

Bayesian robust quantile regression and risk measures

BY

MARCO BOTTONE

supervised by

Lea Petrella

Department of Methods and Models for Economics,
Territory and Finance

SAPIENZA UNIVERSITY OF ROME



SAPIENZA
UNIVERSITÀ DI ROMA

A thesis submitted for the degree of
Doctor of Philosophy

My deep gratitude goes first to my supervisor Prof. Lea Petrella, who expertly guided me through my PhD studies, sharing her expertise and talent. I would also like to thank Dr. Mauro Bernardi for his helpful and constructive comments and his generous support that greatly contributed to improving the thesis.

Contents

1	Quantile regression, Bayesian methods and risk measures	5
1.1	Bayesian quantile regression and the robustness problem . . .	5
1.2	Quantile regression and risk measures	9
2	Bayesian robust quantile regression	13
2.1	Introduction	13
2.2	The Skewed Exponential Power distribution	14
2.3	Robust Bayesian linear quantile regression	17
2.3.1	Model specification	17
2.3.2	Adaptive IMG for linear quantile regression	20
2.4	Nonlinear extension	22
2.4.1	Non-linear model specification	22
2.4.2	Adaptive IMG for quantiles AM	26
2.5	Simulation Studies	27
2.5.1	Simple linear quantile regression	27
2.5.2	Multiple quantile regression	31
2.5.3	Non Linear Model	33
2.6	Empirical applications	34
2.6.1	Boston housing data	35
2.6.2	Munich rental guide	36
2.6.3	Barro growth data	38
2.7	Conclusion	42
3	Bayesian Non-Linear Conditional Autoregressive Risk Measures	45
3.1	Introduction	45
3.2	Conditional Autoregressive Risk Measure models	46
3.3	Bayesian CARM model	49
3.4	Bayesian Non-linear CARM model	50

3.5	Bayesian Methods	52
3.6	The Adaptive Metropolis within Gibbs sampler	53
3.7	Empirical applications	55
3.7.1	CAViaR forecast evaluation	55
3.7.2	CARE forecast evaluation	56
3.7.3	Summary of VaR and ES results	57
3.8	Conclusion	63
4	Bibliography	65

Chapter 1

Quantile regression, Bayesian methods and risk measures

1.1 Bayesian quantile regression and the robustness problem

Quantile regression has become a very popular approach to provide a wide description of the distribution of a response variable conditionally on a set of regressors. While linear regression analysis aims at estimating the conditional mean of a variable of interest, in quantile regression we may estimate any conditional quantile of order τ with $\tau \in (0, 1)$. Since the seminal works of Koenker and Basset (1978) and Koenker and Machado (1999), several papers have emerged in the literature considering quantile regression analysis from both a frequentist and a Bayesian point of view. For the former, following Koenker (2005) and the references therein, the estimation strategy relies on the minimization of a given loss function. Specifically, let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ a random sample of T observations and $\mathbf{X}_t = (1, X_{t,1}, \dots, X_{t,p-1})'$, $t = 1, 2, \dots, T$ the associated set of p covariates. Consider the following linear quantile regression model

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta}_\tau + \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (1.1)$$

where $\boldsymbol{\beta}_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \dots, \beta_{\tau,p-1})'$ is the vector of p unknown regression parameters varying with the quantile τ level. Here, ε_t , for any $t = 1, 2, \dots, T$, are independent random variables which are supposed to have zero τ -th quantile and constant variance. Assuming $\mathbf{y} = (y_1, y_2, \dots, y_T)$ as a realization of \mathbf{Y} , and \mathbf{x}_t as a realization of \mathbf{X}_t , then the unknown quantile regression

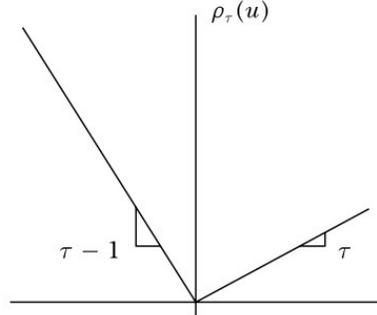


Figure 1.1: Check function used in the frequentist approach to quantile regression

coefficients vector β_τ can be estimated, in the frequentist approach, as the solution to

$$\min_{\beta_\tau} \sum_{t=i}^T \rho_\tau(y_t - \mathbf{x}_t \beta_\tau) \quad (1.2)$$

where

$$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ -(1-\tau)u & \text{if } u < 0, \end{cases} \quad (1.3)$$

which, graphically, assumes the form depicted in figure 1.1.

As can be noted from the figure, the empirical check function is not differentiable at 0. As a consequence, the minimization of 1.2 can be achieved through an algorithm proposed by Koenker and D'Orey (1987) since a closed-form solution is not available.

Moreover, as observed by Koenker and Machado (1999), maximising the likelihood of the Asymmetric Laplace Distribution (ALD) is closely related to minimizing the empirical check function (1.2). In particular, a random variable u has an $ALD(\mu, \sigma, \tau)$ with $\mu = 0$, $\sigma > 0$ and $0 < \tau < 1$ if its pdf is given by

$$f(u; \mu = 0, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\{-\rho_\tau(u)\} \quad (1.4)$$

where μ and σ are the location and the scale parameters respectively and τ is the skewness parameter which is related with the location parameter in a particular way that allows us to use the ALD for quantile regression

models. Specifically, the τ -level quantile corresponds to the natural location parameter, i.e. $P(u \leq \mu) = \tau$, therefore for a given value of τ the estimate of μ represents an estimate of the τ -level quantile for u .

Using this property, Yu and Moyeed (2001) suggested a Bayesian Quantile Regression approach using the ALD as likelihood tool. After the paper of Yu and Moyeed (2001) a wide Bayesian literature followed. Yu et al. (2007) develop a Bayesian framework for Tobit quantile regression and Kobayashi (2017) extended it to accounts for endogeneity. Santos and Bolfarine (2015) propose the use of Bayesian quantile regression for the analysis of response variables limited to the range $(0, 1)$, making use of the ALD in the likelihood calculation. Lum and Gelfand (2012) introduce the asymmetric Laplace process for quantile regression with spatially dependent errors. Reich et al. (2011) develop a Bayesian spatial quantile method for tropospheric ozone accounting for spatial variability by modeling the conditional distribution as a spatial process. Yue and Rue (2011) present a Bayesian quantile inference method based on the integrated nested Laplace approximations (INLA) in additive mixed models assigning appropriate Gaussian Markov random field (GMRF) priors to different types of covariate. Hallin et al. (2010) present a multivariate extension of quantile based on a directional version of Koenker and Bassett's traditional regression quantile using the L_1 optimization ideas. Wang et al. (2016) introduce a quantile structural equation model to provide a comprehensive analysis of the interrelationships among latent variables still using the ALD. Kottas and Genlfand (2001) and Kottas and Krnjajic (2009) propose a Bayesian semiparametric methodology for quantile regression modelling. Hu et al. (2015) introduce a Bayesian quantile regression method for partially linear additive models which explicitly models components that have linear and nonlinear effects while Chen and Yu (2009) propose a nonparametric quantile regression framework using piecewise polynomial functions with number and location of knots inferred through reversible jump Markov chain Monte Carlo. Nonparametric Bayesian quantile regression is also considered in Thomson et al. (2010) that propose to model the dependence of a quantile of one variable on the values of another using a natural cubic spline. Sriram et al. (2013) provide justification for assuming ALD for the response in Bayesian Quantile Regression, even if it can represents a misspecification. Empirical likelihood as a working likelihood for quantile regression in Bayesian quantile inference is considered in Yang and He (2012) while an approach based on the pseudo-joint Asymmetric Laplace Likelihood is implemented in Sriram et al. (2016). Novel application of quantile regression in the risk measure field is considered in Bernardi et al. (2015) and Meligkotsidou et al. (2009).

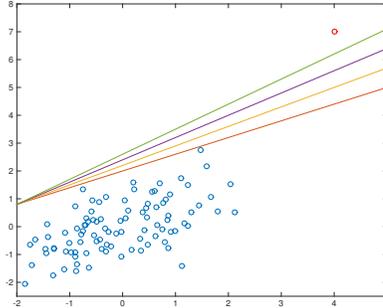


Figure 1.2: The robustness problem in quantile regression analysis

Finally, the problem of variable selection in Bayesian quantile regression models based on the ALD is considered in Yu et al. (2013), Alhamzawi and Yu (2012) (2015), Alhamzawi (2016), Ji et al. (2012).

Although the ALD is widely used in the Bayesian framework, its main disadvantage is displaying medium tails. This may produce misleading information when extreme quantiles are concerned, and in particular, when the data are characterized by the presence of outliers and heavy tails. Absence of a parameter governing tail fatness in the ALD may influence the final inference. Recently, Wichitaksorn et al. (2014) tried to generalize the classical Bayesian quantile regression by using some skew distributions, obtained through a mixture of scaled Normal ones. This class of distributions allows for different degrees of asymmetry of the response variable also imposing a given structure of the tails.

To provide an idea about the robustness problem that may arise in quantile regression analysis let consider the case displayed in figure 1.2 in which we have a set of 100 observations characterized by the presence of the red colored outlier. All the plotted lines are just examples of lines which are compatible with the definition of quantile at $\tau = 0.99$ (since they correctly leave the 99% of observations below) but among them we can observe that the red one is the most conservative. In this case, Bayesian quantile regression based on the ALD tends to estimate a quantile function near to the green line while our goal is to estimate a quantile function as nearest as possible to the red line.

To overcome this drawback, we propose an extension of the Bayesian quantile regression by using the Skew Exponential Power (SEP) distribution

proposed in Zhu and Zinde–Walsh (2009). A useful property of the SEP distribution, similar to the ALD, is that the natural location parameter coincides with the τ -level quantile. Differently from the ALD, the SEP distribution also has a shape parameter governing the decay of the tails. Using the SEP distribution in quantile regression we are able to increase robustness of the inference, in particular when outliers or extreme values are present. One of the potential drawback of the SEP may be that it has one more parameter to estimate than ALD, so the estimation error/variation of this additional parameter may have effect on the quantile regression fitting, beside additional estimating burden. After all, this parameter is crucial to control the decay of the tails. In chapter 2 we show that, for all the estimated parameters, the chain rapidly converges toward the target distribution. In linear regression analysis several works have extensively considered the non-skewed version of the SEP, i.e. the Exponential Power distribution (EP), for its related robustness properties given by the shape parameter. Box and Tiao (1973) first show how to increase robustness of the classical Gaussian linear regression model, introducing the EP as a distribution assumption for the error term. Choy and Smith (1997), explore the robustness properties of posterior moments based on the EP distribution. Choy and Walker (2003) present further extension of the work of Choy and Smith (1997) proposing a case in which the shape parameter assumes values greater than two. Finally, Naranjo et al. (2015) and Kobayashi (2016) suggest the use of the SEP distribution in regression and stochastic volatility models. To the best of our knowledge, this thesis is the first attempt to improve robustness of quantile regression analysis by using the SEP distribution. In chapter 2 linear and Additive Models (AM) with penalized spline are considered to show the flexibility of the SEP in the Bayesian quantile regression context. Moreover, Lasso priors are considered in both cases to account for shrinking parameters problem when the parameters space becomes wide.

1.2 Quantile regression and risk measures

The interest towards a robust model for the conditional quantile function is primary in many fields, particularly in finance, where much attention is posed on methods and models for market risk measurement. Indeed, accurate risk measurement is a primary need for financial institutions and investors especially after the recent financial crisis. Within the instruments for market risk measurement, Value-at-Risk (VaR) (Jorion, 2007) and Expected Shortfall (ES) (Artzner et al, 1999) are certainly the most

popular and used approaches. VaR answers the question on what is the maximum potential loss that will be exceeded with a certain probability in the next days. Therefore, it can be simply understood as a specific conditional quantile of the portfolio returns given the current information i.e. $\mathbb{P}(Y_t < -\text{VaR}_t | \mathcal{F}_t) = \tau$, where Y_t and \mathcal{F}_t denote respectively the return of a portfolio and the information set available at time t , while $\tau \in (0, 1)$ denotes the quantile confidence level associated with the VaR. Even though it is widely used among financial institutions VaR has been criticized because of the absence of the sub-additivity property, namely it does not guarantee that diversified portfolio is less risky than concentrated one. Artzner et al (1999) first recognize this lack of coherency of the VaR and proposed the ES as a possible alternative coherent risk measure which give more information about the distribution of returns in the tails. In particular the ES is defined as the conditional expected loss given that the loss exceed the VaR, i.e. $\mathbb{E}(Y_t | y_t < -\text{VaR})$. But even if ES is a coherent risk measure it is in general more difficult to backtest it, namely, to verify how accurately the strategy or method would have predicted actual results. Moreover, it is not as simple to interpret as the VaR and for these reasons there is not a prevailing risk measure between VaR and ES.

In literature there are several ways to estimate the VaR and the ES, some of them relying on distributional assumptions and some of them just estimate them directly. One of the recent approach is based on assuming conditional autoregressive equation structure, see for example Engle and Manganelli (2004) and Taylor (2008). In chapter 3 of this thesis we focus our attention on the Conditional Autoregressive Value–At–Risk (CAViaR) class of models, introduced by Engle and Manganelli (2004), which belongs to the family of dynamic quantile autoregressive models proposed by Koenker et al. (2006), and the Conditional Autoregressive Expectile (CARE) class of models introduced by Taylor (2008).

The inferential issue for the CAViaR and the CARE class of models has been addressed in literature both from the frequentist and the Bayesian point of view. In the frequentist approach the CAViaR models allow to use the inferential quantile regression methods (Koenker, 2005) by minimizing the loss function introduced by Koenker and Basset (1978) and showed in the previous section (i.e. equation (1.2)). As explained before, the Bayesian approach instead relies on the Asymmetric Laplace distribution (ALD) assumption as a likelihood tool to perform the inferential issue (see e.g. Yu and Moyeed, 2001, Kottas and Gelfand, 2001, Kottas and Krnjajic, 2009, and Sriram et al., 2013, Bernardi et al., 2015).

For the CARE models, which permits to estimate the expectile quantities,

the inferential strategy is instead based on the relation between expectile, quantile and ES (see Newey and Powell, 1987). In particular, the estimation relies on the one-to-one mapping from expectiles to quantiles, and the relationship between VaR and ES. In fact following Efron (1991), the estimator of the τ -th quantile will be the θ -th expectile for which the proportion of observations below it is $\tau\%$, then the one-to-one mapping from expectiles to quantiles is used to obtain VaR and ES using equation (3.8) of Section 2 in chapter 3. The estimation procedure for the generic expectile is addressed in the frequentist approach by using the Asymmetric Least Square (ALS) estimator as in Newey and Powell (1987) while in the Bayesian paradigm the literature relies on the Asymmetric Gaussian distribution assumption (see e.g. Gerlach and Wang, 2015, Gerlach and Chen, 2014; Wichitaksorn et al, 2014; Gerlach et al, 2016, Gerlach and Chen, 2016)

In this thesis we propose to extend existing literature on conditional autoregressive risk measure. First of all, we develop a unified Bayesian Conditional Autoregressive Risk model (B-CARM) which encompass both the CAViaR and the CARE one as particular case, again by using the SEP likelihood tool. As a second result we propose a new Non-Linear and semi-parametric specification of the B-CARM class of models which uses Penalized Splines (see De Boor, 2001, Eilers et al., 1996, Lang and Brazger, 2004) to estimate the relation between the quantile/expectile and the observed variables. The need for a model that allows for non-linearity without assuming any particular restrictions is of clear interest in literature. Indeed, all the specifications of the CAViaR and the CARE models proposed so far emphasize the role of asymmetry and non-linearity in the relation between the observed variables and the current quantile or expectile level, but they all impose the form of this non linearity a priori (see Engle and Manganelli, 2004, Gerlach et al., 2011 and 2012, Chen et al., 2009 and 2012, Gerlach and Chen, 2014). In chapter 3 we will show that the proposed semi-parametric approach allows to obtain different form of nonlinearity in the CAViaR and the CARE models without employ any previous assumption about its structure.

Chapter 2

Bayesian robust quantile regression

2.1 Introduction

In this chapter we propose the SEP distribution to develop a Bayesian robust quantile regression framework. In particular due to the specific characteristics of the SEP distribution we will show how to estimate the quantile function firstly via simple linear regression and secondly by the Additive Models (AM). For the latter, we adopt the Penalized Spline (P-Spline) approach to carry out statistical inference. The Bayesian paradigm is implemented by means of a new adaptive Metropolis MCMC sampling scheme, with a full set of informative priors. In particular, for the AM framework, the proposed algorithm turns into an Adaptive Metropolis within Gibbs MCMC, allowing an efficient estimate of the penalization parameter and the P-Spline coefficients.

When dealing with model building the choice of appropriate predictors and consequently the variable selection issue plays an important role. Here we approach this problem, by considering the Bayesian version of Lasso penalization methodology introduced by Tibshirani (1996). In particular, for the linear quantile regression model, we assume as prior distribution on each regressor, the generalized version of the univariate independent Laplace distribution by Park and Casella (2008), Hans (2009) and Alhamzawi et al. (2012). Using this prior we shrink each parameter separately. When dealing with AM, we generalize the Lang and Brezger (2004) second order random walk prior for the spline coefficients, assuming a multivariate Laplace distribution and accounting for a correlation structure among parameters. This

prior corresponds to the group lasso penalty of Yuan and Lin (2006), Meier et al. (2008) and Li et al. (2010) which in the spline context is immediately interpreted as knots associated with each regressor.

To analyze the performance of the proposed models we perform simulation studies in which we control for the weight of the outliers, the number of the parameters, the shape of the regressors and the presence of heteroschedasticity. Furthermore, we analyze three common real datasets: the corrected version (see Li et al., 2010) of the Boston housing data, first analyzed by Harrison and Rubinfeld (1978); the Munich rental dataset with geoaddivitive spatial effects, considered in Rue and Held (2005) and Yue and Rue (2011) among others; and the Barro growth data, first studied by Barro and Sala-i-Martin (1995) and then extended in the quantile regression framework by Koenker and Machado (1999). The models we propose introduce robustness, variable selection and non-linearity in the estimation process, providing a more flexible framework and a new interpretation of some regression coefficients and, on average, lower posterior standard deviations.

The remainder of the chapter is organized as follows. In Section 2.2, we introduce the SEP distribution and discuss its properties relevant to model conditional quantiles as functions of exogenous covariates. In Section 2.3 we introduce the model specification and the MCMC algorithms proposed. In Section 2.4 we look at the non-linear extension of the linear quantile approach via AM. Section 2.5 explores the sampling performance of the proposed models through simulations. Section 2.6 discusses three well known empirical applications while Section 2.7 concludes.

2.2 The Skewed Exponential Power distribution

Zhu and Zinde-Walsh (2009) have recently proposed a parametrization of the SEP distribution introduced by Fernandez and Steel (1998), particularly convenient when quantiles are the main concern.

Definition 2.2.1. *A random variable $Y \in \mathbb{R}$ is said to be Skewed Exponential Power distributed, i.e. $Y \sim \mathcal{SEP}(\mu, \sigma, \alpha, \tau)$, if its density has the following form:*

$$f_{\text{SEP}}(y; \mu, \sigma, \alpha, \tau) = \begin{cases} \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp\left\{-\frac{1}{\alpha} \left(\frac{\mu-y}{2\tau\sigma}\right)^\alpha\right\}, & \text{if } y \leq \mu \\ \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp\left\{-\frac{1}{\alpha} \left(\frac{y-\mu}{2(1-\tau)\sigma}\right)^\alpha\right\}, & \text{if } y > \mu, \end{cases} \quad (2.1)$$

where $\mu \in \mathfrak{R}$ is the location parameter, $\sigma \in \mathfrak{R}^+$ and $\alpha \in (0, \infty)$ are the scale

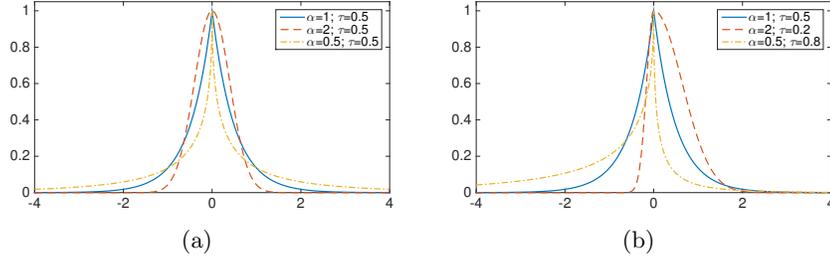


Figure 2.1: Plot of the SEP distribution for different level of the shape parameter (2.1(a)) and skewness parameter (2.1(b)) with $\sigma = 1.0$ and $\mu = 0.0$.

and shape parameters respectively, while $\kappa_{\text{EP}} = \left[2\alpha^{\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right) \right]^{-1}$ with $\Gamma(\cdot)$ is the complete gamma function. Moreover, the parameter $\tau \in (0, 1)$ controls for the skewness of the distribution.

We propose the use of (2.1) for quantile regression inference, as the location parameter μ coincides with the τ -level quantile (demonstrated in the gray box below). We also show (see Zhu and Zinde-Walsh, 2009) that the kurtosis of the SEP is directly determined by its parameter α . In Figure 2.1 we present the *pdf* of the SEP distribution for different values of shape (α) and skewness (τ) parameters, with fixed values for the location and scale parameters $(\mu, \sigma) = (0, 1)$. It is worth noting that, for a fixed value of $\tau = 0.5$ (see subfigure 2.1(a)), we retrieve the Laplace and the Normal distribution when the shape parameter is equal to $\alpha = 1$ and $\alpha = 2$, respectively. Moreover, the smaller the value of α , the fatter the tails of the distribution and, in particular, as $\alpha \rightarrow 0$ the SEP becomes the Cauchy distribution, while as $\alpha \rightarrow \infty$ it becomes equal to the Uniform distribution. Hence, it is evident that the parameter α is important in capturing the behavior of the tails, which may be fundamental when outliers or heavy tailed data are modelled. Furthermore, subfigure 2.1(b) displays the behavior of the SEP for different combinations of α and τ . In this case, the ALD ($\alpha = 1$) and the Skew Normal distribution ($\alpha = 2$) can be obtained thanks to the skewness parameter τ . In the same figure, the relation between τ and the location parameter μ should also be noted. For a fixed μ ($\mu = 0$ in the graph), by varying τ the shape of the distribution changes in such a way that μ becomes its quantile of level τ .

Lemma 2.2.1. Let $Y \sim \mathcal{SEP}(\mu, \sigma, \alpha, \tau)$, then the τ -level quantile of Y coincides with its natural location parameter, i.e. $Q_\tau(Y) = \mu$.

Proof. In order to show that $P(Y \leq \mu) = \tau$ we compute the cdf of a SEP in $y = \mu$

$$\begin{aligned} P(Y \leq \mu) &= \int_{-\infty}^{\mu} \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \left[\exp \left\{ -\frac{1}{\alpha} \left(\frac{\mu - y}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y \leq \mu)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \left(\frac{y - \mu}{2(1 - \tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y > \mu)} \right] dy \\ &= \int_{-\infty}^{\mu} \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \exp \left\{ -\frac{1}{\alpha} \left(\frac{\mu - y}{2\tau\sigma} \right)^\alpha \right\} dy \end{aligned} \quad (2.2)$$

Without loss of generality, let us consider the case when $\mu = 0$ and $\sigma = 1$. The integral reduces to

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \int_{-\infty}^0 \exp \left\{ -\frac{1}{\alpha} \left(\frac{-y}{2\tau} \right)^\alpha \right\} dy. \quad (2.3)$$

By substitute $(-y)^\alpha = x$ we have

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \int_0^{+\infty} \exp \left\{ -\frac{x}{\alpha(2\tau)^\alpha} \right\} \frac{1}{\alpha} (x)^{\frac{1}{\alpha}-1} dx \quad (2.4)$$

Rearranging equation (2.4) and recognizing the kernel of a Gamma pdf with shape $1/\alpha$ and scale $\alpha(2\tau)^\alpha$ the integral becomes

$$\frac{\alpha^{-\frac{1}{\alpha}}}{2\Gamma(1 + \frac{1}{\alpha})} \frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) (\alpha(2\tau)^\alpha)^{\frac{1}{\alpha}}.$$

By using the property $\Gamma(x + 1) = x\Gamma(x)$ all the terms simplify except for τ , concluding the proof. \square

2.3 Robust Bayesian linear quantile regression

In this section we propose the use of the SEP distribution to implement the Bayesian inference for linear quantile regression combined with the prior distributions specification. Since we are interested in Lasso penalization problem in order to achieve sparsity within the quantile regression model, we propose as prior distribution for the regression parameters, a generalized version of the univariate independent Laplace distribution proposed by Park and Casella (2008) and Hans (2009). In line with Alhamzawi et al. (2012), for each quantile regression parameter we assume a Laplace distribution having different scale parameter in order to shrink each regression parameter in a different way. To achieve the Bayesian procedure we provide an adaptive MCMC sampling scheme obtained by running a block-move Independent Metropolis within Gibbs.

2.3.1 Model specification

Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ be a random sample of T observations, and $\mathbf{X}_t = (1, X_{t,1}, \dots, X_{t,p-1})'$, with $t = 1, 2, \dots, T$ equal to the associated set of p covariates. Consider the following linear quantile regression model

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta}_\tau + \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (2.5)$$

where $\boldsymbol{\beta}_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \dots, \beta_{\tau,p-1})'$ is the vector of p unknown regression parameters, varying with the quantile τ level. Here, ε_t , for any $t = 1, 2, \dots, T$, are independent random variables which are supposed to have zero τ -th quantile and constant variance. Assuming $\mathbf{y} = (y_1, y_2, \dots, y_T)$ as a realization of \mathbf{Y} , and \mathbf{x}_t as a realization of \mathbf{X}_t , then the likelihood function for the model (2.5) based on the SEP distribution (2.1) with fixed τ can be written as:

$$\begin{aligned} \mathcal{L}_\tau(\boldsymbol{\beta}_\tau, \sigma, \alpha, | \mathbf{y}, \mathbf{x}_t) &= \prod_{t=1}^T \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \left[\exp \left\{ -\frac{1}{\alpha} \left(\frac{\mathbf{x}_t' \boldsymbol{\beta}_\tau - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \left(\frac{y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right] \\ &= \frac{1}{(2\sigma)^T} \frac{\alpha^{-\frac{T}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})^T} \left[\exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{\mathbf{x}_t' \boldsymbol{\beta}_\tau - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right. \\ &\quad \left. + \exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > \mathbf{x}_t' \boldsymbol{\beta}_\tau)} \right], \quad (2.6) \end{aligned}$$

in this case the parameter μ of equation (2.1) has been replaced by the regression function $\mu = \mathbf{x}_t' \boldsymbol{\beta}_\tau$. As discussed in the previous section, due to the property of the SEP distribution, the regression function $\mathbf{x}_t' \boldsymbol{\beta}_\tau$ corresponds to the conditional τ -level quantile of Y_t , i.e. $Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\beta}_\tau$. In what follows, we omit the subscript τ for the sake of simplicity.

The Bayesian inferential procedure requires the specification of the prior distribution for the unknown vector of parameters $\boldsymbol{\Xi} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \alpha)$. As previously mentioned, we generalize the prior proposed in Park and Casella for the $\boldsymbol{\beta}$ parameter, assuming the hierarchical structure in (2.8) and (2.9); this allows efficient shrinkage of each parameter. The prior distribution is given by:

$$\pi(\boldsymbol{\Xi}) = \pi(\boldsymbol{\beta} | \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma}) \pi(\sigma) \pi(\alpha), \quad (2.7)$$

with

$$\pi(\boldsymbol{\beta} | \boldsymbol{\gamma}) \propto \prod_{j=1}^p L_1(\beta_j | 0, \gamma_j) \quad (2.8)$$

$$\pi(\boldsymbol{\gamma}) \propto \prod_{j=1}^p \mathcal{G}(\gamma_j | \psi, \varpi) \quad (2.9)$$

$$\pi(\sigma) \propto \mathcal{IG}(a, b) \quad (2.10)$$

$$\pi(\alpha) \propto \mathcal{B}(c, d) \mathbb{1}_{(0,2)}(\alpha), \quad (2.11)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$. Here $(\psi, \varpi, a, b, c, d)$ are given positive hyperparameters and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)$ are the parameters of the univariate Laplace distribution:

$$L_1(\beta_j | 0, \gamma_j) = \frac{\gamma_j}{2} \exp\{-\gamma_j |\beta_j|\} \mathbb{1}_{(-\infty, +\infty)}(\beta_j). \quad (2.12)$$

with zero location and γ_j scale parameter. In (2.9)-(2.11) \mathcal{G} , \mathcal{IG} and \mathcal{B} denote the Gamma, Inverse Gamma and Beta distributions, respectively. Given its characteristics, the Laplace distribution is the Bayesian counterpart of the Lasso penalization methodology introduced by Tibshirani (1996) to achieve sparsity within the classical regression framework. The original Bayesian Lasso introduces the same univariate independent Laplace prior distribution for each regression parameter, see Park and Casella (2008) and Hans (2009). Here, as in Alhamzawi et al. (2012), we consider a more general case using the parameters γ_j , $j = 1, 2, \dots, p$. By shrinking each regression parameter in a different way, we overcome problems that may arise in the presence of regressors with different scales of measurement.

As shown in Park and Casella (2008) and Kozumi and Kobayashi (2011),

the Laplace distribution can be expressed as a location–scale mixture of Gaussians which, adapted to our case, becomes

$$L_1(\beta_j | 0, \gamma_j) = \int_0^\infty \frac{1}{\sqrt{2\pi\omega_j}} \exp\left\{-\frac{\beta_j^2}{2\omega_j}\right\} \frac{\gamma_j^2}{2} \exp\left\{-\frac{\gamma_j^2\omega_j}{2}\right\} d\omega_j, \quad (2.13)$$

for $j = 1, 2, \dots, p$, where the mixing variable is exponentially distributed with shape parameter $2/\gamma_j^2$. Furthermore, to retain a parsimonious model parameterization, we introduce a second layer hierarchical prior representation for the vector of shape parameters $\boldsymbol{\gamma}$, in equation (2.9). Using the location–scale representation of the Laplace distribution, the prior structure defined in equations (2.8)–(2.9), can be represented as follows

$$\boldsymbol{\beta} | \boldsymbol{\omega} \sim \mathcal{N}_p(\boldsymbol{\beta} | \mathbf{0}_p, \boldsymbol{\Omega}) \quad (2.14)$$

$$\omega_j | \gamma_j \sim \mathcal{E}(\omega_j | 2/\gamma_j^2) \quad (2.15)$$

$$\gamma_j \sim \mathcal{G}(\gamma_j | \psi, \varpi), \quad (2.16)$$

where $\mathbf{0}_p$ is a column vector of zeros of dimension p , $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_p)'$, $\boldsymbol{\Omega} = \text{diag}\{\omega_j, j = 1, 2, \dots, p\}$ and \mathcal{E} is the exponential distribution. To specify values for the hyperparameters of the prior distributions, vague priors are typically imposed on the scale σ , as it is viewed as a nuisance parameter, see e.g. Yu and Moyeed (2001) and Tokdar and Kadane (2012). We specify the prior for the shape parameter α by imposing a Beta distribution where $c = 2$ and $d = 2$. This allows for a large prior variance avoiding the problem of U-shaped Beta distribution that would cause large probability mass to extreme values. In addition, we extend the Beta distribution to include the case of $\alpha \in (0, 2)$ where $\alpha = 2$ allows consideration of the so called conditional “expectile” of Newey and Powell (1987), while $\alpha = 1$ allows consideration of conditional quantiles based on the ALD introduced by Yu and Moyeed (2001). The parameter α regulates the tail–fatness of the SEP distribution so that smaller values imply larger probabilities of extreme observations (see Section 2.2). Therefore, by choosing $\alpha \in (0, 2)$ we overcome both quantile and expectile regression issues, and we improve robustness by relying on a distribution with fatter tails than the Skew Normal.

In the following Section, we introduce the Bayesian parameter estimation procedure which aims to obtain simulations from the posterior distribution, using an Adaptive Independent Metropolis–Hastings MCMC algorithm.

2.3.2 Adaptive IMG for linear quantile regression

The Bayesian inference is carried out using an adaptive MCMC sampling scheme based on the following posterior distribution

$$\pi(\Xi | \mathbf{y}, \mathbf{x}) \propto \mathcal{L}_\tau(\boldsymbol{\beta}, \sigma, \alpha | \mathbf{y}, \mathbf{x}) \pi(\boldsymbol{\beta} | \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma}) \pi(\sigma) \pi(\alpha), \quad (2.17)$$

where $\mathcal{L}_\tau(\boldsymbol{\beta}, \sigma, \alpha | \mathbf{y}, \mathbf{x})$ indicates the likelihood function specified in equation (3.17). After choosing a set of initial values for the parameter vector $\Xi^{(0)}$, a block-move Independent Metropolis within Gibbs (IMG) is run iteratively to obtain simulations from the posterior distribution at the i -th iteration of $\Xi^{(i)}$, for $i = 1, 2, \dots$. As a first step, the simulation algorithm requires a proposal distribution for the parameters $(\boldsymbol{\beta}, \sigma, \alpha)$.

We choose the following proposal distributions to move each block of the parameters:

$$q(\boldsymbol{\beta}) \sim \mathcal{N}_p\left(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta^{(i)}, \boldsymbol{\Sigma}_\beta^{(i)}\right) \quad (2.18)$$

$$q(\sigma) \sim \mathcal{N}_1\left(\tilde{\sigma} | \mu_{\tilde{\sigma}}^{(i)}, \psi_{\tilde{\sigma}}^{(i)}\right) \left| \frac{\partial \tilde{\sigma}}{\partial \sigma} \right| \quad (2.19)$$

$$q(\alpha) \sim \mathcal{N}_1\left(\mu_\alpha^{(i)}, \psi_\alpha^{(i)}\right) \mathbb{1}_{(0,2)}(\alpha), \quad (2.20)$$

where the scale parameter $\tilde{\sigma} = \log(\sigma)$ is considered on a log-scale and subsequently transformed to preserve positive values. The jacobian term in equation (3.28) is required for the distribution of the transformation $\sigma = \exp(\tilde{\sigma})$. At each iteration i , the IMG algorithm draws a candidate parameter from each parameter block, i.e. $\Upsilon^* = (\xi_1^*, \xi_2^*, \xi_3^*) = (\boldsymbol{\beta}^*, \sigma^*, \alpha^*)$ which is subsequently accepted or rejected. The probability that the candidate parameter ξ_j^* , for $j = 1, 2, 3$ becomes the new state parameter of the chain is evaluated on the basis of the following acceptance probability

$$\lambda\left(\xi_j^{(i-1)}, \xi_j^*\right) = \min \left\{ 1, \frac{\mathcal{L}\left(\xi_j^*, \Xi_{-j}^{(i-1)} | \mathbf{y}, \mathbf{x}\right) \pi\left(\xi_j^*\right) q\left(\xi_j^{(i-1)}\right)}{\mathcal{L}\left(\Xi^{(i-1)} | \mathbf{y}, \mathbf{x}\right) \pi\left(\xi_j^{(i-1)}\right) q\left(\xi_j^*\right)} \right\},$$

for $j = 1, 2, 3$, where $\lambda\left(\xi_j^{(i-1)}, \xi_j^*\right)$ indicates the probability of moving to the new state of the chain, $\pi(\cdot)$ is the prior given in equations (2.8) - (2.11) and $\Xi_{-j}^{(i-1)}$ refers to the whole set of parameters at iteration $i - 1$ without the j -th element of Υ^* . In the last step of the algorithm we sample (ω_j, γ_j) , for $j = 1, 2, \dots, p$ with a Gibbs step, by simulating directly from the respective

full conditional distributions

$$\begin{aligned}\omega_j \mid \beta_j^{(i)}, \gamma_j^{(i-1)} &\sim \mathcal{GIG} \left(\omega_j \mid \frac{1}{2}, \beta_j^{(i)2}, \gamma_j^{(i-1)2} \right) \\ \gamma_j^{2(i)} \mid \omega_j^{(i)} &\sim \mathcal{G} \left(\gamma_j^2 \mid \psi + 1, \varpi + \frac{\omega_j^{(i)}}{2} \right).\end{aligned}$$

where \mathcal{GIG} denotes the Generalized Inverse Gaussian distribution. Since most of the statistical properties of the Markov chain, as well as the performance of the Monte Carlo estimators, depend crucially on the definition of the proposal distribution $q(\cdot)$ (see Andrieu and Moulines, 2006 and Andrieu and Thoms, 2008), we improve the basic IMG–MCMC algorithm with an additional step adapting the proposal parameters using the following equations:

$$\boldsymbol{\mu}_\beta^{(i+1)} = \boldsymbol{\mu}_\beta^{(i)} + \varsigma^{(i+1)} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right), \quad (2.21)$$

$$\boldsymbol{\Sigma}_\beta^{(i+1)} = \boldsymbol{\Sigma}_\beta^{(i)} + \varsigma^{(i+1)} \left(\left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right) \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right)' - \boldsymbol{\Sigma}_\beta^{(i)} \right), \quad (2.22)$$

$$\mu_{\tilde{\sigma}}^{(i+1)} = \mu_{\tilde{\sigma}}^{(i)} + \varsigma^{(i+1)} \left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right), \quad (2.23)$$

$$\psi_{\tilde{\sigma}}^{(i+1)} = \psi_{\tilde{\sigma}}^{(i)} + \varsigma^{(i+1)} \left(\left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right)^2 - \psi_{\tilde{\sigma}}^{(i)} \right), \quad (2.24)$$

$$\mu_\alpha^{(i+1)} = \mu_\alpha^{(i)} + \varsigma^{(i+1)} \left(\alpha - \mu_\alpha^{(i)} \right), \quad (2.25)$$

$$\psi_\alpha^{(i+1)} = \psi_\alpha^{(i)} + \varsigma^{(i+1)} \left(\left(\alpha - \mu_\alpha^{(i)} \right)^2 - \psi_\alpha^{(i)} \right), \quad (2.26)$$

where $\varsigma^{(i+1)}$ denotes a tuning parameter that should be carefully selected at each iteration to ensure the convergence and the ergodicity of the resulting chain (see Andrieu and Moulines, 2006). Roberts and Rosenthal (2007) provide two conditions for the convergence of the chain: the diminishing adaptation condition, which is satisfied if and only if $\varsigma^{(i)} \rightarrow 0$, as $i \rightarrow +\infty$, and the bounded convergence condition, which guarantees that all transition kernels considered have bounded convergence time. Andrieu and Moulines (2006) show that both conditions are satisfied if and only if $\varsigma^{(i)} \propto i^{-d}$ where $d \in [0.5, 1]$. Given this, we choose $\varsigma^{(i)} = \frac{1}{C i^{0.5}}$ where $C = 10$. As argued by Roberts and Rosenthal (2007), these two conditions together ensure for this algorithm asymptotic convergence and a weak law of large numbers respectively.

2.4 Nonlinear extension

In this section, we propose an additive extension of the robust linear quantile regression model to the class of Additive Models (AM) introduced by Hastie and Tibshirani (1986). We set up AM using the SEP likelihood. In order to define the quantile function, we make use of the P-Spline functions resulting in a semiparametric problem. The Bayesian analysis is carried out by generalizing the Lang and Brezger (2004) second order random walk prior for the Spline coefficients, assuming a multivariate Laplace distribution. By doing so we account for a correlation structure among parameters which consider the issue of selection variables.

2.4.1 Non-linear model specification

AM extend multiple linear regression by allowing for the response variable to be modeled as sum of unknown smooth functions of continuous covariates. In this section we set up a robust non linear and semi-parametric framework for quantile regression following a AM approach using the SEP likelihood. In particular, we assume that the τ -level conditional quantile can be modeled as a parametric component jointly with a sum of smooth functions as follows:

$$Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t, \mathbf{Z}_t = \mathbf{z}_t) = \mathbf{x}_t' \boldsymbol{\beta} + \sum_{j=1}^J f_j(z_{j,t}), \quad (2.27)$$

where $\mathbf{x}_t' \boldsymbol{\beta}$ is the parametric component, while each $f_j(z_{j,t})$ is a nonparametric continuous smooth function and $\mathbf{z}_t = (z_{1,t}, z_{2,t}, \dots, z_{J,t})'$ is an additional set of covariates. To implement the statistical analysis we assume that the non-parametric component $f_j(z_{tj})$ can be approximated using a polynomial spline of order d , with $k + 1$ equally spaced knots between $\min(\mathbf{z}_t)$ and $\max(\mathbf{z}_t)$. Using the well known representation of splines in terms of linear combinations of B-splines, we can rewrite equation (2.27) as:

$$Q_\tau(Y_t | \mathbf{X}_t = \mathbf{x}_t, \mathbf{Z}_t = \mathbf{z}_t) = \mathbf{x}_t' \boldsymbol{\beta} + \sum_{j=1}^J \sum_{\nu=1}^{k+d} \theta_{j,\nu} B_{j,\nu}(z_{tj}), \quad (2.28)$$

where $B_{j,\nu}(z_{tj})$ denote B-spline basis functions and $\theta_{j,\nu}$ are the unknown coefficients. In this framework, the value of the estimated coefficients and the shape of the fitted functions strongly depend upon the number and the position of the knots. With respect to the position, in the absence of prior information, we consider equidistant knots as the natural choice. To properly capture the smoothness of the data, careful consideration must be given

to the trade-off between too few and too many knots, which may cause underfitting or overfitting respectively. A possible solution to this problem is known as Penalized Spline (P-Spline) proposed by O’Sullivan (1986 and 1988) and generalized by Eilers and Marx (1996), which relies on the introduction of a penalty element on the first or second differences of the B-Spline coefficients. This setting has been embedded in the Bayesian framework by Lang and Brezger (2004), Brezger and Lang (2006) and Brezger and Steiner (2008) using a second order random walk for all the B-Spline coefficients, i.e.:

$$\theta_{j,\nu} = 2\theta_{j,\nu-1} - \theta_{j,\nu-2} + u_{j,\nu}, \quad \forall j = 1, 2, \dots, J, \quad \forall \nu = 1, 2, \dots, k + d, \quad (2.29)$$

where the generic stochastic component $u_{j,\nu} \sim \mathcal{N}(0, h_j)$, $\theta_{j,1}$ and $\theta_{j,2}$ are initialized with diffuse priors, i.e., $\pi(\theta_{j,\nu}) \propto 1$, for $\nu = 1, 2$. In their work Lang and Brezger (2004) assume that the stochastic components $u_{j,\nu}$ driving the random walk process are independent, i.e. $u_{j,\nu} \perp u_{k,\nu}$, for all $j, k = 1, 2, \dots, J$ with $j \neq k$. As there are no reasons to assume a priori $u_{j,\nu}$ and $u_{k,\nu}$ as being independent ($\forall j, k$), we consider an extension of (2.29) and assume a multivariate Laplace distribution on the vector of regressors accounting for a correlation structure amongst them. It can be showed that, under the assumed prior structure, the original marginal shrinkage effect is preserved because each marginal prior reduces to a univariate Laplace, see, e.g., Kotz et al. (2001).

Moreover, using the Laplace distribution as prior distribution allows the extension of the Lasso approach to the Bayesian paradigm.

Let $\mathbf{u}_j = (u_{j,1}, u_{j,2}, \dots, u_{j,k+d})$, we assume $\mathbf{u}_j \sim \mathcal{AL}_{k+d}(0, \mathbf{I}_{k+d})$, where \mathcal{AL}_{k+d} denotes the multivariate Laplace distribution and \mathbf{I}_{k+d} is the identity matrix of dimension $k + d$. Furthermore, let \mathbf{D}_δ be the difference matrix of dimension $(k + d - \delta) \times (k + d)$, and $\delta = 2$ is the order of the differential operator, such that $\mathbf{D}_\delta \boldsymbol{\theta}_j = \mathbf{u}_j$, then

$$\pi(\boldsymbol{\theta}_j | h_j) \sim \mathcal{AL}_{k+d}(0, h_j (\mathbf{D}'_\delta \mathbf{D}_\delta)), \quad \forall j = 1, 2, \dots, J, \quad (2.30)$$

having density

$$\pi(\boldsymbol{\theta}_j | h_j) = C |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} h_j^{d+k} \exp \left\{ -h_j [\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j]^{\frac{1}{2}} \right\}, \quad (2.31)$$

where $C = \frac{\sqrt{2\pi}}{\Gamma(\frac{d+k+1}{2})}$. As shown in Kotz et al. (2001), the multivariate Laplace distribution can be expressed as a location-scale mixture of Gaus-

sians, where the mixing variable follows a Gamma distribution

$$\pi(\boldsymbol{\theta}_j | \phi_j) \sim \mathcal{N}_{k+d}\left(0, \phi_j (\mathbf{D}'_\delta \mathbf{D}_\delta)^{-1}\right) \quad (2.32)$$

$$\pi(\phi_j | h_j) \sim \mathcal{G}\left(\frac{k+d+1}{2}, \frac{h_j^2}{2}\right), \quad (2.33)$$

for $j = 1, 2, \dots, J$.

It can easily be shown how to retrieve (2.31) from (2.32) - (2.33) by integrating out the augmented variable ϕ_j , i.e.

$$\begin{aligned} \pi(\boldsymbol{\theta}_j | h_j) &= \int_{\mathbb{R}} \frac{|\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} \exp\left\{-\frac{\boldsymbol{\theta}'(\mathbf{D}'_\delta \mathbf{D}_\delta)\boldsymbol{\theta}}{2\phi_j}\right\} \left(\frac{h_j^2}{2}\right)^{\frac{d+k+1}{2}} \phi_j^{\frac{d+k-1}{2}}}{(2\pi\phi_j)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right)} \\ &\quad \times \exp\left\{\frac{-h_j^2\phi_j}{2}\right\} d\phi_j \\ &= \frac{\left(\frac{h_j^2}{2}\right)^{\frac{d+k+1}{2}} |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}}}{(2\pi)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right)} \\ &\quad \times \int_{\mathbb{R}} \phi_j^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\boldsymbol{\theta}'(\mathbf{D}'_\delta \mathbf{D}_\delta)\boldsymbol{\theta}}{\phi_j} + h_j^2\phi_j\right]\right\} d\phi_j, \end{aligned} \quad (2.34)$$

where the integrand in the previous equation (2.34) is proportional to a Generalized Inverse Gaussian distribution $\mathcal{GIG}(p, a, b)$ with parameters $p = \frac{1}{2}$, $a = h_j^2$ and $b = \boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j$ from which we have

$$\int_{\mathbb{R}} \phi_j^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\boldsymbol{\theta}'(\mathbf{D}'_\delta \mathbf{D}_\delta)\boldsymbol{\theta}}{\phi_j} + h_j^2\phi_j\right]\right\} d\phi_j = \frac{2K_{\frac{1}{2}}\left(\sqrt{h_j^2\left(\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j\right)}\right)}{\left(\frac{h_j^2}{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}\right)^{\frac{1}{4}}}, \quad (2.35)$$

where $K_{\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}} \exp\{-z\}$.

Substituting this latter expression into equation (2.34) we obtain

$$\begin{aligned}\pi(\boldsymbol{\theta}_j | h_j) &= \frac{\left(\frac{h_j^2}{2}\right)^{\frac{d+k+1}{2}} |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} \frac{\sqrt{2\pi} \exp\{-h_j \sqrt{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}\}}{[h_j^2 (\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j)]^{\frac{1}{4}}}}{(2\pi)^{\frac{k+d}{2}} \Gamma\left(\frac{d+k+1}{2}\right) \left(\frac{h_j^2}{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}\right)^{\frac{1}{4}}} \\ &= \frac{\sqrt{\pi} |\mathbf{D}'_\delta \mathbf{D}_\delta|^{\frac{1}{2}} h_j^{d+k}}{\Gamma\left(\frac{d+k+1}{2}\right)} \exp\left\{-h_j \sqrt{\boldsymbol{\theta}'_j (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j}\right\},\end{aligned}\quad (2.36)$$

which corresponds to the ALD defined in equations (2.30)–(2.31). The proposed prior distribution for $\boldsymbol{\theta}_j$ corresponds to the group lasso penalty of Yuan and Lin (2006), Meier et al. (2008) and Li et al. (2010), accounting for the group structure when performing variable selection. It is worth emphasizing that, in our context, the group variables have a natural interpretation because they correspond to knots accounting for the smoothness level of the same regressor over different regions of the support. The overall smoothness of the fitted curve is controlled by the variance of the error term h_j , and this corresponds to the inverse of the penalization parameter used by Eilers et al. (1996) in the frequentist framework. We choose a conjugate Gamma prior for h_j^2 , that is $h_j^2 \sim \mathcal{G}(a^{(h)}, b^{(h)})$ with $a^{(h)} = b^{(h)} = 0.001$. A different choice of hyperparameters may be considered, however, they may all bring very similar results. To sum up, the use of a Gamma prior for h_j^2 , and the assumption of the prior structure defined in equations (2.8)–(2.9) for the shape and scale parameters (σ, α) , provides the following hierarchical model

$$y_t = x'_t \boldsymbol{\beta} + \sum_{j=1}^J \mathbf{B}_j^z \boldsymbol{\theta}_j + \epsilon_t, \quad \epsilon_t \sim \mathcal{SEP}(0, \tau, \sigma, \alpha) \quad (2.37)$$

$$\boldsymbol{\beta} | \boldsymbol{\omega} \sim \mathcal{N}_p(\boldsymbol{\beta} | \mathbf{0}_p, \boldsymbol{\Omega}) \quad (2.38)$$

$$\omega_k | \gamma_k \sim \mathcal{E}(\omega_k | 2/\gamma_k^2) \quad (2.39)$$

$$\gamma_k \sim \mathcal{G}(\gamma_k | \psi, \varpi), \quad \forall k = 1, 2, \dots, p \quad (2.40)$$

$$\boldsymbol{\theta}_j | \phi_j \sim \mathcal{N}_{k+d}\left(0, \phi_j (\mathbf{D}'_\delta \mathbf{D}_\delta)^{-1}\right) \quad (2.41)$$

$$\phi_j | h_j \sim \mathcal{G}\left(\frac{d+k+1}{2}, \frac{h_j^2}{2}\right) \quad (2.42)$$

$$h_j^2 \sim \mathcal{IG}\left(\frac{a^{(h)}}{2}, \frac{b^{(h)}}{2}\right) \quad \forall j = 1, 2, \dots, J, \quad (2.43)$$

where $\mathbf{B}_j^z = (B_{j,1}(z_{t_j}), \dots, B_{j,k+d}(z_{t_j}))$.

2.4.2 Adaptive IMG for quantiles AM

In order to perform the Bayesian inference, the Adaptive MCMC algorithm proposed in Section 2.3.2 is slightly modified to deal with the simulation from the posterior distribution of the generalized quantiles' AM parameters. The posterior distribution becomes equal to

$$\begin{aligned} \pi(\Xi \mid \mathbf{y}, \mathbf{x}, \mathbf{z}) &\propto \mathcal{L}_\tau(\beta, \sigma, \alpha, \boldsymbol{\vartheta} \mid \mathbf{y}, \mathbf{x}, \mathbf{z}) \pi(\beta \mid \gamma) \pi(\gamma) \\ &\quad \times \pi(\boldsymbol{\vartheta} \mid \boldsymbol{\phi}) \pi(\boldsymbol{\phi} \mid \mathbf{h}) \pi(\mathbf{h}) \pi(\sigma, \alpha) \end{aligned} \quad (2.44)$$

where the vector Ξ now contains an additional three sets of parameters, namely $\boldsymbol{\vartheta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_J)$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_J)$ and $\mathbf{h} = (h_1^2, h_2^2, \dots, h_J^2)$. The likelihood function $\mathcal{L}_\tau(\beta, \sigma, \alpha, \boldsymbol{\vartheta} \mid \mathbf{y}, \mathbf{x}, \mathbf{z})$ defined in equation (3.17) for the linear model, should be adapted to account for the additional spline coefficients. In order to perform Bayesian analysis, three additional steps to the algorithm described in Section 2.3.2 are required. In particular, having updated all the parameters of the linear part of the model, a candidate for $\boldsymbol{\theta}_j$, for $j = 1, 2, \dots, J$ is drawn from a Gaussian proposal distribution, i.e., $q(\boldsymbol{\theta}_{j,i-1}, \boldsymbol{\theta}_j^*) \sim \mathcal{N}_{k+d}(\boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i)})$ and accepted on the basis of the following acceptance probability

$$\lambda(\boldsymbol{\theta}_j^{(i-1)}, \boldsymbol{\theta}_j^*) = \min \left\{ 1, \frac{\mathcal{L}(\boldsymbol{\beta}^{(i-1)}, \sigma^{(i-1)}, \alpha^{(i-1)}, \boldsymbol{\theta}_j^*, \boldsymbol{\vartheta}_{-j}^{(i-1)} \mid \mathbf{y}, \mathbf{x}, \mathbf{z})}{\mathcal{L}(\boldsymbol{\beta}^{(i-1)}, \sigma^{(i-1)}, \alpha^{(i-1)}, \boldsymbol{\vartheta}_j^{(i-1)} \mid \mathbf{y}, \mathbf{x}, \mathbf{z})} \times \frac{\pi(\boldsymbol{\theta}_j^*)}{\pi(\boldsymbol{\theta}_j^{(i-1)})} \frac{q(\boldsymbol{\theta}_j^{(i-1)})}{q(\boldsymbol{\theta}_j^*)} \right\},$$

where $\boldsymbol{\vartheta}_{-j}$ denote the whole set of B-Spline coefficients without the j -th component. Furthermore, as specified for the regression parameters in Section 2.3.2, an adaptive step for the mean and the variance of the proposal distribution of each $\boldsymbol{\theta}_j$ is implemented using the following equation

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i+1)} = \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)} + \varsigma^{(i+1)} (\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}), \quad (2.45)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i+1)} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i)} + \varsigma^{(i+1)} \left((\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)}) (\boldsymbol{\theta}_j - \boldsymbol{\mu}_{\boldsymbol{\theta}_j}^{(i)})' - \boldsymbol{\Sigma}_{\boldsymbol{\theta}_j}^{(i)} \right), \quad (2.46)$$

for $j = 1, 2, \dots, J$, where ς is the vanishing factor, fixed as discussed above. The hyperparameters $(\boldsymbol{\phi}, \mathbf{h})$ are updated by single-move Gibbs sampling steps, by simulating from the following full conditionals which are proportional to the GIG distribution

$$\begin{aligned}\phi_j \mid \boldsymbol{\theta}_j^{(i)}, h_j^{(i-1)} &\sim \mathcal{GIG} \left(\phi_j \mid \frac{1}{2}, h_j^{(i-1)2}, \boldsymbol{\theta}_j' (\mathbf{D}'_\delta \mathbf{D}_\delta) \boldsymbol{\theta}_j \right) \\ h_j^2 \mid \phi_j^{(i)} &\sim \mathcal{GIG} \left(h_j^2 \mid -\frac{a^h}{2}, \phi_j^{(i)}, \frac{b^h}{2} \right).\end{aligned}$$

2.5 Simulation Studies

We perform simulation studies to highlight the improvements in robustness obtained by implementing SEP-based quantile regression, compared with that obtained by the traditional Bayesian quantile regression, based on the ADL distribution. Our purpose is to illustrate how the SEP misspecified model assumption in the quantile regression framework generates posterior distributions of the regression parameters centered on the true values. The first simulation experiment assesses the robustness properties of the proposed methodology for quantile estimation when the joint distribution of the couple (Y_i, \mathbf{X}_i) , for $i = 1, 2, \dots, T$, is contaminated by the presence of outliers. The second study shows the effectiveness of the shrinkage effect, obtained by imposing the Lasso-type prior, used when the multiple quantile linear model is of key concern. The last experiment aims at highlighting the ability of the model to adapt to non-linear shapes, when data come from heterogeneous fat-tailed distributions. The hyperparameters of the prior distributions are chosen such that the priors are non-informative. In particular, for the nuisance parameter σ we choose $a = b = 0.001$ which corresponds to a proper Inverse Gamma distribution with infinite second moments. When lasso prior is assumed, the hyperparameters (ψ, ϖ) in the Gamma priors for γ_j are set at 0.1.

2.5.1 Simple linear quantile regression

For this experiment we consider a sample of $T = 100$ drawn from the following homoskedastic mixture of distributions

$$f((Y, X)', \boldsymbol{\eta}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_L, \boldsymbol{\Sigma}) = \sum_{l=1}^L \eta_l \varphi(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}), \quad (2.47)$$

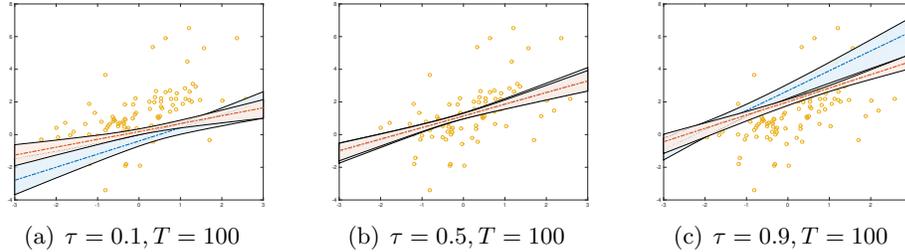


Figure 2.2: Contaminated data example. Comparison between Bayesian quantile regression based on the ALD (blue) and the SEP distribution (red) for different values of the quantile confidence level $\tau = (0.1, 0.5, 0.9)$ and sample sizes $T = 100$. Shaded areas denote 95% HPD credible sets.

where φ denotes the density function of a Gaussian distribution with mean $\boldsymbol{\mu}$ and variance and covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$ is the vector of weights. We set the number of components equal to $L = 3$, with mixture weights $\boldsymbol{\eta} = (0.85, 0.0725, 0.0725)$, locations and scale matrix specified as

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu}_3 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}. \quad (2.48)$$

The quantile regression model used is a simple model with only one exogenous variable i.e. $Y_t = X_t\beta + \varepsilon_t$ for $t = 1, 2, \dots, T$. The aim of this example is to show the performance of the Bayesian quantile linear regression analysis assuming both the ALD and the SEP likelihood when the data are strongly contaminated by the presence of outliers. Since we have only one regressor, for this example we use a simplified version of the sampler proposed in Section 2.3.2, in which a simple Gaussian prior is considered for β . For $\tau = (0.1, 0.5, 0.9)$ we run the MCMC algorithm with $N = 50,000$ iterations and a burn-in of $M = 10,000$. For both the ALD and the SEP distribution assumptions, initial values for the parameters to be estimated, namely $(\boldsymbol{\beta}, \sigma)$, are randomly drawn from $\mathcal{N}(0, 1)$ and $\mathcal{IG}(0.001, 0.001)$, respectively. The additional initial value for the parameter α , required only for the SEP distribution case, is randomly drawn from $\mathcal{B}(2, 2)$.

Figure 2.2 depicts the estimated regression lines as well as the 95% HPD credible sets. The blue line refers to the ALD estimation while the red line to the SEP one. It can be easily observed that the two curves overlap for $\tau = 0.5$ and diverge increasingly for extreme level quantiles i.e. for $\tau = (0.1, 0.9)$. In the case that $\tau = 0.5$ the posterior mean of α is very close to one, implying that the SEP reduces to the ALD distribution.

Parameter	ALD			SEP		
	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
β_0	-0.391 (0.176)	1.149 (0.093)	2.688 (0.237)	0.186 (0.089)	1.144 (0.086)	2.011 (0.100)
β_1	0.801 (0.151)	0.735 (0.106)	1.207 (0.144)	0.428 (0.094)	0.709 (0.093)	0.825 (0.074)
σ	3.844 (0.386)	2.105 (0.212)	3.478 (0.349)	1.049 (0.153)	0.862 (0.112)	0.989 (0.150)
α	-	-	-	0.596 (0.094)	0.832 (0.142)	0.504 (0.068)

Table 2.1: Contaminated data example. Estimated parameters for different levels of the quantile confidence level $\tau = (0.1, 0.5, 0.9)$ and $T = 100$. Standard deviations in parenthesis.

The further we move away from the median, the more e the differences in the estimated regression quantile parameters under the ALD and the SEP assumptions are noticeable. When using the ALD distribution, the intercept and slope of the regression line are strongly influenced by the outliers (7.25% of the total observations) generated by the two external components of the mixture (2.2(a) and 2.2(c)). In both cases, the estimated α of the SEP is considerably smaller than one. As a consequence estimation of the β parameters is made under a distribution with fatter tails than the ALD, strongly penalizing more extreme observations and providing us with more robust results. For the regression parameters, Table 2.1 contains the estimated posterior means and standard deviations under the ALD and the SEP assumptions. In the data generating process, the theoretical slope should always be equal to 0.6. When moving from the median to more extreme quantiles, the ALD-estimated posterior mean of the intercept and the slope is farther from the true value than those obtained with the SEP. It is worth noting that the SEP standard errors are always lower, implying that the estimated parameters are sharper when using the SEP distribution. Finally, figures 2.3-2.5 provide evidence about the efficiency of the MCMC algorithm implemented by showing posterior draws, posterior histograms and Autocorrelation functions for the estimated SEP robust parameters for the three different values of τ .

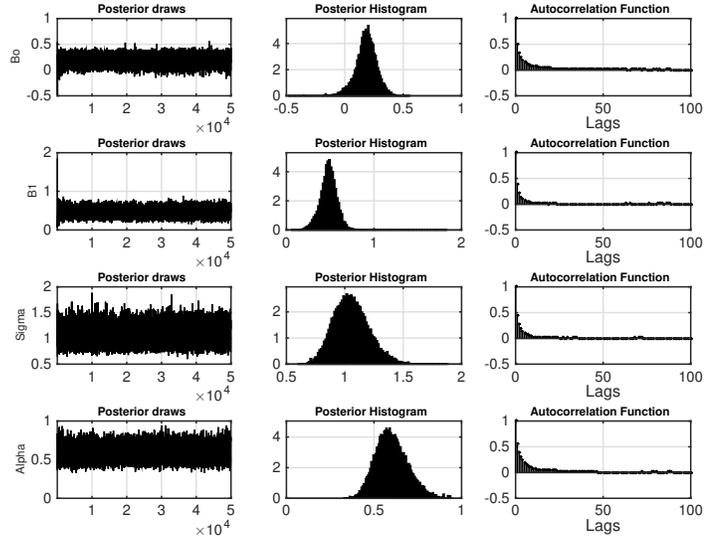


Figure 2.3: Posterior draws, posterior histograms and ACF for estimated robust (SEP) parameters of simple linear quantile regression simulation with $\tau = 0.1$

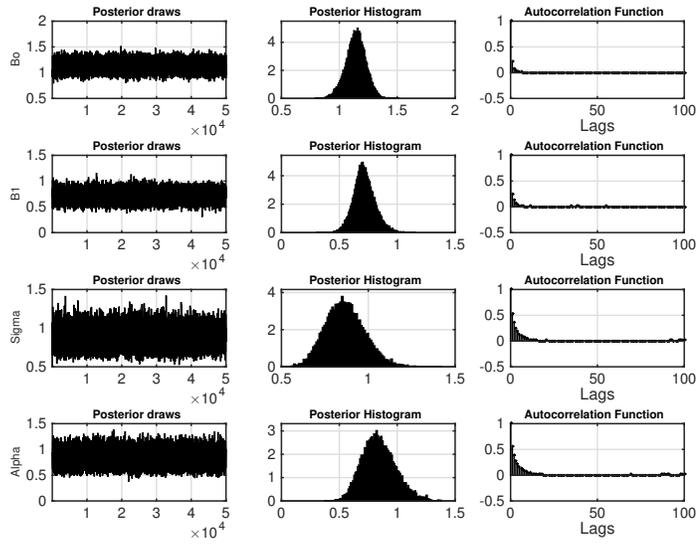


Figure 2.4: Posterior draws, posterior histograms and ACF for estimated robust (SEP) parameters of simple linear quantile regression simulation with $\tau = 0.5$

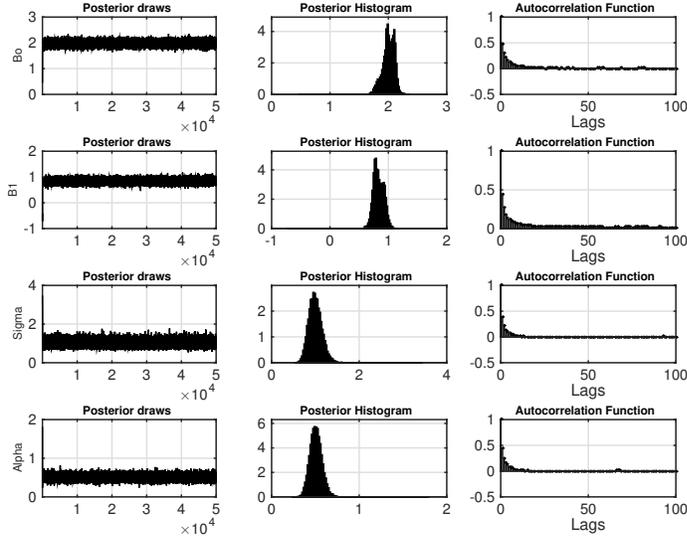


Figure 2.5: Posterior draws, posterior histograms and ACF for estimated robust (SEP) parameters of simple linear quantile regression simulation with $\tau = 0.9$

2.5.2 Multiple quantile regression

In this section, we carry out a Monte Carlo simulation study specifically tailored to evaluate the performance of the model when the Lasso prior (2.8) is considered for the regression parameters. The simulations are similar to the one proposed in Li et al. (2010) and Alhamzawi et al. (2012). In particular, we simulate $T = 200$ observations from the linear model $Y_t = \mathbf{X}_t' \boldsymbol{\beta} + \varepsilon_t$, where the true values for the regressors are set as follows:

- Simulation 1. $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$,
- Simulation 2. $\boldsymbol{\beta} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)'$,
- Simulation 3. $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)'$,

The first simulation corresponds to a sparse regression case, the second to a dense case, and the third to a very sparse case. The covariates are independently generated from a $\mathcal{N}(0, \Sigma)$ with $\sigma_{i,j} = 0.5^{|i-j|}$. Two different distributions for the error terms generating process are considered for each simulation study. The first is a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, with μ set so that the τ -th quantile is 0, while σ^2 is set as 9, as in Li et al. (2010). The second distribution is a Generalized Student't $\mathcal{GS}(\mu, \sigma^2, \nu)$ with two

degrees of freedom, i.e. $\nu = 2$, $\sigma^2 = 9$ and μ set so that the τ -th quantile is 0. For three different quantile levels, $\tau = (0.10, 0.5, 0.9)$ we run 50 simulations for each vector of parameters (β) and each distribution of the error term. Table 2.2 reports the median of mean absolute deviation (MMAD), i.e. $\text{median}\left(\frac{1}{200} \sum_{t=1}^{200} |x'_t \hat{\beta} - x'_t \beta|\right)$, and the median of the parameters $\hat{\beta}$, over 50 estimates. Results for the first simulation are reported, since results from the other two simulations are qualitatively similar. The proposed Bayesian quantile regression method based on the SEP likelihood performs better in terms of MMAD for both distributions of the error term. This is evidence that the presence of the shape parameter α in the likelihood better capture the behavior of the data. The estimated shape parameter is indeed greater and lower than one in the Gaussian and Generalized Student's cases, respectively; this provides a more reliable estimation of the vector β , regardless of the tail weight of the error term distribution. These results are reinforced in the second and third simulation (not reported here) in which we exaggerate the density and the sparsity of the predictors structure. Furthermore, the proposed robust method reduces the bias of estimated β for all quantile confidence levels. Regarding the shrinkage ability of the proposed estimator, when the true parameters are zero, the SEP distribution performs better than the ALD in identifying the parameters .

Error distribution	Par.	ALD			SEP		
		$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.50$	$\tau = 0.90$
Gaussian	MMAD	1.0131	1.1008	1.0579	0.9096	1.0955	0.9708
	β_1	3.1323	3.2209	3.2145	3.0744	3.0036	3.2127
	β_2	1.6408	1.4786	1.6165	1.7656	1.4833	1.6800
	β_3	0.0444	0.0294	0.0267	0.0428	0.0228	0.0186
	β_4	0.0453	0.0243	0.0235	0.0248	0.0191	0.0156
	β_5	1.2731	1.2379	1.3471	1.3969	1.8405	1.4702
	β_6	0.0185	0.0161	0.0205	0.0124	0.0127	0.0128
	β_7	0.0112	0.0106	0.0120	0.0067	0.0063	0.0095
	β_8	0.0073	0.0078	0.0064	0.0038	0.0047	0.0051
Generalized Student t	MMAD	0.5163	0.1807	0.4685	0.4777	0.1789	0.4275
	β_1	3.0630	2.9884	2.9874	3.0826	2.9877	2.9934
	β_2	1.0484	1.3700	1.1366	1.0952	1.3951	1.2110
	β_3	0.0304	0.0144	0.0325	0.0252	0.0135	0.0412
	β_4	0.0258	0.0181	0.0162	0.0263	0.0163	0.0138
	β_5	1.7012	1.9036	1.7701	1.7558	1.9111	1.8052
	β_6	0.0128	0.0085	0.0137	0.0074	0.0072	0.0136
	β_7	0.0055	0.0057	0.0101	0.0052	0.0066	0.0082
	β_8	0.0067	0.0009	0.0002	0.0051	0.0011	-0.0021

Table 2.2: Multiple regression simulated data example 1. MMADs and estimated parameters for Simulation 1 under the SEP and ALD assumption for the quantile error term.

2.5.3 Non Linear Model

In this simulation example we illustrate how well the model assumptions perform when a simple AM is employed with a single continuous smooth non-linear function and where the parametric linear components are set to zero. Following Chen and Yu (2009), we consider two data generating processes $y_t = f_j(x_t) + \epsilon_t$, for $t = 1, 2, \dots, T$ and $j = 1, 2$ where f_1 represents the wave function and f_2 the doppler function, defined as follows

$$f_1(x) = 4(x - 0.5) + 2 \exp\left(-256(x - 0.5)^2\right) \mathbb{1}_{(0,1)}(x) \quad (2.49)$$

$$f_2(x) = (0.2x(1 - 0.2x))^{\frac{1}{2}} \sin\left(\frac{2\pi(1 + \gamma)}{0.2x + \gamma}\right) \mathbb{1}_{(0,1)}(x), \quad (2.50)$$

with $\gamma = 0.15$. These functions are usually used (see also Denison et al. 1998) to check the nonlinear fitting ability of a model. Starting from these two curves, we generate a sample of $T = 200$ and $T = 512$ observations for the wave and the doppler functions respectively, using four different sources of error

$$\text{Gaussian noise,} \quad \epsilon_t \sim \mathcal{N}(0, 1) \quad (2.51)$$

$$y_t = f_1(x_t) + \sigma_1 \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2 \epsilon_t,$$

$$\text{Student-t noise,} \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (2.52)$$

$$y_t = f_1(x_t) + \sigma_1 \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2 \epsilon_t,$$

$$\text{Linear heterogeneity,} \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (2.53)$$

$$y_t = f(x_t) + \sigma_1(1 + x) \epsilon_t,$$

$$y_t = f(x_t) + \sigma_2(1 + x) \epsilon_t,$$

$$\text{Quadratic heterogeneity,} \quad \epsilon_t \sim \mathcal{T}_\nu(0, 1), \quad (2.54)$$

$$y_t = f_1(x_t) + \sigma_1(1 + x_t^2) \epsilon_t,$$

$$y_t = f_2(x_t) + \sigma_2(1 + x_t^2) \epsilon_t,$$

where $\sigma_1 = \sqrt{0.4}$, $\sigma_2 = \sqrt{0.1}$ and $\nu = 2$. The set of model specifications considered are estimated using penalized P-Splines of order 4. imposing a relatively large number of equally spaced knots and a penalization parameter $\delta = 2$, as suggested by Eilers et al. (1996). In particular, we use 20 knots for the wave function and 25 knots for the doppler function due to the presence of many change points. The sampling process is performed using

Model	Noise	Wave			Doppler		
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
ALD	Gaussian	0.0054 (0.0171)	0.0022 (0.0070)	0.0039 (0.0124)	0.0002 (0.0007)	0.0000 (0.0002)	0.0002 (0.0005)
	Student-t	0.0504 (0.1593)	0.0034 (0.0108)	0.0177 (0.0561)	0.0009 (0.0027)	0.0001 (0.0045)	0.0177 (0.0015)
	Lin. Het.	0.1035 (0.3273)	0.0054 (0.0170)	0.0627 (0.1979)	0.0059 (0.0180)	0.0002 (0.0010)	0.0039 (0.0124)
	Quad. Het.	0.0505 (0.1598)	0.0067 (0.0210)	0.0752 (0.2377)	0.0018 (0.0058)	0.0001 (0.0006)	0.0050 (0.0160)
SEP	Gaussian	0.0071 (0.0226)	0.0020 (0.0063)	0.0078 (0.0248)	0.0002 (0.0006)	0.0000 (0.0002)	0.0005 (0.0016)
	Student-t	0.0251 (0.0795)	0.0037 (0.0117)	0.0132 (0.0417)	0.0007 (0.0022)	0.0001 (0.0045)	0.0006 (0.0019)
	Lin. Het.	0.0986 (0.3118)	0.0046 (0.0145)	0.0678 (0.2144)	0.0008 (0.0026)	0.0003 (0.0010)	0.0006 (0.0020)
	Quad. Het.	0.0234 (0.0741)	0.0057 (0.0180)	0.0305 (0.0965)	0.0011 (0.0035)	0.0002 (0.0006)	0.0008 (0.0026)

Table 2.3: Non-linear regression simulated data example. MSE of the fitted curves with four sources of noise evaluated over the 200 synthetic replications. Standard deviations in parenthesis.

50,000 iterations with the first 10,000 as burn-in. For all described curves, Table 2.3 shows the average and the standard errors of the mean squared errors (*mse*) of three different quantile levels (50 repeats). It can be noted that the SEP outperforms the ALD in terms of *mse* almost systematically. The difference between the two curves is less evident in the presence of Gaussian errors, where the ALD also displays a smaller *mse* for the extreme quantiles of the wave function. The improvement in terms of estimation bias becomes larger when looking at more heavy tailed and heteroskedastic error distributions. Concerning the wave function, the SEP shows a *mse* equal to half of that obtained with the ALD at the extreme quantiles. The same conclusions can be drawn for the doppler function that is generally better estimated than the wave function.

2.6 Empirical applications

Three empirical datasets are analyzed in this section: Boston Housing, Munich Rent and Barro growth data. The first dataset is characterized by the presence of many regressors that emphasizes the usefulness of introducing a lasso prior for the regression parameters. The second dataset also has a large set of regressors with some characterized by a non-linear relationship with the response variable. For this dataset we highlight that the assump-

tion of a lasso prior within a robust quantile AM framework leads to a more precise estimation process. Finally we propose the use of our robust quantile lasso AM to study the Barro growth data in a previously unexplored way by assuming a non linear representation for some regressors. We find a new interpretation of regression parameters while maintaining the neoclassical convergence hypothesis.

2.6.1 Boston housing data

In this section we analyze the Boston Housing data first considered by Harrison and Rubinfeld (1978) studying the influence of pollution on house prices. In particular, in this section we consider the corrected data of Li *et al.* (2010). The model is based on the log-transformed corrected median values of owner-occupied housing (values in USD 1000) as the dependent variable while several exogenous variables are taken into account: the point longitudes and latitudes in decimal degrees (LON and LAT respectively); the per capita crime (CRIM); the proportions of residential land zoned and non-retail business acres per town (ZN and INDUS respectively); a dummy equal to 1 if tract borders Charles River (CHAS); the nitric oxide concentration (NOX); the average number of rooms per dwelling (RM); the proportion of owner-occupied units built prior to 1940 (AGE); the weighted distances to five Boston employment centers (DIS); the index of accessibility to radial highways per town (RAD); the full-value property-tax rate per town (TAX); the pupil-teacher ratios per town (PTRATIO); the transformed Black population proportion (B); and, percentage values of lower status population (LSTAT). To provide a description of the conditional distribution of the response variable we consider five values of τ , i.e. 0.10 0.25 0.50 0.75 0.90. Moreover, in order to show the advantage and performance of assuming a Lasso prior for the regressor parameters we also consider a Gaussian prior distribution. Results are reported in Table 2.4. Independently of the choice of prior distribution, all variables have a sign similar to previous studies using the same dataset. Nevertheless, the Lasso prior should be preferred for at least for two reasons. Firstly, it systematically provides smaller posterior standard errors. Secondly, the estimated coefficients appear to be more reliable at extreme quantile levels, i.e. $\tau = 0.1$ or $\tau = 0.9$, where the estimated parameters obtained under Gaussian prior become very unstable for some variables.

Variable	Gaussian Prior					Lasso Prior				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
LON	-0.0614 (0.0364)	-0.0297 (0.0416)	-0.0203 (0.0450)	-0.0114 (0.0555)	0.0072 (0.0434)	-0.0287 (0.0166)	-0.0213 (0.0172)	-0.0258 (0.0187)	-0.0261 (0.0176)	-0.0161 (0.0157)
LAT	-0.0250 (0.0582)	0.0251 (0.0678)	0.0386 (0.0732)	0.0531 (0.0937)	0.0816 (0.0729)	0.0103 (0.0276)	0.0221 (0.0290)	0.0168 (0.0311)	0.0121 (0.0296)	0.0238 (0.0262)
CRIM	-0.0230 (0.0032)	-0.0177 (0.0027)	-0.0093 (0.0015)	-0.0063 (0.0016)	-0.0058 (0.0021)	-0.0241 (0.0028)	-0.0178 (0.0027)	-0.0093 (0.0014)	-0.0059 (0.0017)	-0.0032 (0.0014)
ZN	0.0000 (0.0003)	0.0005 (0.0003)	0.0008 (0.0004)	0.0012 (0.0004)	0.0012 (0.0003)	-0.0000 (0.0003)	0.0006 (0.0003)	0.0009 (0.0004)	0.0010 (0.0003)	0.0009 (0.0002)
INDUS	0.0030 (0.0016)	0.0023 (0.0014)	0.0027 (0.0016)	0.0014 (0.0017)	-0.0013 (0.0013)	0.0016 (0.0015)	0.0017 (0.0014)	0.0016 (0.0016)	0.0010 (0.0017)	-0.0027 (0.0011)
CHAS	0.0482 (0.0264)	0.0464 (0.0197)	0.0597 (0.0264)	0.0737 (0.0308)	0.1134 (0.0302)	0.0183 (0.0190)	0.0377 (0.0205)	0.0411 (0.0236)	0.0448 (0.0275)	0.0834 (0.0312)
NOX	-0.3667 (0.1324)	-0.2642 (0.0946)	-0.3672 (0.1119)	-0.4465 (0.1424)	-0.3204 (0.1191)	-0.0397 (0.0463)	-0.0604 (0.0575)	-0.0480 (0.0551)	-0.0416 (0.0524)	-0.0222 (0.0395)
RM	0.2050 (0.0258)	0.2259 (0.0141)	0.2139 (0.0175)	0.2007 (0.0255)	0.2067 (0.0154)	0.2318 (0.0157)	0.2299 (0.0131)	0.2129 (0.0173)	0.2262 (0.0201)	0.2347 (0.0142)
AGE	-0.0013 (0.0005)	-0.0016 (0.0003)	-0.0009 (0.0004)	-0.0000 (0.0006)	0.0006 (0.0003)	-0.0019 (0.0003)	-0.0018 (0.0003)	-0.0011 (0.0004)	-0.0008 (0.0005)	0.0002 (0.0003)
DIS	-0.0337 (0.0063)	-0.0342 (0.0054)	-0.0330 (0.0062)	-0.0337 (0.0063)	-0.0313 (0.0043)	-0.0279 (0.0050)	-0.0303 (0.0047)	-0.0268 (0.0061)	-0.0267 (0.0054)	-0.0257 (0.0033)
RAD	0.0113 (0.0025)	0.0100 (0.0018)	0.0074 (0.0024)	0.0106 (0.0022)	0.0118 (0.0019)	0.0103 (0.0022)	0.0089 (0.0016)	0.0077 (0.0027)	0.0095 (0.0021)	0.0099 (0.0014)
TAX	-0.0007 (0.0001)	-0.0006 (0.0001)	-0.0005 (0.0001)	-0.0004 (0.0001)	-0.0004 (0.0001)	-0.0007 (0.0001)	-0.0006 (0.0001)	-0.0005 (0.0001)	-0.0005 (0.0001)	-0.0004 (0.0001)
PRATIO	-0.0295 (0.0037)	-0.0292 (0.0032)	-0.0318 (0.0039)	-0.0302 (0.0047)	-0.0253 (0.0033)	-0.0300 (0.0024)	-0.0270 (0.0027)	-0.0280 (0.0037)	-0.0245 (0.0035)	-0.0199 (0.0027)
B	0.0006 (0.0001)	0.0006 (0.0001)	0.0007 (0.0001)	0.0007 (0.0001)	0.0006 (0.0002)	0.0006 (0.0001)	0.0006 (0.0001)	0.0007 (0.0001)	0.0008 (0.0001)	0.0009 (0.0001)
LSTAT	-0.0186 (0.0027)	-0.0172 (0.0019)	-0.0189 (0.0023)	-0.0195 (0.0028)	-0.0191 (0.0015)	-0.0161 (0.0018)	-0.0173 (0.0019)	-0.0205 (0.0023)	-0.0179 (0.0025)	-0.0163 (0.0015)
σ	0.2143 (0.0243)	0.1954 (0.0198)	0.2142 (0.0208)	0.2284 (0.0251)	0.1906 (0.0222)	0.1948 (0.0208)	0.1941 (0.0197)	0.2147 (0.0208)	0.2105 (0.0237)	0.1722 (0.0202)
α	0.7565 (0.0603)	0.7825 (0.0558)	0.8440 (0.0620)	0.7929 (0.0627)	0.6039 (0.0390)	0.7027 (0.0470)	0.7766 (0.0541)	0.8403 (0.0602)	0.7401 (0.0569)	0.5674 (0.0335)

Table 2.4: Linear regression model results for Boston dataset. Standard deviations in parenthesis.

2.6.2 Munich rental guide

In section 2.5.3 we provide empirical evidence that the SEP distribution produces more reliable estimates of the conditional quantile when there is heteroskedasticity and heavy tails. To provide example based on real data we analyze the very well known 2003 Munich rental dataset, characterized by the presence of heterogeneous variability. Several analyses of this dataset (see for example Kneib et al, 2011 and Mayr et al, 2012) showed the presence of spatial effects, modeled by considering a parameter for each of the 380 districts of Munich. For this reason, the parameter space handled is

quite wide, highlighting the need for a variable selection approach. Here therefore, we assume a Lasso prior distribution on the unknown parameters in line with that proposed in (2.31) and we compare its performance with a Gaussian prior assumption. The response variable is the rent in Euro per square meters for a flat in Munich. Two sets of covariates describe linear and non linear relationships between the rent and its determinants. The linear predictors are a set of 13 dummies for the goodness of location, the goodness of rooms and the number of rooms in the flat. The floor size and the year of construction have instead a non-linear impact on the response variable. Finally, the spatial location of the flat allows the implementation of a geoaddivitive model of the kind introduced by Kammann and Wand (2003). To this end, we use a Bayesian semi-parametric quantile regression model with a spatial effect similar to the one considered in Rue and Held (2005) and Yue and Rue (2011) among the others. A complete description of the dataset can be found in Rue and Held (2005).

We estimate the τ -th conditional quantile for the rent r_t , i.e., $Q_\tau(r_t | \mathbf{x}_t, \mathbf{z}_t)$ from the following model:

$$\begin{aligned} r_t &= q_{t,\tau} + \epsilon_t \\ q_{t,\tau} &= \mathbf{x}_t' \boldsymbol{\beta}_\tau + f_{s,\tau}(z_{s,t}) + f_{y,\tau}(z_{y,t}) + f_{l,\tau}(z_{l,t}), \end{aligned} \quad (2.55)$$

where $t = 1, 2, \dots, 2035$; ϵ_t is the error term with zero τ -th quantile and constant variance, \mathbf{x}_t is the whole set of dummies treated as linear parametric predictors, $\mathbf{z}_t = (z_{s,t}, z_{y,t}, z_{l,t})$ are the predictor variables for “size”, “year” and “spatial” effect while $f_{s,\tau}$, $f_{y,\tau}$ and $f_{l,\tau}$ are their non-linear functions. The estimation procedure of three quantile confidence levels $\tau = (0.25, 0.5, 0.75)$ was performed using the Adaptive MCMC procedure for AM described in subsection 2.4.2.

Figures 2.6 and 2.7 show the estimated non-linear effect for the year of construction and the floor size using Gaussian and Lasso priors respectively. The points below each sub-figure represent the available observations for each value of the covariates, while dotted lines represent the 95% posterior credible intervals. We can observe that both priors provide similar estimated splines for the effect of floor size on house prices. Small flats (less than $40m^2$) have a very high rent per square meter, while for big flats the rent remains almost unchanged. For the year of construction, the estimated splines are relatively different under the two prior specifications but they actually contain similar information. When looking at the level and confidence intervals of the variable, the effect of this covariate is almost null until the 1990s, when a clear positive and increasing effect is shown under both

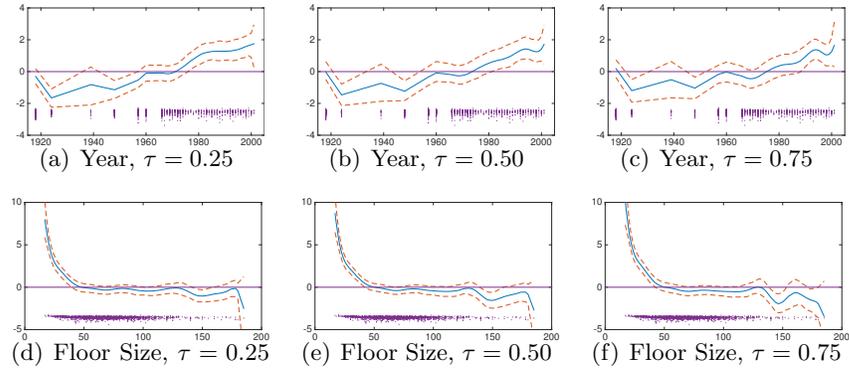


Figure 2.6: Estimated non-parametric effect using Gaussian prior with 95% credible bands for Munich data.

prior specifications. Figure 2.8 displays the estimated spatial effects of 380 subquarters in Munich. Values are normalized to be in the range $(0, 1)$. As expected, for both Gaussian and Lasso priors, rents are high in the centre of Munich and some well-known districts, becoming lower on the margins. Finally, estimated posterior means and standard deviations for the linear parametric effects are shown in Table 2.5. The signs of the variables are in line with previous work; however, under Lasso prior some new findings are associated with the effect of the following variables: No hot water, No central heating and 6 Rooms. Interestingly, Lasso prior differentiate the effect of these variables for each quantile. We find that the absence of hot water and the presence of 6 Rooms have a statistically significant negative effect only on expensive houses, i.e. for $\tau = 0.75$ while the opposite occurs for the absence of central heating. We argue that these results highlight the variety of the consumption choices due to different budget constraints. It is worth noting that Lasso prior correctly shrinks the effect of Special bathroom interior, that is not very significant when estimated using Gaussian prior.

2.6.3 Barro growth data

Our final application, is an analysis of the dataset underlying the international economic growth model, firstly considered by Barro and Sala i-Martin (1995) and extended to the quantile regression framework by Koenker and

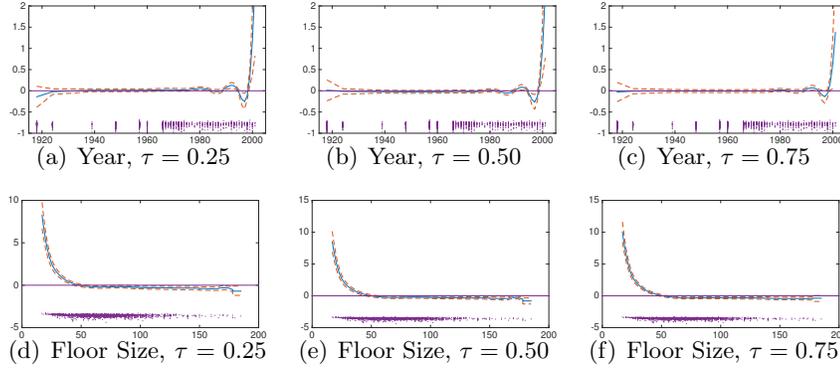


Figure 2.7: Estimated non-parametric effect using Lasso prior with 95% credible bands for Munich data.

Machado (1999). Since the standard OLS model does not provide a clear result about the convergence hypothesis of neoclassical growth models, several papers have analyzed growth equations using the quantile regression technique with noteworthy results. Barreto and Hughes (2004) show that the determinants of economic growth for countries in the left or right tails of the distribution are very different from those in the mean. Using the Barro growth model (Barro, 1991), Mello and Perrelli (2003) use quantile regression to find evidence in favor of the convergence hypothesis for countries in the upper quantile of the conditional distribution of the response variable. Finally, Laurini (2007) uses spline functions to test the convergence hypothesis with a dataset of Brazilian municipalities. To the best of our knowledge, this thesis is the first attempt to propose a Bayesian quantile Lasso AM in order to study the impact of both linear and non-linear effects of covariates on the cross-country GDP growth using the Barro and Sala i-Martin (1995) model. The dataset consists of 161 countries and includes 13 covariates covering the two periods 1965-75 and 1975-85. With a quantile AMl we are able to combine the theory of non-linear return to education with that of economic convergence, using spline functions to model the variables: Male secondary school (MSS), Female Secondary school (FSS), Male Higher Education (MHE) and Female Higher Education (FHE), adopting a linear representation for the remaining variables.

The parameter estimates of the linear covariates (Table 2.6) are consistent with previous studies based on quantile regression methods. In par-

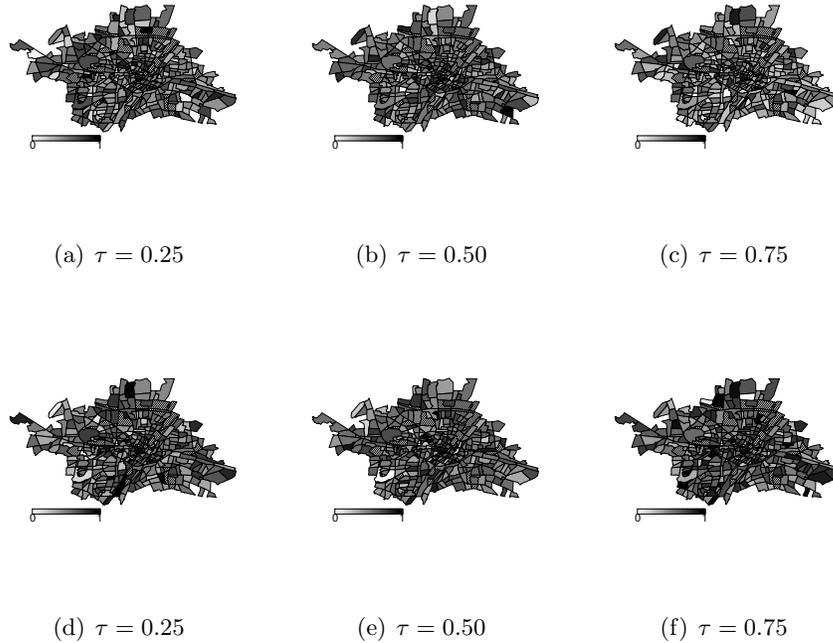


Figure 2.8: Estimated spatial effects using Gaussian (first row) and Lasso (second row) prior for the 380 subquarters of Munich

particular, it is worth noting that the coefficients related to the initial per-capita GDP are always negative, confirming the neoclassical theory about conditional convergence. Regarding the coefficients related to the public consumption as a share of GDP, previous studies on this dataset find an increasing but always negative effect of this variable on the economic growth. Looking at this evidence Koenker and Machado (1999) conclude that this variable tends to help faster-growing countries proportionally more than those countries in the lower tail of the growth distribution, thus tending to accentuate the inequality among nations. Our result increases the evidence of Koenker and Machado (1999) since our estimates suggest that there is a statistically significant effect of this variable on growth only for countries in the lower tail of the growth distribution, thus exasperating inequality among nations. Figure 2.9 displays the estimated spline functions along with their credible sets, for three quantile levels $\tau = (0.25, 0.5, 0.75)$. A noticeable non-linear path is present for almost all the selected covariates. For a given

Variable	Gaussian Prior			Lasso Prior		
	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Good location	0.6466 (0.0925)	0.7454 (0.0880)	0.7606 (0.0857)	0.6304 (0.1230)	0.7042 (0.1124)	0.5922 (0.1039)
Excellent location	1.4213 (0.2770)	1.6999 (0.2879)	1.9381 (0.2829)	1.4136 (0.2376)	1.6305 (0.2454)	1.8450 (0.2527)
No hot water	-1.3361 (0.2410)	-1.8499 (0.2664)	-2.2199 (0.2738)	-0.0353 (0.0475)	-0.0335 (0.0454)	-2.7652 (0.3731)
No central heating	-1.5449 (0.1759)	-1.4206 (0.1957)	-1.0610 (0.1867)	-1.9830 (0.1872)	-2.0557 (0.2076)	-0.1316 (0.1193)
No tiles in bathroom	-0.4260 (0.1079)	-0.5792 (0.1091)	-0.5942 (0.1140)	-0.1597 (0.1143)	-0.3277 (0.1453)	-0.2576 (0.1575)
Special bathroom interior	0.3926 (0.1462)	0.3803 (0.1580)	0.4897 (0.1489)	0.0598 (0.0606)	0.0824 (0.0779)	0.0598 (0.0581)
Special kitchen interior	0.9145 (0.1787)	1.1405 (0.1564)	1.2480 (0.1740)	1.0824 (0.3355)	1.3077 (0.2239)	1.3153 (0.1989)
1 Room	7.1564 (0.1754)	8.3633 (0.1700)	9.3637 (0.1751)	6.8372 (0.1926)	8.6886 (0.1729)	10.0425 (0.1652)
2 Rooms	6.9968 (0.1002)	8.4530 (0.1024)	10.0062 (0.0922)	6.5432 (0.1154)	8.5664 (0.1126)	10.3823 (0.1084)
3 Rooms	6.7542 (0.0998)	8.1964 (0.0927)	9.7554 (0.0947)	6.2149 (0.1159)	8.1500 (0.1101)	10.0117 (0.1076)
4 Rooms	6.2745 (0.1459)	7.7603 (0.1404)	9.2060 (0.1490)	5.7041 (0.1557)	7.5529 (0.1644)	9.3650 (0.1601)
5 Rooms	6.0948 (0.2659)	7.6821 (0.3047)	9.6398 (0.2956)	5.0121 (0.4154)	7.0744 (0.3538)	9.2734 (0.3514)
6 Rooms	6.3496 (0.4450)	7.6293 (0.4479)	9.1707 (0.4661)	0.4008 (0.0658)	0.4462 (0.0752)	8.4068 (0.5960)

Table 2.5: Posterior Means and standard errors (in parenthesis) of the linear regressors for three different quantile levels.

variable, the sign of each estimated spline varies among different quantile levels, suggesting that the importance of different types of education is not the same for countries in the lower and upper tails of the growth conditional distribution. This result is of particular interest since it isolates the positive and negative contributions of each type of education to the rate of economic growth. There are two opposite paths characterizing the effect of secondary schooling and higher education on growth: the first is increasing in the quantile level τ , and the second is decreasing. In particular, our estimates suggest that relatively low education levels help countries in the upper tail of the grow distribution, while higher education levels boost the rate of growth

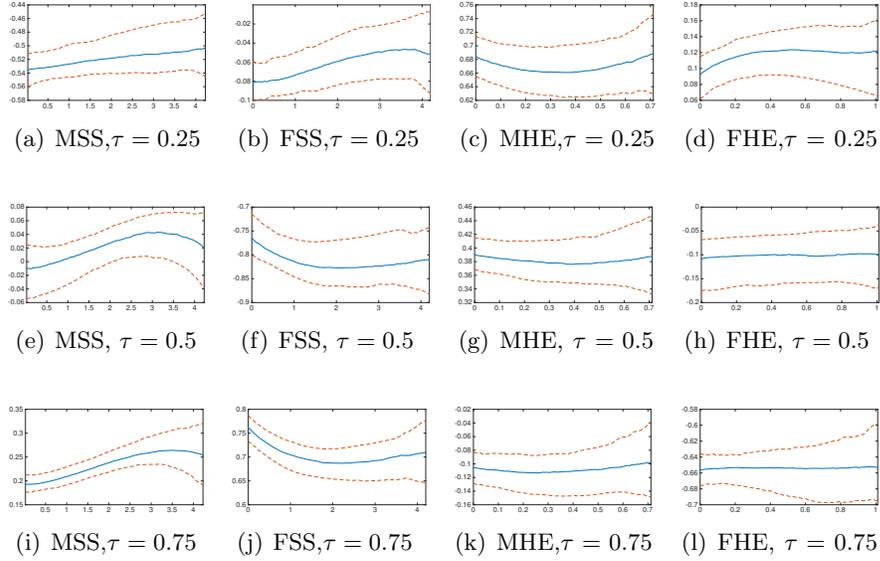


Figure 2.9: Barro dataset. Penalized splines with 95% HPD credible sets for the variables: “Male secondary school” (MSS, first column), “Female secondary school ” (FSS, second column), “Male Higher education ” (MHE, third column) and Female Higher education ” (FHE, fourth column) for five different quantile levels.

for countries in the lower tail. These results can be interpreted in view of the fact that, high and low GDP growth levels are linked with emerging and developed nations respectively. Basic schooling is a key factor for developing countries which base their economies on high labor intense activities. For these countries, costs resulting from higher levels of schooling outweigh the returns while the opposite is true for developed countries. The latter, indeed, have the possibility to take advantage of skilled labor forces to exploit the higher returns derived from available technology.

2.7 Conclusion

In this chapter we show how the SEP distribution provides a flexible tool to model the conditional quantile of a response variable as a function of exogenous covariates in a Bayesian quantile regression context. In particular, extreme observations are properly accounted for by the shape parameter governing the tails decay of the distribution, efficiently handling data with outliers or with fat tail-decay. Moreover, we extend the linear quantile

Variable	Quantile levels				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Initial Per Capita GDP	-0.0266 (0.0043)	-0.0270 (0.0048)	-0.0307 (0.0055)	-0.0293 (0.0045)	-0.0322 (0.0047)
Life Expectancy	0.0324 (0.0091)	0.0127 (0.0104)	0.1947 (0.0110)	0.2021 (0.0097)	0.1866 (0.0092)
Human Capital	-0.0024 (0.0011)	-0.0037 (0.0017)	-0.0010 (0.0020)	-0.0023 (0.0019)	-0.0018 (0.0017)
Education/GDP	-0.2707 (0.1246)	-0.1127 (0.1579)	-0.2956 (0.1690)	-0.2156 (0.1721)	-0.0993 (0.1732)
Investment/GDP	0.0999 (0.0268)	0.0930 (0.0291)	0.0359 (0.0337)	0.0343 (0.0277)	0.0482 (0.0264)
Public Consumption/GDP	-0.1628 (0.0387)	-0.1723 (0.0443)	0.0052 (0.0463)	0.0251 (0.0384)	-0.0206 (0.0280)
Black Market Premium	-0.0227 (0.0063)	-0.0267 (0.0074)	-0.0360 (0.0077)	-0.0319 (0.0064)	-0.0316 (0.0072)
Political Instability	-0.0264 (0.0083)	-0.0302 (0.0088)	-0.0153 (0.0103)	-0.0053 (0.0098)	-0.0042 (0.0073)
Growth Rate Terms Trade	0.1220 (0.0366)	0.1250 (0.0504)	0.2274 (0.0652)	0.2478 (0.0639)	0.2744 (0.0569)

Table 2.6: Posterior Means and standard errors (in parenthesis) of the linear regressors for three different quantile levels.

regression framework to the AM when quantile functions are approximated with splines. In both cases we provide a new adaptive Metropolis within Gibbs algorithm in order to implement statistical inference. In this chapter, we accommodate for variable selection problems and shrinking parameters by using the Bayesian version of Lasso penalization methods. In the linear case, we suggest the use of generalized independent Laplace priors on the regressor parameters to shrink each parameter separately, and a multivariate Laplace distribution on the spline coefficients generalizing the Lang and Brezger (2004) second order random walk prior. Finally, we demonstrate the power of these models through simulation and real dataset application where the flexibility of the quantile methodology proposed in terms of robustness and sparsity is evident.

Chapter 3

Bayesian Non–Linear Conditional Autoregressive Risk Measures

3.1 Introduction

In this chapter we introduce three main innovations on the existing literature on conditional autoregressive risk measure. First of all, we develop a unified Bayesian Conditional Autoregressive Risk model (B-CARM) which encompass both the CAViaR and the CARE one as particular case by using the Skew Exponential Power (SEP) likelihood tool (see e.g. Zhu and Zinde-Walsh, 2009 and Bernardi et al., 2016). Using the properties and the parametrization of the SEP presented in Zhu and Zinde-Walsh (2009) and extended in the quantile regression framework by Bernardi et al. (2016), we show how to estimate different class of models by varying a single parameter of the distribution. A similar idea is developed by Kobayashi (2015) which proposes a unified framework to analyse skewness, tail heaviness, quantiles and expectiles of the return distribution based on a stochastic volatility model using the SEP distribution. Differently from Kobayashi (2015) our model uses the properties of the SEP distribution in the field of risk models for Conditional Autoregressive model. As a second result of this chapter we propose a new Non–Linear and semi–parametric specification of the B-CARM class of models which uses Penalized Splines (see De Boor, 2001, Eilers et al., 1996, Lang and Brazger, 2004) to estimate the relation between the quantile/expectile and the observed variables. The need for a model that allows for non-linearity without assuming any particular restric-

tions is of clear interest in literature. Indeed, all the specifications of the CAViaR and the CARE models proposed so far emphasize the role of asymmetry and non-linearity in the relation between the observed variables and the current quantile or expectile level, but they all impose the form of this non linearity a priori (see Engle and Manganelli, 2004, Gerlach et al., 2011 and 2012, Chen et al., 2009 and 2012, Gelach and Chen, 2014). Finally, a new Adaptive-Independent Metropolis-Hastings (AIMH) algorithm is implemented to efficiently estimate the model parameters. This approach is effective in handling the non linearity of our model specification. Adaptive MCMC (AMCMC) methods extend standard MCMC algorithms since they allow the proposal parameters to be updated at each iteration to tailor the shape of the proposal distribution to that of the target, see e.g. Liang (2010). These methods do not require the prior specification of the proposal parameters and their theoretical properties are now well understood, see e.g. Andrieu and Thoms (2008) and Liang *et al.* (2010), Atchadé and Rosenthal (2005), Atchadé *et al.* (2011).

Ultimately we applied the B-CARM to five stock market indices and we compare its forecasting performances with the ones of some competitive known models. In particular we conduct a backtesting procedure showing that the B-CARM performances are in line with those of the competitors without imposing any restrictive assumption on the relations among variables. By that we can consider the B-CARM a valid and more general alternative to the extended models.

The remaining of the chapter is organized as follows: Section 3.2 briefly reviews conditional autoregressive risk models; Section 3.3 introduces the SEP as likelihood tool for the Bayesian Conditional Autoregressive Risk Model with particular case the CAViaR and CARE ones while section 3.4 proposes a non linear and semi parametric extension of it; Section 3.5 discusses the implemented Bayesian methods; Section 3.6 presents the Adaptive-Independent-Metropolis-Hastings algorithm; Section 3.7 shows results from real datasets and section 3.8 provides concluding remarks.

3.2 Conditional Autoregressive Risk Measure models

The literature on VaR and ES estimation is very wide but it is usually classified into three broad categories: non-parametric, semi-parametric and fully parametric methods. Non-parametric methods estimate the VaR without any assumption on the distribution of the portfolio returns. The most pop-

ular non-parametric VaR method is called historical simulation and provides the VaR estimation based on the quantile of the empirical distribution of historical returns. Once obtained the VaR, in this approach the ES can be estimated as the mean of the returns that exceed the VaR estimate. On the opposite side, fully parametric VaR and ES methods are based on a specific assumption of the return distribution as well as the model dynamics. A common approach is based on a Generalized Autoregressive Conditional Heteroskedastic (GARCH) models (see Bollerslev, 1986; Engle, 1982) to forecast the volatility using a Gaussian or Student's-t distribution assumption. Moreover, semi-parametric methods only assume a given model dynamics and estimates the VaR and the ES using quantile or expectile regression techniques. While the quantile estimate represents a VaR estimate itself, the expectile is not an estimate either of the τ quantile nor the θ ES, but these quantities can be recovered using the known relations between expectile, quantile and ES (Efron, 1991; Taylor, 2008). The most important VaR and ES semiparametric models are a class of conditional autoregressive risk measure models known as the CAViaR and the CARE models introduced by Engle and Manganelli (2004) and Taylor (2008) respectively.

The CAViaR class of models attempt to compute the τ -level VaR by estimating the τ -level quantile of the portfolio returns through a conditional autoregressive equations structure. Let y_t be the return at time t , the CAViaR models have the following form:

$$y_t = q_t + \varepsilon_t \tag{3.1}$$

$$q_t = \omega + \gamma q_{t-1} + \beta l(y_{t-1}) \tag{3.2}$$

where q_t is the τ level quantile of y_t defined as the value of q_t that minimizes the function $E[(\theta - I(y_t - q_t))(y_t - q_t)]$, (ω, γ, β) are parameters and $l(\cdot)$ is an unknown function of the past returns. Here, ε_t , for any $t = 1, 2, \dots, T$, are independent random variables which are supposed to have zero τ -th quantile and constant variance. As noted by Engle and Manganelli (2004), $l(\cdot)$ can be interpreted as the News Impact Curve (NIC) introduced by Engle and Ng, (1993) for ARCH-type models. The form of the function $l(\cdot)$ is one of the most addressed tasks in the risk modeling literature. Indeed, we can recognize different CAViaR models by considering different form of $l(\cdot)$ such

as:

$$l(y_t) = |y_t| \quad \text{Symmetric Absolute Value} \quad (3.3)$$

$$l(y_t) = \beta_1 (y_t)^+ + \beta_2 (y_t)^- \quad \text{Asymmetric Slope} \quad (3.4)$$

$$l(y_t) = \begin{cases} \beta_1 |y_t|, & z_t \leq r \\ \beta_2 |y_t|, & z_t > r \end{cases} \quad \text{Threshold CAViaR} \quad (3.5)$$

where, z_t is an observed threshold variable that could be exogenous or self-exciting, i.e. $z_t = y_t$ and r is the threshold variable, typically set equal to zero, i.e. $r = 0$ (Gerlach, Chen and Chan, 2011). In the frequentist approach CAViaR estimate is obtained through quantile regression methods (Koenker 2005 book, (2004)) by minimizing the loss function introduced by Koenker and Basset (1978). The Bayesian approach instead relies on the Asymmetric Laplace distribution (ALD) assumption as a likelihood tool to perform the inferential issue (see e.g. Yu and Moyeed, 2001, Kottas and Gelfand, 2001, Kottas and Krnjajic, 2009, and Sriram et al., 2013, Bernardi et al., 2015). The same structure of the CAViaR models is used to build the CARE models defined as:

$$y_t = \mu_t + \varepsilon_t \quad (3.6)$$

$$\mu_t = \omega + \gamma\mu_{t-1} + \beta l(y_{t-1}) \quad (3.7)$$

where μ_t is the θ -th expectile of y_t defined as the value of μ_t that minimizes the function $E \left[|(\theta - I(y_t - \mu_t))| (y_t - \mu_t)^2 \right]$. In this model ε_t , for any $t = 1, 2, \dots, T$, are independent random variables now supposed to have zero θ -th expectile and constant variance. The specifications of $l(\cdot)$ in equation (3.3)–(3.5) remain valid to define different CARE models. The estimation procedure for the generic expectile is addressed in the frequentist approach by using the Asymmetric Least Square (ALS) estimator as in Newey and Powell (1987) while in the Bayesian paradigm the literature relies on the Asymmetric Gaussian distribution assumption (see e.g. Gerlach and Wang, 2015, Gerlach and Chen, 2014; Wichitaksorn et al, 2014; Gerlach et al, 2016, Gerlach and Chen, 2016). Finally, since the expectile do not represent a risk measure itself an other passage is required to obtain VaR and ES from the estimated expectile. In the first case we simply use the expectile as an estimator of the quantile by iteratively searching for the θ -th expectile for which we observe $\tau\%$ observations below it. This procedure was suggested by Efron (1991) and allows to obtain the quantile of interest and consequently the associated VaR level. The ES is instead obtained using the one to one

mapping between expectile and ES suggested in Taylor (2008) given by:

$$ES_t(\tau) = \left(1 + \frac{\theta}{(1-2\theta)\tau}\right) \mu_t(\theta) - \frac{\theta}{(1-2\theta)\tau} E(y_t) \quad (3.8)$$

In the next two sections we present how to implement a unified framework for CAViaR and CARE estimation and provide a completely new way to model the function $l(\cdot)$ that extend and generalize upon all the current models.

3.3 Bayesian CARM model

We develop a unified Bayesian Conditional Autoregressive Risk model (B-CARM) which encompass both the CaViaR and the CARE one as particular case by using the Skew Exponential Power (SEP) likelihood tool (see e.g. Zhu and Zinde-Walsh, 2009 and Bernardi et al., 2016). Using the properties and the parametrization of the SEP presented in Zhu and Zinde-Walsh (2009) and extended in the quantile regression framework by Bernardi et al. (2016), we show how to estimate different class of models by varying a single parameter of the distribution. We call this unified framework for a Bayesian estimation of different risk measure, the Bayesian Conditional Autoregressive Risk Measure (B-CARM) class of models. This is specified by the following equation:

$$y_t = g_t + \varepsilon_t \quad (3.9)$$

$$g_t = \omega + \gamma g_{t-1} + \beta l(y_{t-1}) \quad (3.10)$$

where g_t in this model can be for example either the quantile or the expectile and ε_t for any $t = 1, 2, \dots, T$, are independent random variables than can have either zero τ -th quantile or θ -th expectile according g_t . This can be obtained basing the Bayesian estimation strategy on the SEP likelihood tool. We make use of the parametrization of the SEP distribution proposed by Zhu and Zinde-Walsh (2009), that has the following form:

$$f_{\text{SEP}}(y_t; g_t, \sigma, \tau, \alpha) = \begin{cases} \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp\left\{-\frac{1}{\alpha} \left(\frac{g_t - y_t}{2\tau\sigma}\right)^\alpha\right\}, & \text{if } y_t \leq g_t \\ \frac{1}{\sigma} \kappa_{\text{EP}}(\alpha) \exp\left\{-\frac{1}{\alpha} \left(\frac{y_t - g_t}{2(1-\tau)\sigma}\right)^\alpha\right\}, & \text{if } y_t > g_t, \end{cases}$$

where g_t is the location parameter and $\tau \in (0, 1)$ is the skewness parameter. Moreover, $\sigma \in \mathfrak{R}^+$ and $\alpha \in (0, \infty)$ are the scale and shape parameters, respectively, while $\kappa_{\text{EP}} = \left[2\alpha^{\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)\right]^{-1}$ and $\Gamma(\cdot)$ is the complete

gamma function. In their papers Ayebo and Kozubowski (2004) and Zhu and Zinde-Walsh (2009) presented the most important properties of the SEP distribution. They highlight that the SEP can be seen as a generalization of the Asymmetric Laplace (AL) and the Asymmetric Gaussian (AG) distributions that can be recovered for $\alpha = 1$ and $\alpha = 2$ respectively. This imply that if we do not estimate the parameter α from the data, g_t becomes either the quantile (for $\alpha = 1$) or the expectile (for $\alpha = 2$) of y_t obtaining the specification of the CAViaR and the CARE models respectively, i.e.:

$$g_t = q_t \quad \text{for } \alpha = 1 \quad (3.11)$$

$$g_t = \mu_t \quad \text{for } \alpha = 2 \quad (3.12)$$

As before, since the expectile do not represent a risk measure itself, when α is chosen to be equal to two, i.e. $\alpha = 2$, we estimate the VaR by iteratively searching for the expectile for which we observe a given percentage of observations below it, while the ES is obtained using the one to one mapping between expectile and ES given in equation (3.8).

3.4 Bayesian Non-linear CARM model

To keep as general as possible the model specified in equation (3.10) we present a new Bayesian Non-linear CARM (BNL-CARM) model that accounts for any possible non-linearity between the risk measure of interest and the observed information by using a spline approach, in particular B-spline functions.

To this aim in equation (3.10) we approximate $l(\cdot)$ by a B-spline of order d with k equally spaced knots. Using the properties of the B-spline, we can define the BNL-CARM model as follows:

$$y_t = g_t + \varepsilon_t \quad (3.13)$$

$$g_t = \omega + \gamma g_{t-i} + \sum_{\nu=1}^{k+d} \beta_{\nu} B_{\nu}(y_{t-1}) \quad (3.14)$$

where $B_{\nu}(y_{t-1})$ denote B-spline basis functions and β_{ν} are unknown coefficients to be determined. The proposed B-CARM equation (3.14) allows to answer the need for non linearity assuming a more general structure. In fact, the form of the non linearity between innovations and the risk measure, introduced by the existing models, is always imposed a priori by specifying a given function for the past returns as in equation (3.3)-(3.5). Using

B-Spline instead, we model the function $l(\cdot)$ that is implicit in the data, without assuming any a priori structure for it.

As it is well known in equation (3.14) it is important to take into account for the sensitivity of the values of the estimated coefficients and the shape of the fitted Spline functions upon the number and the position of the knots. In absence of any prior information about their position, equidistant knots is a natural choice and it becomes crucial to establish the number of equispaced knots. To catch properly the smoothness of the data a trade off arise since few knots or too many knots may cause underfitting or overfitting respectively. A possible solution of this problem is known as Penalized Spline (P-Spline) and proposed by O’Sullivan (1986 and 1988) and generalized by Eilers and Marx (1996). To ensure enough flexibility without incurring in the overfitting problem they propose to use a relatively large number of knots jointly with some form of penalization to smooth sufficiently the fitted curve. Specifically, the approach followed by Eilers and Marx (1996) relies on the introduction of a penalty element on the first or second differences of the B-Spline coefficients. This setting was translated in a Bayesian framework by Lang and Brezger (2004), Brezger and Lang (2006) and Brezger and Steiner (2008), where they use a random walk as stochastic analogous of the frequentist approach. In this work we assume a second order random walk for all the B-Spline coefficients, that is

$$\beta_\nu = 2\beta_{\nu-1} - \beta_{\nu-2} + u_\nu, \quad \forall \nu = 1, 2, \dots, k + d, \quad (3.15)$$

where the generic stochastic component u_ν has a Gaussian distribution with zero mean and variance equal to ϕ^2 , i.e. $u_\nu \sim \mathcal{N}(0, \phi^2)$ and $\beta_{\nu-1}$ and $\beta_{\nu-2}$ are initialized with diffuse priors (i.e. $\propto 1$). The smoothness of the fitted curve is controlled by the variance of the error term which correspond to the inverse of the penalization parameter used by Eilers and Marx (1996) in the frequentist framework. We choose a conjugate Inverse Gamma prior for ϕ^2 , that is $\phi^2 \sim \mathcal{IG}(a^{(\phi)}, b^{(\phi)})$ with $a^{(\phi)} = b^{(\phi)} = 0.001$. Different choices of hyper parameters are allowed but they all bring to very similar results. Finally, it is possible to write the prior distribution for the B-Spline coefficients that is

$$\pi(\boldsymbol{\beta} | \phi) \propto \mathcal{N}_{k+d} \left(0, \phi^2 (D_2' D_2)^{-1} \right), \quad (3.16)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{k+d})'$ and D_2 is the difference matrix of dimension $(k + d - 2) \times (k + d)$ and order 2 derived by the order of the random walk in equation (3.15)

3.5 Bayesian Methods

Bayesian inference requires the specification of the likelihood function as well as the prior distribution for all the parameters of interest. Using the parametrization of the SEP distribution showed in section 3.3 we can write the likelihood function of the model, for a given τ as:

$$\begin{aligned}
\mathcal{L}_\tau(\Omega, \sigma, | \alpha, \mathbf{y}) &= \prod_{t=1}^T \frac{1}{2\sigma} \frac{\alpha^{-\frac{1}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})} \left[\exp \left\{ -\frac{1}{\alpha} \left(\frac{g_t(\Omega) - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq g_t(\Omega))} \right. \\
&\quad \left. + \exp \left\{ -\frac{1}{\alpha} \left(\frac{y_t - g_t(\Omega)}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > g_t(\Omega))} \right] \\
&= \frac{1}{(2\sigma)^T} \frac{\alpha^{-\frac{T}{\alpha}}}{\Gamma(1 + \frac{1}{\alpha})^T} \left[\exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{g_t(\Omega) - y_t}{2\tau\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t \leq g_t(\Omega))} \right. \\
&\quad \left. + \exp \left\{ -\frac{1}{\alpha} \sum_{t=1}^T \left(\frac{y_t - g_t(\Omega)}{2(1-\tau)\sigma} \right)^\alpha \right\} \mathbb{1}_{(y_t > g_t(\Omega))} \right], \tag{3.17}
\end{aligned}$$

where the vector $\Omega = (\omega, \gamma, \beta)$ collects all the BNL-CARM parameters. The parameter α is fixed equal to 1 or 2, i.e. $\alpha = 1$ or $\alpha = 2$, depending on the model we want to estimate, as showed in previous section. Concerning the prior specification, we assume the following hierarchical prior structure independent on the value of τ :

$$\pi(\Xi) = \pi(\beta | \phi^2) \pi(\phi^2) \pi(\omega) \pi(\gamma) \pi(\sigma), \tag{3.18}$$

with

$$\pi(\beta | \phi^2) \propto \mathcal{N}_{k+d} \left(0, \phi^2 (D_2' D_2)^{-1} \right), \tag{3.19}$$

$$\pi(\phi^2) \propto \mathcal{IG} \left(a^{(\phi)}, b^{(\phi)} \right) \tag{3.20}$$

$$\pi(\gamma) \propto \mathcal{N} \left(0, \sigma_\gamma^2 \right) \tag{3.21}$$

$$\pi(\omega) \propto \mathcal{N} \left(0, \sigma_\omega^2 \right) \tag{3.22}$$

$$\pi(\sigma) \propto \mathcal{IG} \left(a, b \right), \tag{3.23}$$

where $\Xi = (\beta, \phi^2, \omega, \gamma, \sigma)$, $(a^{(\phi)}, b^{(\phi)}, \sigma_\gamma^2, \sigma_\omega^2, a, b)$ are given positive hyperparameters, while $\mathcal{N}(\cdot)$ and $\mathcal{IG}(\cdot)$ denote the Normal and the Inverse Gamma distributions respectively.

3.6 The Adaptive Metropolis within Gibbs sampler

The Bayesian inference is carried out using an adaptive MCMC sampling scheme based on the following posterior distribution

$$\pi(\Xi | \mathbf{y}) \propto \mathcal{L}(\Omega, \sigma | \alpha, \mathbf{y}) \pi(\beta | \phi^2) \pi(\phi^2) \pi(\omega) \pi(\gamma) \pi(\sigma), \quad (3.24)$$

where $\mathcal{L}(\Omega, \sigma | \alpha, \mathbf{y})$ indicates the likelihood function specified in equation (3.17). After choosing a set of initial values for the parameter vector $\Xi^{(0)}$, simulations from the posterior distribution at the i -th iteration of $\Xi^{(i)}$, for $i = 1, 2, \dots$, are obtained by running iteratively a block-move Independent Metropolis within Gibbs (IMG). The simulation algorithm requires as first step the specification of a proposal distribution for parameters (Ω, σ) . To propose a move for each block of the parameters, we choose the following proposal distributions:

$$q(\beta_{i-1}, \beta_i^*) \sim \mathcal{N}^{k+d}(\mu_\beta^{(i)}, \Sigma_\beta^{(i)}) \quad (3.25)$$

$$q(\omega_{i-1}, \omega_i^*) \sim \mathcal{N}(\mu_\omega^{(i)}, \psi_\omega^{(i)}) \quad (3.26)$$

$$q(\gamma_{i-1}, \gamma_i^*) \sim \mathcal{N}(\mu_\gamma^{(i)}, \psi_\gamma^{(i)}) \quad (3.27)$$

$$q(\sigma_{i-1}, \sigma_i^*) \sim \mathcal{N}_{\tilde{\sigma}}(\mu_{\tilde{\sigma}}^{(i)}, \psi_{\tilde{\sigma}}^{(i)}) 1/\sigma_i^* \quad (3.28)$$

where the scale parameter $\tilde{\sigma} = \log(\sigma)$ is proposed on a log-scale and subsequently transformed to preserve positiveness. The jacobian term in equation (3.28) is required to get the distribution of the transformation $\sigma = \exp(\tilde{\sigma})$. At each iteration $i = 1, 2, \dots$, the IMG algorithm proceeds by simulating a candidate draw from each parameter block, i.e. $\Upsilon^* = (\xi_1^*, \xi_2^*, \xi_3^*, \xi_4^*) = (\beta^*, \omega^*, \gamma^*, \sigma^*)$ which is subsequently accepted or rejected. The generic probability that the proposed candidate parameter ξ_j^* , for $j = 1, 2, 3, 4$ becomes the new state of the chain is evaluated on the basis of the following acceptance probability

$$\lambda(\xi_j^{(i-1)}, \xi_j^*) = \min \left\{ 1, \frac{\mathcal{L}(\xi_j^*, \Xi_{-j}^{(i-1)} | \mathbf{y}, \mathbf{x})}{\mathcal{L}(\Xi^{(i-1)} | \mathbf{y}, \mathbf{x})} \frac{\pi(\xi_j^*)}{\pi(\xi_j^{(i-1)})} \frac{q(\xi_j^{(i-1)})}{q(\xi_j^*)} \right\},$$

for $j = 1, 2, 3, 4$, where $\lambda(\xi_j^{(i-1)}, \xi_j^*)$ indicates the probability to move from the old to the proposed state of the chain, $\pi(\cdot)$ is the generic prior given in

section 3.5 and $\Xi_{-j}^{(i-1)}$ refers to the whole set of parameters at iteration $i-1$ without the j -th element of Υ^* .

Following Brezger and Steiner (2008) the variance parameters $\phi_1^2, \dots, \phi_J^2$ are updated by single-move Gibbs sampling steps since the full conditional distributions for these parameters is still proportional to an Inverse Gamma with shape $a_i^{(\phi)} = a_{i-1}^{(\phi)} + \text{rank}(D^{2'} D^2) / 2$ and scale $b_i^{(\phi)} = b_{i-1}^{(\phi)} + \beta' D^{2'} D^2 \beta$. Since most of the statistical properties of the Markov chain as well as the performance of the Monte Carlo estimators crucially depend on the definition of the proposal distribution $q(\cdot)$ (see Andrieu and Moulines, 2006 and Andrieu and Thoms, 2008) we improve the basic IMG-MCMC algorithm with an additional step adapting the proposal parameters using the following equations:

$$\boldsymbol{\mu}_\beta^{(i+1)} = \boldsymbol{\mu}_\beta^{(i)} + \delta^{(i+1)} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right), \quad (3.29)$$

$$\boldsymbol{\Sigma}_\beta^{(i+1)} = \boldsymbol{\Sigma}_\beta^{(i)} + \delta^{(i+1)} \left(\left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right) \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta^{(i)} \right)^\top - \boldsymbol{\Sigma}_\beta^{(i)} \right), \quad (3.30)$$

$$\mu_\omega^{(i+1)} = \mu_\omega^{(i)} + \zeta^{(i+1)} \left(\omega - \mu_\omega^{(i)} \right), \quad (3.31)$$

$$\psi_\omega^{(i+1)} = \psi_\omega^{(i)} + \zeta^{(i+1)} \left(\left(\omega - \mu_\omega^{(i)} \right)^2 - \psi_\omega^{(i)} \right), \quad (3.32)$$

$$\mu_\gamma^{(i+1)} = \mu_\gamma^{(i)} + \zeta^{(i+1)} \left(\gamma - \mu_\gamma^{(i)} \right), \quad (3.33)$$

$$\psi_\gamma^{(i+1)} = \psi_\gamma^{(i)} + \zeta^{(i+1)} \left(\left(\gamma - \mu_\gamma^{(i)} \right)^2 - \psi_\gamma^{(i)} \right), \quad (3.34)$$

$$\mu_{\tilde{\sigma}}^{(i+1)} = \mu_{\tilde{\sigma}}^{(i)} + \zeta^{(i+1)} \left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right), \quad (3.35)$$

$$\psi_{\tilde{\sigma}}^{(i+1)} = \psi_{\tilde{\sigma}}^{(i)} + \zeta^{(i+1)} \left(\left(\tilde{\sigma} - \mu_{\tilde{\sigma}}^{(i)} \right)^2 - \psi_{\tilde{\sigma}}^{(i)} \right), \quad (3.36)$$

where $\zeta^{(i+1)}$ denotes a tuning parameter that should be carefully selected at each iteration to ensure the convergence and the ergodicity of the resulting chain (see Andrieu and Moulines, 2006). Roberts and Rosenthal (2007) provide two conditions for the convergence of the chain: the diminishing adaptation condition, which is satisfied if and only if $\zeta^{(i)} \rightarrow 0$, as $i \rightarrow +\infty$, and the bounded convergence condition, which essentially guarantees that all transition kernels considered have bounded convergence time. Andrieu and Moulines (2006) show that both conditions are satisfied if and only if $\zeta^{(i)} \propto i^{-d}$ where $d \in [0.5, 1]$. For those reasons we choose $\zeta^{(i)} = \frac{1}{C i^{0.5}}$ where C is set to 10, i.e. $C = 10$. As argued by Roberts and Rosenthal (2007), together these two conditions ensure asymptotic convergence and a weak

law of large numbers for this algorithm.

3.7 Empirical applications

To assess the goodness of the proposed models we evaluate their forecasting performance over five stock market indices: Nasdaq (US); Straits Times Index (STI, Singapore); Hang Seng Index (Hong Kong); Corea SE (Corea) and AEX (Holland). Daily closing stock prices (P_t) from January 1, 1988 to June 1, 2015 are obtained from the Thomson Reuters (Datastream) database. The percentage returns are computed as $y_t = (\log(P_t) - \log(P_{t-1})) \times 100$ and the full sample is divided into an in-sample period (from January 1, 1988 to December 30, 2005) and a forecast (out-of-sample) period of 2456 observations (from January 1, 2006 to June 1, 2015), which covers both the 2008-2009 General Financial Crisis and the following Sovereign Debt Crisis. Summary statistics are displayed in Table 3.1. All the series present the typical stylized facts of financial returns such as positive kurtosis and negative skewness. Moreover, the p-value of the Jarque-Bera test (9-th column) always reject the null hypothesis of normality. The following two subsections aim to evaluate the forecast ability of the BNL-CAViaR, i.e. BNL-CARM with $\alpha = 1$, and BNL-CARE model, i.e. BNL-CARM with $\alpha = 2$, for all the series considered. Tables 3.2 and 3.3 show the results of the out-of-sample estimation exercise conducted on all the series specified above.

	Mean	Median	Std.	Min	Max	Q1	Q3	Skewness	Normality test	Kurtosis
NASDAQ	0.0482	0.1189	1.7123	-11.1149	17.2030	-0.7369	0.8655	0.0535	0.0010	9.0442
STI	0.0204	0.0225	1.2346	-10.5446	12.8738	-0.5361	0.6005	-0.0883	0.0010	11.8515
Hang Seng	0.0347	0.0000	1.5914	-24.5202	17.2471	-0.6120	0.7641	-0.5602	0.000	20.5566
Corea SE	0.0194	0.0000	1.6768	-12.8047	11.2844	-0.7112	0.7781	-0.1339	0.000	8.4068
AEX	0.0258	0.0362	1.3050	-9.5903	10.0283	-0.5399	0.6472	-0.1622	0.000	10.0313

Table 3.1: Summary statistics.

3.7.1 CAViaR forecast evaluation

CAViaR models allow to obtain dynamic VaR measures. The accuracy of the VaR estimate is evaluated toward back-testing procedure to verify how accurately the strategy or method predict actual results. Therefore we consider some common criteria to evaluate actual returns with their 1 day ahead VaR forecast and compare the results obtained using the BNL-CAViaR model with those from four very known competing Bayesian models: the Symmetric Absolute Value (SAV) model; The Asymmetric Slope (AS) model; the Threshold CAViaR (T-CAViaR) and the Inverse Garch (IG) model. Table

3.2 shows the results of the most common back-testing methods applied to all the series for two τ levels 0.01 and 0.05. The first column reports the ratio between the actual and the expected number of violations for a given coverage level τ , i.e. $A/E = \frac{1}{m\tau} \sum_{t=n+1}^{n+m} I(y_t < -VaR_t)$, where $m = 2456$ is the length of the forecast window. The last three columns report p-values for three very common back testing methods: the unconditional (UC) and conditional (CC) coverage tests of Kupiec (1995) and Christoffersen (1998) and the CAViaR Dynamic Quantile (DQ) test of Engle and Manganelli (2004). The first two are likelihood ratio tests based on the assumption that the hit variable $I_t(\tau)$ defined as

$$I_t(\tau) = \begin{cases} 1, & \text{if } y_t < -VaR_{t|t-1} \\ 0, & \text{else} \end{cases}$$

has a Bernoulli distribution with probability τ . Under this hypothesis, the UC test verifies that the violation probability is equal to the coverage rate, i.e. $P(I_t(\tau)) = E(I_t(\tau)) = \tau$, while the CC test, in addition to UC test, also examines the independence hypothesis between violations observed at two different dates. The DQ test is instead a regression type test based on the de-meaned process associated to $I_t(\tau)$, namely $H_t(\tau) = I_t(\tau) - \tau$. This test uses a regression model to assess the hypothesis of a linear relation between H_t , its lagged values and other relevant regressors and is known to be more powerful than the UC and the CC tests.

3.7.2 CARE forecast evaluation

As described in section 3.2, CARE models are used to estimate the θ -th expectile for which the proportion of observations below it is $\tau\%$. The value of θ that satisfies this condition is obtained by estimating models for different values of θ , over a grid with step size 0.0001, stopping the process when the condition stated above becomes true. The θ values, founded in this way and referred to the same set of models used to assess CAViaR forecast ability, are reported in table 3.3. As said in section 3.2, unlike CAViaR models, CARE models are not a risk measure themselves, hence another step is required to produce VaR and ES estimates. For the former, it is sufficient to replace $\mu_t(\theta)$ with $q_t(\tau)$ since, by construction, $\mu_t(\theta)$ represents an estimate of the τ quantile. For the latter, equation (3.8) allows to map the θ -th conditional expectile, $\mu_t(\theta)$, in to the τ -th expected shortfall, $ES_t(\tau)$. A complete and recent review for backtesting VaR and ES is proposed by Roccioletti (2016). VaR results are evaluated using the same techniques showed in the previous subsection, i.e. CC, UC and DQ test. ES results are instead more difficult

to evaluate and optimal assessment for ES forecasts is still an issue under investigation. Here we follow the approach of McNeil et al. (2005) and Taylor (2008) based on a direct test of the residuals, i.e. the difference between the observations and the ES level, for only those observations beyond the quantile VaR prediction. The test assess whether the residuals, standardized by the conditional volatility (the conditional quantile estimate in our case), are i.i.d with zero mean. Specifically, a bootstrap test is implemented (as in Efron and Tibshirani, 1993 page 224) in order to avoid distributional assumptions. Moreover, we implement the three test proposed by Acerbi and Szekely (2014) and further described in Roccioletti (2016), to which we refer for details. Shortly, the first test is based on the same idea of the residual test proposed by McNeil et al. (2005), the second test derives from the representation of the ES as an unconditional expectation and, finally, the third test relies on the possibility to backtest the tails of a model by checking if the observed ranks are i.i.d. $U(0, 1)$ as they should if the model distribution is correct.

3.7.3 Summary of VaR and ES results

Overall, the BNL-CARM method proposed here for VaR and ES estimation shows a forecast performance in line with the competitors. For a given series, the VRate/ α levels are similar among the models and consequently it is not possible to define a model that clearly outperforms the others in terms of number of violations. Moreover, following the traffic light approach suggested by the Basel Committee (1996), all the models considered can be classified as acceptable (green zone) or at least disputable (yellow zone) but they never appear to be seriously flawed (red zone). Also looking at the p-values we can see that the proposed BNL-CARM model has similar performances to that of competitors. The results printed in tables 3.2 and 3.3 allow us to consider the proposed BNL-CARM as a valid alternative method for VaR and ES estimation, despite the very long out of sample period that exasperate the possibility that the markets moved in a way that could not be anticipated. In figures 3.1 and 3.2 we plot the posterior estimate of the NICs from the BNL-CAViaR (BNL-CARM with $\alpha = 1$) and BNL-CARE (BNL-CARM with $\alpha = 2$) models respectively. For each figure, left panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.05)$, while right panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.01)$.

Although the performance comparisons do not reveal a uniformly superior

model, our non linear specification based on the P-Spline functions should be preferred since it has at least two important advantages. First, from figures 3.1 and 3.2 we can see that the proposed BNL-CARM provides an estimate of the entire shape of NIC differently from the other models known in the literature which impose a given structure of the NIC and provide only a partial estimate of its shape. From the figures it can be seen different NIC naturally arise from each model giving us unconstrained information on the relation between the risk measure at a given time t and the past observation y_{t-1} .

Second, the P-Spline approach used to model the NIC, also allows us to obtain an estimate of the threshold level of return from which an asymmetric response to the risk is observed. A common stylized fact of financial time series is that positive and negative returns have a different impact on volatility. For this reason models like Asymmetric Slope and Threshold CAViaR/CARE with $r = 0$ (see equation (3.4) and (3.5)) consider two different parameters for positive and negative returns. A drawback of these models is represented by the fact that a threshold equal to zero is assumed while in most cases this does not happen and a threshold slightly higher than zero can be observed. In our model, the threshold level of return that determines an asymmetric response to the risk is naturally estimated through the NIC. In fact it corresponds to the point of (local) minimum of the NIC that, in figures 3.1 and 3.2, is in general slightly greater than zero. This evidence is besides consistent with the recent financial crisis which has increased the degree of risk aversion among traders.

Name	$\tau = 0.01$							$\tau = 0.05$						
	A/E	AD Viol.		Viol.	LR _{uc}	LR _{cc}	DQ	A/E	AD Viol.		Viol.	LR _{uc}	LR _{cc}	DQ
		Mean	Max						Mean	Max				
NASDAQ														
BNL-CAViaR	0.9294	8.1873	16.8951	22.0000	0.7285	0.7659	0.1647	1.1153	4.9302	15.0650	132.0000	0.2042	0.1617	0.0081
SAV	0.8450	7.8827	18.4128	20.0000	0.4372	0.6236	0.0021	1.1238	4.8217	14.5786	133.0000	0.1736	0.2368	0.0073
AS	0.9717	6.8868	16.6069	23.0000	0.8910	0.7904	0.7459	1.1491	4.7502	15.2862	136.0000	0.1026	0.0753	0.0054
T-CAViaR	1.0562	6.5224	16.5475	25.0000	0.7838	0.7374	0.0166	1.1660	4.7278	15.3603	138.0000	0.0699	0.0486	0.0061
IG	0.7605	7.5507	16.3418	18.0000	0.2221	0.4134	0.1445	1.0646	4.7801	15.1120	126.0000	0.4721	0.5928	0.0002
STI														
BNL-CAViaR	1.6779	5.5242	12.4027	40.0000	0.0024	0.0093	0.0402	1.4178	3.7918	15.8468	169.0000	0.0000	0.0000	0.0001
SAV	1.1326	6.1693	12.2386	27.0000	0.5228	0.4902	0.4106	1.1997	4.0479	15.2211	143.0000	0.0295	0.0926	0.3659
AS	1.4262	5.9951	13.0748	34.0000	0.0491	0.0360	0.0357	1.1829	4.1119	16.8616	141.0000	0.0457	0.0903	0.0284
T-CAViaR	1.3842	5.8490	12.6181	33.0000	0.0744	0.0462	0.0393	1.1913	4.1962	17.2827	142.0000	0.0368	0.1111	0.0469
IG	1.2584	6.2053	13.4765	30.0000	0.2220	0.3236	0.4469	1.1326	4.2092	14.8665	135.0000	0.1443	0.3031	0.5170
HANG SENG														
BNL-CAViaR	0.8554	7.0580	16.8589	21.0000	0.4615	0.6361	0.2374	1.1568	5.0442	24.2631	142.0000	0.0808	0.1508	0.2853
SAV	0.9776	6.9059	17.3296	24.0000	0.9124	0.7842	0.2460	1.0509	5.3408	21.5952	129.0000	0.5627	0.8029	0.6653
AS	0.7739	7.3706	18.0361	19.0000	0.2419	0.4347	0.9304	1.2057	5.2312	24.2745	148.0000	0.0230	0.0412	0.1990
T-CAViaR	1.1405	7.4650	17.8806	28.0000	0.4924	0.4901	0.7147	1.2464	5.3132	25.7873	153.0000	0.0068	0.0169	0.0452
IG	0.9776	7.2290	17.7004	24.0000	0.9124	0.7842	0.3500	1.0998	5.3182	21.4609	135.0000	0.2619	0.4527	0.4413
COREA SE														
BNL-CAViaR	0.8554	9.0909	19.2306	21.0000	0.4615	0.6361	0.0285	1.0916	5.1322	16.4552	134.0000	0.3021	0.2357	0.0899
SAV	0.9369	8.8752	21.0103	23.0000	0.7522	0.7653	0.0471	1.2464	4.8307	17.0654	153.0000	0.0068	0.0031	0.0000
AS	0.8961	8.3305	17.3556	22.0000	0.5999	0.7142	0.0009	1.2464	4.8185	17.8016	153.0000	0.0068	0.0134	0.0206
T-CAViaR	0.8961	8.7924	21.6213	22.0000	0.5999	0.7142	0.0005	1.3116	4.5555	17.7849	161.0000	0.0007	0.0023	0.0034
IG	0.7739	9.5466	21.4519	19.0000	0.2419	0.4347	0.0391	1.1813	4.9746	17.5884	145.0000	0.0444	0.0250	0.0002
AEX														
BNL-CAViaR	1.1813	6.3214	15.9358	29.0000	0.3791	0.4801	0.5302	1.1976	4.5020	16.3289	147.0000	0.0288	0.0840	0.2291
SAV	1.3035	6.7077	17.6953	32.0000	0.1482	0.2304	0.0503	1.2301	4.6909	15.7095	151.0000	0.0113	0.0266	0.0123
AS	1.1405	6.6591	18.0329	28.0000	0.4924	0.5719	0.5881	1.2464	4.6245	16.5510	153.0000	0.0068	0.0183	0.1031
T-CAViaR	1.1405	6.7541	17.7252	28.0000	0.4924	0.5719	0.7629	1.2546	4.6429	17.0975	154.0000	0.0052	0.0183	0.0852
IG	1.5071	6.4840	18.6003	37.0000	0.0187	0.0358	0.0232	1.1813	4.6609	16.4337	145.0000	0.0444	0.0923	0.1995

Table 3.2: Summary statistics for CAViaR models. The columns, denoted by LR_{uc}, LR_{cc} and DQ, report the p-values of unconditional and conditional coverage tests of Kupiec (1995) and Christoffersen (1998) and those of the CAViaR Dynamic Quantile (DQ) test of Engle and Manganelli (2004).

Name	$\tau = 0.01$							$\tau = 0.05$						
	A/E	AD Viol.		Viol.	LR _{uc}	LR _{cc}	DQ	A/E	AD Viol.		Viol.	LR _{uc}	LR _{cc}	DQ
		Mean	Max						Mean	Max				
NASDAQ	0.5495	9.7970	17.0521	13.0000	0.0161	0.0515	0.0148	0.9552	5.4016	15.0456	113.0000	0.6179	0.4426	0.0023
BNL-CARE	0.5917	8.8546	16.4286	14.0000	0.0309	0.0896	0.0755	1.1243	4.9644	15.1434	133.0000	0.1721	0.1344	0.0058
SAV	0.7185	8.4857	16.4863	17.0000	0.1479	0.3104	0.1938	1.1158	4.9261	15.6430	132.0000	0.2025	0.0222	0.0019
AS	0.6762	8.5946	16.5967	16.0000	0.0931	0.2189	0.1550	1.0651	4.9908	15.9801	126.0000	0.4692	0.0644	0.0011
T-CAViaR	0.5495	9.1913	16.3818	13.0000	0.0161	0.0515	0.0456	1.0482	5.0311	15.7148	124.0000	0.5903	0.7025	0.0043
IG														
STI														
BNL-CARE	1.1514	7.2479	16.7349	27.0000	0.4706	0.4689	0.0006	1.3049	4.1243	14.8784	153.0000	0.0012	0.0048	0.0007
SAV	1.1087	7.5478	16.2270	26.0000	0.6015	0.5025	0.0000	1.1940	4.3083	13.9796	140.0000	0.0358	0.0968	0.0565
AS	1.2367	7.5753	17.2724	29.0000	0.2657	0.3614	0.0000	1.2793	4.3044	14.9382	150.0000	0.0028	0.0114	0.0000
T-CAViaR	1.2793	7.0945	14.8624	30.0000	0.1918	0.2994	0.0021	1.2623	4.3091	15.2040	148.0000	0.0050	0.0192	0.0000
IG	0.9382	7.5553	16.2511	22.0000	0.7626	0.7756	0.0373	0.7591	5.0063	14.0010	89.0000	0.0053	0.0143	0.0007
HANG SENG														
BNL-CARE	0.6897	10.4212	23.0460	16.0000	0.1120	0.2530	0.0232	1.1638	5.5433	22.0030	135.0000	0.0766	0.0241	0.0528
SAV	0.5172	7.0467	18.0908	12.0000	0.0101	0.0343	0.0284	0.9914	5.5299	23.4589	115.0000	0.9278	0.9475	0.9531
AS	0.6034	6.0396	10.0313	14.0000	0.0384	0.1077	0.4178	1.2414	5.3341	25.0204	144.0000	0.0099	0.0194	0.0194
T-CARE	0.7328	8.2293	12.4659	17.0000	0.1750	0.3516	0.8843	1.1724	5.4944	15.6011	136.0000	0.0626	0.0868	0.0141
IG	0.7759	7.1842	17.4372	18.0000	0.2596	0.4601	0.0976	1.1466	5.4345	22.1726	133.0000	0.1121	0.2295	0.5137
COREA SE														
BNL-CARE	0.7290	9.6112	19.1914	17.0000	0.1676	0.3406	0.1873	1.0978	5.2079	16.0184	128.0000	0.2838	0.0547	0.0735
SAV	0.8148	9.3384	20.9001	19.0000	0.3540	0.5567	0.1052	1.0377	5.3919	17.3147	121.0000	0.6742	0.0392	0.0035
AS	0.6861	8.6568	17.7857	16.0000	0.1067	0.2438	0.2074	1.0806	5.2674	18.2211	126.0000	0.3751	0.1068	0.0692
T-CARE	0.7719	8.5588	17.7789	18.0000	0.2495	0.4480	0.0012	1.1578	4.9948	18.0399	135.0000	0.0868	0.0234	0.0678
IG	0.7719	9.9378	21.3648	18.0000	0.2495	0.4480	0.0206	1.0720	5.3474	17.5268	125.0000	0.4271	0.0223	0.0001
AEX														
BNL-CAViaR	1.1647	7.6630	17.1928	28.0000	0.4277	0.5248	0.2678	1.2313	4.7551	15.6930	148.0000	0.0117	0.0339	0.0733
SAV	1.4559	7.9988	16.7691	35.0000	0.0352	0.0649	0.0000	1.2562	4.9312	15.0831	151.0000	0.0054	0.0183	0.0033
AS	1.1647	7.7921	16.1030	28.0000	0.4277	0.5248	0.0530	1.2146	4.9075	15.8170	146.0000	0.0191	0.0489	0.0119
T-CARE	1.4143	7.9595	17.4213	34.0000	0.0544	0.0964	0.0000	1.3062	4.7539	15.5217	157.0000	0.0010	0.0042	0.0020
IG	1.4143	7.5696	16.6859	34.0000	0.0544	0.0964	0.0000	1.1897	4.8737	15.8861	143.0000	0.0376	0.1000	0.1301

Table 3.3: Summary statistics for CARE models. The columns, denoted by LR_{uc}, LR_{cc} and DQ, report the p-values of unconditional and conditional coverage tests of Kupiec (1995) and Christoffersen (1998) and those of the CAViaR Dynamic Quantile (DQ) test of Engle and Manganelli (2004).

Name	$\tau = 0.01$					$\tau = 0.05$				
	BNL-CARE	SAV	AS	T-CARE	IG	BNL-CARE	SAV	AS	T-CARE	IG
Nasdaq	0.0130	0.0150	0.0180	0.0170	0.0140	0.1139	0.1339	0.1329	0.1279	0.0010
STI	0.0290	0.0280	0.0320	0.0340	0.0240	0.1548	0.1449	0.1518	0.1489	0.0909
Hang Seng	0.0200	0.0210	0.0010	0.0250	0.0240	0.1439	0.1349	0.1558	0.1548	0.1419
Corea SE	0.0190	0.0230	0.0200	0.0200	0.0200	0.1369	0.1379	0.1359	0.1429	0.1339
AEX	0.0310	0.0380	0.0300	0.0360	0.0370	0.1518	0.1548	0.1538	0.1598	0.0010

Table 3.4: Bootstrap test p-values for zero mean of the standardized residuals. Test based on 1000 post sample estimates of the conditional 1% and 5% ES.

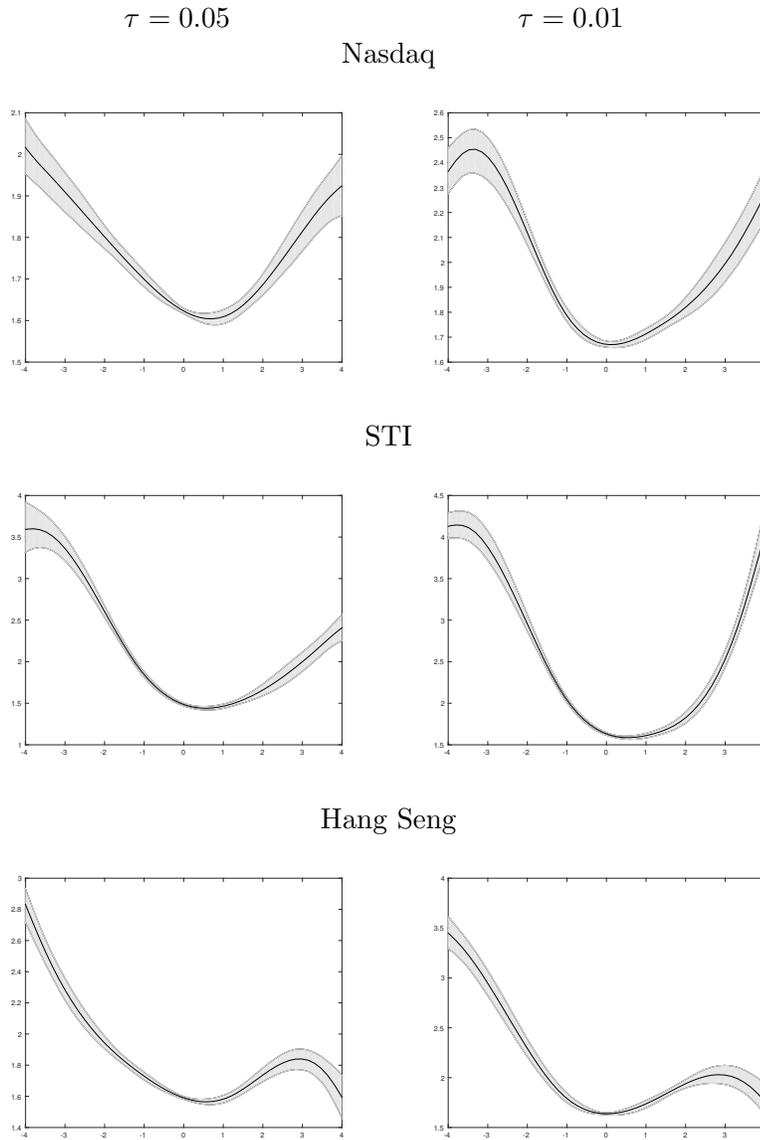


Figure 3.1: Posterior estimate of the Non-Linear CAViaR NIC, for three representative series, from the empirical application of section 3.7. Left panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.05)$. Right panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.01)$.

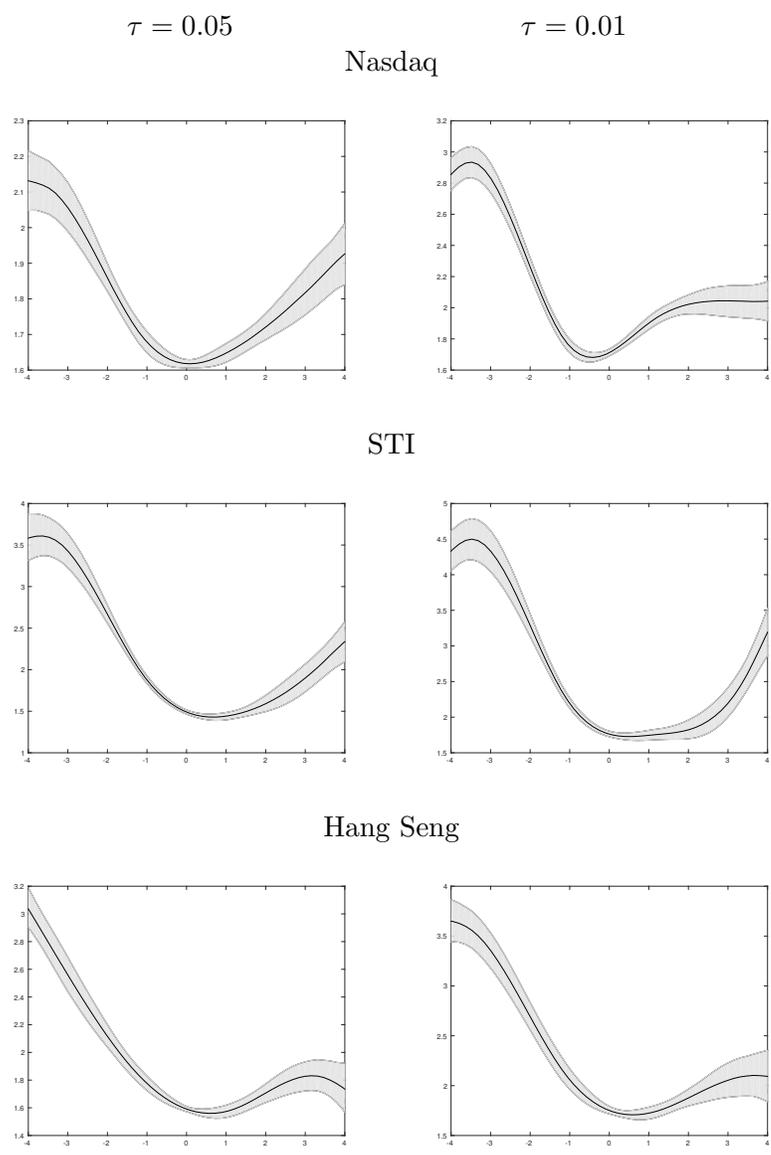


Figure 3.2: Posterior estimate of the Non-Linear CARE NIC, for three representative series, from the empirical application of section 3.7. Left panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.05)$. Right panels exhibit the NIC (black line) along with the 95% HPD regions (grey areas) at the quantile confidence levels $\tau = (0.01)$.

3.8 Conclusion

In this chapter we present a new extension of the CAViaR and CARE models embedded in a Bayesian quantile regression framework. We combine the P-spline approximation of the models with the Skew Exponential Power (SEP) likelihood tool. The introduction of these two innovations allows to capture the well known stylized facts about financial time series. Moreover the SEP enable to estimate both CAViaR and CARE models using a unique statistical framework. The Bayesian estimation methodology is carried out using a new adaptive MCMC technique. The results from five stock market indexes show that the model and the estimation methodology effectively capture the nonlinear relation between the unobserved τ -level quantile and its determinants.

4

Bibliography

- Acerbi, C. and Szekely, B. (2014). Backtesting expected shortfall. *MSCI Inc.*
- Alhamzawi, R. (2016). Bayesian elastic net tobit quantile regression. *Communications in Statistics - Simulation and Computation*, 45(7):2409–2427.
- Alhamzawi, R. and Yu, K. (2012). Variable selection in quantile regression via gibbs sampling. *Journal of Applied Statistics*, 39(4):799–813.
- Alhamzawi, R. and Yu, K. (2015). Bayesian tobit quantile regression using g-prior distribution with ridge parameter. *Journal of Statistical Computation and Simulation*, 85(14):2903–2918.
- Alhamzawi, R., Yu, K., and Benoit, D. F. (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling*, 12(3):279–297.
- Andrieu, C. and Moulines, . (2006). On the ergodicity properties of some adaptive mcmc algorithms. *The Annals of Applied Probability*, 16(3):1462–1505.
- Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Atchad, Y., Fort, G., Moulines, E., and Priouret, P. (2011). Adaptive markov chain monte carlo: theory and methods. In Barber, D., Cemgil, A. T., and Chiappa, S., editors, *Bayesian Time Series Models*, pages 32–51. Cambridge University Press. Cambridge Books Online.

- Atchad, Y. F. and Rosenthal, J. S. (2005). On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828.
- Ayebo, A. and Kozubowski, T. J. (2004). An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science*, 1:187–210.
- Barreto, R. A. and Hughes, A. W. (2004). Under performers and over achievers: A quantile regression analysis of growth. *Economic Record*, 80(248):17–35.
- Barro, R. and i Martin, X. S. (1995). *Economic Growth*. McGraw-Hill, New York.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443.
- Bernardi, M., Bottone, M., and Petrella, L. (2016). Bayesian robust quantile regression. *Working Paper*.
- Bernardi, M., Gayraud, G., and Petrella, L. (2015). Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis*, 10(3):pp. 553–603.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327.
- Boor, C. D. (2001). *A Pratical Guide to Splines*. Springer, New York, rev. ed. edition.
- Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley series in behavioral science: quantitative methods. Addison-Wesley Pub. Co.
- Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on bayesian p-splines. *Computational Statistics and Data Analysis*, 50(4):967 – 991.
- Brezger, A. and Steiner, W. J. (2008). Monotonic regression based on bayesian p-splines: An application to estimating price response functions from store-level scanner data. *Journal of Business and Economic Statistics*, 26(1):pp. 90–104.

- Chen, C., Lin, S., and Yu, P. (2012). Smooth transition quantile capital asset pricing models with heteroscedasticity. *Computational Economics*, 40:19–48.
- Chen, C. and Yu, K. (2009). Automatic bayesian quantile regression curve fitting. *Statistics and Computing*, 19(3):271–281.
- Chen, C. W. S., So, M. K. P., and Lin, E. M. H. (2009). Volatility forecasting with double markov switching garch models. *Journal of Forecasting*, 28(8):681–697.
- Choy, S. and Walker, S. G. (2003). The extended exponential power distribution and bayesian robustness. *Statistics & Probability Letters*, 65(3):227 – 232.
- Choy, S. T. B. and Smith, A. F. M. (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):pp. 463–474.
- Christoffersen, P. (1998). Evaluating interval forecasts. *Journal of Empirical Finance*, 39:841–862.
- Committee, B. (1996). Amendment to the capital accord to incorporate market risks.
- Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):333–350.
- Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, a:93–125.
- Efron, B. and Tibshirani, J. (1993). *An introduction to the Bootstrap*. Chapman and Hall, New York.
- Eilers, P. H. C., Rijnmond, D. M., and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.
- Engle, R. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22:367–381.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.
- Faming Liang, Chuanhai Liu, R. C. (2010). *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley.
- Fernandez, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):pp. 359–371.
- Gerlach, R., Chen, C., and Chan, N. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business & Economic Statistics*, 29:481–492.
- Gerlach, R. and Chen, C. W. S. (2014). Bayesian expected shortfall forecasting incorporating the intraday range. *Journal of Financial Econometrics*.
- Gerlach, R. and Chen, C. W. S. (2016). Semi-parametric expected shortfall forecasting in financial markets. *Journal of Statistical Computation and Simulation*, 0(0):1–23.
- Gerlach, R., Chen, C. W. S., and Lin, L.-Y. (2012). Bayesian semi-parametric expected shortfall forecasting in financial markets. *University of Sydney. Working Paper*.
- Gerlach, R., Walpole, D., and Wang, C. (2016). Semi-parametric bayesian tail risk forecasting incorporating realized measures of volatility. *Quantitative Finance*, 0(0):1–17.
- Gerlach, R. and Wang, C. (2015). Bayesian semi-parametric realized-care models for tail risk forecasting incorporating range and realized measures. *BA Working Paper No: BAWP-2015-07*.
- Hallin, M., Paindaveine, D., and iman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From l_1 optimization to halfspace depth. *Ann. Statist.*, 38(2):635–669.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.
- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):pp. 297–310.
- Hu, Y., Zhao, K., and Lian, H. (2015). Bayesian quantile regression for partially linear additive models. *Statistics and Computing*, 25(3):651–668.

- Ji, Y., Lin, N., and Zhang, B. (2012). Model selection in binary and tobit quantile regression using the gibbs sampler. *Computational Statistics & Data Analysis*, 56(4):827 – 839.
- Jorion, P. (2007). *Value-at-Risk: The new benchmark for managing financial risk*. McGraw-Hill, Chicago, 3rd edition.
- Jr., D. H. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1):81 – 102.
- Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):1–18.
- Karthik, S., R. V., R., and Pulak, G. (2016). On bayesian quantile regression using a pseudo-joint asymmetric laplace likelihood. *Sankhya A: The Indian Journal of Statistics*, 78(1):87–104.
- Kobayashi, G. (2015). Skew exponential power stochastic volatility model for analysis of skewness, non-normal tails, quantiles and expectiles. *Computational Statistics*, 31(1):49–88.
- Kobayashi, G. (2016). Skew exponential power stochastic volatility model for analysis of skewness, non-normal tails, quantiles and expectiles. *Computational Statistics*, 31(1):49–88.
- Kobayashi, G. (2017). Bayesian endogenous tobit quantile regression. *Bayesian Anal.*, 12(1):161–191.
- Koenker, B. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.
- Koenker, B. and Basset, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.
- Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310.
- Koenker, R., Xiao, Z., Fan, J., Fan, Y., Knight, M., Hallin, M., Werker, B. J. M., Hafner, C. M., Linton, O. B., and Robinson, P. M. (2006). Quantile autoregression [with comments, rejoinder]. *Journal of the American Statistical Association*, 101(475):pp. 980–1006.

- Kottas, A. and Gelfand, A. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96:1458–1468.
- Kottas, A. and Krnjajic, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36:297–319.
- Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). Asymmetric multivariate laplace distribution. In *The Laplace Distribution and Generalizations*, pages 239–272. Springer.
- Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81:1565–1578.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3:73–84.
- Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):pp. 183–212.
- Laurini, M. P. (2007). A note on the use of quantile regression in beta convergence analysis. *Economics Bulletin*, 3(52):1–8.
- Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Anal.*, 5(3):533–556.
- Lum, K. and Gelfand, A. (2012). Spatial quantile multiple regression using the asymmetric laplace process. *Bayesian Analysis*, 7:235–258.
- Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data: a flexible approach based on boosting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):403–427.
- McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, Techniques, Tools*. Princeton Series in Finance, Princeton.
- Meier, L., Van De Geer, S., and Bhlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71.

- Meligkotsidou, L., Vrontos, I. D., and Vrontos, S. D. (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance*, 16(2):264 – 279.
- Mello, M. and Perrelli, R. (2003). Growth equations: a quantile regression exploration. *The Quarterly Review of Economics and Finance*, 43(4):643 – 667. Capital Accumulation and Allocation in Economic Growth.
- Naranjo, L., Pérez, C. J., and Martín, J. (2015). Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing*, 25(3):497–514.
- Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science*, 1:505–527.
- O’Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, 9(2):363–379.
- Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686.
- Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106(493):6–20. PMID: 23459794.
- Roberts, G. O. and Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive markov chain monte carlo algorithms. *Journal of Applied Probability*, 44(2):pp. 458–475.
- Roccioletti, S. (2016). *Backtesting Value at Risk and Expected Shortfall*. BestMasters. Gabler Verlag, Springer.
- Roger W. Koenker, V. D. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):383–393.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

- Santos, B. and Bolfarine, H. (2015). Bayesian analysis for zero-or-one inflated proportion data using quantile regression. *Journal of Statistical Computation and Simulation*, 85(17):3579–3593.
- Sriram, K., Ramamoorthi, R., and Ghosh, P. (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis*, 8:479–504.
- Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252.
- Thomas Kneib, Susanne Konrath, L. F. (2011). High dimensional structured additive regression models: Bayesian regularization, smoothing and predictive performance. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 60(1):51–70.
- Thompson, P., Cai, Y., Moyeed, R., Reeve, D., and Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis*, 54(4):1138 – 1150.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.
- Tokdar, S. and Kadane, J. (2012). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Economics*, 6:382–406.
- Wang, Y., Feng, X.-N., and Song, X.-Y. (2016). Bayesian quantile structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2):246–258.
- Wichitaksorn, N., Choy, S. T. B., and Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics*, 42(4):579–596.
- Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *The Annals of Statistics*, 40(2):1102–1131.
- Yu, K., Chen, C. W., Reed, C., and Dunson, D. B. (2013). Bayesian variable selection in quantile regression. *Statistics and Its Interface*, 6:261–274.
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.

- Yu, K. and Stander, J. (2007). Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics*, 137(1):260 – 276.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.
- Yue, Y. R. and Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Stat. Data Anal.*, 55(1):84–96.
- Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148(1):86 – 99.