j|^\kappa}$ and fixing $\kappa = 2$. Hence, the solution to (24) with one iteration of the coordinate descent algorithm simplifies to:

$$\hat{\alpha}^*_{p,j} = sign(\overline{\beta}_{p,j})||X_j||^{-2}(|\overline{\beta}_{p,j}| \cdot ||X_j||^2 - \phi_j)_+, \tag{25}$$

where $sign(y) = 1$ if $y \geq 0$ and -1 otherwise, and let $y_+ = max\{y, 0\}$. See the appendix for a full derivation of the coordinate descent algorithm.

While the original SAVS procedure proposed by Ray and Bhattacharya (2018) provides good sparsification properties by fixing $\kappa_j = 2$ across all coefficients for conditional mean models, a uniform value of κ across quantiles can be inappropriate if the degree of sparsity varies across the conditional distribution. Quantile specific sparsity can arise when the coefficients β_p have a quantile profile which crosses the zero line for any $p \in (0, 1)$. Hence, to accommodate this, we make two contributions to the SAVS algorithm which we call the BQR_{SAVS} algorithm. Firstly, we treat κ_j as a parameter to be estimated from the data. And secondly, since standard score functions such as the log-score employed in projective variable selection (Piironen et al., 2020) or other functionals (Kowal, 2021) would entail variable selection consistent with conditional mean models, we use quantile specific score functions. In particular, we make use of the quantile BIC (qBIC) of Lee et al. (2014) and choose $\hat{\kappa}_p$ such that it minimises the qBIC criterion:

$$
\begin{aligned}
\hat{\kappa}_p &= \underset{\kappa}{argmin} \quad qBIC(\kappa) \\
&= \underset{\kappa}{argmin} \quad log(\sum_{t=1}^{T} \rho_p(y_t - x'_t \overline{\beta}_{p,\kappa})) + |\hat{\mathcal{S}}_\kappa| \frac{log(T)}{2T} C_K,
\end{aligned}
\tag{26}
$$

where $|\hat{\mathcal{S}}_\kappa|$ is the cardinality of sparsified $\overline{\beta}_{p,\kappa}$ conditional on parameter $\hat{\kappa}_p$, and C_T, is a positive constant which diverges to infinity with K. We follow Lee et al. (2014) by setting $C_T = log(p)$. Lee et al. (2014) show that estimating the level of penalisation in ℓ_1 penalised quantile regression through the qBIC, results in model selection consistency under expanding K and T.[4] While model selection consistency is hard to prove in the Bayesian paradigm and is an active field of investigation, it stands to reason that we get close to model selection consistency in the frequentist sense when $\overline{\beta}_p$ is consistently estimated. We leave this proof for future research.

---

[4]Model selection consistency is also dependent on specifying an upper bound, s, for $\mathcal{S}$ when $K >> T$. Not specifying s still leads to model selection consistency, however, with mild extra assumptions on the rate of K.
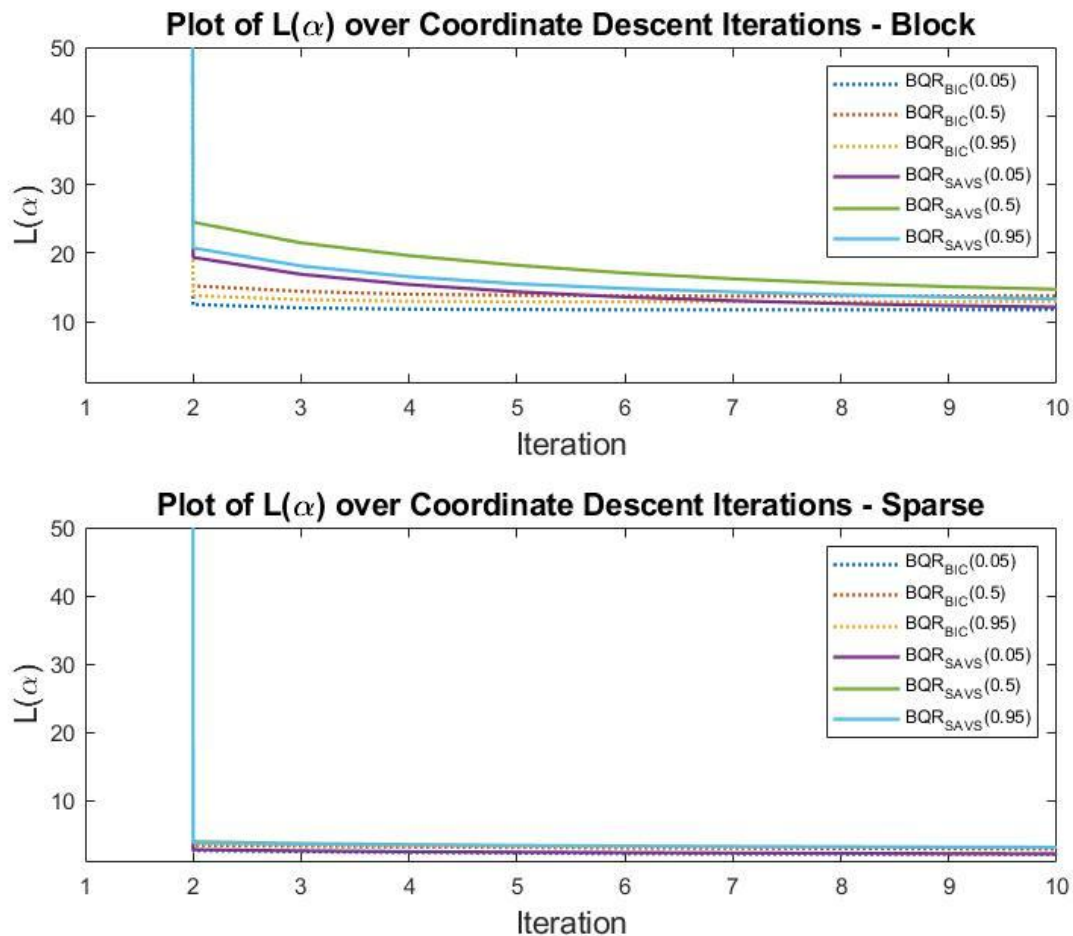
Figure 1: Coordinate Descent paths for objective function (24) for a block-sparsity and sparse DGP taken from the Monte Carlo simulations below. The convergence paths are an average over 50 simulations with 100 in-sample observations with 100 covaraites, not including the constant.

Alternatively to finding $\hat{\kappa}_p$ through minimisation of (26), one can embed a decision rule that incorporates uncertainty in $\hat{\kappa}_p$, such as choosing the smallest $\hat{\kappa}_p$ (and therefore the smallest model size), that is no worse as measured by the qBIC, to the un-sparsified posterior with probability ζ. Obtaining a standard error estimate for the difference between the qBICs of the sparsified and un-sparsified posteriors are readily available from the the output of the Markov chain. We found, however, throughout simulations as well as the application that either method gave similar results, such that we use the minimum qBIC procedure as in (26). We will denote the BQR_{SAVS} using $\hat{\kappa}_p$ that minimises the qBIC as BQR_{BIC}.

The importance of selecting $\hat{\kappa}_p$ in a data informed manner, even for simple data generating

processes without quantile varying sparsity, can be gauged from the convergence path of the coordinate descent algorithms, shown in figure (1).

Taking average solution paths over 50 simulations based on a simple high dimensional block and sparse DGP, which are further elaborated in section 3, figure (1) shows clearly that convergence is reached for the qBIC penalised SAVS objective function after only one iteration, while the plain SAVS takes usually more than 10 iterations on DGPs which are non-sparse. The level of the objective function is additionally much lower for qBIC compared to SAVS. Hence, stopping the coordinate descent algorithm with the plain SAVS bears the risk of inefficient penalisation.

Finally, since with correlated designs, the mean of the posterior distribution of β, can perform badly in terms of forecasting due to multimodality, we implement sparsification (25) as well as penalisation parameter choice (26) on an iteration basis. This bears the added advantage of being able to quantify variable selection uncertainty through the frequency of how often a coefficient is selected into model $\mathcal{M}^s$, where $\mathcal{M}^s$ denotes a quantile regression model indexed by a vector of binary indicators of dimension K per iteration s. The percentage of times coefficient j appears in $\mathcal{M}^s$, can thus be interpreted akin to posterior variable inclusion probabilities such as $\mathbb{E}_\gamma(\gamma_j|\bullet)$. Alongside posterior credible intervals for β_p, this gives an interpretable, Bayesian decision theoretically motivated way to conduction variable selection for any continuous prior for the BQR framework.

3   Monte Carlo Experiment

3.1   Monte Carlo Setup

Simulated experiments are conducted in order to investigate the potential benefits of sparsification for the BQR in terms of its variable selection properties, coefficient bias and robustness to different data generating processes (DGPs). Another purpose of this simulation experiment is to compare variable selection of the proposed methods to the standard way of conducting Bayesian variable selection, namely via the SSVS prior. In particular, we are interested in the BQR_{SAVS}' and BQR_{BIC}'s ability to adapt to different degrees of sparsity and error distributions. We generate data from the following model:

$$y_t = \beta_0 + x'_{t,c}\beta_c + x'_{t,q}\beta_q + (x'_t\vartheta)u_t, \ \ t = 1,\cdots,T, \tag{27}$$

where $\{u_t\}_{t=1}^T$ are assumed to follow some otherwise unspecified cumulative density function (CDF), F, and the subscripts c and q refer to covariate groups whose regression coefficients don't vary and vary across quantiles respectively. From (27), one can see that the quantile profile of the quantile varying coefficients is in turn dictated by the correlation between the covariate and error process u_t which is enforced through ϑ being a K-dimensional binary vector which is non-zero if $x_{t,i} \in x_{t,q}$ (as well as for the constant coefficient). Assuming a tick-loss function and solving for quantile regression coefficients $\beta(p)$ for any p, Koenker (2005) has shown that quantile regression coefficients have a random coefficient interpretation, where the quantile profile is proportional to the quantiles of the error process, with its centre on β_c and β_q respectively:

$$\beta_0(p) = \beta_0 + \vartheta_0 F^{-1}(p)$$
$$\beta_q(p) = \beta_q + \vartheta_q F^{-1}(p)^5 \qquad (28)$$
$$\beta_c(p) = \beta_c$$

The constant coefficient β_0 by default always has a quantile profile so as to enforce location effects. Solution (28) to model (27) therefore motivates the investigation of sparsity detection for both the β_c and β_q vectors, and also whether the assumption of the error distribution effects this.

In particular, consider a homoskedastic design in which ϑ is populated by zeros such that $\beta(p) = \beta \; \forall p$ where $\beta = (\beta_c', \beta_q')'$. These offer baseline DGPs in which, as per (28), only the constant has a non-constant quantile function. Sparsity patterns considered are:

- $\beta_{sparse} = (1, 1.5, 1, 0.5, 0.33, 0.25, 0_{1\times(K-4)})$

- $\beta_{block} = (1, 0.5_{1\times\frac{K}{5}}, 0_{1\times 2\frac{K}{5}}, 0.5_{1\times\frac{K}{5}})$

To investigate how well the BQR_{SAVS} and BQR_{BIC} detect sparsity specific to a given quantile, we consider additionally a heteroskedastic set of DGPs where β_q is zero only for some p. In particular, we let the constant coefficient be zero at the median and an additional covariate be zero only for $0.15 < p < 0.85$.[6] This is equivalent to making the ϑ be quantile varying as well. In particular ϑ_q for the quantile specific sparsity is defined as:

---

[6]Centering the quantile function of β_0 on 0 helps identifying quantile profiles on other covariates on account of tails of coefficients being 'aligned'

$$\vartheta_q^* = \vartheta_q[I(u_t \leq F^{-1}(0.15)) + I(u_t \geq F^{-1}(0.85))] \tag{29}$$

Sparsity patterns that are considered for the quantile specific sparse Monte Carlo designs are:

- $\beta_{sparse,c} = (0, 0.5, 0.33, 0.25, 0_{1\times(K-4)})$, $\beta_q = 0$ and $\beta_0 = 0$

- $\beta_{block,c} = (0, 0.5_{1\times\frac{K}{5}}, 0_{1\times2\frac{K}{5}}, 0.5_{1\times\frac{K}{5}})$, $\beta_q = 0$ and $\beta_0 = 0$

For the homoskedastic DGPs, we consider standard normal and student-t-distributed errors with 3 degrees of freedom. For heteroskedastic DGPs, we also consider standard normal error and one further variant. Since it may also be plausible to find that there is sparsity across p, but with constant non-zero coefficients for non-sparse quantiles, we consider as a final heteroskedastic design the quantile varying coefficient to be -0.5 for $p \in (0, 0.15]$ and 0.5 for $p \in [0.85, 1)$. For clarity, the quantile profiles for the quantile specific sparse coefficient are presented in figure (2). The quantile profile of the quantile varying beta coefficients for the 4 designs are summarised in table (1).
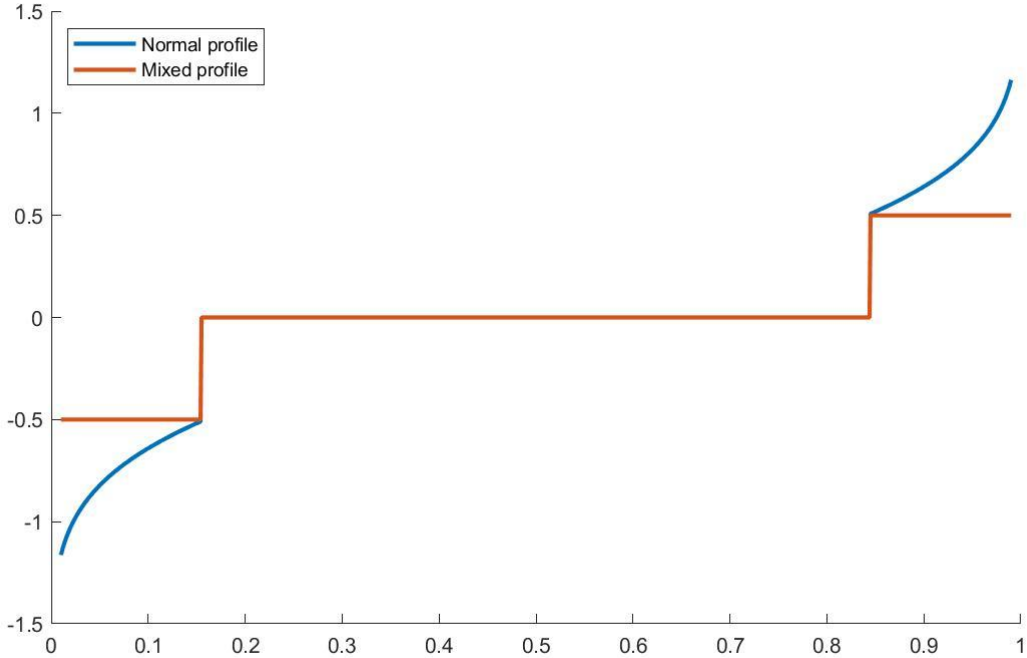

Figure 2: Quantile specific sparsity profiles

16

Design	β profiles
y_1	$\beta_0(p) = \beta_0 + F^{-1}_{N(0,1)}(p)$
y_2	$\beta_0(p) = \beta_0 + F^{-1}_{T(3)}(p)$
y_3	$\beta_0(p) = \beta_0 + F^{-1}_{N(0,1)}(p)$ $\beta_1(p) = \beta_1 + [I(u_t \leq F^{-1}_{N(0,1)}(0.15)) + I(u_t \geq F^{-1}_{N(0,1)}(0.85))]F^{-1}_{N(0,1)}(p)$
y_4	$\beta_0(p) = \beta_0 + F^{-1}_{N(0,1)}(p)$ $\beta_1(p) = \beta_1 + [-\frac{1}{2}I(u_t \leq F^{-1}_{N(0,1)}(0.15)) + \frac{1}{2}I(u_t \geq F^{-1}_{N(0,1)}(0.85))]$

Table 1: Summary of simulation setups

For all DGPs, we generate 50 synthetic data sets and set $K = 100$ (constant not included). The design matrix, X, is simulated using a multivariate Gaussian with zero mean and Toeplitz covariance structure, Ω, where $\Omega_{i,j} = \rho^{|i-j|}$ and $\rho = 0.5$. There are two sample sizes considered $T_i \in \{100, 500\}$.

Performance of the priors are gauged in terms of coefficient bias and measures of correct variable selection. Coefficient bias is calculated as the average (over Monte Carlo iterations) root mean deviation to the true quantile regression coefficients:

$$\frac{1}{iter}\sqrt{||\hat{\beta}(p) - \beta(p)||^2_2} \tag{30}$$

where *iter* is the number of Monte Carlo iterations, and $\hat{\beta}(p)$ is the mean posterior vector of the estimator.

Variable selection will be evaluated with Matthew's correlation coefficient (MCC) and the hit rate. While the MCC gives an overall measure of goodness of variable selection based on the true and false positives and negatives, the hit rate tells us how well we identify the true positives. The MCC and the hit rate can be calculated from the confusion matrix with the following formulas:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{31}$$

$$Hit\ rate = \frac{TP}{TP + FN} \tag{32}$$

where T stands for "True", F stands for "False", N stands for "Negative", and P stands for "Positive". The MCC and hit rate are commonly used for binary classification problems but

high-dimensional problems can be easily amended to fit the framework: non-zero coefficients have a value of 1 and zero coefficients have a value of 0. The same way coefficients shrunk to 0 receive a value of 0, while coefficients that are not 0 get a value of 1. This way a confusion matrix can be constructed from which the measures can be calculated. The MCC is between -1 (worst) and +1 (best), while the hit rate will range between 0 (worst) and 1 (best).

3.2   Monte Carlo Results

The coefficient bias results of the designs for a selection of quantiles are presented in table (2), the results for the MCC are shown in table (3), and the results of for the hit rates can be seen in table (4) [7]. The tables tell a very clear story: sparsification via the BQR_{SAVS} and BQR_{BIC} improves or yields similar coefficient bias compared to the unsparsified posteriors, while yielding more often than not better variable selection accuracy than the $SSVSBQR$, independent of sparsity, in-sample length and error-distribution. The BQR_{BIC} procedure stands out particularly in the sparse DGPs in which this sparsification procedure can reduce bias close to half for some quantiles in the larger sample setting compared to the unsparsified posterior and offers further bias reduction compared to the BQR_{SAVS}. Results including the adjustment term $\overline{Z}_p$, presented in the appendix shows that the results are largely unaffected as expected.

3.2.1   Baseline DGPs

Focusing first on bias for the unsparsified posterior, one can see from the first two columns of table (2), that the $HSBQR$ and $LBQR$ do similarly well in the sparse DGPs, while clearly outperforming the $SSVSBQR$. Although, all priors reduce bias with more data, the overall tendency doesn't change in the larger T=500 DGP. The largest difference among the SSVS, Lasso and horseshoe priors manifests itself in the block DGP's where the $HSBQR$ does consistently better, especially in the tails and in the smaller sample size of T=100. These findings corroborate the findings of Kohns and Szendrei (2020) that the adaptive shrinkage from global local priors are able to capture different sparsity blocks among the covariates and adapt to low data settings.

---

[7]Note that the results presented in tables (2)-(4) do not include the ALD adjustment term $\overline{Z}_p$.

	y_1					y_2					y_3					y_4				
	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
T=100																				
Sparse																				
$SSVSBQR$	0.206	0.210	0.217	0.212	0.221	0.227	0.209	0.195	0.201	0.238	0.235	0.241	0.227	0.253	0.243	0.193	0.219	0.223	0.222	0.194
$LBQR$	0.132	0.079	0.068	0.084	0.143	0.173	0.082	0.067	0.084	0.164	0.158	0.081	0.063	0.087	0.153	0.133	0.084	0.065	0.087	0.140
$LBQR_{SAVS}$	0.132	0.076	0.065	0.082	0.143	0.173	0.080	0.065	0.082	0.164	0.158	0.082	0.061	0.088	0.154	0.133	0.086	0.062	0.087	0.141
$LBQR_{BIC}$	0.132	0.076	0.065	0.082	0.143	0.173	0.080	0.065	0.081	0.164	0.158	0.081	0.064	0.088	0.153	0.133	0.085	0.062	0.086	0.141
$HSBQR$	0.119	0.094	0.085	0.097	0.127	0.172	0.099	0.087	0.097	0.176	0.165	0.104	0.075	0.100	0.157	0.131	0.099	0.077	0.099	0.131
$HSBQR_{SAVS}$	0.108	0.074	0.064	0.080	0.117	0.164	0.080	0.066	0.081	0.169	0.157	0.095	0.061	0.090	0.148	0.123	0.089	0.061	0.088	0.124
$HSBQR_{BIC}$	0.101	0.063	0.058	0.072	0.109	0.159	0.067	0.056	0.073	0.164	0.149	0.096	0.060	0.090	0.140	0.119	0.090	0.060	0.088	0.119
Block																				
$SSVSBQR$	0.320	0.325	0.325	0.336	0.336	0.339	0.333	0.322	0.328	0.360	0.358	0.371	0.362	0.359	0.353	0.337	0.340	0.353	0.339	0.332
$LBQR$	0.369	0.328	0.327	0.323	0.382	0.338	0.329	0.326	0.328	0.342	0.414	0.352	0.339	0.355	0.419	0.410	0.348	0.342	0.346	0.411
$LBQR_{SAVS}$	0.368	0.326	0.325	0.321	0.381	0.337	0.327	0.324	0.326	0.341	0.413	0.349	0.336	0.353	0.418	0.410	0.346	0.339	0.344	0.411
$LBQR_{BIC}$	0.368	0.326	0.326	0.321	0.381	0.337	0.327	0.324	0.326	0.341	0.413	0.349	0.336	0.352	0.418	0.410	0.346	0.339	0.344	0.410
$HSBQR$	0.250	0.248	0.251	0.253	0.244	0.260	0.240	0.244	0.249	0.272	0.288	0.293	0.290	0.286	0.278	0.258	0.267	0.269	0.263	0.255
$HSBQR_{SAVS}$	0.252	0.244	0.244	0.247	0.243	0.263	0.236	0.239	0.244	0.271	0.289	0.287	0.282	0.280	0.277	0.261	0.263	0.262	0.258	0.256
$HSBQR_{BIC}$	0.259	0.246	0.246	0.247	0.247	0.271	0.241	0.244	0.247	0.275	0.287	0.287	0.282	0.280	0.277	0.258	0.263	0.263	0.259	0.254
T=500																				
Sparse																				
$SSVSBQR$	0.133	0.095	0.087	0.095	0.131	0.165	0.096	0.083	0.095	0.170	0.159	0.084	0.069	0.087	0.156	0.134	0.088	0.073	0.083	0.132
$LBQR$	0.092	0.053	0.042	0.054	0.091	0.138	0.060	0.044	0.060	0.147	0.119	0.056	0.039	0.056	0.119	0.102	0.053	0.036	0.055	0.105
$LBQR_{SAVS}$	0.092	0.046	0.033	0.049	0.090	0.139	0.053	0.035	0.056	0.147	0.119	0.056	0.029	0.056	0.119	0.102	0.052	0.027	0.055	0.105
$LBQR_{BIC}$	0.092	0.041	0.029	0.046	0.090	0.138	0.048	0.032	0.053	0.147	0.119	0.056	0.031	0.056	0.119	0.101	0.051	0.030	0.054	0.105
$HSBQR$	0.099	0.072	0.062	0.074	0.103	0.152	0.074	0.061	0.078	0.153	0.137	0.066	0.047	0.068	0.134	0.110	0.065	0.047	0.067	0.113
$HSBQR_{SAVS}$	0.086	0.051	0.038	0.054	0.089	0.145	0.057	0.037	0.060	0.145	0.125	0.057	0.025	0.059	0.122	0.096	0.056	0.026	0.058	0.101
$HSBQR_{BIC}$	0.074	0.043	0.029	0.047	0.076	0.139	0.050	0.030	0.053	0.138	0.106	0.063	0.030	0.064	0.107	0.083	0.060	0.030	0.062	0.090
Block																				
$SSVSBQR$	0.150	0.101	0.094	0.104	0.145	0.177	0.101	0.091	0.101	0.180	0.190	0.097	0.079	0.098	0.180	0.162	0.097	0.084	0.099	0.156
$LBQR$	0.305	0.100	0.079	0.094	0.306	0.290	0.097	0.079	0.099	0.288	0.340	0.102	0.074	0.098	0.327	0.322	0.100	0.073	0.098	0.320
$LBQR_{SAVS}$	0.304	0.101	0.080	0.096	0.305	0.289	0.099	0.081	0.100	0.287	0.340	0.104	0.074	0.100	0.326	0.321	0.103	0.073	0.100	0.319
$LBQR_{BIC}$	0.304	0.100	0.079	0.094	0.305	0.289	0.097	0.079	0.099	0.287	0.340	0.102	0.073	0.098	0.326	0.321	0.100	0.073	0.098	0.319
$HSBQR$	0.142	0.104	0.094	0.105	0.143	0.173	0.103	0.091	0.104	0.175	0.189	0.099	0.083	0.097	0.184	0.155	0.100	0.084	0.099	0.154
$HSBQR_{SAVS}$	0.137	0.092	0.082	0.093	0.137	0.172	0.092	0.078	0.093	0.171	0.186	0.090	0.072	0.089	0.182	0.151	0.091	0.073	0.091	0.150
$HSBQR_{BIC}$	0.138	0.095	0.085	0.096	0.137	0.173	0.094	0.080	0.095	0.171	0.187	0.090	0.072	0.088	0.182	0.153	0.091	0.073	0.090	0.152

Table 2: Root Mean Coefficient Bias Results

	y_1					y_2					y_3					y_4				
	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
T=100																				
Sparse																				
$SSVSBQR$	0.243	0.275	0.269	0.292	0.325	0.266	0.273	0.305	0.304	0.353	0.225	0.188	0.180	0.184	0.217	0.262	0.210	0.172	0.205	0.266
$LBQR_{SAVS}$	0.589	0.710	0.789	0.781	0.660	0.570	0.712	0.797	0.791	0.680	0.428	0.581	0.558	0.546	0.432	0.442	0.551	0.545	0.538	0.399
$LBQR_{BIC}$	0.589	0.720	0.829	0.814	0.656	0.574	0.717	0.824	0.821	0.680	0.433	0.586	0.518	0.528	0.436	0.439	0.548	0.572	0.548	0.404
$HSBQR_{SAVS}$	0.390	0.532	0.560	0.550	0.513	0.382	0.523	0.554	0.556	0.503	0.350	0.422	0.442	0.415	0.353	0.394	0.426	0.429	0.411	0.396
$HSBQR_{BIC}$	0.552	0.720	0.782	0.772	0.695	0.541	0.722	0.793	0.789	0.689	0.468	0.495	0.542	0.488	0.472	0.510	0.504	0.539	0.493	0.506
Block																				
$SSVSBQR$	0.420	0.386	0.381	0.384	0.405	0.416	0.375	0.392	0.382	0.423	0.374	0.319	0.314	0.330	0.394	0.410	0.352	0.311	0.376	0.417
$LBQR_{SAVS}$	0.255	0.429	0.488	0.473	0.261	0.252	0.437	0.496	0.471	0.256	0.230	0.415	0.439	0.410	0.217	0.232	0.424	0.427	0.426	0.222
$LBQR_{BIC}$	0.255	0.424	0.478	0.471	0.261	0.254	0.437	0.492	0.471	0.256	0.229	0.411	0.436	0.409	0.217	0.232	0.421	0.424	0.423	0.222
$HSBQR_{SAVS}$	0.387	0.408	0.421	0.410	0.393	0.406	0.430	0.440	0.420	0.401	0.337	0.342	0.343	0.351	0.344	0.383	0.391	0.385	0.395	0.383
$HSBQR_{BIC}$	0.423	0.442	0.459	0.452	0.436	0.436	0.458	0.471	0.462	0.444	0.251	0.327	0.350	0.333	0.255	0.276	0.352	0.369	0.355	0.272
T=500																				
Sparse																				
$SSVSBQR$	0.275	0.308	0.305	0.309	0.314	0.294	0.311	0.320	0.324	0.318	0.284	0.331	0.304	0.334	0.293	0.301	0.316	0.291	0.348	0.325
$LBQR_{SAVS}$	0.704	0.745	0.755	0.789	0.790	0.740	0.707	0.758	0.792	0.809	0.571	0.797	0.718	0.773	0.593	0.589	0.783	0.757	0.776	0.571
$LBQR_{BIC}$	0.708	0.903	0.924	0.926	0.805	0.745	0.878	0.907	0.924	0.820	0.594	0.836	0.835	0.839	0.603	0.598	0.848	0.852	0.823	0.590
$HSBQR_{SAVS}$	0.401	0.538	0.563	0.539	0.470	0.393	0.570	0.579	0.555	0.470	0.395	0.665	0.635	0.642	0.416	0.422	0.658	0.634	0.636	0.441
$HSBQR_{BIC}$	0.711	0.899	0.899	0.901	0.817	0.713	0.909	0.901	0.900	0.837	0.699	0.810	0.847	0.805	0.722	0.724	0.819	0.845	0.797	0.747
Block																				
$SSVSBQR$	0.682	0.725	0.711	0.713	0.696	0.694	0.730	0.722	0.726	0.708	0.638	0.753	0.768	0.768	0.657	0.679	0.760	0.744	0.753	0.707
$LBQR_{SAVS}$	0.584	0.937	0.959	0.959	0.595	0.595	0.951	0.963	0.955	0.591	0.552	0.935	0.956	0.945	0.549	0.556	0.941	0.958	0.937	0.563
$LBQR_{BIC}$	0.584	0.930	0.956	0.950	0.595	0.595	0.937	0.957	0.946	0.591	0.552	0.925	0.952	0.938	0.549	0.556	0.935	0.950	0.930	0.563
$HSBQR_{SAVS}$	0.691	0.800	0.810	0.799	0.696	0.708	0.804	0.818	0.812	0.712	0.612	0.830	0.863	0.847	0.616	0.679	0.823	0.855	0.843	0.677
$HSBQR_{BIC}$	0.745	0.929	0.947	0.930	0.766	0.749	0.934	0.953	0.941	0.783	0.433	0.867	0.906	0.880	0.441	0.439	0.833	0.887	0.847	0.449

Table 3: MCC Results

	y₁					y₂					y₃					y₄				
	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95	0.05	0.25	0.50	0.75	0.95
T=100																				
Sparse																				
$SSVSBQR$	0.622	0.804	0.854	0.848	0.818	0.637	0.796	0.867	0.862	0.831	0.573	0.650	0.688	0.676	0.570	0.591	0.695	0.676	0.687	0.581
$LBQR_{SAVS}$	0.399	0.647	0.788	0.742	0.512	0.362	0.646	0.772	0.768	0.524	0.226	0.445	0.501	0.424	0.233	0.240	0.428	0.458	0.417	0.211
$LBQR_{BIC}$	0.400	0.568	0.759	0.712	0.509	0.364	0.597	0.747	0.755	0.524	0.227	0.410	0.333	0.353	0.233	0.235	0.385	0.392	0.378	0.214
$HSBQR_{SAVS}$	0.576	0.814	0.844	0.835	0.769	0.567	0.792	0.827	0.822	0.759	0.530	0.589	0.611	0.573	0.550	0.544	0.600	0.609	0.578	0.558
$HSBQR_{BIC}$	0.483	0.638	0.706	0.706	0.700	0.505	0.645	0.727	0.725	0.695	0.437	0.367	0.385	0.353	0.449	0.455	0.382	0.384	0.360	0.473
Block																				
$SSVSBQR$	0.575	0.652	0.682	0.666	0.593	0.570	0.650	0.699	0.669	0.601	0.557	0.620	0.638	0.634	0.557	0.573	0.645	0.652	0.660	0.571
$LBQR_{SAVS}$	0.120	0.325	0.387	0.372	0.129	0.124	0.333	0.406	0.369	0.122	0.113	0.332	0.370	0.329	0.105	0.121	0.334	0.359	0.340	0.111
$LBQR_{BIC}$	0.121	0.318	0.371	0.366	0.129	0.123	0.332	0.400	0.368	0.122	0.112	0.326	0.359	0.324	0.105	0.121	0.330	0.349	0.334	0.111
$HSBQR_{SAVS}$	0.554	0.651	0.686	0.672	0.583	0.569	0.665	0.692	0.671	0.585	0.548	0.631	0.649	0.639	0.555	0.566	0.648	0.666	0.655	0.568
$HSBQR_{BIC}$	0.362	0.414	0.443	0.439	0.389	0.369	0.423	0.445	0.436	0.387	0.686	0.654	0.632	0.664	0.694	0.718	0.702	0.685	0.709	0.721
T=500																				
Sparse																				
$SSVSBQR$	0.875	0.973	0.984	0.973	0.964	0.905	0.980	0.989	0.984	0.949	0.857	0.963	0.990	0.976	0.874	0.845	0.968	0.966	0.985	0.881
$LBQR_{SAVS}$	0.569	0.919	0.950	0.938	0.698	0.608	0.880	0.937	0.929	0.712	0.413	0.886	0.918	0.869	0.442	0.427	0.872	0.931	0.872	0.430
$LBQR_{BIC}$	0.573	0.851	0.886	0.896	0.700	0.611	0.819	0.874	0.888	0.712	0.410	0.848	0.733	0.838	0.428	0.417	0.858	0.786	0.852	0.430
$HSBQR_{SAVS}$	0.773	0.950	0.962	0.952	0.926	0.752	0.959	0.958	0.955	0.917	0.824	0.927	0.945	0.908	0.840	0.846	0.934	0.947	0.925	0.856
$HSBQR_{BIC}$	0.637	0.845	0.844	0.855	0.813	0.648	0.864	0.851	0.855	0.827	0.648	0.752	0.755	0.746	0.664	0.652	0.796	0.768	0.768	0.674
Block																				
$SSVSBQR$	0.969	1.000	1.000	0.999	0.983	0.966	0.999	1.000	0.999	0.980	0.938	0.991	0.998	0.992	0.938	0.957	0.992	0.998	0.992	0.962
$LBQR_{SAVS}$	0.500	0.974	0.996	0.986	0.499	0.509	0.980	0.993	0.984	0.500	0.473	0.963	0.991	0.964	0.476	0.482	0.964	0.992	0.967	0.487
$LBQR_{BIC}$	0.500	0.975	0.996	0.987	0.499	0.509	0.981	0.993	0.985	0.500	0.474	0.965	0.990	0.968	0.476	0.482	0.968	0.993	0.970	0.487
$HSBQR_{SAVS}$	0.964	0.999	0.999	0.997	0.971	0.964	0.999	0.999	0.998	0.972	0.932	0.986	0.995	0.989	0.927	0.955	0.988	0.995	0.989	0.950
$HSBQR_{BIC}$	0.951	0.986	0.991	0.985	0.956	0.955	0.989	0.993	0.990	0.954	0.968	0.984	0.994	0.986	0.965	0.983	0.988	0.994	0.989	0.980

Table 4: Hit Rate Results

21

With sparsification, one sees generally that the BQR_{SAVS} and BQR_{BIC} reduce bias nearly uniformly across DGPs. However, while the $LBQR$ and $HSBQR$ had similar bias in sparse DGPs, the $HSBQR_{BIC}$ has the lowest bias among all models, while trading blows with the $HSBQR_{SAVS}$ in the block DGPs. The $HSBQR$ coming ahead with sparsification indicates that signals are better identified through the horseshoe than the lasso prior.

The variable selection results in the first two columns of (3) and (4), mirror the bias results, while highlighting that the reason for the better performance from the BQR_{BIC} models comes from lower false positive rates. This can be deduced from the fact that while the hit rates in table (4) are higher for the other models, the MCC are lower. This also holds for the SSVS prior. Hence, both the $SSVSBQR$ and BQR_{SAVS} models tend to select too many noise variables.

3.2.2 Quantile specific Sparse DGPs

Similar to the baseline DGPs, the bias tendencies carry mostly also over to the quantile specific simulations, with the difference that the $HSBQR$ does exceptionally well in the tails of the block DGPs. This can be understood from the fact that adaptive shrinkage plays an even larger role with quantile specific sparsity, particularly in the more extreme tails which are data sparse (Koenker, 2005).

Adding sparsification, the BQR_{BIC} does better than BQR_{SAVS} in sparse DGPs, however, similarly to the baseline results, does not offer a systematic improvement over the BQR_{SAVS} in the block DGPs. This makes sense in the high-dimensional settings considered where capturing the quantile profile of the single covariate has a larger impact on fit in the sparse DGP than the block DGP, where many more covariates have non-zero coefficients.

In terms of variable selection, we find that the continuous priors considered outperform the $SSVSBQR$ in sparse DGPs and the $HSBQR_{BIC}$ does especially well in the tails, consistently having the highest MCC. The results in the block design deviate from the baseline results: here we see that the $SSVSBQR$ and BQR_{SAVS} models tend to do better than the $HSBQR_{BIC}$ in terms of the MCC, yet have similar or higher bias. Considering the hit rates, this difference indicates that the BQR_{BIC} now offers lower penalisation in order to capture the quantile effect from the quantile specific variable which the tick-loss function heavily tilts the predicted quantile towards. Allowing for proper ALD correction by inlcuding $\overline{Z}_p$ as in (21) for these DGPs, however, allows the $HSBQR_{BIC}$ to again outperform in the tails or come close to the best performer in terms of variable selection. Looking at the hit rates, it is suggestive that

the $\overline{Z}_p$ correction implies some additional penalisation particularly in the tails which increases sparsity and decreases the amount of false positives from the sparse parts. Nevertheless, in terms of bias, the correction induces very small differences and vanishes with more data.

All in all, the simulations have shown that sparsification often significantly lowers bias, or at worst holds bias stable, while improving variable selection compared to the SSVS. The BQR_{BIC} outperforms the BQR_{SAVS}, especially in sparse designs and in the tails where the extra flexibility of determining the penalisation from the data allows to capture the quantile profile. The results have also shown that although sparsification often helps the priors' performance, care should be taken in selecting the shrinkage prior apriori.

4 Growth at Risk application

The initial motivation for growth-at-risk, as introduced by Adrian et al. (2019), is to measure the downside risks, or vulnerabilities, to real GDP growth, which are captured as the response of forecasted conditional quantiles to changes in summary measures of financial conditions. Such models provide useful information for policymakers, as it has been empirically and theoretically shown that crises stemming from the financial sector carry the risk of creating negative feedback loops between the real economy and the financial system (Jordà et al., 2015). However, as seen from recent macroeconomic developments, vulnerabilities to GDP growth may stem from various, possibly a priori unknown parts of the economy which entails that a larger information set is needed. Recent contributions to the GaR literature indeed conduct real time nowcasting exercises (Carriero et al., 2020; Ferrara et al., 2020) and forecast density combination (Kohns and Szendrei, 2020; Korobilis, 2017) using larger macro data sets in combination with shrinkage priors in order to deal with parameter proliferation.

However, shrinkage priors, such as the very popular global-local priors, are continuous and are therefore hard to interpret in higher dimensions. For this application, similar to Kohns and Szendrei (2020), we are interested in creating quantile and combined density forecasts for real GDP growth based on the entire McCracken and Ng (2020) data base. The logic of combining forecasts of multiple quantiles is to allow for the full heterogeneity of the data's effects to permeate to the conditional forecast distribution, which has been shown in the previous literature to often outperform models focused on only location effects (Korobilis, 2017) [8]. However, unlike

---

[8]This step should also be natural from the standpoint that conditional quantiles characterise empirical CDFs which in turn provide direct links to the latent forecasted probability density function.

previous contributions, we apply the introduced sparsification methods to understand which variables drive the forecasting results in an effort to better communicate which subsets of the data set explain particular parts of the forecast distribution.

The McCracken and Ng (2020) database[9], including real GDP, consists of 248 macroeconomic time series, starting from 1959Q1 at a quarterly frequency and is updated monthly. We take the quarter-on-quarter growth rate of annualized real GDP as our dependent variable and all others as independent covariates. These variables include a wide variety of macroeconomic effects which cover real, financial as well as national accounts data. Since not all series start at 1959Q1, for the growth at risk application, only variables that are available from 1970Q1 on wards were considered which gives 229 explanatory variables.

Predictions for our quantile methodology are generated from a direct forecasting model:

$$y_{t+h} = x_t' \beta + \epsilon_{t+h} \tag{33}$$

for $t = 1, \cdots, T-h$, where h refers to the forecast horizon. We consider one- to four-quarter ahead forecast horizons ($h = 1, \cdots, 4$). Following Yu and Moyeed (2001), the quantile specific predictive distribution can be obtained by marginalising out the uncertainty of the quantile regression parameter posterior $p(\beta_p | \bullet)$:

$$p(y_{t+h,p} | y_{1:t}) = \int p(y_{1:t} | \beta_p) p(\beta_p | y_{1:t}) d\beta_p, \tag{34}$$

where in (34) $y_{1:t}$ refers to in-sample observations for $(y_1, \cdots, y_t)$. (34) is conveniently approximated through Monte Carlo integration, as posterior draws β_p^s are available from MCMC algorithms detailed in the appendix:

$$p(y_{t+h,p} | y_{1:t}) \approx \frac{1}{S} \sum_{s=1}^{S} p(y_{1:t} | \beta_p^s). \tag{35}$$

Note that for each MCMC iteration $p(\beta_p | \bullet)$, β_p^s, we obtain a sparsified vector through the BQR_{SAVS} and BQR_{BIC} procedure respectively, which are used again in (35) to generate separate predictive distributions. Predictive distributions are estimated for plain, BQR_{SAVS} sparsified and BQR_{BIC} sparsified posteriors of the $HSBQR$ and $LBQR$ as described in section (2.4) and the $SSVSBQR$ serves as a benchmark for the large data models. To further contrast our results to the well established 'vulnerable growth' model of Adrian et al. (2019), we estimate a Bayesian quantile regression with a fairly uninformative standard normal prior on the

---

[9]https://research.stlouisfed.org/econ/mccracken/fred-databases/

regression coefficients of only lagged real GDP and the NFCI index. This model is abbreviated in the tables below as ABG_{BQR}. Finally, to obtain a combined predictive density, we follow Gaglianone and Lima (2012) and Korobilis (2017) by estimating quantile predictive distributions for 19 equidistant quantiles, $p \in [0.05, 0.10, \cdots, 0.90, 0.95]$, sort the stacked $vec(S \times 19)$ vector, and smooth this stacked vector via a Gaussian kernel[10]. Although stacking quantile predictive distributions will inevitably create some quantile crossing, the combination of multiple neighbouring quantiles helps in providing probability density in tail regions, which, due to the tick-loss function in the likelihood, are statistically data sparse relative to the median (Koenker, 2005).

To provide direct reference also to the original work of Adrian et al. (2019), we construct quantile and density forecasts based on their two-step estimation procedure where conditional quantiles are smoothed via the skew-t-distribution (Azzalini and Capitanio, 2003).[11] Forecasts for all models are computed on an expanding window basis where the initial in-sample period uses the first 50 observations of the sample, which makes for 149-h forecast windows.[12]

To evaluate forecasts, we employ both measures of overall density accuracy and quantile specific fit. For the overall density, we consider two strictly proper scoring rules in the sense of Gneiting and Raftery (2007), the average log-predictive density score (LPDS) and average cumulative rank probability score (CRPS). Define T_0 as the initial in-sample period, taken to be (1970Q1-1982Q2), and T as the final forecast period (Q42019), then the LPDS and CRPS are calculated as follows:

$$
\begin{aligned}
LPDS &= \frac{1}{T - T_0 - h + 1} \sum_{t=1}^{T-T_0-h+1} log\Big(\frac{1}{S} \sum_{t=1}^{S} p(y_{T_0+t+h-1} | y_{1:T_0+t})\Big) \\
CRPS &= \frac{1}{T - T_0 - h + 1} \sum_{t=1}^{T-T_0-h+1} |y_{T_0+t+h-1} - \hat{y}_{T_0+t+h-1}^A| - \frac{1}{2}|\hat{y}_{T_0+t+h-1}^A - \hat{y}_{T_0+t+h-1}^B|,
\end{aligned}
\tag{36}
$$

where $\hat{y}_{T_0+t+h-1}^A$ and $\hat{y}_{T_0+t+h-1}^B$ are two independent draws from the forecast distribution. To facilitate the discussion, the objective is to maximise the LPDS and minimise the CRPS. Since the LPDS and CRPS do not elucidate forecast calibration across different parts of the overall forecast distribution, we further make use of Rossi and Sekhposyan (2019)'s calibration test. If a model is well calibrated, then the forecasted CDF of the probability integral transforms

---

[10]This can be implemented for example via the "kdensity" function in Matlab

[11]Results for the skew-t big data models are reported in the appendix

[12]The ABG model contains 135-h rolling forecasts due to the availability of the NFCI index.

(PITs), $g_{t+h} = \int_0^{y_{t+h}} = p(u|y_{t+h})du$, should be statistically indistinguishable from a 45 degree line (Diebold et al., 1998). Following Rossi and Sekhposyan (2019), we report confidence bands around the 45 degree line to account for sampling uncertainty [13].

Lastly, as a measure of local fit, we make use of the quantile weighted CRPS (qwCRPS) to asses tail forecast performance. This metric is based on the quantile score (QS), as proposed by Gneiting and Ranjan (2011) and is calculated as follows: let $\hat{y}_{t+h,p}$ denote the expected quantile for real GDP growth h steps ahead, then the QS is computed as:

$$QS_{t+h,p} = (y_{t+h} - \hat{y}_{t+h,p}) \times (p - \mathbb{I}(y_{t+h} - \hat{y}_{t+h,p})). \tag{37}$$

The qwCRPS is thus given by:

$$qwCRPS_{t+h} = \int_0^1 w_p QS_{t+h,p}dp, \tag{38}$$

where w_p denotes a weighting scheme to evaluate specific parts of the forecast density. Since quantile regression is often used to capture downside risks, we use $w_p = (1 - p)^2$, as suggested by Gneiting and Ranjan (2011) for left tail evaluation.

4.1 Forecast Density Evaluation Results

Point and density scoring results for all forecast horizons are reported in table (5), where stars indicate statistical significance as per the Diebold et al. (1998) test and calibration results are graphed in figure (4). The density scores show generally that there is a clear benefit to big data, as average LPDS are often several order of magnitudes higher, and CRPS significantly lower, especially for the $HSBQR$ compared to either of the ABG models. The cumulative log-scores over time in graph (3) highlight that the vulnerable growth models underperform particularly during the financial crisis where the realisations of GDP growth seem to fall far into the tails of their forecast distributions, thereby being heavily penalised [14] [15]. In line with

---

[13]The confidence bands of Rossi and Sekhposyan (2019) should be taken as general guidance, as strictly speaking, they are derived using a rolling window of estimation, while we use an expanding window. As recommended by Rossi and Sekhposyan (2019), uncertainty intervals for higher order forecasts $(h > 1)$, are computed using bootstrapping.

[14]The large performance discrepancy compared to the CRPS, therefore makes sense given that the LPDS penalises low mass event very heavily.

[15]Fitting the skew-t distribution to the ABG model, markedly improves overall density scoring results and calibration at the one-quarter-ahead forecast horizon, where gains are largest for the middle to left parts of the density, as seen in figure (4). This gives an indication that the low-dimensional information set cannot provide

	MSFE	LPDS	CRPS	qwCRPS	MSFE	LPDS	CRPS	qwCRPS
				$h=1$				$h=2$
ABG_{BQR}	0.504	-11.607	0.313	1.049	0.404	-7.870	0.254	0.891
$ABG_{BQR-Skt}$	0.519	-1.727	0.301	1.026	0.426	-4.102	0.277	0.970
$SSVSBQR$	0.660***	-0.955***	0.337***	1.492	0.433***	-0.622**	0.235***	1.106
$HSBQR$	0.533*	-0.858***	0.296***	0.988*	0.411***	-0.558**	0.222***	0.804*
$HSBQR_{BIC}$	0.519	-0.895***	0.291***	0.948**	0.422***	-1.064**	0.227***	0.794*
$HSBQR_{SAVS}$	0.536	-0.976***	0.296***	0.968	0.468**	-2.856*	0.254***	0.855
$LBQR$	0.551*	-4.603**	0.305***	1.091	0.441***	-2.522*	0.238***	0.863
$LBQR_{BIC}$	0.552	-4.628**	0.305***	1.088	0.441***	-2.517*	0.238***	0.863
$LBQR_{SAVS}$	0.572	-4.832***	0.316***	1.085	0.467***	-3.006*	0.250***	0.859
				$h=3$				$h=4$
ABG_{BQR}	0.378	-6.825	0.237	0.847	0.351	-4.387	0.235	0.844
$ABG_{BQR-Skt}$	0.397	-1.006	0.249	0.857	0.367	-1.252	0.251	0.875
$SSVSBQR$	0.391***	-0.797	0.219***	1.027	0.394***	-0.796	0.221***	1.038
$HSBQR$	0.390***	-0.649	0.215***	0.833	0.376**	-0.657	0.213***	0.773
$HSBQR_{BIC}$	0.397***	-1.589	0.219***	0.817	0.374**	-0.933	0.213***	0.761
$HSBQR_{SAVS}$	0.435*	-2.913	0.237***	0.830	0.391	-3.605	0.216***	0.757
$LBQR$	0.422***	-2.820	0.230***	0.872	0.382***	-3.498	0.216***	0.809
$LBQR_{BIC}$	0.423***	-2.920	0.230***	0.870	0.383***	-3.603	0.216***	0.809
$LBQR_{SAVS}$	0.442**	-4.323	0.239***	0.860	0.409**	-4.949	0.230***	0.823

Table 5: Forecast Evaluation Results

the simulation results, we find that the $HSBQR$ outperforms both the $SSVSBQR$ and $LBQR$, and that the BQR_{BIC} is clearly preferred for sparsification compared to the BQR_{SAVS}, where BQR_{BIC} sparsification often increases performance, as gauged by the CRPS, or only slightly worsens it as per the LPDS. Also similar to the simulations, the quality of sparsification is again dependent on the quality of shrinkage. This can be seen from the generally worse density scores of the $LBQR$ models. Interestingly, for all models considered, forecast accuracy increases with the forecast horizon. This may stem the fact that changes in parts of the economy only affect aggregate output with a lag. Similar to much of the previous macro forecasting literature (Stock and Watson, 1998; Banerjee and Marcellino, 2006; Edge et al., 2010) we also find that

---

enough information between quantiles, for which the skew-t provides extra spread. This in particular aids log-score penalisation. We find that applying the skew-t to the big-data models, doesn't not change performance qualitatively vis-a-vis ABG models. The table with the skew-t results are provided in the appendix.

in terms of point predictions, simple autoregressive models such as the ABG models are hard to beat.

In terms of capturing downside risks, the qwCRPS results show that the $HSBQR_{BIC}$ consistently produces the best left tail performance, and significantly outperforms the ABG models at one-quarter and two-quarters ahead. The BQR_{BIC} sparsification procedure in fact increases tail performance both for the $HSBQR$ and $LBQR$.

Calibration based on PITs are able to shed more light on the differing performances. Figure (4) shows the models' PIT CDFs with confidence intervals computed as suggested by Rossi and Sekhposyan (2019). Corroborating the good density scores of the $HSBQR$ variants and the $SSVSBQR$, figure (4) shows that these models are also well calibrated, especially in the left tail at all horizons. The ABG models in turn tend to underestimate the lower tail which helps explain the bad log-score performance during the financial crisis in figure (3). Among the big-data models, the general tendency from the calibration results is that the $SSVSBQR$ slightly overpredicts the left tail and underpredicts the right tail, while the $LBQR$ tends to underpredict both tails. The $HSBQR$ in turn falls somewhere in the middle. In line with the density scores, sparsification via the BQR_{BIC} is able to preserve calibration, while the BQR_{SAVS} tends to overshrink. This is especially prevalent in the left to middle quantiles of the $HSBQR_{SAVS}$ at one-year ahead predictions, where more observations fall below those quantiles.
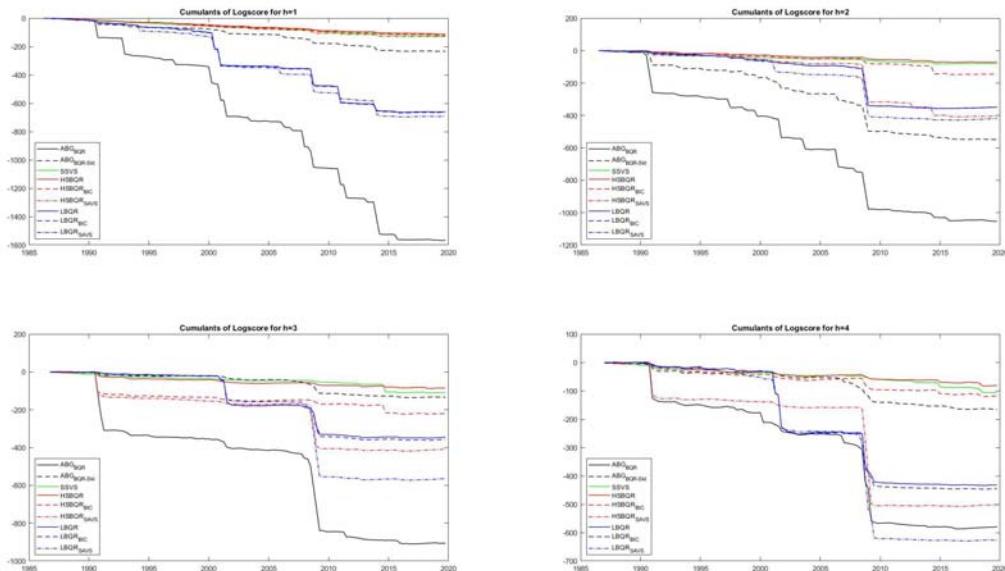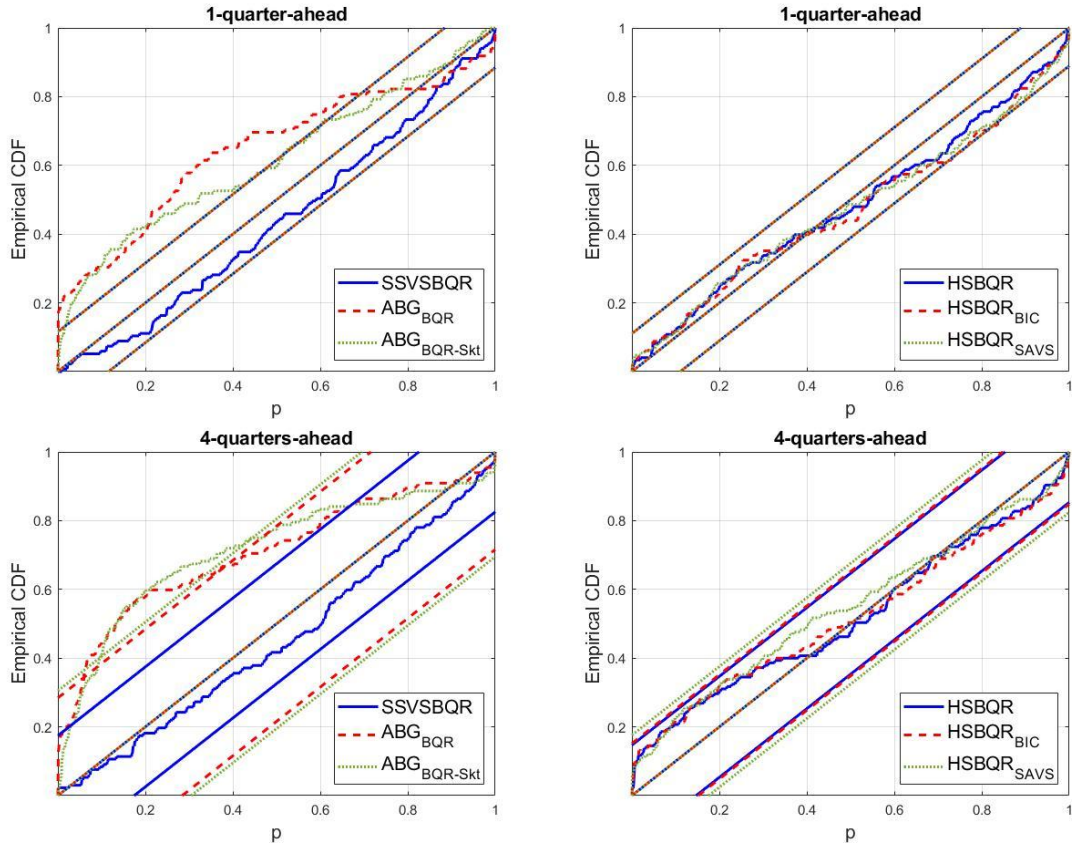

Figure 3: Cumulative log-scores.

Figure 4: Rossi and Sekhposyan (2019) calibration test.

4.2 Quantile specific variable inclusion

The results from the density and calibration tests have shown that the BQR_{BIC} variable selection procedure is able to preserve the fit of the unsparsified forecast distributions, and even improve left tail accuracy, which gives credibility to the fact that most salient features of the models have been captured despite sparsification. In order to analyse variable importance both over time and across quantiles, we provide in figures (5), (6), (7), and (9) heatmaps of variable inclusion probability for each prior and sparsification variant. While figure (5), (6), and (7) give a general static picture of variable selection across quantiles, averaged over time, figure (9) shows variable inclusion probabilities for left tail (average of 5th-15th quantiles), right tail (average of 85th to 95th quantiles) as well as the middle regions (average of 45th-55th quantiles) across time to showcase time variation in variable selection. Since it is of interest to understand differences in variable selection compared to widely used conditional mean models, we present at the bottom of figures (5) and (7) variable inclusion probabilities with each of the priors considered

for the Bayesian quantile regression, however applied to a normal observational likelihood[16]. To make the graphs more easily interpretable, we present the covariates on the horizontal axis in terms of general variable groupings given by the FRED-QD data base. For example, series in the NIPA group refer to "National Income and Product Account" information that spans information such as aggregate consumption, investment and government expenditures. Exact variable definitions for each group can be found in the appendix of McCracken and Ng (2020).

4.2.1 Variable Inclusion Across Quantiles

For all quantile models under consideration, figures (5) - (7) clearly show that there is substantial quantile varying sparsity. For instance, one can see for the $HSBQR_{BIC}$ model in figure (5) at one-quarter ahead, that "Employment and Unemployment" data and "Household Balance Sheets" information are more important for the right tail, while playing only a marginal role in the left tail. At one-year ahead forecasts, variable importance shifts in the left tail toward "Interest Rates" and "Consumer Sentiment", while the right tail has higher inclusion probability for "Money and Credit" and "Non-Household Balance Sheets". The former information can be interpreted as important sectors that predict both the 2001 as well as great financial crisis. As expected, middle quantiles draw information from either tails, displaying variable inclusion patterns which overlap both with the left and right tail.[17]

A commonality among, especially, the one-quarter ahead forecasts is the high inclusion probability of NIPA variables at all quantiles. Since the NIPA variable group contains national account information which gets aggregated via the national account identity to GDP, it is natural to interpret this variable category as capturing information inherent in lagged GDP. In fact taking the first principal component of the four most included NIPA variables by the $HSBQR_{BIC}$ yields a series that is highly correlated with GDP growth. The fact that these series are included for the majority of the quantiles presents a departure from the previous literature such as Adrian et al. (2019) who find that current economic conditions, as captured by lagged GDP only forecast the median of the conditional distribution.

Contrasting these results to conditional mean models, we find that these tend to select variable groups which have high inclusion probability not only in the median, but also the tails. To underscore this, figure (8) plots the correlation coefficient between left, right and

---

[16]Details on the conditional mean models for the different priors is provided in the appendix.

[17]Note that variable groups overlapping need not signify that the fitted quantiles contain the same information since coefficients may vary across quantiles.
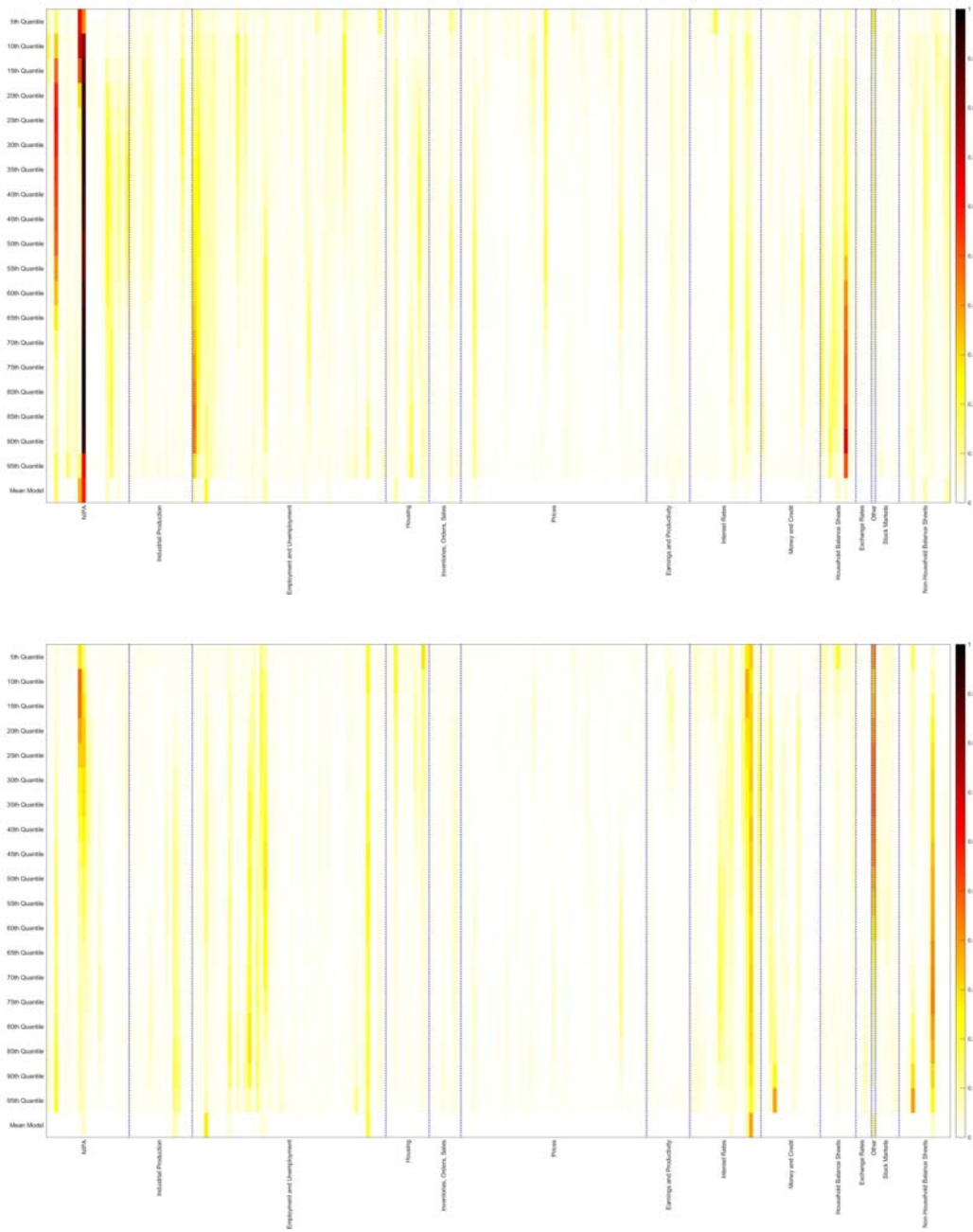
Figure 5: Average inclusion probability for each quantile of the $HSBQR_{BIC}$ for h=1(top) and h=4 (bottom). HS with SAVS for mean model included for reference
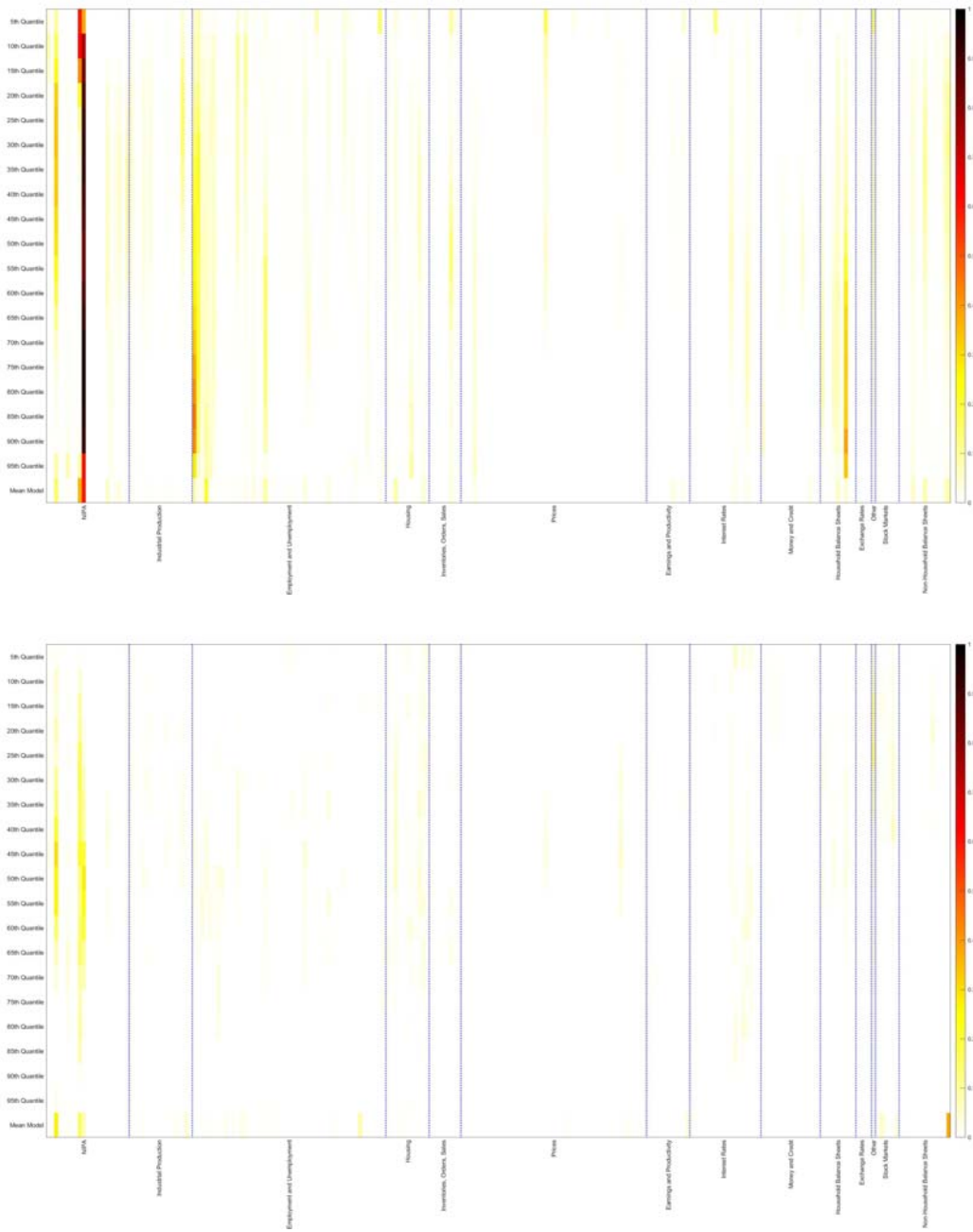
Figure 6: Average inclusion probability for each quantile of the $HSBQR_{SAVS}$ (top) and the $LBQR_{SAVS}$ (bottom) for h=1
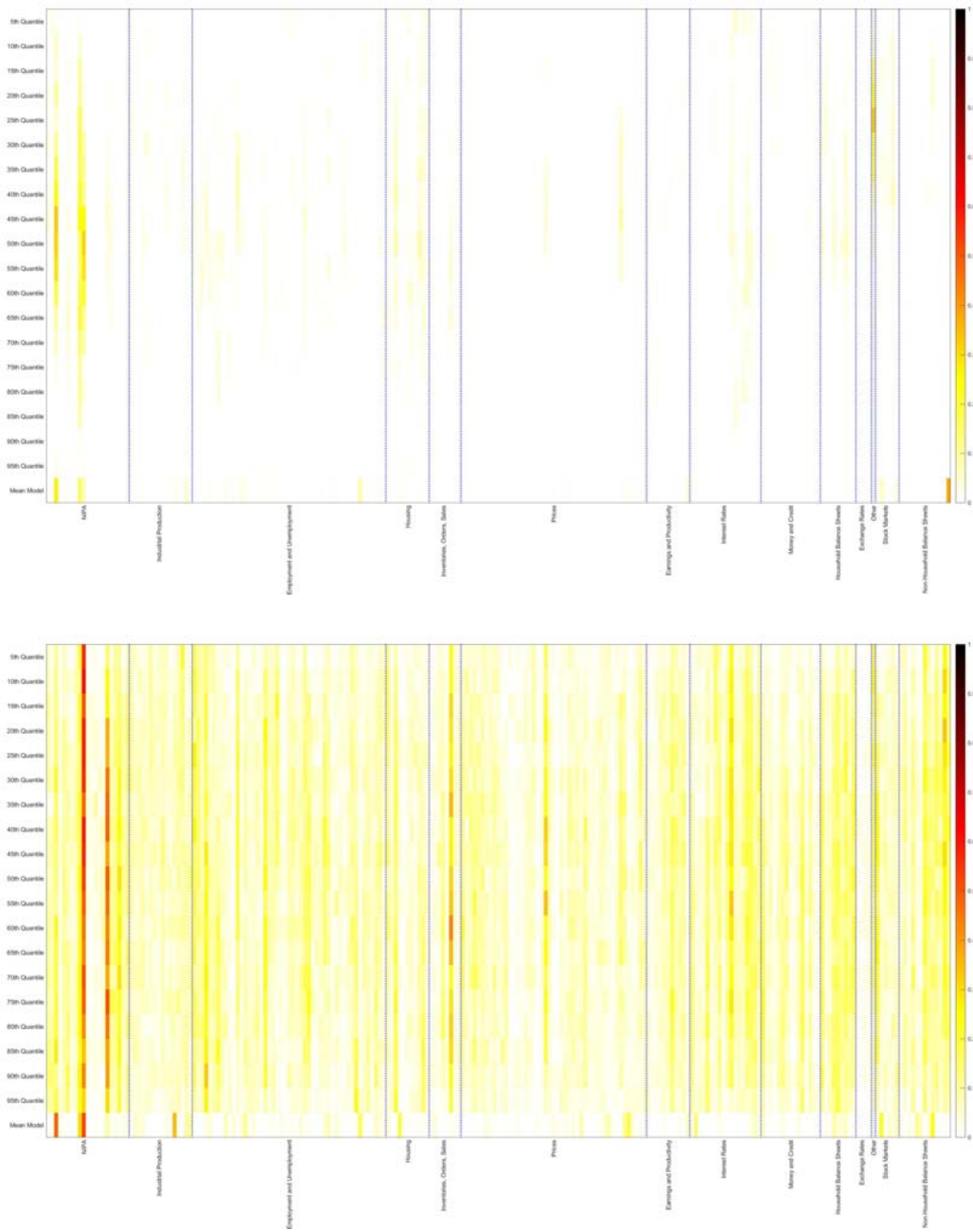
Figure 7: Average inclusion probability for each quantile of the $LBQR_{BIC}$ (top) and the $SSVSBQR$ (bottom) for h=1

middle quantiles of the $HSBQR_{BIC}$ in a rolling 30 quarter window with point forecasts of the conditional mean horseshoe prior model with SAVS sparsification. While this figure is not a conclusive test of which quantiles influence the mean the most, it does highlight that correlation is highest with the median and right tail during tranquil times, while in recession periods, as indicated by grey bars, correlation with the lower quantiles increases, hence signifying that information spills over from the left tail to the mean. This figure, also highlights that there are larger differences in correlation between the upper, middle and lower quantiles in the one-year ahead forecasts after the 2001 crisis compared to the one-quarter ahead forecasts. This may reflect the fact that interest rate and credit information propagate only with a lag to aggregated economic activity (Romer, 2018) and therefore are increasingly picked up in higher order forecasts after periods in which monetary policy had changed a lot to counteract the recession. Lastly, in line with the recent macroeconomic forecasting literature such as Giannone et al. (2017), we also find that there is evidence for model uncertainty which can be deduced from the amount of yellow shaded areas at all forecasting horizons.

Comparing variable selection among priors and sparsification techniques, some further interesting patterns emerge. First, the BQR_{SAVS} algorithm with fixed penalisation induces much sparser models than under the more adaptive BQR_{BIC} penalisation for both continuous shrinkage priors, and in doing so, eliminates near all model uncertainty, as seen by the large white spaces in figure (6). Notably though, variables which are identified as strong signals by the $HSBQR_{BIC}$, tend to also appear with higher inclusion probability in the $HSBQR_{SAVS}$. As seen in the previous section, this heavy sparsification, however, comes at a large penalty on density scores and calibration, hurting particularly tail forecasting performance, as underscored in figure (3) around the financial crisis. Second, in line with the simulation results, we see also in the empirical application that the $LBQR$ tends in general to exert stronger shrinkage than the $HSBQR$, resulting in far sparser models than under the horseshoe prior, while the $SSVSBQR$ tends to select larger models (lower panel of figure 7). This is corroborated by table (6) which shows average model sizes of selected estimators. Lastly, the $SSVSBQR$ prior also induces far greater variable inclusion uncertainty, as the heatmap shows denser yellow regions compared to the $HSBQR_{BIC}$. Nevertheless, variables with the highest inclusion probabilities tend to overlap with those selected by the $HSBQR_{BIC}$.

The fact that the $SSVSBQR$ prior increases model uncertainty without outperforming the $HSBQR$ and $HSBQR_{BIC}$ might indicate that the discretisation of the model-space along with the less flexible slab distribution employed by the $SSVSBQR$ prior induces higher uncertainty
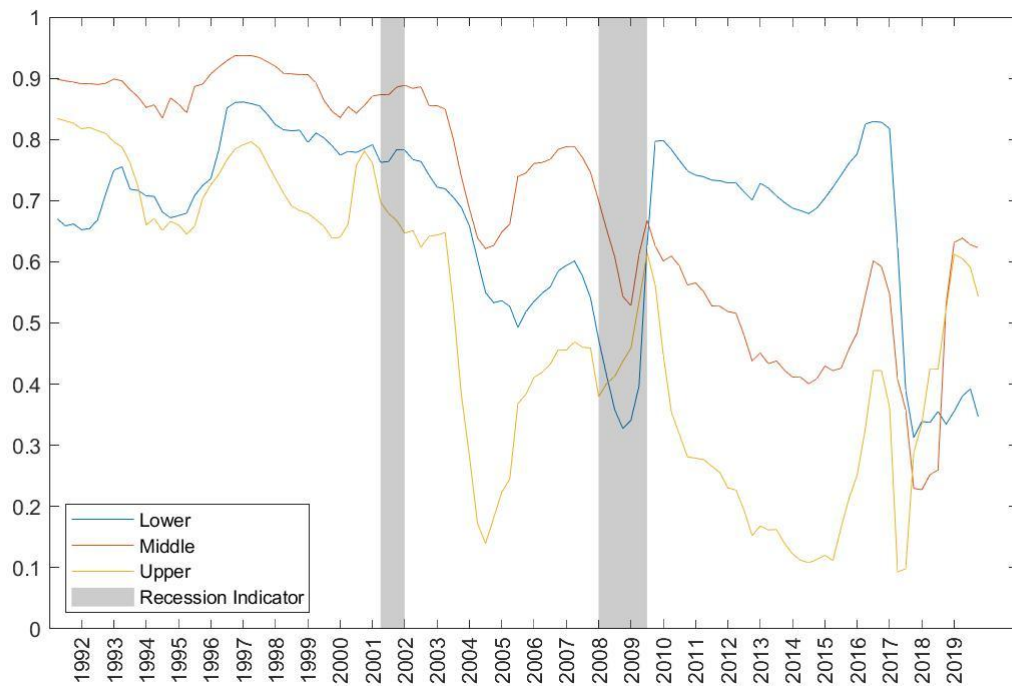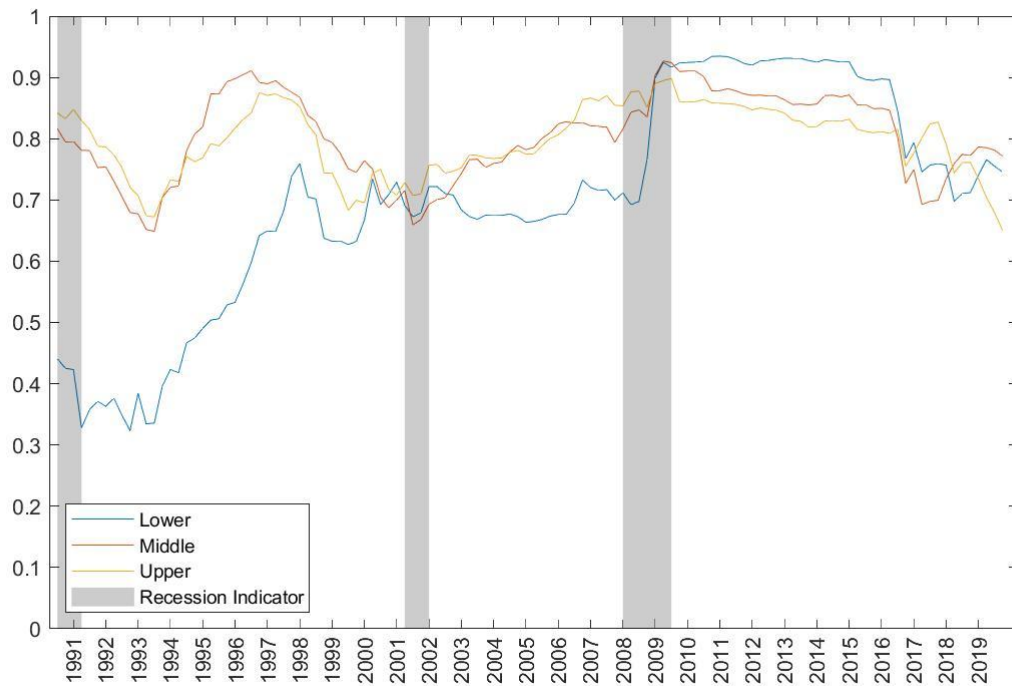
Figure 8: Correlation coefficients computed on a rolling 30 quarter window

	Left	Mid	Right	Left	Mid	Right
		h=1			h=4	
$SSVSBQR$	13.023	19.195	14.668	9.063	13.024	8.505
$HSBQR_{BIC}$	9.003	9.640	8.202	7.474	7.314	7.148
$LBQR_{BIC}$	1.541	3.055	1.107	1.439	2.748	1.119

Table 6: Average Model sizes of the different models

over variable inclusion, especially with correlated data. We leave exploration of whether global-local priors can illicit different answers to the 'illusion of sparsity' found by Giannone et al. (2017) for future investigation.

4.2.2 Variable Inclusion Across Time

While variable inclusion averaged across time revealed that there is support for quantile varying sparsity in the high dimensional GaR, figures (9) and (10) clearly show that variable selection has substantial time variation as well. Taking the $HSBQR_{BIC}$ as the benchmark, we can see in figure (9) that the time variation in variable selection is in line with what one would expect given the economic history of the US. There are two periods that stand out at one-quarter ahead forecast: the financial crisis of 2008 and the mid to late 80's. Around the financial crisis the left tail increasingly includes employment and financial data such as interest rates and non-household balance sheet information. During the mid-to late 80's, a period which is often referred to as the Volcker period in the macro literature[18], price variables have high inclusion probability. The middle and right tail share many common variable inclusion patterns over time, where employment and housing data impact the selected models during the 2000's and the financial crisis.

Comparing to the $LBQR$ and $SSVSBQR$ in figure (10), the contrasts in variable selection over time are even starker. The $LBQR_{BIC}$ produces very sparse and the $SSVSBQR$ models produce denser models, and the inclusion probabilities for both models are very erratic. Seldom is a variable included with high probability for either priors for more than a quarter which can

---

[18]When Paul Volcker became chairman of the Federal Reserve Board in 1979, the annual average inflation rate in the United States was 9%. The Federal Reserves was able to bring inflation down to 4% by the end of 1983. However, this disinflation came at a cost as during this period, the US experienced two recessions that are attributed to disinflationary monetary policy (Goodfriend and King, 2005).

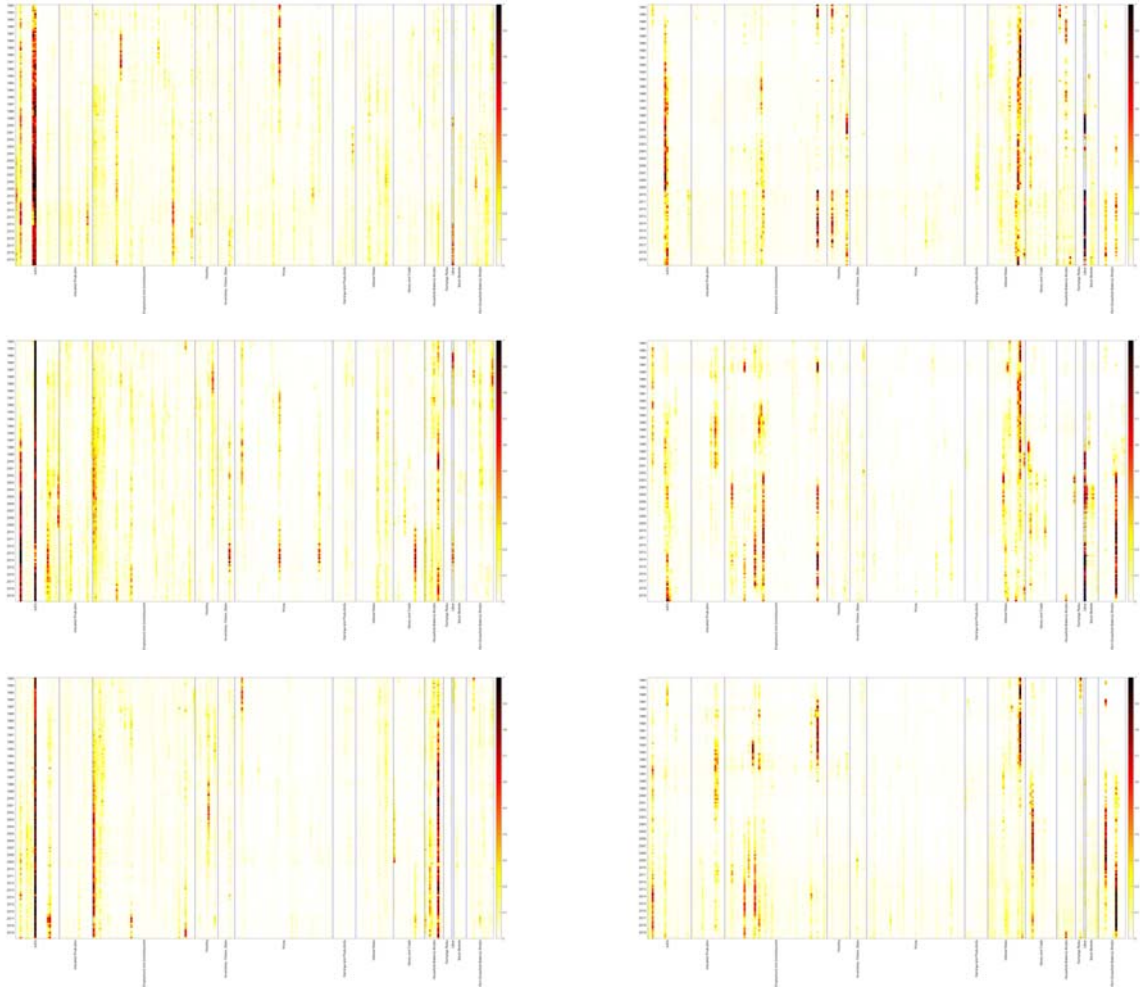Figure 9: Average inclusion probability each year for the $HSBQR_{BIC}$ for its 3 left quantiles (top), 3 central quantiles (middle), and 3 right quantiles (bottom) for h=1 (left) and h=4 (right)
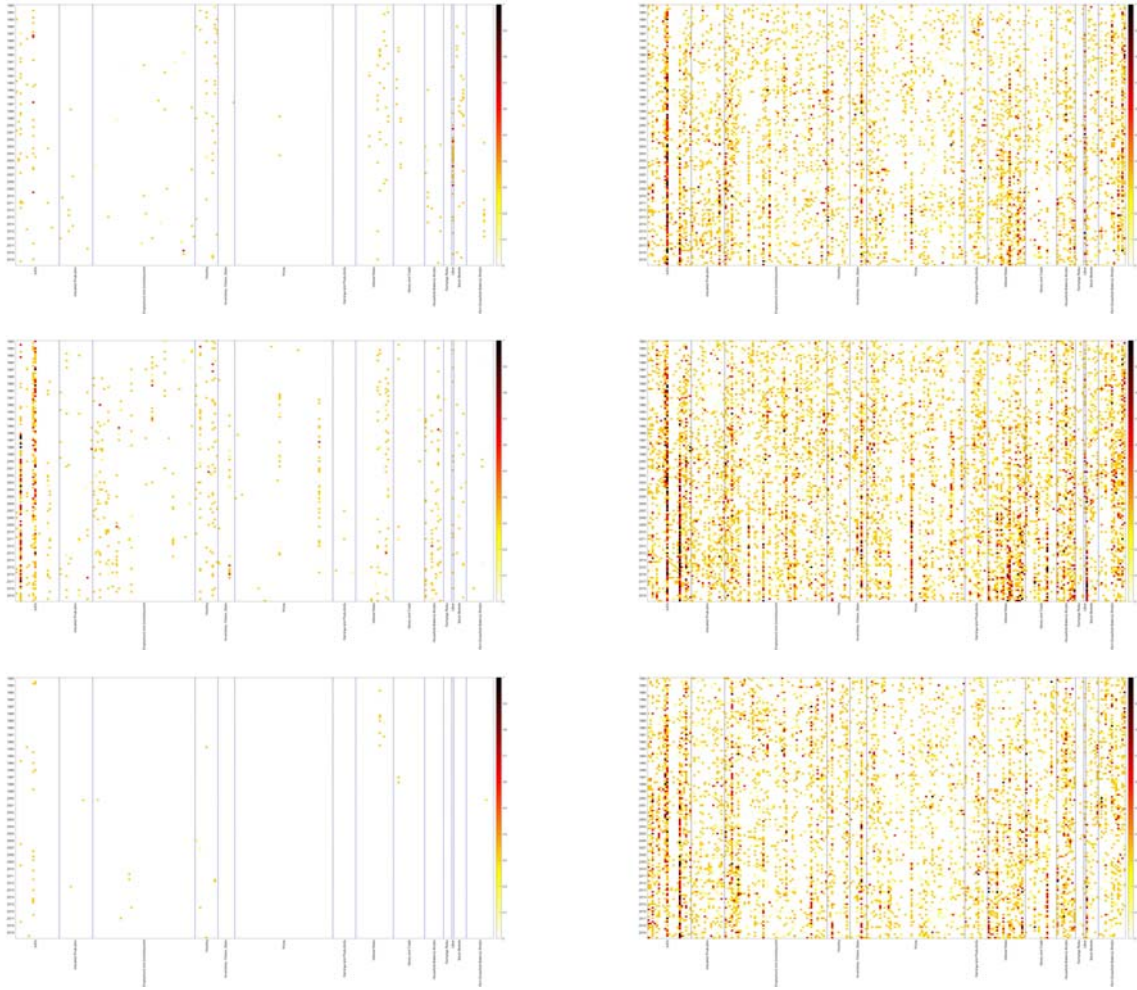
Figure 10: Average inclusion probability each year for the $LBQR_{BIC}$ (left) and $SSVSBQR$ (right) for their 3 left quantiles (top), 3 central quantiles (middle), and 3 right quantiles (bottom) for h=1

be seen from the very dotted heatmaps. The $HSBQR$ models on the other hand produce far smoother inclusion patterns which indicates that variables with high inclusion probability impact forecasts for longer periods. While in figure (10) we present $LBQR_{BIC}$, the same is true for $LBQR_{SAVS}$.

The general finding that there is time variation in the importance of variables for aggregate macroeconomic time-series has already been extensively studied for conditional mean models such as in Koop and Korobilis (2012), Billio et al. (2013) and McAlinn and West (2019). Our findings motivate that similar methods be also considered for the Bayesian quantile regression moving forward. It is however not surprising that quantile regression models are more sensitive to varying variable importance, even with static parameters. Namely, as the time-series accumulates, crisis periods identify new information for the tails which the tick-loss function will heavily tilt the fitted quantiles toward. This explains the large increase in the importance of variable groups which help predict the 2000 and financial crisis, especially for the left tail and the median.

5   Conclusion

This paper has, in the spirit of Hahn and Carvalho (2015) and Ray and Bhattacharya (2018), proposed to conduct variable selection for continuous shrinkage priors within the Bayesian Quantile Regression by decoupling shrinkage from sparsity as derived from a Bayesian decision theoretic perspective. The resultant easy to implement SAVS procedure for the BQR selects variables on a quantile specific basis where the degree of sparsity is estimated from the data via the quantile BIC in order to allow for heterogeneous levels of penalisation across the conditional distribution.

Applying the proposed method to both simulated experiments and an empirical application, we have found that the qBIC augmented version, the BQR_{BIC}, retains or even improves the fit of the unsparsified posterior, while giving easy to interpret variable selection results.

Simulated experiments have tested a variety of sparsity patterns and error distributions, both for quantile constant and quantile varying coefficient vectors and have found that sparsification often significantly lowers bias, or at worst holds bias stable, while improving variable selection compared to the traditional Bayesian variable selection priors such as the SSVS. The proposed selection procedure via the qBIC outperforms the BQR_{SAVS} especially in sparse DGPs and more generally in the tails where the extra flexibility of determining the penalisation from

the data allows to capture the quantile profile.

The high dimensional growth-at-risk application to the US verifies the findings of the Monte Carlo exercise, namely that the BQR_{BIC} retains density forecast performance and even improves downside risk estimation. Compared to the benchmark 'vulnerable growth' model of Adrian et al. (2019), the big data models provide significant gains in density forecasts. In particular, the proposed sparsification techniques applied to the $HSBQR$ provide well calibrated densities with much improved left tail performance. Variable selection results of the proposed methods revealed that there is substantial quantile varying sparsity for US macroeconomic data. As intuition would suggest variable groups which predict crises periods well, such as interest rate and housing data appear with higher inclusion probability in the left tail while right and middle quantiles tend to be affected by many overlapping variable groups.

Contrasting these results to variable selection of more commonly used conditional mean models, we find that mean models tend to select variable groups which have high inclusion probability not only in the median, but also the tails. This highlights the importance of quantile models as it allows us to decompose which variables impact different parts of the distribution. The general finding that there is time variation in the importance of variables for aggregate macroeconomic time-series seems to hold true for quantile models as well. We feel that it is a worthy endeavour to further research methods that allow for quantile as well as time variation for variable selection, to gain a better understanding of which variables are important for the different quantiles across time.

References

Adams, P., T. Adrian, N. Boyarchenko, and D. Giannone (2020). Forecasting macroeconomic risks.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Alhamzawi, R. and K. Yu (2013). Conjugate priors and variable selection for bayesian quantile regression. *Computational Statistics & Data Analysis 64*, 209–219.

Alhamzawi, R., K. Yu, and D. F. Benoit (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling 12*(3), 279–297.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 367–389.

Banerjee, A. and M. Marcellino (2006). Are there any reliable leading indicators for us inflation and gdp growth? *International Journal of Forecasting 22*(1), 137–151.

Bhadra, A., J. Datta, N. G. Polson, B. Willard, et al. (2019). Lasso meets horseshoe: A survey. *Statistical Science 34*(3), 405–427.

Bhattacharya, A., A. Chakraborty, and B. K. Mallick (2016). Fast sampling with gaussian scale mixture priors in high-dimensional regression. *Biometrika 103*(4), 985–991.

Bhattacharya, A., D. Pati, N. S. Pillai, and D. B. Dunson (2015). Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association 110*(512), 1479–1490.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics 177*(2), 213–232.

Carriero, A., T. E. Clark, and M. G. Marcellino (2020). Nowcasting tail risks to economic activity with many indicators.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

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

Cross, J. L., C. Hou, and A. Poon (2020). Macroeconomic forecasting with large bayesian vars: Global-local priors and the illusion of sparsity. *International Journal of Forecasting 36*(3), 899–915.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Edge, R. M., M. T. Kiley, and J.-P. Laforte (2010). A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a dsge model. *Journal of Applied Econometrics 25*(4), 720–754.

Ferrara, L., M. Mogliani, and J.-G. Sahuc (2020). High-frequency monitoring of growth-at-risk.

Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters*, 109126.

Friedman, J., T. Hastie, H. Höfling, and R. Tibshirani (2007). Pathwise coordinate optimization. *The annals of applied statistics 1*(2), 302–332.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software 33*(1), 1.

Friedman, J., T. Hastie, R. Tibshirani, et al. (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.

Gaglianone, W. P. and L. R. Lima (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking 44*(8), 1589–1607.

George, E. I. and R. E. McCulloch (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

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

George, E. I., D. Sun, and S. Ni (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics 142*(1), 553–580.

Giannone, D., M. Lenza, and G. E. Primiceri (2017). Economic predictions with big data: The illusion of sparsity.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association 102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Goodfriend, M. and R. G. King (2005). The incredible volcker disinflation. *Journal of Monetary Economics 52*(5), 981–1015.

Guérin, P. and M. Marcellino (2013). Markov-switching midas models. *Journal of Business & Economic Statistics 31*(1), 45–56.

Hahn, P. R. and C. M. Carvalho (2015). Decoupling shrinkage and selection in bayesian linear models: a posterior summary perspective. *Journal of the American Statistical Association 110*(509), 435–448.

Hasenzagl, T., L. Reichlin, and G. Ricco (2020). Financial variables as predictors of real growth vulnerability.

Huber, F., G. Koop, and L. Onorante (2019). Inducing sparsity and shrinkage in time-varying parameter models. *arXiv preprint arXiv:1905.10787*.

Hubrich, K. and R. J. Tetlow (2015). Financial stress and economic dynamics: The transmission of crises. *Journal of Monetary Economics 70*, 100–115.

Ishwaran, H., J. S. Rao, et al. (2005). Spike and slab variable selection: frequentist and bayesian strategies. *Annals of statistics 33*(2), 730–773.

Jordà, Ò., M. Schularick, and A. M. Taylor (2015). Leveraged bubbles. *Journal of Monetary Economics 76*, S1–S20.

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

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

Koenker, R., V. Chernozhukov, X. He, and L. Peng (2017). Handbook of quantile regression.

Kohns, D. and A. Bhattacharjee (2020). Developments on the bayesian structural time series model: Trending growth. *arXiv preprint arXiv:2011.00938*.

Kohns, D. and T. Szendrei (2020). Horseshoe prior bayesian quantile regression. *arXiv preprint arXiv:2006.07655*.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting 33*(1), 11–20.

Kowal, D. R. (2021). Fast, optimal, and targeted predictions using parametrized decision analysis. *Journal of the American Statistical Association*, 1–28.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation 81*(11), 1565–1578.

Lee, E. R., H. Noh, and B. U. Park (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association 109*(505), 216–229.

Li, Q., R. Xi, N. Lin, et al. (2010). Bayesian regularized quantile regression. *Bayesian Analysis 5*(3), 533–556.

McAlinn, K. and M. West (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics 210*(1), 155–169.

McCracken, M. and S. Ng (2020). Fred-qd: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

McCracken, M. W., J. T. McGillicuddy, and M. T. Owyang (2021). Binary conditional forecasts. *Journal of Business & Economic Statistics*, 1–13.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association 83*(404), 1023–1032.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Piironen, J., M. Paasiniemi, A. Vehtari, et al. (2020). Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics 14*(1), 2155–2197.

Piironen, J., A. Vehtari, et al. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics 11*(2), 5018–5051.

Polson, N. G. and J. G. Scott (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics 9*, 501–538.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*(437), 179–191.

Ray, P. and A. Bhattacharya (2018). Signal adaptive variable selector for the horseshoe prior. *arXiv preprint arXiv:1810.09004*.

Romer, D. (2018). *Advanced macroeconomics*. Mcgraw-hill.

Rossi, B. and T. Sekhposyan (2019). Alternative tests for correct specification of conditional predictive densities. *Journal of Econometrics 208*(2), 638–657.

Stock, J. H. and M. W. Watson (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

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*(4), 437–447.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.

A   Appendix

A.1   Expected Loss function for the BQR

In order to decouple shrinkage from sparsity, loss function (18) necessitates integration over the uncertainty in the latent predictions $\tilde{Y}$. This involves two integration steps: 1) over $\tilde{Y}$ conditional on θ, and 2) over θ, conditional on $\tilde{Y}$. The conditional predictive likelihood, based on in-sample values of X, is:

$$y_t|\beta, Z, \sigma \sim N(x_t'\beta + \xi z_t, \tau^2 z_t \sigma) \tag{39}$$

Note that for the sake of readability, the dependence on a given quantile is suppressed.

Integration over $\tilde{Y}|\theta$

$$
\begin{aligned}
\mathcal{L}(\tilde{Y}, \alpha) &= E_{\tilde{Y}|\bullet}[\tilde{Y}, \alpha|\beta, \sigma, Z] \\
&= \phi||\alpha||_0 + \int T^{-1}||X\alpha - \tilde{Y}||_2^2 \, p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} \\
&= \phi||\alpha||_0 + T^{-1}\alpha'X'X\alpha - T^{-1}2\alpha'X' \int \tilde{Y} \, p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} + T^{-1} \int \tilde{Y}'\tilde{Y} \, p(\tilde{Y}|\beta, \sigma, Z)d\tilde{Y} \\
&= \phi||\alpha||_0 + T^{-1}[\alpha'X'X\alpha - 2\alpha X'(X\beta + \xi Z) + \Sigma_{\tilde{Y}} + ||X\beta + \xi Z||_2^2] \\
&= \phi||\alpha||_0 + T^{-1}[||X\alpha - X\beta||_2^2 - 2\alpha X'(X\beta + \xi Z) + \Sigma_{\tilde{Y}} + A]
\end{aligned}
\tag{40}
$$

Where A is defined as $\xi^2 Z'Z + 2\beta'X'\xi Z$, and the integral of $\tilde{Y}'\tilde{Y}$ follows from:

$$\int \tilde{Y}'\tilde{Y} \, p(\tilde{Y}|\beta, \sigma, Z) = \sum_{t=1}^T \sigma_{\tilde{y}_t}^2 + \sum_{t=1}^T (x_t'\beta + \xi z_t)^2, \tag{41}$$

where $\sigma_{\tilde{y}_t}^2 = \int (\tilde{y}_t - \tilde{\mu})^2 p(\tilde{y}_t|\beta, \sigma, Z)$.

Integration over $\beta, \sigma, Z|\bullet$

$$
\begin{aligned}
\mathcal{L}(\alpha) &= E_{\beta, \sigma, Z|\bullet}[\mathcal{L}(\beta, \sigma, \mathcal{Z}, \alpha)] \\
&\phi||\alpha||_0 + T^{-1} \int ||X\alpha - X\beta||_2^2 \, p(\beta.\sigma, Z|\tilde{Y})d(\beta, \sigma, Z) \\
&+ T^{-1} \int \Sigma_{\tilde{y}} \, p(\beta.\sigma, Z|\tilde{Y})d(\beta, \sigma, Z) \\
&+ T^{-1}A \, p(\beta.\sigma, Z|\tilde{Y})d(\beta, \sigma, Z) + T-1 \int \alpha'X'\xi Z \, p(\beta.\sigma, Z|\tilde{Y})d(\beta, \sigma, Z)
\end{aligned}
\tag{42}
$$

Since all the terms involving $\Sigma_{\tilde{y}}$ and A don't involve α, they do not affect the minimisation of the loss function and can therefore be dropped.

$$
\begin{aligned}
&= \phi ||\alpha||_0 + T^{-1}[||X\alpha - X\beta||_2^2 - 2\alpha' X' X\overline{\beta} + \overline{\beta}' X' X\overline{\beta} + tr(X' X\Sigma_{\overline{\beta}})] \\
&\quad - T^{-1}\alpha X' \xi \int Z\, p(\beta, \sigma, Z|\tilde{Y})d(\beta, \sigma, Z) \\
&= \phi ||\alpha||_0 + T^{-1}||X\alpha - X\beta||_2^2 + T^{-1}tr(X' X\Sigma_{\overline{\beta}}) - T^{-1}\alpha' X' \xi \overline{Z}.
\end{aligned}
\tag{43}
$$

where $\overline{Z}_p = \left(\frac{|y_1 - x_1'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2}, \cdots , \frac{|y_T - x_T'\beta_p|}{\sqrt{\xi^2 + 2\tau^2}} + \frac{\sigma\tau^2}{\xi^2 + 2\tau^2} \right)'$. Entries of $\overline{Z}_p$ follow from the definition of the expectation of the inverse Gaussian with location μ_t and scale λ_t: since $1/z_t \sim iG(\mu_t, \lambda_t)$, then $z_t \sim iG(\frac{1}{\mu_t} + \frac{1}{\lambda_t}, \frac{1}{\lambda_t \mu_t} + \frac{1}{\lambda_t^2})$ (see for example Khare and Hobert (2012)).

A.2 Derivation of SAVS-BQR

As described in section 2.4, instead of directly minimising (21), we use ℓ_1-norm as well as adaptive penalisation akin to Zou (2006):

$$
\mathcal{L}(\alpha) = \operatorname*{argmin}_{\alpha} \{ \frac{1}{2}||X\overline{\beta} - X\alpha||_2^2 + \sum_{j=1}^{K} \phi_j |\alpha_j| - \alpha' X' \xi \overline{Z} \}.
\tag{44}
$$

Notice that for the first term, we use the simple fact that $||Xa - Xb|| = ||Xb - Xa||$, and have introduced another factor $\frac{1}{2}$ as well as dropped T^{-1} for notational convenience. We suppressed dependence of the latent quantities on quantile p for the same reason.

While the LARS algorithm (see Friedman et al. (2001)) can be applied to (44), we follow Ray and Bhattacharya (2018) in making use of the efficient coordinate descent algorithm introduced by Friedman et al. (2007). The coordinate descent algorithm reduces computational complexity by updating the entry in solution vector α_j, conditional on all other α, which is iterated until convergence.

For a given state of the algorithm, $\tilde{\alpha}$, the objective function is recast as a function of the jth variable:

$$
\tilde{\mathcal{L}}(\alpha_j) = \{ \frac{1}{2}||X\beta - X_{-j}\tilde{\alpha}_{-j} - X_j\alpha_j||_2^2 + \sum_{k \neq j} \phi_k |\alpha| - \tilde{\alpha}'_{-j} X'_{-j} \xi \overline{Z} - \alpha'_j X'_j \xi \overline{Z} \},
\tag{45}
$$

where, likewise, X_{-j} denotes columns in X which are not the jth column. Since the optimisation problem is thus broken down to a single covariate basis, the objective function (45) is the

classical thresholding problem. Taking first order conditions for any j:

$$\frac{\partial \tilde{L}_j(\alpha_j)}{\partial \alpha_j} = X_j'X_j + X_j'(X_{-j}\tilde{\alpha}_{-j} - X\overline{\beta}) + \phi_j s_j + d_j = 0$$

$$\alpha_j = \frac{1}{X_j'X_j} sign(X_j'\tilde{R}_j - d_j - \phi_j s_j) \tag{46}$$

$$= \frac{1}{X_j'X_j} sign(X_j'\tilde{R}_j - d_j)(|X_j'\tilde{R}_j - d_j| - \phi_j)_+,$$

where $\tilde{R}_j$ is the partial residual vector between $X\overline{\beta}$ and $X_{-j}\tilde{\alpha}_j$ at a given state and s_j is defined as the subgradient, $s_j \in \partial|\alpha_j|$. Notice that we collect $-X_{-j}\xi\overline{Z}$ into d_j.

Now, since the coordinate descnet algorithm is stopped after the first iteration, as suggested in Ray and Bhattacharya (2018), $R_j = \overline{\beta}_j X_j$, and therefore $X_j R_j = \overline{\beta}_j ||X||_2^2$. Hence, by neglecting d_j as suggested in section (3.2), the solution to theorem 1 is achieved.

A.3 Sampling Algorithms

With conditional posteriors presented in section 2.2 at hand, we utilise standard Gibbs samplers to obtain draws from the posterior distributions. Collecting posterior hyperparameters for β, which will differ for each prior under investigation, in Λ_*, the dynamics of the of the Markov chain $\{\beta^m, \sigma^m, \Lambda_*^m, Z^m\}_{m=0}^{\infty}$ are implicitly defined through the following steps

1. Draw $z_t \sim \pi(.|\beta, \sigma, \Lambda_*, \theta, \tau, Y)$ from $1/iG(\overline{c}_t, \overline{d}_t)$ for all t and stack to a T x 1 vector Z_{n+1}

2. Draw $\sigma_{n+1} \sim \pi(.|\beta, \Lambda_*, \theta, \tau, Y, Z_{n+1})$ from $IG(\overline{a}, \overline{b})$

3. Draw $\beta_{n+1} \sim \pi(.|\sigma_{n+1}, \Lambda_*, \theta, \tau, Y, Z_{n+1})$ from $N(\overline{\beta}, \overline{\Lambda}_*)$

4. Draw Λ_{n+1} according to each prior in section 2.2

5. Iterate (1-4) until convergence is achieved

Note that for the horseshoe prior, we use a slice sampler to sample the elements of Λ_*, as suggested in Kohns and Szendrei (2020).

A.3.1 Fast BQR Sampler

The most costly operation in the Gibbs sampler in (A.3), is the inversion of the possibly large dimensional posterior covariance of the regression coefficients, step 4. To speed up computation, we make use of the fast BQR sampler proposed in Kohns and Szendrei (2020):

As derived section (7), using the scale mixture representation in, the conditional posterior of β given all other parameters can be written as:

$$\beta|\bullet \sim N(A^{-1}X'\Sigma y, A^{-1}), \quad A = (X'\Sigma X + \Lambda_*^{-1}) \tag{47}$$

Suppose, we want to sample from $N_K(\mu, \Sigma)$, where

$$\Sigma = (\Phi'\Phi + D)^{-1}, \quad \mu = \Sigma\Phi'\alpha. \tag{48}$$

Assume $D \in \mathbb{R}^{K \times K}$ is a positive definitive matrix, $\phi \in \mathbb{R}^{T \times K}$, and $\alpha \in \mathbb{R}^{T \times 1}$. Then (47) is a special case of of (48) when setting $\Phi = \sqrt{\Sigma}X$, $D = \Lambda_*$ and $\alpha = \sqrt{\Sigma}y$. An exact algorithm to sample from (47) is thus given by:

Algorithm 1. *Fast HS-BQR sampler*

  1. *Sample independently $u \sim N(0, D)$ and $\delta \sim N(0, I_T)$*

  2. *Set $\xi = \Phi u + \delta$*

  3. *Solve $(\Phi D\Phi' + I_T)w = (\alpha - \xi)$*

  4. *Set $\theta = u + D\Phi'w$*

Proposition *Suppose θ is obtained through algorithm 1. Then $\theta \sim N(\mu, \Sigma)$.*

For a proof, we refer to Kohns and Szendrei (2020).

A.3.2   Conditional Mean Models

Assuming a Gaussian likelihood for linear model $y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_T)$, the prior for regression coefficients, β, is assumed to be multivariate norman, and the error variance, σ^2, inverse-Gamma:

$$p(\beta|V) \sim N(0, V))$$
$$p(V) \sim f \tag{49}$$
$$p(\sigma^2) \sim IG(\underline{a}, \underline{b}).$$

The horseshoe, SSVS and lasso priors differ in how the hierarchical prior variance on β, $p(V)$, is specified. For all priors, we set $\underline{a} = \underline{b} = 0.1$.

The conditional posteriors due to conditional conjugacy are:

$$\beta|\bullet \sim N(\overline{\beta}, \overline{V})$$
$$\sigma^2|\bullet \sim IG(\overline{a}/2, \overline{b}/2) \tag{50}$$
$$\overline{V}|\bullet \sim g,$$

where $\overline{a} = \underline{a} + T$ and $\overline{b} = \underline{b} + \frac{(y-X\beta)'(y-X\beta)}{T-K} + \overline{\beta}'[V + (X'X)^{-1}]^{-1}\overline{\beta}$. The posterior of the variance hyperparameters and of the regression parameters are prior specific and detailed below. These conditional posteriors are sampled in a standard Gibbs sampler with 5000 burnin draws and 5000 retained MCMC samples.

Horseshoe

For the horseshoe prior, set $V = \sigma^2 \nu^2 diag(\lambda_1^2, \cdots, \lambda_K^2)$, where

$$\nu \sim C_+(0, 1)$$
$$\lambda_j \sim C_+(0, 1) \forall j. \tag{51}$$

Then, by standard calculations (see Bhattacharya et al. (2016))):

$$\overline{\beta} = \overline{V}^{-1} X' y,$$
$$\overline{V} = \sigma^2 (X'X + V^{-1}) \tag{52}$$

Instead of computing the large dimensional inverse A^{-1}, we rely on a data augmentation technique introduced by Bhattacharya et al. (2016).

SSVS

For the SSVS prior, we use the same hierarchy as presented in (16), which is printed below for convenience:

$$\beta_{j,p}|\gamma_j, \lambda_j \sim (1 - \gamma_j)N(0, c\lambda_j^2) + \gamma_j N(0, \lambda_j^2) \forall j \in \{1, \cdots, K\}$$
$$\lambda_j^2 \sim IG(a_2, b_2)$$
$$\gamma_j|\pi_0 \sim Bern(\pi_0) \tag{53}$$
$$\pi_0 \sim B(a_3, b_3),$$

where $V = diag(\lambda_1^2, \cdots, \lambda_K^2)$ for all j if $\gamma_j = 1$, and $diag(c\lambda_1^2, \cdots, c\lambda_k^2)$ otherwise. The conditional posteriors are standard and derived for example in George and McCulloch (1993) and

Ishwaran et al. (2005). The difference to the prior of George and McCulloch (1993) lies in the additional prior on δ_j^2 which is assumed to be inverse gamma. It can be shown that this implies a mixture of student-t distributions for β_j marginally (Konrath et al., 2008):

$$
\begin{aligned}
\overline{\beta} &= \overline{V}^{-1} X'y/\sigma^2 \\
\overline{V} &= (X'X/\sigma^2 + V^{-1})^{-1}, \ V = diag(\lambda_j \gamma_j) \\
\gamma_j|\bullet &\sim N(1-\pi_0)N(\beta_j|0, c \times \lambda_j^2)I_{\gamma_j=0} + \pi_0 N(\beta_j|0, \lambda_j^2)I_{\gamma_j=0} \\
\pi_0|\bullet &\sim Beta(a_3 + n_1, b_3 + K - n_1), \text{where } n_1 = \sum_j I_{\gamma_j=1}
\end{aligned}
\tag{54}
$$

To be agnostic about the degree of sparsity, we set $a_3 = b_3 = 0$ as for the quantile SSVS model.

Lasso

Similar to the SSVS, the lasso prior is the same as in (11):

$$
\begin{aligned}
\beta_j|\phi &\sim N(0, \lambda_j), \\
\lambda_j|\phi &\sim exp(\frac{\phi}{2}) \\
\phi &\sim G(a_1, b_1)
\end{aligned}
\tag{55}
$$

where $\Lambda_* = diag(\lambda_1, \cdots, \lambda_K)$. The conditional posteriors for the hyperparameters are standard:

$$
\begin{aligned}
p(\lambda_j^{-1}|\bullet) &\sim iG(\sqrt{\frac{\phi}{\beta_{j,p}^2}}, \phi) \\
p(\phi|\bullet) &\sim G(K + a_1, \frac{1}{2}\sum_{j=1}^{K} \lambda_j + b_1) \\
\overline{\beta} &= \overline{V}^{-1} X'y/\sigma^2 \\
\overline{V} &= (X'X/\sigma^2 + V^{-1})^{-1}, \ V = diag(\lambda_1, \cdots, \lambda_K)
\end{aligned}
\tag{56}
$$

A.4   Tables

	MSFE	LPDS	CRPS	qwCRPS	MSFE	LPDS	CRPS	qwCRPS
			h=1				h=2	
ABG_{BQR}	0.504	-11.607	0.313	1.049	0.404	-7.870	0.254	0.891
$ABG_{BQR-Skt}$	0.519	-1.727	0.301	1.026	0.426	-4.102	0.277	0.970
$SSVSBQR$	0.539*	-0.889**	0.320***	1.061	0.435**	-0.607	0.221***	0.746**
$HSBQR$	0.542**	-3.695	0.313***	0.952*	0.473**	-5.101	0.236***	0.776**
$HSBQR_{BIC}$	0.522	-3.020	0.300***	0.918**	0.428**	-5.775*	0.239***	0.779***
$HSBQR_{SAVS}$	0.537	-2.838	0.304***	0.925*	0.416	-3.959	0.259***	0.847
$LBQR$	0.570	-1.313	0.289***	0.949	0.480	-0.760*	0.229***	0.748***
$LBQR_{BIC}$	0.544	-1.316	0.289***	0.949	0.450	-0.776*	0.230***	0.750***
$LBQR_{SAVS}$	0.544	-2.021	0.301**	0.999	0.449	-1.219*	0.245***	0.808
			h=3				h=4	
ABG_{BQR}	0.378	-6.825	0.237	0.847	0.351	-4.387	0.235	0.844
$ABG_{BQR-Skt}$	0.397	-1.006	0.249	0.857	0.367	-1.252	0.251	0.875
$SSVSBQR$	0.431**	-0.542	0.211***	0.723	0.392*	-0.657*	0.214***	0.710
$HSBQR$	0.437**	-5.224	0.231***	0.777	0.392**	-5.258*	0.224***	0.731
$HSBQR_{BIC}$	0.400*	-5.958**	0.230***	0.775	0.377*	-4.815**	0.220***	0.723
$HSBQR_{SAVS}$	0.395	-4.714	0.240***	0.810	0.380	-3.568	0.216***	0.729
$LBQR$	0.431*	-0.620	0.222***	0.743**	0.413*	-0.667	0.213***	0.716**
$LBQR_{BIC}$	0.422*	-0.630	0.223***	0.745**	0.387*	-0.675	0.214***	0.718**
$LBQR_{SAVS}$	0.421*	-0.964	0.233***	0.797	0.386*	-1.557	0.234***	0.804

Table 7: Forecast Evaluation Results (Skewed-t distribution)

	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
	y_1—					y_2—					y_3—					y_4				
Coefficient Bias																				
T=100																				
Sparse																				
$HSBQRZ$	0.121	0.092	0.086	0.098	0.126	0.172	0.099	0.088	0.097	0.176	0.156	0.098	0.077	0.094	0.148	0.128	0.095	0.079	0.094	0.127
$HSBQRZ_{SAVS}$	0.110	0.072	0.064	0.081	0.116	0.165	0.081	0.066	0.080	0.169	0.146	0.089	0.062	0.086	0.138	0.119	0.085	0.063	0.085	0.119
$HSBQRZ_{BIC}$	0.103	0.062	0.058	0.072	0.107	0.159	0.067	0.056	0.072	0.163	0.141	0.087	0.060	0.085	0.132	0.117	0.084	0.061	0.084	0.116
Block																				
$HSBQRZ$	0.250	0.247	0.252	0.254	0.244	0.260	0.240	0.244	0.250	0.272	0.289	0.293	0.291	0.285	0.277	0.257	0.266	0.268	0.263	0.256
$HSBQRZ_{SAVS}$	0.252	0.242	0.246	0.248	0.243	0.262	0.236	0.239	0.245	0.272	0.290	0.287	0.283	0.279	0.276	0.260	0.262	0.262	0.258	0.256
$HSBQRZ_{BIC}$	0.259	0.246	0.247	0.247	0.247	0.271	0.241	0.244	0.248	0.276	0.292	0.286	0.283	0.278	0.277	0.262	0.262	0.262	0.258	0.258
T=500																				
Sparse																				
$HSBQRZ$	0.099	0.073	0.062	0.073	0.103	0.152	0.074	0.061	0.075	0.152	0.140	0.066	0.048	0.067	0.135	0.113	0.065	0.047	0.066	0.111
$HSBQRZ_{SAVS}$	0.086	0.053	0.038	0.054	0.089	0.145	0.057	0.037	0.059	0.144	0.127	0.058	0.026	0.058	0.123	0.100	0.056	0.026	0.057	0.098
$HSBQRZ_{BIC}$	0.076	0.045	0.029	0.047	0.077	0.140	0.050	0.030	0.054	0.137	0.106	0.061	0.025	0.062	0.106	0.084	0.058	0.024	0.059	0.085
Block																				
$HSBQRZ$	0.143	0.101	0.094	0.103	0.144	0.174	0.104	0.091	0.103	0.176	0.187	0.098	0.082	0.096	0.184	0.156	0.101	0.085	0.099	0.154
$HSBQRZ_{SAVS}$	0.139	0.089	0.081	0.092	0.138	0.173	0.093	0.077	0.092	0.171	0.185	0.091	0.072	0.087	0.181	0.152	0.092	0.074	0.091	0.150
$HSBQRZ_{BIC}$	0.146	0.096	0.085	0.099	0.151	0.180	0.099	0.080	0.097	0.183	0.190	0.091	0.072	0.088	0.186	0.156	0.092	0.074	0.091	0.154
MCC																				
T=100																				
Sparse																				
$HSBQRZ_{SAVS}$	0.392	0.535	0.555	0.548	0.509	0.383	0.516	0.552	0.556	0.504	0.338	0.446	0.446	0.430	0.324	0.382	0.453	0.433	0.424	0.365
$HSBQRZ_{BIC}$	0.555	0.725	0.780	0.772	0.713	0.561	0.722	0.793	0.791	0.707	0.423	0.496	0.540	0.464	0.405	0.446	0.485	0.523	0.448	0.426
Block																				
$HSBQRZ_{SAVS}$	0.386	0.412	0.420	0.411	0.393	0.405	0.429	0.439	0.421	0.398	0.336	0.342	0.344	0.351	0.345	0.384	0.392	0.386	0.394	0.382
$HSBQRZ_{BIC}$	0.420	0.445	0.457	0.452	0.435	0.435	0.457	0.471	0.462	0.441	0.372	0.371	0.350	0.376	0.383	0.413	0.405	0.372	0.405	0.411
T=500																				
Sparse																				
$HSBQRZ_{SAVS}$	0.397	0.540	0.567	0.543	0.462	0.389	0.568	0.576	0.574	0.471	0.362	0.679	0.641	0.653	0.408	0.398	0.668	0.648	0.654	0.434
$HSBQRZ_{BIC}$	0.681	0.898	0.904	0.898	0.840	0.702	0.905	0.906	0.903	0.850	0.659	0.783	0.836	0.792	0.665	0.709	0.785	0.827	0.785	0.697
Block																				
$HSBQRZ_{SAVS}$	0.691	0.804	0.810	0.807	0.691	0.707	0.809	0.820	0.815	0.710	0.613	0.837	0.865	0.850	0.617	0.676	0.821	0.855	0.844	0.677
$HSBQRZ_{BIC}$	0.779	0.946	0.947	0.943	0.828	0.781	0.946	0.951	0.952	0.839	0.714	0.889	0.907	0.899	0.717	0.767	0.861	0.883	0.875	0.765
Hit rate																				
T=100																				
Sparse																				
$HSBQRZ_{SAVS}$	0.578	0.817	0.842	0.835	0.768	0.567	0.790	0.828	0.822	0.762	0.538	0.608	0.612	0.565	0.511	0.540	0.618	0.608	0.571	0.518
$HSBQRZ_{BIC}$	0.481	0.640	0.707	0.703	0.690	0.495	0.642	0.728	0.724	0.684	0.479	0.546	0.484	0.507	0.460	0.501	0.586	0.509	0.536	0.481
Block																				
$HSBQRZ_{SAVS}$	0.554	0.652	0.685	0.671	0.583	0.570	0.665	0.691	0.672	0.583	0.549	0.631	0.650	0.639	0.556	0.566	0.649	0.667	0.655	0.567
$HSBQRZ_{BIC}$	0.358	0.414	0.443	0.439	0.388	0.368	0.423	0.444	0.436	0.388	0.454	0.570	0.632	0.581	0.461	0.499	0.624	0.686	0.632	0.498
T=500																				
Sparse																				
$HSBQRZ_{SAVS}$	0.769	0.956	0.966	0.951	0.924	0.745	0.950	0.957	0.963	0.916	0.818	0.935	0.949	0.920	0.830	0.834	0.940	0.956	0.927	0.845
$HSBQRZ_{BIC}$	0.651	0.831	0.853	0.843	0.763	0.632	0.846	0.862	0.848	0.786	0.656	0.889	0.887	0.868	0.653	0.658	0.905	0.926	0.891	0.655
Block																				
$HSBQRZ_{SAVS}$	0.963	0.999	0.998	0.998	0.969	0.966	0.998	1.000	0.998	0.972	0.932	0.986	0.995	0.989	0.929	0.954	0.988	0.995	0.990	0.953
$HSBQRZ_{BIC}$	0.905	0.982	0.989	0.980	0.875	0.918	0.983	0.994	0.985	0.876	0.852	0.982	0.993	0.985	0.854	0.915	0.986	0.994	0.988	0.913

Table 8: Nuisance Parameter correction results for the HSBQR