
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Kohns, D; Szendrei , Tibor

Horseshoe prior Bayesian quantile regression

Published in:

Journal of the Royal Statistical Society. Series C: Applied Statistics

DOI:

[10.1093/jrssc/qlad091](https://doi.org/10.1093/jrssc/qlad091)

Published: 01/01/2024

Document Version

Publisher's PDF, also known as Version of record

Published under the following license:

CC BY

Please cite the original version:

Kohns, D., & Szendrei , T. (2024). Horseshoe prior Bayesian quantile regression. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 73(1), 193-220. Article qlad091. <https://doi.org/10.1093/jrssc/qlad091>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Horseshoe prior Bayesian quantile regression

David Kohns^{1,2} and Tibor Szendrei²

¹Department of Computer Science, Aalto University, Espoo, Finland

²Department of Economics, Heriot-Watt University, Edinburgh, UK

Address for correspondence: David Kohns, Department of Computer Science, Aalto University, Konemiehentie 2, Espoo, Finland. Email: david.kohns@aalto.fi

Abstract

This paper extends the horseshoe prior to Bayesian quantile regression and provides a fast sampling algorithm for computation in high dimensions. Compared to alternative shrinkage priors, our method yields better performance in coefficient bias and forecast error, especially in sparse designs and in estimating extreme quantiles. In a high-dimensional growth-at-risk forecasting application, we forecast tail risks and complete forecast densities using a database covering over 200 macroeconomic variables. Quantile specific and density calibration score functions show that our method provides competitive performance compared to competing Bayesian quantile regression priors, especially at short- and medium-run horizons.

Keywords: global–local prior, growth-at-risk, Monte Carlo, quantile regression, sampling method

1 Introduction

Quantile regression has been an important tool in the econometricians' toolkit when estimating heterogeneous effects across the conditional response distribution, since the seminal work of [Koenker and Bassett \(1978\)](#). In contrast to least squares methods, it estimates quantiles of the dependent variables' conditional distribution directly, which allows for richer inference than solely focusing on the conditional mean. While highly influential in the risk management and finance literature in calculating risk measures such as value-at-risk (VaR) (i.e. the loss a portfolios value incurs at a specified probability level), it has experienced a recent surge in use especially in the macroeconomic literature to quantify downside risks of the aggregate economy to financial conditions ([Adams et al., 2021](#); [Adrian et al., 2019](#); [Carriero et al., 2020](#); [Figueres & Jarociński, 2020](#); [Hasenzagl et al., 2020](#); [Korobilis, 2017](#); [Mazzi & Mitchell, 2019](#); [Prasad et al., 2019](#)).

A challenge for these purposes is that sources of risk can be numerous such that simple quantile regression is often rendered imprecise or infeasible in high dimensions. While a variety of regularisation and dimension reduction techniques have been proposed for macroeconomic forecasting, [Stock and Watson \(2002, 2012\)](#), [Kim and Swanson \(2014\)](#), [Bai and Ng \(2008\)](#), and [De Mol et al. \(2008\)](#), extensions of high-dimensional methods, in particular Bayesian methods, applied to quantile regression, remain relatively scant.

The Bayesian quantile regression (BQR) approach, as popularised by [Yu and Moyeed \(2001\)](#), is based on the asymmetric Laplace likelihood (ALL), which has a special connection to the frequentist quantile regression solution, in that its maximum likelihood estimates are equivalent to traditional quantile regression with a check-loss function ([Koenker, 2005](#)). A hurdle in the Bayesian literature has been that ALL-based methods result in improper posteriors with any but non-informative or exponential Laplace priors, where the latter results in the popular Bayesian Lasso quantile regression ([Alhamzawi & Yu, 2013](#); [Alhamzawi et al., 2012](#); [Chen et al., 2013](#); [Li et al., 2010](#)). The broader Bayesian shrinkage literature has shown, however, that global–local shrinkage priors such as the horseshoe ([Carvalho et al., 2010](#)) and Dirichlet–Laplace prior

Received: July 4, 2022. Revised: August 13, 2023. Accepted: August 23, 2023

© The Royal Statistical Society 2023.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Bhattacharya et al., 2016) offer asymptotic as well as computational advantages over the former methods (Bhadra et al., 2019). These methods have not yet been considered for the BQR. The aim of this paper is to bridge this gap and extend the global–local prior to quantile regression.

This paper’s primary contribution is twofold. First, we derive the horseshoe prior of Carvalho et al. (2010) for the BQR framework of Yu and Moyeed (2001). Second, we develop an efficient posterior sampler for the quantile specific regression coefficients based on data augmentation akin to Bhattacharya et al. (2016) which speeds up computation significantly for high-dimensional quantile problems.

To showcase the performance of the horseshoe BQR (HS-BQR), we provide a large-scale Monte Carlo study as well as a high-dimensional VaR application to US GDP (often called growth-at-risk, GaR, in the literature). In the Monte Carlo study, we show that the proposed estimator provides more stable and at worst, similar performance compared to a variety of Bayesian Lasso quantile regression methods in terms of coefficient bias and forecast accuracy. We find that, particularly, tails of the distributions are consistently better estimated by the HS-BQR which echos findings from the Bayesian VaR literature (Chen et al., 2012). In the GaR application, we show that the HS-BQR produces better calibrated forecast densities compared to the Bayesian alternatives and importantly provides the best performance for lower and upper tails which makes it a powerful tool for recession probability monitoring. The framework provided in this paper has the additional advantage that the derived algorithms can be directly applied to other global–local priors that can be expressed as scale mixture of normals.¹

In what follows, we first present the methodological framework of the proposed model and sampling algorithm. Following this, we will provide evidence from Monte Carlo simulations² and an empirical application of the favourable performance of the HS-BQR compared to alternative methods. We conclude with further generalisations of the algorithms provided and a discussion of our results. Estimation and replication code is publicly available at <https://github.com/davkoh/Horseshoe-Bayesian-Quantile-Regression>.

2 Methodology

2.1 Bayesian quantile regression

Assuming a linear model such as

$$y_t = x_t' \beta + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (1)$$

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 $p \in (0, 1)$:

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

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 $\{\epsilon_t\}_{t=1}^T$ (Koenker, 2005). We will maintain the assumption throughout the paper that the design X is known. The loss function $\rho_p(\cdot)$ is often expressed as a tick loss function of the form $\rho_p(y) = [p - I(y < 0)]y$, where $I(\cdot)$ 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 $\mathcal{ALD}(p, \varrho, \sigma)$, where ϱ is the location parameter set to

¹ For an overview of global–local priors, see Polson and Scott (2010).

² We thank two anonymous referees for suggesting robustness and computational efficiency analyses for the proposed methods. The paper has benefited greatly by this extension.

0 and σ denotes the scale of the \mathcal{ACD} . Assuming an \mathcal{ACD} 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{ACD} 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{ACD} error process can be recovered as a mixture between an exponentially distributed variable z_t with mean σ , $z_t \sim \exp(\sigma)$, and a standard normal variable, $u_t, u_t \sim N(0, 1)$:

$$\begin{aligned} \epsilon_t &= \zeta z_t + \tau \sqrt{\sigma z_t} u_t, \\ \zeta &= \frac{1 - 2p}{p(1 - p)}, \\ \tau^2 &= \frac{2}{p(1 - p)}, \end{aligned} \tag{4}$$

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

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

where $Y = (y_1, \dots, y_T)'$, $X = (x'_1, \dots, x'_T)'$, $Z = (z_1, \dots, z_T)'$, and $\Sigma = \text{diag}(1/(\tau^2 z_1 \sigma), \dots, 1/(\tau^2 z_T \sigma))$. 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:

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

with a prior mean of zero, the prior variance parameters, Λ_* control the amount of shrinkage towards sparsity.

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

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

The conditional posterior of the scale parameter is

$$\begin{aligned} p(\sigma | \cdot) &\sim \text{IG}(\bar{a}, \bar{b}), \\ \bar{a} &= \underline{a} + \frac{3T}{2}, \\ \bar{b} &= \underline{b} + \sum_{t=1}^T \frac{(y_t - x'_t \beta_p - \zeta z_t)^2}{2z_t \tau^2} + \sum_{t=1}^T z_t, \end{aligned} \tag{8}$$

where IG stands for the inverse-Gamma distribution with density $p(x) = \frac{\beta}{\Gamma(\alpha)} (1/x)^{\alpha+1} \exp(-\beta/x)$ and \underline{a} and \underline{b} are priors' guesses for shape and rate (both set to relatively uninformative values of 0.1). Finally, the conditional posterior for z_t is

$$p(z_t | \cdot) \sim 1/\text{iG}(\bar{c}_t, \bar{d}_t)$$

$$\bar{c}_t = \frac{\sqrt{\xi^2 + 2\tau^2}}{|y_t - x_t' \beta_p|} \quad (9)$$

$$\bar{d}_t = \frac{\xi^2 + 2\tau^2}{\sigma^2},$$

where iG stands for the inverse-Gaussian density with pdf $p(x) = \sqrt{\frac{\bar{d}_t}{2\pi x^3}} \exp(-\frac{\bar{d}_t(x - \bar{c}_t)^2}{2\bar{c}_t^2 x})$ with location (\bar{c}_t) and shape (\bar{d}_t) parameters.

The posteriors (7)–(9) allow for efficient Gibbs sampling algorithms which for the independent prior set-up 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.

2.2 Shrinkage priors

In order to efficiently estimate the posterior of a large-dimensional coefficient vector in small samples, informative priors are needed. Ideally, these priors are able to separate noise variables from signals such that the noise is shrunk towards zero and signals attain their unrestricted parameter values. The global–local prior framework, initially formalised in [Polson and Scott \(2010\)](#), follows a particularly suitable hierarchy for such estimation problems, in which global scales dictate the overall level of sparsity, while local scales, specific to each covariate, allow to reduce or tighten the pull towards zero depending on the signal:

$$\beta_{p,j} | \lambda_j^2, \quad v^2 \sim N(0, \lambda_j^2 v^2), \quad j \in (1, \dots, K),$$

$$\lambda_j \sim \pi(\lambda_j), \quad j \in (1, \dots, K), \quad (10)$$

$$v \sim \pi(v).$$

The horseshoe prior of [Carvalho et al. \(2010\)](#), in particular, employs two Half-Cauchy distributions for λ and v , with density $p(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, $x > 0$:

$$\lambda_j \sim C_+(0, 1),$$

$$v \sim C_+(0, 1), \quad (11)$$

such that $\Lambda_* = v^2 \text{diag}(\lambda_1^2, \dots, \lambda_K)$. Notice that in our application of the horseshoe prior to the BQR, we have formulated the prior hierarchy independent of σ , while the global–local literature often uses a dependent prior. It is easily verified that the assumption of prior independence is needed so that the conditional posterior distributions are analytically available as derived above. Additionally, [Moran et al. \(2018\)](#) have shown that in high-dimensional settings, the independence assumption aids inference of the error variance. This is due to the fact that conjugate priors act mathematically as additional observations which artificially bias the error variances downwards when $K \gg T$.

Due to the fact that Half-Cauchy distributions have high mass on 0 with fat tails, they are well suited for regularisation and variable selection tasks as they encapsulate the idea that only a few covariates are of importance. One way to analyse this behaviour is via shrinkage coefficients, $\kappa_{p,j}$, which dictate how much the posterior means of $\beta_p | \cdot$ are pulled towards the prior mean of zero. We establish their implied probability density functions with the two following propositions:

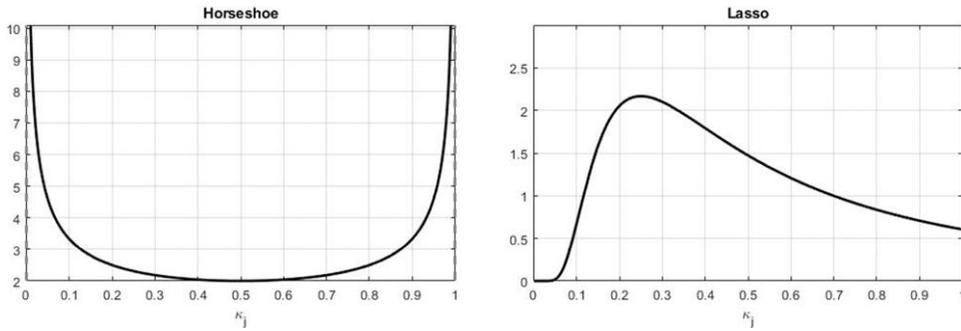


Figure 1. Distribution of $\kappa_{p,j}$, the shrinkage coefficient implied by the horseshoe prior and the Lasso prior, setting $a_j = 1$ and $T = 100$.

Proposition 1 Assume that we can approximate $X'X \approx \text{diag}(\sum_{t=1}^T x_{1,t}^2, \dots, \sum_{t=1}^T x_{K,t}^2)$, then the posterior mean (7), $\bar{\beta}_p$, may be decomposed as

$$\bar{\beta}_{p,j} = (1 - \kappa_{p,j})\hat{\beta}_{p,j}, \tag{12}$$

where $\kappa_{p,j} = \frac{1}{1+v^2\lambda_j^2 s_j} \in (0, 1)$, $s_j = \sum_{t=1}^T \frac{x_{j,t}^2}{t^2 \sigma_{z_t}^2}$, $\hat{\beta}_p = (X'\Sigma X)^{-1} X'\Sigma(Y - \zeta Z)$, and $\hat{\beta}_{p,j}$ is the j th entry in $\hat{\beta}_p$.

Equation (12) gives an intuitive understanding of the impact of the scales $\{\lambda_j, v\}$ on $\kappa_{p,j}$. In particular, $\hat{\beta}_{p,j}$ may be regarded as a conditional maximum likelihood estimate and $\kappa_{p,j} \rightarrow 0$ as $\lambda_j \rightarrow \infty$ or $v \rightarrow \infty$.

Proposition 2 For the horseshoe prior in equations (10) and (11), by change of variables, we establish the density $\kappa_{p,j} | \sigma, v, \{z_t\}_{t \in T}$ is given by

$$p(\kappa_{p,j} | \sigma, v, \{z_t\}_{t \in T}) = \frac{1}{\pi} \frac{a_j}{(a_j^2 - 1)\kappa_{p,j} + 1} \frac{1}{\sqrt{\kappa_{p,j}} \sqrt{1 - \kappa_{p,j}}}, \tag{13}$$

where $a_j = v\sqrt{s_j}$.

Notice that if $a_j = 1$, then this distribution reduces to a Beta(0.5,0.5) which is typical for the horseshoe prior when applied to generalised linear models (Pironen & Vehtari, 2017) and favours either shrinkage towards zero or leaving the posterior relatively un-regularised. Double exponential-based priors, on other hand, such as the Lasso prior of Park and Casella (2008) extended by Li et al. (2010) to the BQR model do not, as shown in Figure 1.

With $\kappa_{p,j}$ favoured close to 0 or 1, the summation over all coefficients can thus be seen as model-based approximate measure of sparsity, which we denote as the effective model size m_{eff} :

$$m_{\text{eff}} = \sum_{j=1}^K (1 - \kappa_{p,j}). \tag{14}$$

At this point, however, we note that although equation (14) may constitute a measure of active coefficients, it does not directly inform on model selection akin to spike-and-slab priors, post-processing techniques, or further simplifying assumptions to the posterior as seen in Alhamzawi and Yu (2013), Pironen et al. (2020), Bondell and Reich (2012), and Li and Pati (2017).³ We view it as a tool to diagnose detection of sparsity and the effect of changing distributions for the global scale v and the quantile index as will be further elaborated in the Monte Carlo simulation (Section 3).

³ Model sizes are often better identified than the inclusion of specific variables, particularly when the data are highly correlated, which aids the interpretation of m_{eff} as an approximate measure of sparsity.

2.3 Gibbs sampler

With the conditional posteriors at hand, we utilise a standard Gibbs sampler. The dynamics of the Markov chain $\{(\beta_m, \sigma_m, \lambda_m^2, v_m^2, z_m)\}_{m=0}^\infty$ are implicitly defined through the following steps:

1. Draw $Z \sim \pi(\cdot \mid \beta, \sigma, \lambda^2, v^2, \theta, \tau, X, Y)$ from $1/iG(\bar{c}_t, \bar{d}_t)$ for all t and call the $T \times 1$ vector z_{n+1} .
2. Draw $\sigma_{n+1} \sim \pi(\cdot \mid \beta, \lambda^2, v^2, \theta, \tau, X, Y, z_{n+1})$ from $IG(\bar{a}, \bar{b})$.
3. Draw $\beta_{n+1} \sim \pi(\cdot \mid \sigma_{n+1}, \lambda^2, v^2, \theta, \tau, X, Y, z_{n+1})$ from $N(\bar{\beta}, \bar{V})$.
4. Simulate λ_{n+1}^2 and v_{n+1}^2 through slice sampling given in Appendix A.
5. Iterate (1–4) until convergence is achieved.

In step 4, we make use of the fact that due to the assumption of independence on the scales (λ, v) , the posteriors also follow independent Half-Cauchy distributions. Since the Cauchy has no defined moments which would enable sampling, the literature has proposed Gibbs samplers which rely either on slice sampling (Polson et al., 2014) or mixture representations (Makalic & Schmidt, 2015). Because slice sampling does not involve rejections and or extra mixing variables, we follow Polson et al. (2014) by formulating a block slice sampling algorithm for $\lambda = (\lambda_1, \dots, \lambda_K)'$ given in Appendix A.

Khare and Hobert (2012) show that the Markov chain of this sampler is geometrically ergodic and also valid in $K \gg T$ settings which gives theoretical justification to apply this sampler to high-dimensional settings. However, a computational bottleneck is present in very high dimensions in evaluating the $K \times K$ -dimensional inverse for the conditional posterior of β . Cholesky decomposition-based methods will generally be of order $O(K^3)$. Taking into consideration that in quantile settings, one is usually interested in obtaining more than one expected quantile, this can result in prohibitively long computation times. We therefore provide a more efficient sampling algorithm for β which leverages data augmentation similar to the algorithm developed by Bhattacharya et al. (2016) which is of order $O(T^2K)$ and especially beneficial in high-dimensional settings.

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

$$\Theta = (\Phi' \Phi + D^{-1})^{-1}, \quad \mu = \Theta \Phi' (\alpha - \zeta Z). \quad (15)$$

Assume $D \in \mathbb{R}^{K \times K}$ is a positive definite matrix and diagonal for simplicity, $\phi \in \mathbb{R}^{T \times K}$, and $\alpha \in \mathbb{R}^{T \times 1}$. Then equation (7) is recovered when setting $\Phi = U^{1/2} X$, $D = \Lambda_*$, and $\alpha = U^{1/2} y$. An exact algorithm to sample from equation (7) 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 $\zeta = \Phi u + \delta$
 3. Solve $(\Phi D \Phi' + I_T) w = (\alpha - \zeta - \zeta Z)$
 4. Set $\theta = u + D \Phi' w$
-

Suppose θ is obtained through Algorithm 1. Then $\theta \sim N(\mu, \Theta)$. A full proof which closely follows Bhattacharya et al. (2016) is given in Appendix A.

3 Simulation set-up

In order to verify the theoretical advantages of the HS-BQR over the exponential-based quantile regression priors laid out above, we conduct a variety of high-dimensional Monte Carlo simulations that test the priors' ability to adapt to different degrees of sparsity and error distributions in the data generating processes. We consider three variants of the original Lasso prior which have been adapted to the BQR:

1. Bayesian Lasso QR (LBQR): The Lasso prior is derived by noticing that the ℓ_1 -norm penalised check-loss function

$$\min_{\beta} \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) + \lambda \sum_{j=1}^K |\beta_{p,j}|, \tag{16}$$

where λ is a penalisation coefficient, can be obtained as the maximum a posteriori (MAP) estimate of the ALL with a Laplace prior on the regression coefficients, $\pi(\beta_p | \sigma, \lambda) = (\sigma\lambda/2)^K \exp(-\sigma\lambda \sum_{j=1}^K |\beta_{p,j}|)$. The posterior takes the following form:

$$\beta_p | y, X, \sigma, \lambda \propto \exp\left(-\sigma \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) - \sigma\lambda \sum_{j=1}^K |\beta_{p,j}|\right). \tag{17}$$

To estimate equation (17), we use the Gibbs sampler of Li et al. (2010) with their recommended hyper-priors. Due to the shrinkage coefficient profile discussed above, we expect the LBQR to do well in sparse designs with well-identified signal and noise.

2. Bayesian elastic net QR (BQRENET): The elastic net estimator quantile regression differs from the Lasso in that it adds an ℓ_2 -norm of the regression coefficients to the minimisation problem. This is the ridge component which allows to shrink coefficients in a less aggressive manner than the ℓ_1 -norm. This makes it useful when dealing with correlated or dense designs. Assuming the elastic net estimator for the quantile regression, as

$$\min_{\beta_p} \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) + \lambda_1 \sum_{j=1}^K |\beta_{p,j}| + \lambda_2 \sum_{k=1}^K \beta_{p,j}^2 \tag{18}$$

the prior can, similarly to above, be formulated as an exponential prior, $\pi(\beta_{k,p} | \lambda_1, \lambda_2, \sigma) \propto \frac{\sigma^k}{2} \exp(-\sigma\lambda_1 |\beta_{p,j}| - \sigma\lambda_2 \beta_{p,j}^2)$. The posterior is then

$$\beta_p | y, X, \sigma, \lambda \propto \exp\left(-\sigma \sum_{t=1}^T \rho_p(y_t - x'_t \beta_p) - \sigma\lambda_1 \sum_{j=1}^K |\beta_{p,j}| - \sigma\lambda_2 \sum_{j=1}^K \beta_{p,j}^2\right). \tag{19}$$

We use the same hyper-priors as recommended by Li et al. (2010)

3. Bayesian adaptive Lasso QR (BALQR): The adaptive Lasso as proposed by Alhamzawi et al. (2012) uses the same set-up as the LBQR, but allows for the shrinkage coefficient to vary with each covariate. The prior can then be formulated as follows: $\pi(\beta_p | \sigma, \lambda_j) = (\sigma\lambda_j/2)^K \exp(-\sigma \sum_{j=1}^K \lambda_j |\beta_{p,j}|)$. Since this estimator allows for coefficient specific shrinkage, we expect it to outperform the LBQR.

Three sample sizes are considered: $T_{1,2} = \{200, 500\}$.⁴ In total, 100 Monte Carlo data sets were generated⁵ for which the last 100 observations are constructed to be the same for each T_i in order to make forecast errors comparable. The remainder of the observations are used as training samples to retrieve the mean posterior $\hat{\beta}_p$ vector to calculate bias.⁶

⁴ For the HS-BQR, we also consider a third sample size of $T_3 = 1,000$. This was done to get a better understanding whether the estimator can identify quantile varying parameters in the block design.

⁵ Except for $T = 500$ block case where only 20 Monte Carlo experiments were done due to the time it takes to run the estimator on such large dimensions.

⁶ Alternatively, one could also use the MAP estimate of the regression posterior as the point estimate. This might seem more natural when comparing BQR methods to frequentist quantile estimators due to their equivalence as discussed in Kozumi and Kobayashi (2011). We found that since the conditional posteriors are normal, there is no practical difference between the posterior mean and MAP.

We consider 12 designs in total which vary along two different dimensions: the degree of sparsity and the error generating process. We test the following sparsity patterns:

- Sparse with $\beta = (1, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, 0_{1 \times 2T_1})$,
- Dense with $\beta = (1, 0.85_{1 \times T_1})$,
- Block structure with $\beta = (1, 0.85_{1 \times T_1}, 0_{1 \times T_1}, 0.85_{1 \times T_1})$.

Notice that for T_1 there are always more coefficients than observations.

Consider a linear model as in equation (1). To retrieve the true quantile regression coefficients, β_p , we make use of Koenker's (2005) alternative representation of the quantile regression:

$$y_t = x_t' \beta + (x_t' \vartheta) u_t, \quad (20)$$

where u_t is assumed to be i.i.d. having some Cumulative Distribution Function (CDF), F . The dimensionality of ϑ is $K \times 1$ and determines which covariates have non-constant quantile functions. This can be seen from the solution for β_p to equation (20):

$$\beta_p = \beta + \vartheta F^{-1}(p). \quad (21)$$

Hence, the true β_p profile of a quantile regression model has a random coefficient model interpretation, where the vector of coefficients can be decomposed into a fixed plus a random component. In particular, the random component depends on the inverse CDF of the error, $F^{-1}(p)$. One can therefore think of ϑ as determining which variable is correlated with the error, where by default the first entry, ϑ_0 , is set to 1. This entails that location effects will always be present.⁷

From a frequentist' perspective equation (21) is our oracle estimator for β_p for a given quantile p , which, given that the ALL approximation in equation (5) holds, can be compared to the mean of the posterior of equation (7) (Kozumi & Kobayashi, 2011). With this in mind, it is trivial to calculate the true β_p 's for the error generating processes considered.

The second dimension along which the Data Generating Processes (DGP) differ is in their error process. The proposed DGPs can be grouped into two broad cases: (1) i.i.d. errors (y_1 and y_2) and (2) heteroskedastic errors (y_3 and y_4). In y_1 , we assume that the error distribution follows a standard normal distribution and in y_2 the error has Student's t distributed errors with 3 degrees of freedom. For the other cases, we assume simple heteroskedasticity caused by correlation between the second covariate (whose coefficient we denote as $\beta_{p,1}$) and ϵ . Lastly, y_4 can be thought of as containing a mixture between a uniform and a standard normal error distribution. In all simulations, the design matrix is simulated using a multivariate normal distribution with mean 0 and a covariance matrix with its (i, j) th element defined as $0.5^{|i-j|}$.

Relating the assumed error processes to the random coefficient representation (21), it is clear that, under i.i.d. errors, only the constant has a non-constant quantile function caused by F^{-1} (hereinafter called location shifters). Under the heteroskedastic designs, apart from the constant, $\beta_{p,1}$ will have a non-constant quantile function as well. Hence, $\beta_{p,1}$ in y_3 is determined by $F_{N(0,1)}^{-1}$ across p , and $\beta_{p,1}$ in y_4 follows $F_{U(0,2)}^{-1}$, i.e. increases linearly with p . The simulation designs (and the true quantile functions) are summarised in Table 1.

We evaluate the performance of the estimators in terms of bias in the coefficients, forecast error, implied degree of sparsity, and sampling diagnostics. Using the true quantile profile in β_p in equation (21), we calculate root mean coefficient bias (RMCB) and root mean squared forecast error (RMSFE) as

1. Root mean coefficient bias = $\sqrt{\frac{1}{\text{iter}} \|\hat{\beta}_p - \beta_p\|_2^2}$,
2. Root mean squared forecast error = $\sqrt{\frac{1}{\text{iter}} \|X\hat{\beta}_p - X\beta_p\|_2^2}$,

⁷ While it is possible for ϑ to take on any value, for simplicity we assume that the elements of ϑ only to take on the values $\{0, 1\}$.

Table 1. Summary of simulation set-ups

Data generating process	Error distributions	Quantile functions
$y_1 = X\beta + \epsilon$	$\epsilon \sim N(0, 1)$	$\beta_{p,0} = \beta_0 + F_{N(0,1)}^{-1}(p)$
$y_2 = X\beta + \epsilon$	$\epsilon \sim T(3)$	$\beta_{p,0} = \beta_0 + F_{T(3)}^{-1}(p)$
$y_3 = X\beta + (1 + X_2)\epsilon$	$\epsilon \sim N(0, 1)$	$\beta_{p,0} = \beta_0 + F_{N(0,1)}^{-1}(p)$
		$\beta_{p,1} = \beta_1 + F_{N(0,1)}^{-1}(p)$
$y_4 = X\beta + \epsilon_1 + X_2\epsilon_2$	$\epsilon_1 \sim N(0, 1)$	$\beta_{p,0} = \beta_0 + F_{N(0,1)}^{-1}(p)$
	$\epsilon_2 \sim U(0, 2)$	$\beta_{p,1} = \beta_1 + F_{U(0,2)}^{-1}(p)$

where iter is the number of Monte Carlo experiments. For most cases, iter = 100, except for Block T_2 , where it is set to 20.⁸ Since the root mean squared forecast errors show the same tendencies as the root mean coefficient bias. The table summarising the forecast results can be found in the [online supplementary material](#).

Finally, we conduct three sets of robustness checks. To gauge the influence of the hyper-prior choices on the approximate posterior degree of sparsity, we calculate the average posterior effective model sizes (14)⁹ as

$$\widehat{m}_{\text{eff}} = \frac{1}{\text{iter}} \sum_{i=1}^{\text{iter}} \frac{1}{S} \sum_{s=1}^S \sum_{j=1}^K (1 - \kappa_{p,j}^{(s)}), \tag{22}$$

where S are the number of Markov Chain Monte Carlo (MCMC) draws after burn-in. [Piironen and Vehtari \(2017\)](#) point out that \widehat{m}_{eff} may also be used to diagnose the informativeness of the scale on the prior for v . Namely, when the likelihood only weakly identifies the degree of sparsity, a standard Half-Cauchy distribution for v causes the implied level of sparsity to diverge to K .

In normal observation models with very low sample sizes, [Polson and Scott \(2010\)](#) additionally show that using our prior $v \sim C^+(0, 1)$ may induce multi-modality in the posterior of the regression coefficients. Although multi-modality is often a feature of aggressive shrinkage by construction ([Piironen et al., 2020](#)), our second robustness statistic, \hat{R} ([Gelman & Rubin, 1992](#)), checks whether any difficult geometries in the posteriors cause issues for convergence in the posteriors of β_p (due to multi-modality or otherwise). We compute \hat{R} based on 4 chains with 5,000 samples each after burn-in:

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\beta_{p,j} | Y)}{W}}, \tag{23}$$

where $\widehat{\text{var}}^+(\beta_{p,j} | Y) = \frac{N+1}{N} W + \frac{1}{N} B$ is an estimator for the *within* (W) and *between* (B) chain sample variances.¹⁰ Under diffuse starting points for the parameters, [Gelman and Rubin \(1992\)](#) recommend to choose $\hat{R} = 1.1$ as an upper bound for convergent chains, although [Vehtari et al. \(2021\)](#) recommend an upper bound of 1.05 and even 1.01 for high-dimensional vectors.

Finally, we compute the number of effective independent samples from our and competing samples to gauge overall sampling efficiency:

$$\widehat{N}_{\text{eff}} = 1 - \frac{W - \frac{1}{M} \sum_{m=1}^M \rho_{l,m}}{\widehat{\text{var}}^+(\beta_{p,j})}, \tag{24}$$

where $\rho_{l,m}$ is the auto-correlation at lag l from chain $m \in (1, \dots, M)$.¹¹

⁸ The only estimator where there is a deviation from this is the BALQR where the variance–covariance matrix of the posterior coefficients was not invertible for some of the cases. This is indicative that the BALQR prior did not shrink enough

⁹ Since \widehat{m}_{eff} and all other diagnostic test statistics were computed based on multiple parallel MCMC chains, we have found that averaging over 20 simulations was sufficient.

¹⁰ See [Gelman and Rubin \(1992\)](#) for the estimators of W and B .

¹¹ The measures \widehat{N}_{eff} and \hat{R} are further broken down to zero and non-zero coefficients to get a better picture of the estimators performance. For further details, see the [online supplementary material](#).

3.1 I.i.d. distributed random error simulation results

The bias results for the three designs (sparse, dense, block) across a selection of quantiles are presented in [Table 2](#) and the results of the forecast performance can be found in the [online supplementary material](#). Here and in the following tables relating to coefficient bias in this section, the best performers are highlighted in bold for readability. To shed light on whether the estimators capture the variable's quantile function appropriately, we additionally show plots for variables with non-constant quantile curves for each quantile. The HS-BQR's plots are presented in [Figure 4](#). The line in the plots shows the average, while the shaded region highlights the 95% coverage of β_p values across the Monte Carlo runs.

[Table 2](#) shows that the HS-BQR performs competitively compared to the considered estimators in all i.i.d designs regardless of what type of sparsity structure is considered. In particular, for the sparse case, the HS-BQR provides the lowest coefficient bias for both y_1 and y_2 for all quantiles. The forecast results corroborate these findings with the HS-BQR providing the lowest root mean squared forecast errors among the estimators considered.

Table 2. Root mean coefficient bias: homoskedastic designs

p	y_1					y_2				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
T = 100										
Sparse										
HS-BQR	0.045	0.036	0.034	0.038	0.050	0.061	0.047	0.044	0.048	0.069
LBQR	0.051	0.044	0.050	0.074	0.146	0.073	0.052	0.063	0.090	0.170
BQRENET	0.046	0.042	0.053	0.080	0.113	0.067	0.048	0.055	0.083	0.136
BALQR	0.075	0.049	0.043	0.052	0.080	0.161	0.144	0.145	0.144	0.164
Dense										
HS-BQR	0.711	0.710	0.709	0.716	0.722	0.721	0.722	0.721	0.727	0.738
LBQR	0.780	0.731	0.728	0.773	0.816	0.782	0.741	0.721	0.773	0.849
BQRENET	0.739	0.676	0.679	0.716	0.781	0.746	0.700	0.694	0.735	0.790
BALQR	1.271	1.233	1.250	1.246	1.265	1.276	1.245	1.240	1.267	1.286
Block										
HS-BQR	0.747	0.752	0.754	0.760	0.766	0.752	0.754	0.760	0.762	0.769
LBQR	0.821	0.737	0.716	0.783	0.870	0.831	0.743	0.704	0.773	0.879
BQRENET	0.776	0.690	0.696	0.730	0.835	0.790	0.706	0.689	0.739	0.847
BALQR	0.682	0.669	0.671	0.670	0.687	0.680	0.668	0.666	0.670	0.686
T = 400										
Dense										
HS-BQR	0.136	0.116	0.112	0.115	0.133	0.216	0.158	0.151	0.158	0.219
LBQR	0.118	0.105	0.100	0.103	0.118	0.184	0.143	0.132	0.143	0.184
BQRENET	0.106	0.100	0.096	0.100	0.110	0.177	0.140	0.129	0.139	0.179
BALQR	0.111	0.100	0.096	0.100	0.110	0.195	0.151	0.138	0.150	0.196
Block										
HS-BQR	0.487	0.486	0.486	0.490	0.490	0.498	0.498	0.498	0.501	0.504
LBQR	0.560	0.536	0.544	0.549	0.554	0.558	0.543	0.536	0.549	0.570
BQRENET	0.513	0.504	0.507	0.518	0.514	0.527	0.526	0.506	0.519	0.534
BALQR	0.837	0.847	0.830	0.846	0.846	0.845	0.850	0.828	0.852	0.832

Note. HS-BQR = horseshoe Bayesian quantile regression; LBQR = Bayesian Lasso QR; BQRENET = Bayesian elastic net QR; BALQR = Bayesian adaptive Lasso QR.

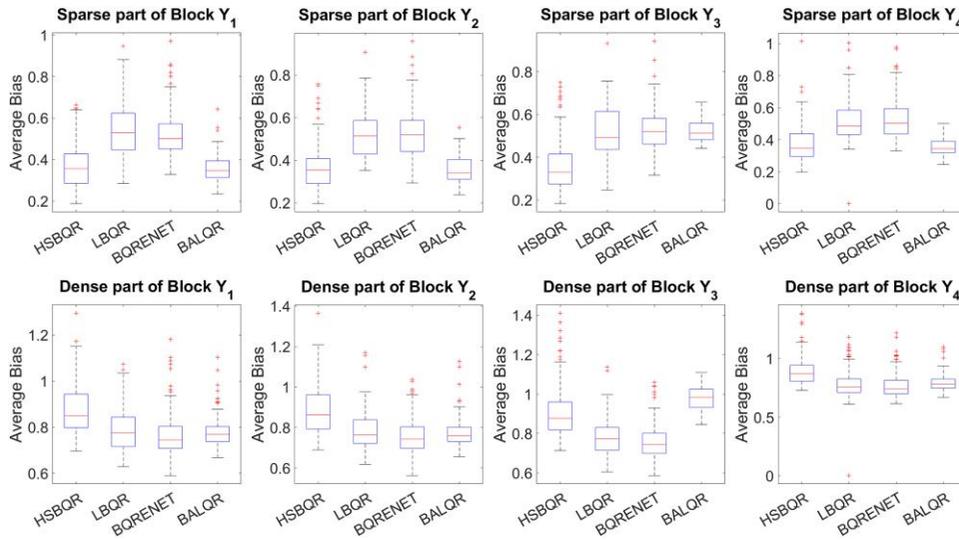


Figure 2. Average coefficient bias in the different parts of the block design ($T = 100$).

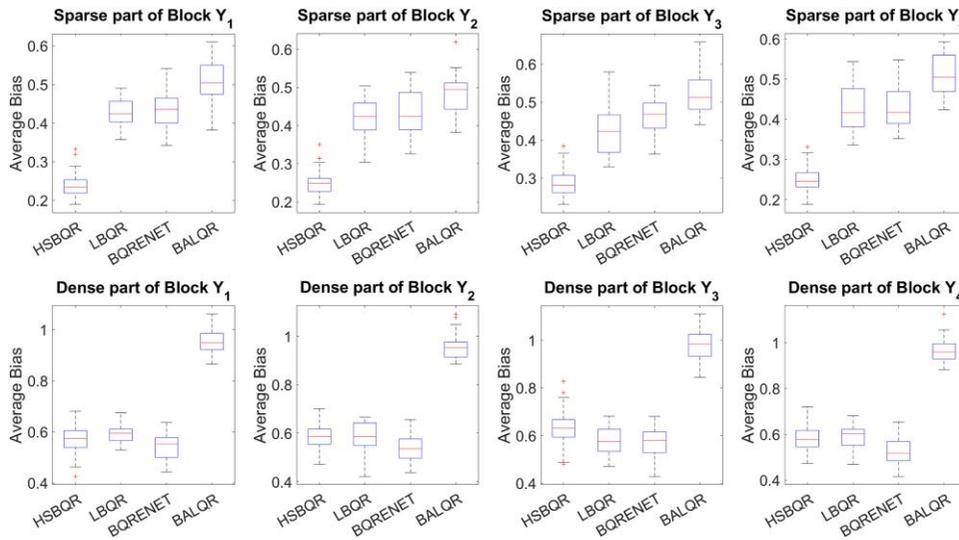


Figure 3. Average coefficient bias in the different parts of the block design ($T = 400$).

The HS-BQR’s performance is competitive for the dense and block cases as well, as can be seen in Table 2, however falls slightly short for the central quantiles to the BQRENET in the dense and to the BALQR in the block cases for T_1 . Forecast errors confirm these results. This coheres with the theoretical properties of the priors. The ridge component in the BQRENET provides better inference for dense designs, while the BALQR benefits in block structures from adaptive shrinkage without having to identify a global shrinkage parameter.

Figures 2 and 3 show the performance of the estimators at different parts of the block design for T_1 and T_2 , respectively. It reveals how the HS-BQR does extremely well in the sparse regions of the data for y_1 and y_2 , while not being able to replicate this performance in the dense regions of the data for T_1 . This is not to say that it performs poorly: while the HS-BQR yields higher average bias than the competing estimators, this is not statistically different from the bias of the other estimators. When more data are introduced in T_2 , the difference in bias for the dense parts become

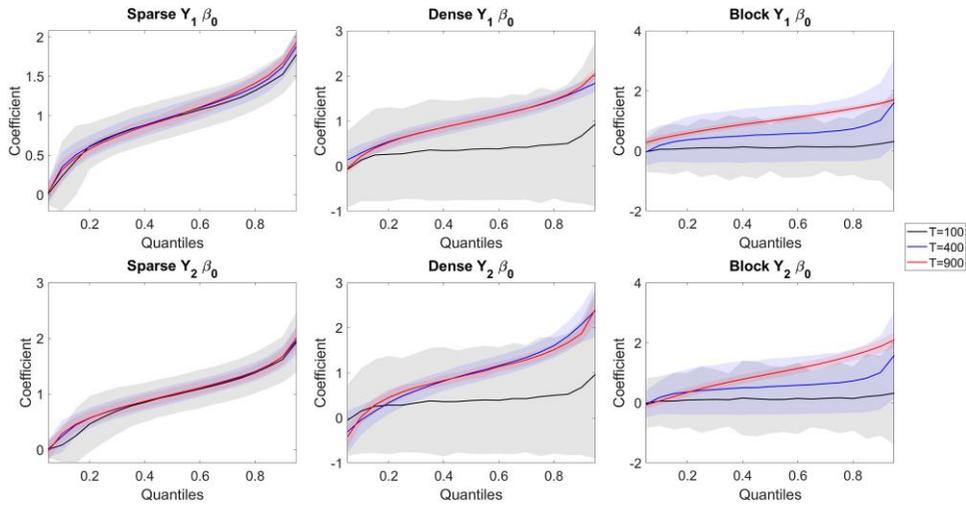


Figure 4. β_0 profiles for y_1 and y_2 across quantiles for the different sparsity settings.

even smaller among the different priors, while the sparse parts are estimated with considerably more accuracy for the HS-BQR.

Generally, as more data are introduced, the performance across the estimators converge to similar bias and forecast results, which confirms asymptotic validity of the priors and their samplers. An exception is presented by the BALQR which seems to fare worse with more data for the block design.

Both the normally distributed y_1 and t -distributed y_2 showcase a situation where the extreme quantiles (0.1 and 0.9) have higher bias than the central quantile (0.5) for all the estimators considered. This is a common finding in quantile regressions which is on account of more extreme quantiles being ‘data sparse’ as a few observations get large weights. While it is expected that there is a U-shape in the coefficient bias as we move across the quantiles, the slope of this shape is not uniform across the estimators. In particular, it can be seen in Table 2 that the HS-BQR’s bias does not increase as much as the other estimators.¹² Similarly, extreme quantiles generally tend to have higher forecast errors for all estimators, but the HS-BQR’s extreme quantiles do not suffer as much as its competitors, as shown in online supplementary Table A1. This property cannot be overstated, as quantile regression is often employed for extreme quantiles. The only case where the HS-BQR’s extreme quantiles performance are less accurate is for the dense design of T_2 , where the BQRENET’s performance does not suffer as much as the HS-BQR’s when considering the extreme quantiles.

Figure 4 underpins the findings of the tables: the HS-BQR captures the normal inverse CDF shape for y_1 and inverse t -distribution for y_2 very well in the sparse design for all T_i , however in the dense design, it only identifies location shifts for the more extreme quantiles for T_1 . Nevertheless, this property is fixed when more data is available. The figure also highlights how the HS-BQR struggles the most with block designs: It only captures the quantile profiles correctly for T_3 . This finding underpins, that in designs with unmodelled block structures and, hence, badly identified global shrinkage, quantile effects might be shrunk away. Implementation of group-level shrinkage along with prior information about the sparsity pattern in the data might be able to alleviate this problem, which we leave for future research.

3.2 Heteroskedastic error simulation results

As with the homoskedastic DGPs, we see that for all estimators, the error rate increases when moving away from the central quantiles and that coefficient bias as well as forecast accuracy worsens

¹² Apart from the HS-BQR in the block design of T_1 , where the estimators have lower coefficient bias and forecast error for its extreme low quantiles than its central quantiles.

Table 3. Root mean coefficient bias: heteroskedastic designs

p	y_3					y_4				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
T=100										
Sparse										
HS-BQR	0.061	0.069	0.084	0.101	0.132	0.043	0.043	0.059	0.082	0.119
LBQR	4.795	2.909	7.457	3.713	2.640	4.899	2.843	7.609	3.813	2.626
BQRENET	0.053	0.046	0.074	0.130	0.186	0.053	0.060	0.084	0.114	0.176
BALQR	0.515	0.513	0.525	0.512	0.584	0.281	0.301	0.275	0.300	0.320
Dense										
HS-BQR	0.767	0.763	0.764	0.771	0.773	0.764	0.766	0.774	0.780	0.786
LBQR	0.811	0.759	0.753	0.807	0.871	0.764	0.726	0.742	0.778	0.838
BQRENET	0.752	0.714	0.684	0.772	0.815	0.733	0.683	0.678	0.703	0.791
BALQR	1.307	1.287	1.283	1.287	1.309	1.268	1.260	1.254	1.264	1.287
Block										
HS-BQR	0.760	0.756	0.764	0.757	0.773	0.668	0.665	0.666	0.670	0.677
LBQR	0.803	0.717	0.708	0.750	0.858	0.766	0.713	0.708	0.799	0.863
BQRENET	0.700	0.693	0.692	0.742	0.845	0.749	0.706	0.699	0.744	0.818
BALQR	0.687	0.677	0.679	0.678	0.699	0.687	0.679	0.682	0.682	0.703
T=400										
Dense										
HS-BQR	0.409	0.328	0.313	0.332	0.406	0.136	0.122	0.132	0.154	0.188
LBQR	0.313	0.267	0.255	0.267	0.316	0.180	0.161	0.156	0.161	0.182
BQRENET	0.289	0.245	0.231	0.243	0.291	0.168	0.150	0.144	0.150	0.171
BALQR	0.355	0.290	0.272	0.291	0.359	0.185	0.163	0.156	0.163	0.188
Block										
HS-BQR	0.537	0.540	0.541	0.542	0.541	0.498	0.497	0.498	0.498	0.502
LBQR	0.567	0.566	0.533	0.574	0.576	0.532	0.557	0.543	0.559	0.558
BQRENET	0.534	0.558	0.537	0.527	0.547	0.505	0.497	0.497	0.509	0.526
BALQR	0.878	0.866	0.857	0.863	0.863	0.857	0.843	0.841	0.832	0.844

Note. HS-BQR = horseshoe Bayesian quantile regression; LBQR = Bayesian Lasso QR; BQRENET = Bayesian elastic net QR; BALQR = Bayesian adaptive Lasso QR.

for dense and block designs compared to the sparse design. Further, the bias results in Table 3 show that the HS-BQR provides competitive performance to the alternative estimators, where it consistently outperforms the other priors for y_4 in sparse designs.¹³ Similar to the previous discussion, the HS-BQR stands out in that it provides consistently more stable inference of extreme quantiles independent of the sparsity structure, with the exception of T_2 dense.

In dense designs, as for the homoskedastic simulation results, the BQRENET aided by the ridge component in the prior provides lower coefficient bias and forecast error than the HS-BQR, whereas in block DGPs, the BALQR outperforms the HS-BQR for y_3 in T_1 but not in T_2 . A different picture emerges for y_4 . Here, the HS-BQR's performance in coefficient bias is only rivaled by the BQRENET for both T_1 and T_2 for both dense and block cases.

Consulting Figures 2 and 3 for y_3 and y_4 shows how the HS-BQR performs particularly well in the sparse regions of the data, as was the case for y_1 and y_2 . Just like in the homoskedastic designs,

¹³ The LBQR does surprisingly poorly in the sparse T_1 heteroskedastic cases. This is on account of the estimator completely missing the quantile profiles for both y_3 and y_4 (see respective figures in the online supplementary material).

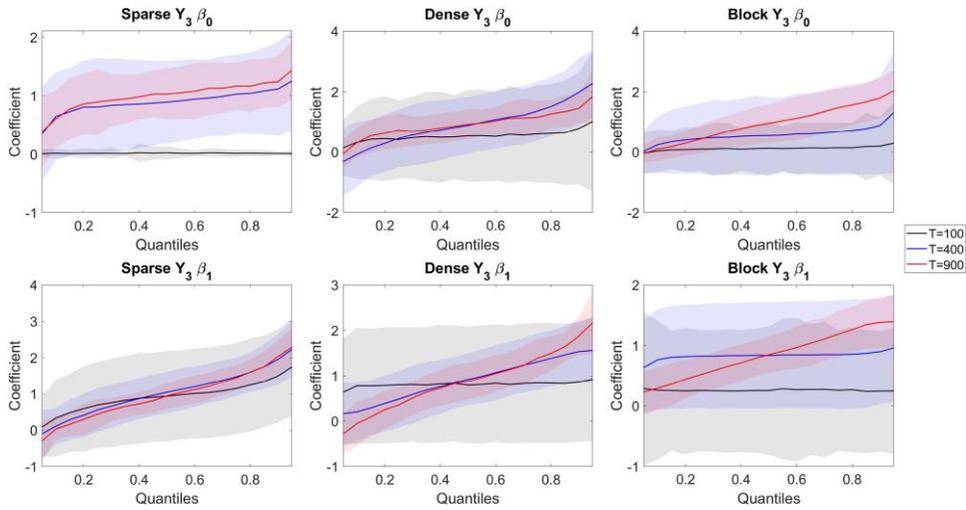


Figure 5. β_0 and β_1 profiles for y_3 across quantiles for the different sparsity settings.

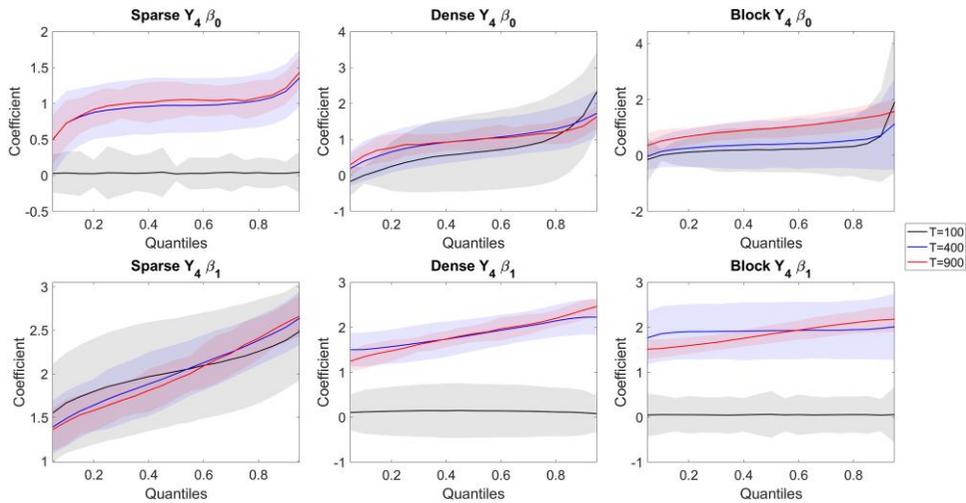


Figure 6. β_0 and β_1 profiles for y_4 across quantiles for the different sparsity settings.

the HS-BQR is not able to replicate its performance in the dense regions of the data for T_1 , but it does not do much worse than the competing estimators. Similarly, the HS-BQR's bias for sparse parts of the block DGP's are far smaller, while for the dense parts, it is on par with the other estimators.

The plots in Figure 6 provide another explanation as to why the HS-BQR's forecast performance is much better for the block case of y_4 , which is that it captures some aspects of the quantile function for $\beta_{p,0}$, even for the smallest data setting T_1 . The plots in Figures 5 and 6 also highlight why the estimators have lacklustre performance for y_3 and y_4 for T_1 even for the sparse designs: The estimators have difficulties identifying the quantile profiles of $\beta_{p,0}$ and $\beta_{p,1}$ simultaneously. This deficiency is amended with more data as shown by the plots for T_2 and T_3 : The HS-BQR captures the quantile profiles for both the sparse and dense DGPs; however, its performance on the block design only gets better for T_3 . This shows the scale at which the methods require data to identify the correct quantile profiles of the variables when the DGP contains mixed sparsity structures. This shows the scale at which the methods require data to identify the correct quantile profiles of the variables when the DGP contains mixed sparsity structures.

The simulations have shown that the HS-BQR provides competitive results but also that all quantile methods under consideration have difficulty simultaneously identifying the true regressors and partialling out the location ($\beta_{p,0}$) and scale ($\beta_{p,1}$) effects in high-dimensional setting especially when data are not abundant.

3.3 Robustness measures

The diagnostic check measures are presented in Table 4 for y_1 which is shown here as representative for the other simulations given in the [online supplementary material](#).

First, the posterior estimates for \widehat{m}_{eff} clearly show that the horseshoe prior approximates indicates the degree of sparsity of the DGPs well for all simulations designs. In accordance with the trends observed for in Sections 3.1 and 3.2, we see that that HS-BQR provides the most accurate \widehat{m}_{eff} estimates for the sparse designs and is close to the best performers for the other designs. The effective model size estimates for Lasso type shrinkage priors, however, need to be interpreted with caution as the prior densities for $\kappa_{p,j}$ do not favour mass on 0 or 1.¹⁴

Second, we do not observe any difficulties in convergence of the MCMC chains for the HS prior, \hat{R} is well below even 1.01 for all quantiles and sparsity settings. Indeed, the proposed sampling scheme for the horseshoe prior (Section 2.3) is the only one that clears the benchmark thresholds for convergence in every simulation.

Lastly, \widehat{N}_{eff} results highlight that the proposed sampler is very efficient compared to the other samplers, particularly for sparse designs.

3.4 Computation time

In this section, we compare computational efficiency to publicly available implementations of the competing priors.¹⁵ The main computational burden in sampling the posterior of the BQR, is inverting the variance–covariance matrix $\bar{\Lambda}_*$, whose computational complexity can be reduced with the fast HS-BQR sampler (Section 2.3) to $O(T^2K)$ from $O(K^3)$. As such, we expect our algorithm to scale well with the dimension of the covariates.

As an exercise, we report in Figure 7 the time it takes to iterate once through the entire Gibbs sampler sequentially for 19 equally spaced quantiles for increasing dimensions, K . For simplicity, we consider Monte Carlo simulation design y_1 , sparse and take the average time from 100 Gibbs loops.¹⁶ These experiments were conducted on an AMD Ryzen 9 4900HS@3 GHz and 16 GB of RAM. To account for speed differences across programming languages, we will present results for HS-BQR executed in MATLAB as well as in R to ensure that the choice of the coding language does not have an undue influence on the results. The results in MATLAB are denoted as HS-BQR_M, while the results in R are represented as HS-BQR_R. Further computation time experiments can be found in the [online supplementary material](#). Note that for readability, we take the \log_{10} of the run times.

As expected, the figure reveals that the proposed sampling algorithm scales much better with the number of covariates, independent of the implementation in R of MATLAB. Hence, not only is the proposed sampler and prior combination statistical efficient, but also computationally efficient as the covariate dimension increases.

4 GaR application

We now compare the HS-BQR to the same set of competing estimators as above in estimating forecast densities of US quarterly GDP growth as well as its down- and upside risks, commonly referred to as GaR. The need for GaR was highlighted by the global financial crisis which showed how downside risks, so the lower quantiles of the density of GDP growth, evolve with the state of credit and financial market (Adrian et al., 2019; Prasad et al., 2019). Quantifying this vulnerability is of key interest of policymakers, as it is well known that recessions caused by financial crises are often more severe than ordinary recessions (Jordà et al., 2015).

¹⁴ See Kohns and Szendrei (2021), who show how to obtain actual model sizes for (Bayesian) quantile regression with any continuous prior, using a decision theoretic approach.

¹⁵ See Alhamzawi and Ali (2020) for an R package implementing the adaptive Lasso, and see <https://sites.wustl.edu/nanlin/publications/> for codes implementing the priors of Li et al. (2010).

¹⁶ Computation time per iteration does not change significantly as more iterations were are run with our CPU.

Table 4. Robustness measure results for γ_1 for $T = 100$

p	\hat{m}_{eff}														
	Sparse ($K = 6$)				Dense ($K = 201$)				Block ($K = 401$)						
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS-BQR	5	7	8	9	9	148	138	136	138	148	305	268	268	272	308
LBQR	25	24	29	33	29	128	155	160	155	128	365	443	458	447	376
BQRENET	14	29	47	69	83	160	179	182	179	161	471	526	537	530	471
BALQR	155	134	129	133	156	136	128	126	127	137	410	382	375	382	410
\hat{R}															
HS-BQR	1.003	1.002	1.001	1.001	1.002	1.002	1.003	1.003	1.003	1.002	1.006	1.007	1.007	1.006	1.006
LBQR	1.207	1.195	1.193	1.195	1.190	1.648	1.644	1.646	1.646	1.640	1.495	1.489	1.474	1.481	1.496
BQRENET	1.247	1.237	1.239	1.238	1.242	1.738	1.736	1.740	1.734	1.747	1.586	1.579	1.575	1.582	1.584
BALQR	1.015	1.013	1.013	1.013	1.013	1.451	1.464	1.453	1.473	1.439	1.021	1.020	1.017	1.019	1.021
\hat{N}_{eff}															
HS-BQR	5510	7626	6987	7004	6728	2916	2585	2514	2617	3064	1540	1424	1421	1487	1502
LBQR	53	55	55	54	54	13	10	10	10	10	41	40	43	42	40
BQRENET	259	277	274	273	268	11	9	9	9	9	41	41	41	41	40
BALQR	538	561	569	546	553	57	21	22	21	22	393	409	463	414	369

Note. HS-BQR = horseshoe Bayesian quantile regression; LBQR = Bayesian Lasso QR; BQRENET = Bayesian elastic net QR; BALQR = Bayesian adaptive Lasso QR.

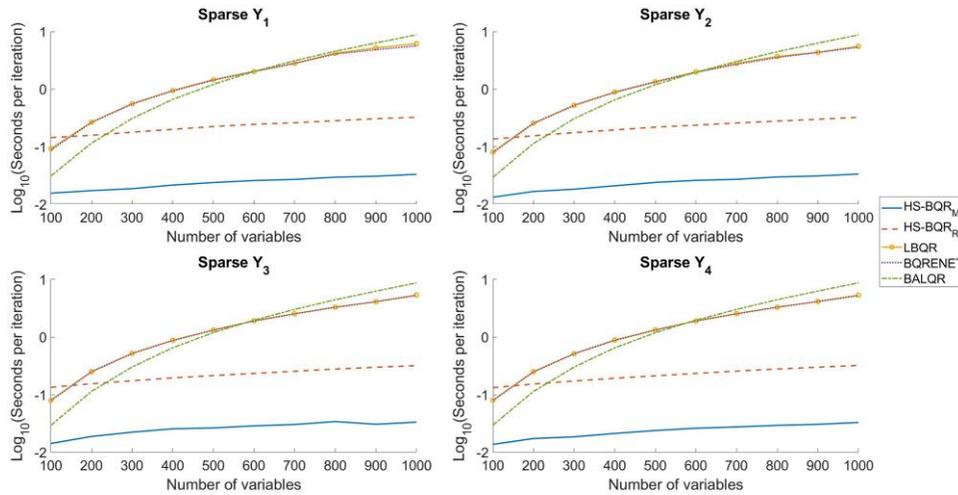


Figure 7. Computation time as a function of variable numbers. Seconds per Gibbs iteration are averages over 100 cycles. $T = 100$.

Unlike much of the previous GaR literature which focuses on GDP growth density forecasts based on only one indicator of financial distress, we apply the HS-BQR to forecasting the entire conditional GDP density using the McCracken database, a large macroeconomic data set. The ability to produce well-calibrated density forecasts in the face of large data contexts is important for nowcasting applications, in which the information flow is necessarily high-dimensional, or variable selection of large amount of competing uncertainty indexes. The latter purpose has been suggested by Adams et al. (2021) and Figueres and Jarociński (2020) who have argued that is not a priori clear which index of market frictions impacts GDP growth the most.

The McCracken and Ng (2020) database¹⁷ consists of 248 variables (including GDP) from 1959 Q1 at a quarterly frequency and is updated monthly. We take the quarter-on-quarter growth rate of annualised 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 variables start at 1959 Q1, for the GaR application, only variables that are available from 1970 Q1 were considered which gives 229 explanatory variables.

To obtain the forecasts, we use the general linear model:

$$y_{t+h} = x_t' \beta_p + \epsilon_{t+h} \tag{25}$$

for $t = 1, \dots, T - h$, where h refers to the forecast horizon. We consider one- to four-quarter ahead forecast horizons ($h = 1, \dots, 4$). Using the quantile set-up, forecasts from each quantile are denoted as $y_{T+h|T}^p$. Note that these h -step-ahead forecasts are equivalent to the h -step-ahead p th VaR. Forecasts are computed on a rolling basis where the initial in-sample period uses the first 50 observations of the sample, which makes for 149-h rolling forecast windows. We estimate a grid of 19 equidistant quantiles to construct the predictive density $p(\hat{y}_{T+h|T})$ via a normal kernel.¹⁸

Forecast densities are evaluated along Kolmogorov–Smirnov (KS) statistics based on (unsorted) probability integral transforms (PIT) and average log-scores.¹⁹ The PIT is often used when evaluating density forecasts and provides a measure of calibration which is independent of the

¹⁷ <https://research.stlouisfed.org/econ/mccracken/fred-databases/>

¹⁸ Alternatively, one could follow the popular density construction approach by Adrian et al. (2019) who fit their quantiles to a skewed t -distribution. We argue when discussing the results that this approach is less flexible than the proposed approach.

¹⁹ There are a plethora of tests to evaluate distributions based on QQ-plot of the PIT. The choice of the KS was based solely on its simplicity to compute and any other test would suffice for evaluation.

econometricians loss function. In particular, the PIT is the corresponding CDF of the density function evaluated at the actual observation of the out-of-sample periods, y_{t+h} :

$$g_{t+h} = \int_{-\infty}^{y_{t+h}} p(u | y_{t+h}) du = P(y_{t+h} | y_t). \quad (26)$$

The estimated predictive density is consistent with the true density when the CDF of g_{t+h} form a 45° line (Diebold et al., 1998), i.e. forms the CDF of a uniform distribution. Deviation from uniformity is tested via the KS test.

The unsorted PIT-based test is very close to what is referred to as hit-rate test in the VaR literature. An additional advantage of this test is that it allows one to gauge, whether quantile-crossing is a serious problem of the estimator. Ideally, we want the forecasted quantiles to be monotonically increasing. When this monotonicity is violated, our estimated density is invalid. Any violation of this monotonicity in our QQ-plot is a clear indication that the estimator yields quantiles that frequently cross.

Second, we compare density fit via average log-scores. Log-scores provide a strictly proper scoring rule in the sense of Gneiting and Raftery (2007) and take into account location, skewness, and kurtosis of the forecast distribution (Gelman et al., 2013). Since quantile-crossing may lead to non-sensical density forecasts, before calculating the log-scores we sort the estimated quantiles and perform kernel smoothing to obtain $p(y_{t+h})$. Average log-scores are then calculated as follows:

$$\log S_b = \frac{1}{T-b-1} \sum_{t=1}^{T-b-1} \log p(y_{t+h} | y_t). \quad (27)$$

We break from the forecast density literature a bit, by not exclusively focusing on testing the whole density, but also evaluating specific quantiles' performance as well. To appraise the HS-BQR compared to the alternative estimators, the pseudo- R^2 for the quantiles is computed,²⁰ following Koenker and Machado (1999). The pseudo- R^2 of the following regression is obtained from:

$$Q_{y_{t+h}}(p | V_{t+h,p}) = \beta_0 + \beta_1 V_{t+h,p}, \quad (28)$$

where $V_{t+h,p}$ is the fitted value of of the estimator for the p th quantile. Running the regression in equation (28) for the p th quantile gives an intuitive test for the ability of the estimated fitted value to capture the dynamics we are interested in. In particular, the pseudo- R^2 is calculated the following way:

$$R^2 = 1 - \frac{\text{RASW}}{\text{TASW}} \quad (29)$$

where RASW is the residual absolute sum of weighted differences, so the residuals of equation (28) and TASW is the total absolute sum of weighted differences, so the residuals of equation (28), where β_1 is constrained to 0. In essence, the pseudo- R^2 shows how much information $V_{t,p}$ adds to the regression compared to a quantile regression with only a constant.

To gain a visual understanding of how the forecast densities perform over time, Figure 8 plots in its left panel the one-step-ahead forecast densities of the HS-BQR, and the right panel shows all other competing estimators. The figure highlights that the HS-BQR provides better calibration especially in the beginning period of the forecast evaluation during which upper, lower, and middle quantiles span a reasonable range of values despite the relative scarcity of observations to number of covariates. It is clear from the right panel that precisely in the early forecast periods, the Lasso-based priors offer too little regularisation, yielding far too extreme upper and lower quantile

²⁰ Since GaR is meant to be a VaR of growth, utilising tests designed to test the adequacy of VaR models is a natural extension for evaluation. Two popular tests to verify the performance of a specific quantile are the DQ test of Engle and Manganelli (2004) and the VQR test of Gaglianone et al. (2011). These tests provide a principled way of testing the null hypothesis of the selected quantile being correct. However, they do not offer a comparative measure as to how much better the proposed method provides better fit for a specific quantile.

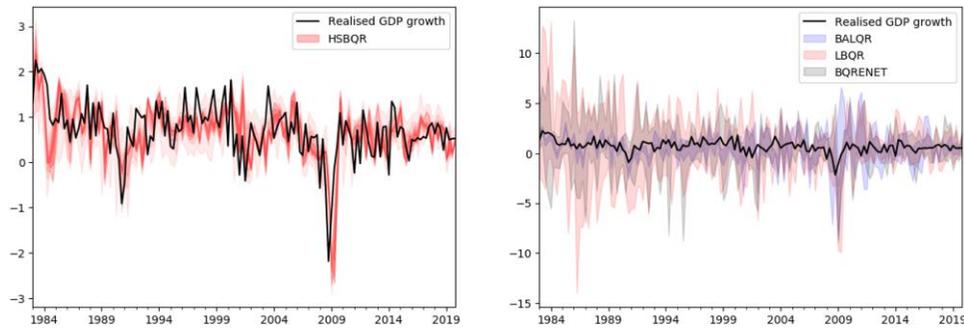


Figure 8. One-step-ahead forecast distributions for the L1QR, BQR, BALQR, and HS-BQR. Shaded areas correspond to plots of all 19 quantiles.

growth forecasts. As more data comes in, the right panel shows that the extreme variability of the Lasso-based estimators decreases somewhat over time, but the HS-BQR provides good fit throughout the entire evaluation period. The two-quarter, three-quarter, and four-quarter ahead forecast densities are presented in the [online supplementary material](#) and confirm these findings.

The visual inspection is corroborated by the more formal PIT-based KS statistics and average forecast log-scores in [Table 5](#): the KS statistics show that the HS-BQR is the only estimator to provide forecasts densities whose PIT are statistically indistinguishable from a uniform CDF at the 10% significance level, and whose log-scores are highest for all but the two-quarter ahead horizon. As expected, the test statistics as well as the QQ-plots of the PITs plotted in [online supplementary material](#) indicate that as the forecast horizon increases to two and three quarters, density calibration deteriorates somewhat for all estimators. Contrary to [Carriero et al. \(2020\)](#) and [Mazzi and Mitchell \(2019\)](#), however, we find that density fit increases again at the fourth horizon, which suggests that the HS-BQR is useful not only for short-term density forecasts but also for medium-term forecasts.

An additional feature of the HS-BQR forecasts is that they exhibit limited quantile-crossing problem. The HS-BQR’s forecasted quantiles exhibit very little quantile-crossing, especially when comparing it to the alternative estimators. In fact, in the one-step-ahead case, the HS-BQR is the only estimator that yields non-crossing quantiles.

To quantify the relative performance of the estimators in capturing tail risks, we show in the third panel of [Table 5](#) estimates of the pseudo- R^2 which are calculated as in equation (28) for the extreme and middle quantiles. It is apparent that not only does the HS-BQR provide better quantile fit at all shown quantiles, but that the largest margin (compared to the other estimators) is at the lowest and highest quantiles at all horizons which echos the results from the simulations. This is corroborated by the PIT graphs, which show that the HS-BQR’s tail quantiles are consistently the closest to the ideal 45° line.

The proposed estimator also provides competitive point forecasts which are shown in the right panel of [Table 6](#) for the 50th quantile. As shown by [He et al. \(1990\)](#), median quantile forecasts are more robust to outliers than conditional mean forecasts. [Table 6](#) clearly shows that the HS-BQR offers sizeable improvements in root mean squared forecast error over the competing quantile models of 25%–66%, which are all statistically significant as per the [Diebold and Mariano \(2002\)](#) test. [Table 7](#) additionally shows that with the whole sample, the HS-BQR has no problems and high effective sample sizes, which are similarly to the simulation results much more efficient than the competing priors. Since the model sizes clearly also do not diverge to K , we can also safely dismiss any problems coming from scaling of the global shrinkage parameter, τ .

To showcase how these improvements translate to actual events of importance to policymakers, we plotted density forecasts at all horizons right before NBER marked recession or trough dates. We concentrate on the quarters before the height of each individual crisis, as the recent [Gar literature](#) highlights the usefulness of quantile methods to detect vulnerabilities to parts of the economy before these vulnerabilities materialise ([Adrian et al., 2019](#)). Representative for all other

Table 6. Average log-scores and RMSFE on the median of the different estimators for h -step-ahead quantile forecasts

	Average log-scores				Median RMSFE			
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
HS-BQR	3.432	3.282	3.472	3.534	0.006	0.004	0.004	0.004
LBQR	3.235	3.294	3.396	3.431	0.010***	0.012***	0.008***	0.008***
BQRENET	2.999	3.369	3.426	3.430	0.009***	0.007***	0.007***	0.008***
BALQR	1.839	2.069	2.154	2.133	0.015***	0.012***	0.010***	0.007***
SPF	3.083	3.276	3.285	3.185	0.005	0.004	0.004	0.004

Note. Density is approximated by a normal kernel of the 19 forecasted quantiles. For the RMSFE stars indicate statistical difference to the HS-BQR median forecasts based on the Diebold–Mariano test (1998) at 1% (***) significance, respectively. HS-BQR = horseshoe Bayesian quantile regression; LBQR = Bayesian Lasso QR; BQRENET = Bayesian elastic net QR; BALQR = Bayesian adaptive Lasso QR; SPF = Survey of Professional Forecasters.

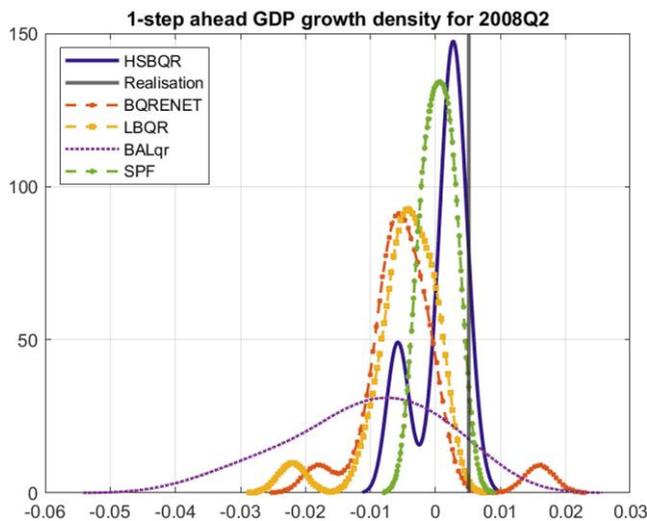


Figure 9. Q2 2008: Smoothed forecast densities of all competing estimators and the Survey of Professional Forecasters. Densities are estimated via a Gaussian kernel of 19 equidistant forecasted quantiles. The growth realisation is marked by a vertical thick line.

pre-crisis period shown in [online supplementary material](#), [Figure 9](#) shows forecast densities for Q2 2008. The actual realisation is marked by a vertical grey line. Two points emerge from this graph: the HS-BQR provides the largest mass at the actual realisation of growth (which translates to the highest density fit for this realisation as measured by the log-score) and it provides a bi-modal distribution which yields a policy relevant characterisation of forecasted risk. The second mode hovers over negative growth outcomes, thereby giving a clear indication of risks of a recession. Compared to the HS-BQR, the competing quantile methods do provide mass on negative growth outcomes which is corroborated by [Carriero et al. \(2020\)](#) and [Mazzi and Mitchell \(2019\)](#) however, provide little, or close to no mass on the actual realisation. In fact, consulting [Figure 8](#), one can see that the Lasso-based BQR methods throughout the entire forecast evaluation period provide mass on negative growth outcomes, in other words forecast positive probability of recessions. This is less confidence inspiring than the forecast densities of the HS-BQR which are more conservative with mass on negative growth outcomes. To argue that this is not an artefact of the kernel smoothing, we provide forecast densities for relatively ‘tranquil’ economic times, namely 2005 Q1, in the [online supplementary material](#). For these forecast densities, the HS-BQR combines to a uni-modal,

non-skewed, normal looking forecast density with high mass on the realisation. This highlights an advantage of quantile smoothing compared to [Adrian et al. \(2019\)](#) approach of fitting the quantiles to a t -distribution. By smoothing the 19 forecasted quantiles via a kernel, we impose no restrictions on the number of modes or degree of skewness of the combined density. Finally, to compare the utility of the HS-BQR approach to a widely used forecast density constructed by survey expectations, we plotted the Survey of Professional Forecasters distribution (SPF) into the same density graphs.²¹ From [Figure 9](#), one can see that the HS-BQR not only outperforms the SPF but provides a better indication of the looming recession indicated by larger mass on negative growth outcomes.

Lastly, [Table 7](#) shows that while none of the priors showcase bad convergence of the MCMC chains, the horseshoe model clearly outperforms the other models in terms of efficiency of the algorithm, which corroborates the findings from the simulation study. Likewise, we do not observe any problems with respect to the effect of the global shrinkage parameters as effective model sizes point clearly to a sparse set of indicators having non-zero coefficients. Similar to [Kohns and Szendrei \(2021\)](#), we observe that model sizes vary across quantiles.

The ability to produce well-calibrated density forecasts and, especially accurate downside risk measures in the face of large data contexts makes the HS-BQR a powerful tool for nowcasting applications or variable selection of large amount of competing uncertainty indexes.

5 Conclusion

In this paper, we have extended the widely popular horseshoe prior of [Carvalho et al. \(2010\)](#) to the BQR and provided a new algorithm to sample the shrinkage coefficients via slice sampling for the independent prior and a fast sampling algorithm that speeds up computation significantly in high dimensions.

In our simulations, we considered a variety of sparse, dense, and block designs with different error distributions which revealed three points about the HS-BQR. First, the HS-BQR provides better or comparable performance in terms of both coefficient bias and forecast risk where best performance can be expected for sparse designs. Second, the aggressive shrinkage profile of the HS-BQR leads to especially good performance in tail estimation (0.1 and 0.9). Finally, an issue that all BQR methods share is simultaneously identifying the correct location and scale effects in high-dimensional setting.

Our empirical application shows that the HS-BQR provides considerable gains in calibration, density fit and even point estimates compared to double exponential-based priors at all horizons, especially so at short-term, $h = 1$, and medium-term, $h = 4$, horizons. Local measures of fit confirmed that HS-BQR's fitted quantiles provide the best goodness of fit. The HS-BQR proved especially useful right before NBER marked recession and trough dates, providing forecast densities foreshadowing crises. This shows that the HS-BQR is an adequate method to give credible VaR estimates. We expect therefore that the HS-BQR performs well in nowcasting settings such as [Carriero et al. \(2020\)](#) and [Mazzi and Mitchell \(2019\)](#) which we leave for future research.

The results show that the HS-BQR is a competitive estimator for which especially good behaviour can be expected in sparse designs with few observations. However, there are multiple fronts on which the proposed HS-BQR can be improved upon. For instance, the simulations highlighted that in dense and block designs, the HS prior tends to shrink the constant too aggressively. Hence, extensions which allow for differing shrinkage terms for subsets of the regressors might be able to alleviate this problem. Extensions to the HS-BQR should also address the problems of simultaneously estimation location and scale effects as this is needed to attain oracle properties in quantile regression. Further, the analysis of the approximate degree of sparsity via effective model sizes prompts an extension to make priors about effective model sizes explicit via a regularised horseshoe prior akin to [Pironen and Vehtari \(2017\)](#) for Generalized Linear Models (GLM). Such a prior would scale $\pi(v)$ appropriately with a prior guess on the number of active coefficients. And lastly, it would be useful to extend the analysis of this paper to the plethora of other global–local priors such as [Bhadra et al. \(2019\)](#) and [Cadonna et al. \(2020\)](#).

²¹ Smooth densities have been estimated based on a normal kernel over 19 equidistant quantiles of the survey.

Table 7. Robustness measures

p	\widehat{m}_{eff}					\widehat{R}					\widehat{N}_{eff}				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS-BQR	5	11	9	8	10	1.002	1.002	1.002	1.004	1.002	6196	6978	6787	5929	7156
LBQR	8	17	20	19	12	1.022	1.008	1.007	1.010	1.043	201	503	566	428	107
BQNET	6	12	23	30	13	1.017	1.020	1.021	1.031	1.037	518	1200	811	378	150
BALQR	216	214	214	214	216	1.002	1.001	1.001	1.001	1.002	2159	3521	4251	3661	2158

Note. \widehat{m}_{eff} , \widehat{R} , and \widehat{N}_{eff} are based on 4 MCMC chains with 5,000 iterations as burn-in and 5,000 kept for inference.

Acknowledgments

The authors thank Arnab Bhattacharjee, Atanas Christev, Laurent Ferrara, Gary Koop, Aubrey Poon, Giovanni Ricco, Mark Schaffer and all participants of the International Symposium on Forecasting (2020) in Rio de Janeiro, all participants of the Scottish Graduate Program in Economics conference in Crieff (2020), as well as all participants of the PhD Conference at Panmure House in Edinburgh (2020) for their invaluable feedback. The authors would also like to thank Nial Friel and two anonymous referees for their comments and suggestions. The usual disclaimer applies.

Conflict of interest: None declared.

Funding

Tibor Szendrei is grateful to the Economic and Social Research Council (ESRC, UK) for awarding him a PhD studentship.

Data availability

Data are publicly available from the projects GitHub page at <https://github.com/davkoh/Horseshoe-Bayesian-Quantile-Regression>. Data for the empirical application have been retrieved from the [McCracken and Ng \(2020\)](#) database.

Supplementary material

[Supplementary material](#) is available online at *Journal of the Royal Statistical Society: Series C*.

Appendix A

A.1 Derivation of Algorithm 1

Proof. Using the Sherman–Morrison–Woodbury identity and some further algebra, we get $\mu = D\Phi'(\Phi D\Phi' + I_T)^{-1}(\alpha - \zeta Z)$. Plugging in the identity from step 2 into step 3, we obtain $\theta = u + D\Phi'(\Phi D\Phi' + I_T)^{-1}(\alpha - \zeta - \zeta Z)$. Since by definition $\zeta \sim N(0, \Phi D\Phi' + I_K)$, θ follows a normal distribution with mean $D\Phi'(\Phi D\Phi' + I_K)(\alpha - \zeta Z) = \mu$. As $\text{cov}(u, \zeta) = D\Phi'$, it follows that $\text{cov}(\theta) = D - D\Phi'(\Phi D\Phi' + I_K)^{-1}\Phi D$ which by the Sherman–Morrison–Woodbury identity is equal to Θ . The provided algorithm is not specific to the horseshoe prior and follows through for any prior of the form in equation (10). The computational advantage provided in Algorithm 1 compared to Cholesky-based decompositions is that we can cheaply sample from $(u, \zeta)'$ which via data augmentation yields samples from the desired distributions. \square

We now give further details on the derivation of Algorithm 1. The goal of the algorithm is to circumvent having to compute large $K \times K$ matrices by redefining auxiliary variables which under certain linear combination result in draws of the desired distribution $N(\bar{\beta}_p, \bar{\Lambda}_*^{-1})$. As above, by the Sherman–Morrison–Woodbury theorem (see, e.g. [Hager, 1989](#)), Θ and μ can be expanded as

$$\begin{aligned}\Theta &= (\Phi'\Phi + D^{-1})^{-1} = D - D\Phi'(\Phi D\Phi' + I_T)^{-1}\Phi D, \\ \mu &= D\Phi'(\Phi D\Phi' + I_T)^{-1}\alpha,\end{aligned}$$

where the second equality follows from applying the push-through identity after some re-arranging of terms. This expansion per se would not help in sampling from $N(0, \Theta)$. Letting ζ and u being defined as above, $\omega = (\zeta', u')' \in \mathbb{R}^{T+K}$ follows a multivariate normal distribution centred on 0 with covariance

$$\Omega = \begin{pmatrix} P & S \\ S' & D \end{pmatrix},$$

where it is easily verified that $P = (\Phi D \Phi' + I_T)$ and S' can be derived as:

$$\begin{aligned} \text{Cov}(\zeta, u) &= \text{Cov}(D^{1/2}\epsilon, \Phi u) \\ &= E(D^{1/2}\epsilon u' X' U^{1/2}) \\ &= E(D^{1/2}\epsilon \epsilon' D^{1/2} X' U^{1/2}) \\ &= D X' U^{1/2} \\ &= D \Phi', \end{aligned}$$

where $\epsilon \sim N(0, I_K)$. The second equality follows from the fact that both quantities are independently centred on 0. Rewriting Ω into its LDU decomposition (see, e.g. Hamilton, 1994) as

$$\begin{pmatrix} P & S \\ S' & D \end{pmatrix} = \underbrace{\begin{pmatrix} I_T & 0 \\ S' P^{-1} & I_K \end{pmatrix}}_L \underbrace{\begin{pmatrix} P & 0 \\ 0 & D - S' P^{-1} S \end{pmatrix}}_\Gamma \underbrace{\begin{pmatrix} I_T & P^{-1} S \\ 0 & I_K \end{pmatrix}}_{L'}$$

where the lower $K \times K$ block in Γ is equal to Θ using the Sherman–Morrison–Woodbury identity. To retrieve the lower part, we isolate Γ which is easily obtained because L is lower triangular and thus the inverse is readily available as

$$L^{-1} = \begin{pmatrix} I_T & 0 \\ -S' P^{-1} & I_K \end{pmatrix}.$$

Since ω has already been sampled from $N(0, \Omega)$ in steps 2 and 3 of the algorithm, the transformation $\omega_* = L^{-1}\omega$ is distributed $N(0, \Gamma)$. Collecting the lower block of ω_* yields a sample from $N(0, \Theta)$. Finally, by adding the α term in step 3 of the algorithm, we centre the combined distribution on μ which completes the algorithm.

A.2 Slice sampling

Slice sampling generates pseudo-random numbers from any distribution function $f(y)$ by sampling uniformly from horizontal slices through the PDF. Advantages of the algorithm include its simplicity, that it involves no rejections, and that it requires no external parameters to be set. Define $\eta_j = 1/\lambda_j^2$ and $\psi_j = \beta_j/\nu$. The conditional posterior distribution of η_j , given all other parameters is given by

$$p(\eta_j | \nu, \Theta, \psi_j, \theta, \tau, X, Y, Z) \propto \exp\left(-\frac{\psi_j^2}{2} \eta_j\right) \frac{1}{1 + \eta_j}.$$

Slice sampling can now be implemented to draw from equation (14):

1. Sample $(u_j | \eta_j)$ uniformly in the interval $(0, 1/(1 + \eta_j))$.
2. Sample $\eta_j | \psi_j, u_j \sim \text{Ex}(2/\psi_j^2)$ from an exponential density truncated to have zero probability outside $(0, (1 - u_j)/u_j)$.

Taking the inverse square root of the sample of (2), one receives back the estimate for λ_j . By replacing $\eta = 1/\nu$ and ψ_j^2 by $\sum_{j=1}^K (\beta_j/\lambda_j)^2/2$, ν can be sampled in a similar manner.

A.3 Proofs for propositions

Here, we give further details for the proofs for Propositions 1 and 2.

A.3.1 Proposition 1

With a few steps of algebra, write $(X'\Sigma X + \Lambda^{-1})^{-1} X'\Sigma X(Y - \xi Z)$ as

$$v^2 \Lambda (v^2 \Lambda + (X' \Sigma X)^{-1})^{-1} (X' \Sigma X)^{-1} X' \Sigma (Y - \xi Z).$$

Notice that $(X' \Sigma X)^{-1} X' \Sigma (Y - \xi Z)$ can be viewed as a conditional maximum likelihood estimator for β_p , which is well defined whenever X is full column rank. To arrive at equation (1), assume that

$$X' \Sigma X \approx \text{diag} \left(\sum_{t=1}^T \frac{x_{1,t}^2}{\tau^2 \sigma_{z_t}^2}, \dots, \sum_{t=1}^T \frac{x_{k,t}^2}{\tau^2 \sigma_{z_t}^2} \right), \text{ then } \bar{\beta}_{p,j} \approx (1 - \kappa_{p,j}) \hat{\beta}_{p,j}.$$

Under the simplifying assumption that $z_t \approx \bar{z} = \frac{1}{T} \sum_{t=1}^T z_t$, then $\kappa_{p,j} = \frac{1}{T v^2 \lambda_j^2 \bar{s}_j}$, where $\bar{s}_j = \frac{1}{\tau^2 \sigma_z^2} \text{var}(X_j)$. This makes the connection to shrinkage coefficients for GLMs as derived by Piironen and Vehtari (2017) more apparent. Since we have found that the above approximation holds well in any of the simulations and empirical applications, we implement this in our code.

A.3.2 Proposition 2

From Proposition 1, $\kappa_{p,j} = \frac{1}{1+v^2 \lambda_j^2 \bar{s}_j}$. Now apply change of variables $\lambda_j \mapsto \kappa_{p,j}$, where the probability density function for λ_j is $\frac{1}{\pi} \frac{1}{1+\lambda_j^2}$ for $\lambda_j > 0$:

$$\frac{1}{a_j} \times \frac{\frac{1}{\kappa_{p,j}} + \frac{1 - \kappa_{p,j}}{\kappa_{p,j}}}{\sqrt{\frac{1 - \kappa_{p,j}}{\kappa_{p,j}}}} \times \frac{1}{\pi} \frac{1}{1 + \sqrt{\frac{1 - \kappa_{p,j}}{\kappa_{p,j}}}} \frac{1}{a_j^2},$$

where $a_j = v \sqrt{\bar{s}_j}$. Then equation (31) simplifies to (2).

References

- Adams P., Adrian T., Boyarchenko N., & Giannone D. (2021). Forecasting macroeconomic risks. *International Journal of Forecasting*, 37(3), 1173–1191. <https://doi.org/10.1016/j.ijforecast.2021.01.003>
- Adrian T., Boyarchenko N., & Giannone D. (2019). Vulnerable growth. *American Economic Review*, 109(4), 1263–89. <https://doi.org/10.1257/aer.20161923>
- Alhamzawi R., & Ali H. T. M. (2020). Brq: An R package for Bayesian quantile regression. *Metron*, 78(3), 313–328. <https://doi.org/10.1007/s40300-020-00190-6>
- Alhamzawi R., & Yu K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis*, 64, 209–219.
- Alhamzawi R., Yu K., & Benoit D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, 12(3), 279–297. <https://doi.org/10.1177/1471082X1101200304>
- Bai J., & Ng S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2), 304–317. <https://doi.org/10.1016/j.jeconom.2008.08.010>
- Bhadra A., Datta J., Polson N. G., & Willard B. (2019). Lasso meets horseshoe: A survey. *Statistical Science*, 34(3), 405–427. <https://doi.org/10.1214/19-STS700>
- Bhattacharya A., Chakraborty A., & Mallick B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika*, 103(4), 985–991. <https://doi.org/10.1093/biomet/asw042>
- Bondell H. D., & Reich B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *Journal of the American Statistical Association*, 107(500), 1610–1624. <https://doi.org/10.1080/01621459.2012.716344>
- Cadonna A., Frühwirth-Schnatter S., & Knaus P. (2020). Triple the gamma—A unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics*, 8(2), 20. <https://doi.org/10.3390/econometrics8020020>
- Carriero A., Clark T. E., & Marcellino M. G. (2020). Nowcasting tail risks to economic activity with many indicators. FRB of Cleveland Working Paper No. 20-13R2. <https://doi.org/10.26509/frbc-wp-202013>
- Carvalho C. M., Polson N. G., & Scott J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465–480. <https://doi.org/10.1093/biomet/asq017>
- Chen C. W., Dunson D. B., Reed C., & Yu K. (2013). Bayesian variable selection in quantile regression. *Statistics and its Interface*, 6(2), 261–274. <https://doi.org/10.4310/SII.2013.v6.n2.a9>
- Chen C. W., Gerlach R., Hwang B. B., & McAleer M. (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, 28(3), 557–574. <https://doi.org/10.1016/j.ijforecast.2011.12.004>

- De Mol C., Giannone D., & Reichlin L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2), 318–328. <https://doi.org/10.1016/j.jeconom.2008.08.011>
- Diebold F. X., Gunther T. A., & Tay A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863–883. <https://doi.org/10.2307/2527342>
- Diebold F. X., & Mariano R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. <https://doi.org/10.1198/073500102753410444>
- Engle R. F., & Manganelli S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381. <https://doi.org/10.1198/073500104000000370>
- Figueres J. M., & Jaročinski M. (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters*, 191, 109126. <https://doi.org/10.1016/j.econlet.2020.109126>
- Gaglianone W. P., Lima L. R., Linton O., & Smith D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, 29(1), 150–160. <https://doi.org/10.1198/jbes.2010.07318>
- Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., & Rubin D. B. (2013). *Bayesian data analysis*. CRC Press.
- Gelman A., & Rubin D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Hager W. W. (1989). Updating the inverse of a matrix. *SIAM Review*, 31(2), 221–239. <https://doi.org/10.1137/1031049>
- Hamilton J. (1994). *Time series econometrics*. Princeton University Press, Princeton, NJ.
- Hasenzagl T., Reichlin L., & Ricco G. (2020). *Financial variables as predictors of real growth vulnerability*. CEPR Discussion Paper No. DP14322. <https://ssrn.com/abstract=3526047>
- He X., Jurečková J., Koenker R., & Portnoy S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica: Journal of the Econometric Society*, 58(5), 1195–1214. <https://doi.org/10.2307/2938306>
- Jordà Ò., Schularick M., & Taylor A. M. (2015). Leveraged bubbles. *Journal of Monetary Economics*, 76, S1–S20. <https://doi.org/10.1016/j.jmoneco.2015.08.005>
- Khare K., & Hobert J. P. (2012). Geometric ergodicity of the gibbs sampler for Bayesian quantile regression. *Journal of Multivariate Analysis*, 112, 108–116. <https://doi.org/10.1016/j.jmva.2012.05.004>
- Kim H. H., & Swanson N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178, 352–367. <https://doi.org/10.1016/j.jeconom.2013.08.033>
- Koenker R. (2005). *Quantile regression*. Cambridge University Press.
- Koenker R., & Bassett G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. <https://doi.org/10.2307/1913643>
- Koenker R., Chernozhukov V., He X., & Peng L. (2017). *Handbook of quantile regression* (1st). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315120256>
- Koenker R., & Machado J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448), 1296–1310. <https://doi.org/10.1080/01621459.1999.10473882>
- Kohns D., & Szendrei T. (2021). ‘Decoupling shrinkage and selection for the Bayesian quantile regression’, arXiv, arXiv:2107.08498, preprint: not peer reviewed.
- Korobilis D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting*, 33(1), 11–20. <https://doi.org/10.1016/j.ijforecast.2016.07.005>
- Kozumi H., & Kobayashi G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11), 1565–1578. <https://doi.org/10.1080/00949655.2010.496117>
- Li H., & Pati D. (2017). Variable selection using shrinkage priors. *Computational Statistics & Data Analysis*, 107, 107–119. <https://doi.org/10.1016/j.csda.2016.10.008>
- Li Q., Xi R., & Lin N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3), 533–556. <https://doi.org/10.1214/10-BA521>
- Makalic E., & Schmidt D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1), 179–182. <https://doi.org/10.1109/LSP.2015.2503725>
- Mazzi G. L., & Mitchell J. (2019). *Nowcasting euro area gdp growth using quantile regression*. Manuscript.
- McCracken M., & Ng S. (2020). *Fred-*qd*: A quarterly database for macroeconomic research* (Technical Report). National Bureau of Economic Research.
- Moran G. E., Ročková V., & George E. I. (2018). Variance prior forms for high-dimensional Bayesian variable selection. *Bayesian Analysis*, 14(4), 1091–1119. <https://doi.org/10.1214/19-BA1149>

- Park T., & Casella G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681–686. <https://doi.org/10.1198/016214508000000337>
- Piironen J., Paasiniemi M., & Vehtari A. (2020). Projective inference in high-dimensional problems: Prediction and feature selection.
- Piironen J., & Vehtari A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2), 5018–5051. <https://doi.org/10.1214/17-EJS1337SI>
- Polson N. G., & Scott J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In J. M. Bernardo et al. (Eds.), *Bayesian statistics* (Vol. 9501–538). <https://doi.org/10.1093/acprof:oso/9780199694587.003.0017>
- Polson N. G., Scott J. G., & Windle J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4), 713–733. <https://doi.org/10.1111/rssb.12042>
- Prasad M. A., Elekdag S., Jeasakul M. P., Lafarguette R., Alter M. A., Feng A. X., & Wang C. (2019). *Growth at risk: Concept and application in IMF country surveillance*. International Monetary Fund.
- Stock J. H., & Watson M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179. <https://doi.org/10.1198/016214502388618960>
- Stock J. H., & Watson M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4), 481–493. <https://doi.org/10.1080/07350015.2012.715956>
- Vehtari A., Gelman A., Simpson D., Carpenter B., & Bürkner P.-C. (2021). Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Yu K., & Moyeed R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437–447. [https://doi.org/10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9)