Bayesian CV@R / Super-quantile Regression

Mike G. Tsionas* and Marwan Izzeldin†

October 7, 2017

Abstract

In this paper we provide a Bayesian interpretation of the conditional value at risk, CV@R, or super-quantile regression recently developed by Rockafellar, Royset and Miranda (2014). Computations are based on particle filtering using a special posterior distribution consistent with the super-quantile concept. An empirical application to data used by RRM as well to another data set on energy prices confirms their results and shows the applicability of the new techniques.

Keywords: CV@R; super-quantile regression; risk measures; Bayesian analysis; particle filtering.

Acknowledgments: The authors are grateful to an anonymous referee for valuable comments on an earlier version.

*Lancaster University Management School, LA1 4YX, U.K & Athens University of Economics and Business, Greece, m.tsionas@lancaster.ac.uk
†Lancaster University Management School, LA1 4YX, U.K, m.izzeldin@lancaster.ac.uk

1 Introduction

The search for a coherent measure of risk has been convincingly set by R.T. Rockafellar, see Rockaffellar (2007) for example. In his words:

> "What is risk? Everyone agrees that risk is associated with having to make a decision without fully knowing its consequences, due to future uncertainty, but also knowing that some of those consequences might be bad, or at least undesirable relative to others. Still, how might the *quantity* of risk be evaluated to construct a model in which optimization can be carried out?" (Rockafellar, 2007, p. 44).

A single measure of risk is clearly desirable not only in regression but in wider contexts. "The core of the difficulty in optimization under uncertainty is the fact that a random variable is not, itself, a single quantity. The key to coping with this will be to condense the random variable into a single quantity by *quantifying the risk of loss*, rather than the degree of uncertainty, in it." (Rockafellar, 2007, p. 44).

The conditional value-at-risk has been found to be a coherent measure of risk in the sense of Rockafellar and Uryasev (2000, 2002). Rockafellar, Royset and Miranda (2014, RRM) proposed an ingenious super-quantile or CV@R regression concept. Although the concept is very appealing, computations turn out to be non-trivial for moderate sample sizes in regression. Computations rely mainly on large linear programming problems. In this paper we provide a Bayesian interpretation of super-quantile regression that does not require the use of large-scale linear programs. In fact, our posterior can be explored using a basic Metropolis-Hastings algorithm in a, more or less, automatic way. For better performance we use a particle filtering approach (Chopin, 2002). In an empirical application provided by RRM we obtain similar results but, in addition, we are able to compute marginal posterior densities and posterior standard deviations. In RRM standard errors are not provided as the asymptotic theory of super-quantile regression is not well developed.[1]

In section 2 we present the V@R and CV@R concepts and the regression model. In section 3 we describe our Particle Filtering methodology for static models in the context of Bayesian super-quantile regression. The empirical results are presented in section 4.

2 CV@R and regression models

For a random variable $\mathbb{X}$, the value-at-risk (V@R) is the $\alpha-$quantile:

$$\text{V@R}_\alpha(\mathbb{X}) := q_\alpha(\mathbb{X}) = \min\{z|F(z) \geq \alpha\}, 0 < \alpha < 1. \tag{1}$$

The conditional value-at-risk (CV@R) or super-quantile of $\mathbb{X}$ is the expectation of $\mathbb{X}$ in the conditional distribution of its upper α-tail:

$$\text{CV@R}_\alpha(\mathbb{X}) := \overline{q}_\alpha(\mathbb{X}) = \frac{1}{1-\alpha}\int_\alpha^1 q_\gamma(\mathbb{X})d\gamma. \tag{2}$$

CV@R is a coherent measure of risk as has been shown by Rockafellar and Uryasev (2000, 2002). Clearly, it is an "average" of quantiles of the random variable. Another often useful formula is the following:

$$\overline{q}_\alpha(\mathbb{X}) := \min_{C\in\Re}\{C + \tfrac{1}{1-\alpha}\mathcal{E}[\max\{\mathbb{X}-C, 0\}]\}. \tag{3}$$

CV@R is a continuous function of α with $\lim_{\alpha\to 1}\overline{q}_\alpha(\mathbb{X}) = \sup\mathbb{X}$ (the essential supremum) and $\lim_{\alpha\to 0}\overline{q}_\alpha(\mathbb{X}) = \mathcal{E}(\mathbb{X})$.

The concept has appeared in the literature before, see for example, Cai and Wang (2008), Chernozhukov and Umantsev (2001), Fermanian and Scaillet (2005), Leorato, Peracchi, and Tanase (2012), Peracchi and Tanase (2008), Rockafellar and

[1] It must be acknowledged that RRM do suggest the bootstrapping as a possible way to obtain standard errors. It is likely that the best bootstrap approach would be a re-centered parametric bootstrap. The problem is that huge linear programs need to be solved for each bootstrap replication and this can be difficult, especially when the sample size is large.

Uryasev (2002), Rockafellar, Uryasev and Zabarankin M (2008), Scaillet (2004, 2005), Wozabal and Wozabal N (2009). A conditional value-at-risk concept known as CAViaR has been proposed by Engle and Manganelli (2004) but its meaning is different compared to Rockafellar and Uryasev (2002), Rockafellar, Uryasev and Zabarankin M (2008), and Rockafellar, Royset and Miranda (2014, RRM).

RRM proposed an (admittedly ingenious) super-quantile or CV@R regression concept. Suppose the regression model is:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t,\ t = 1, \cdots, n, \tag{4}$$

where $\mathbf{x}_t$ is a $k \times 1$ vector of covariates and $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters. The extension to models of the form: $y_t =< \boldsymbol{\beta}, \mathbf{h}(\mathbf{x}_t) >$for a vector of basis function $\mathbf{h}(\mathbf{x}) = [h_1(x_1), \cdots, h_k(x_k)]'$ is trivial.

The error term is assumed to have zero mean, that is $\mathcal{E}(u_t) = 0, t = 1, \cdots, T$, so we agree to always include a column of ones in $\mathbf{X} = [\mathbf{x}_t, t = 1, \cdots, n]$, i.e. *a constant term is included in the regression.* This is standard practice in most applications. The objective function of RRM (p. 149), after omitting a long discussion related to their error function which can be found in their paper, is:

$$\min_{\boldsymbol{\beta} \in \Re^k} \mathcal{D}_\alpha^T(\boldsymbol{\beta}) := \frac{1}{1-\alpha} \int_\alpha^1 \overline{q}_\gamma(u_t(\boldsymbol{\beta}))d\gamma - \mathcal{E}(u_t(\boldsymbol{\beta})), \tag{5}$$

where $u_t(\boldsymbol{\beta}) = y_t - \mathbf{x}_t'\boldsymbol{\beta}, t = 1, \cdots, n$, and $\overline{q}_\gamma(u) := \min_{C \in \Re} \left\{ C + \frac{1}{1-\gamma}\mathcal{E}\left[\max\{u - C, 0\}\right] \right\}$. Since the error terms have zero mean we can set $\mathcal{E}[u_t(\boldsymbol{\beta})] = 0$. The reason is that we always include a constant term in the regression equation.

It is well known that quantile regression in the sense of Koenker and Bassett (1978) solves the following problem:

$$\min_{\boldsymbol{\beta} \in \Re^k} \sum_{t=1}^n \varrho_\alpha(y_t - \mathbf{x}_t'\boldsymbol{\beta}), \tag{6}$$

where

$$\varrho_\alpha(z) = \begin{cases} \alpha z,\ z \geq 0 \\ -(1-\alpha)z,\ z < 0. \end{cases} \tag{7}$$

Bayesian treatments of the problem were offered by Tsionas (2003) and Kozumi and Kobayashi (2009). The Bayesian approach utilizes the asymmetric Laplace distribution. This approach cannot be used here as finding an error density for which the maximum likelihood estimator coincides with the super-quantile regression estimator, seems very hard to find.[2]

To proceed, we propose the following density:

$$f(u|\alpha, \sigma) \propto \sigma^{-1} \exp\left\{-\sigma^{-1}\overline{\varrho}_\alpha(u)\right\}, -\infty < u < \infty, \sigma > 0, \tag{8}$$

where σ is a scale parameter, and

$$\overline{\varrho}_\alpha(u) = \frac{1}{1-\alpha} \int_\alpha^1 \overline{q}_\gamma(u)d\gamma, \tag{9}$$

which is the equivalent of $q_\gamma(u)$ in super-quantile regression. The standard form of the density in (8) is: $f(u|\alpha, 1) \propto \exp\left\{-\overline{\varrho}_\alpha(u)\right\}$.[3]

[2]This is mentioned explicitly in Chun, Shapiro and Uryasev (2012, p.742) who suggest an M-estimator. Their M-estimator includes a constraint and linear programming is used to perform the computations. Whether the procedure is consistent with some Bayesian approach to inference remains to be seen as the hard part is to incorporate the constraint. Presumably, a variation of empirical likelihood can be used whose Bayesian interpretation is the exponentially tilted empirical likelihood.

[3]Notably the same construction is used in OLS and L_p-regression for any $p > 0$ and even $p = \infty$ known as Chebyshev regression. The construction is based on the statistical fact that if $l(u)$ is a "loss" function for the error u, then a density for u, say $f(u|\sigma)$ can be always constructed as $f(u|\sigma) \propto \sigma^{-1} \exp\left(-\sigma^{-1}u\right)$.

It is clear that $f(u|\alpha,\sigma)=\mathcal{K}(\alpha,\sigma)^{-1}\sigma^{-1}\exp\left\{-\sigma^{-1}\overline{\varrho}_\alpha(u)\right\}$, where $\mathcal{K}(\alpha,\sigma)=\int_{-\infty}^{\infty}f(u|\alpha,\sigma)du$. Given

$$\mathcal{K}(\alpha,1):=\mathcal{K}_\alpha=\int_{-\infty}^{\infty}f(u|\alpha,1)du, \tag{10}$$

viz. the integrating constant of the standard form of the density, we have:

$$f(u|\alpha,\sigma)=\mathcal{K}_\alpha^{-1}\sigma^{-1}\exp\left\{-\sigma^{-1}\overline{\varrho}_\alpha(u)\right\}. \tag{11}$$

Therefore, the integrating constant does not depend on σ (or $\boldsymbol{\beta}$) and can be dropped in subsequent discussions. The likelihood function, under the assumption of independence of $\{u_t,t=1,\cdots,T\}$, can be written as:

$$\mathscr{L}(\boldsymbol{\beta},\sigma;\mathbf{y},\mathbf{X},\alpha)=\sigma^{-T}\mathcal{K}_\alpha^{-n}\exp\left\{-\sigma^{-1}\sum_{t=1}^{n}\overline{\varrho}_\alpha(u_t(\boldsymbol{\beta}))\right\}, \tag{12}$$

where $\mathbf{y},\mathbf{X}$ denotes the data. Maximizing the likelihood with respect to $\boldsymbol{\beta}$ is clearly equivalent to minimizing $\sum_{t=1}^{n}\overline{\varrho}_\alpha(u_t(\boldsymbol{\beta}))$ for any fixed σ. In turn, this is the objective function of so-called problem D^ν in RRM (page 149). Given the prior:

$$p(\boldsymbol{\beta},\sigma)\propto\sigma^{-1}, \tag{13}$$

after combining with the likelihood function in (12) and using Bayes' theorem we obtain the kernel posterior distribution:

$$p(\boldsymbol{\beta},\sigma|\mathbf{y},\mathbf{X},\alpha)\propto\sigma^{-(n+1)}\exp\left\{-\sigma^{-1}\sum_{t=1}^{n}\overline{\varrho}_\alpha(u_t(\boldsymbol{\beta}))\right\}. \tag{14}$$

We use the prior in (13) for convenience but also because this serves as a benchmark prior. Our numerical techniques can, effortlessly, accommodate any prior for $\boldsymbol{\beta}$ and any inverted *gamma* prior for σ.[4] Analytical integration with respect to σ is possible and we have:

$$p(\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\alpha)\propto\int_0^{\infty}p(\boldsymbol{\beta},\sigma|\mathbf{y},\mathbf{X},\alpha)d\sigma\propto\left\{\sum_{t=1}^{n}\overline{\varrho}_\alpha(u_t(\boldsymbol{\beta}))\right\}^{-n/2}, \tag{15}$$

by using the integral of the inverted *gamma* distribution. It is, again, clear that maximizing the likelihood with respect to $\boldsymbol{\beta}$ is equivalent to minimizing $\sum_{t=1}^{T}\overline{\varrho}_\alpha(u_t(\boldsymbol{\beta}))$. Before implementing our numerical techniques, based on Sequential Monte Carlo (SMC) we must resolve two problems:

(i) Computing $\overline{q}_\gamma(u)$.

(ii) Computing $\int_\alpha^1\overline{q}_\gamma(u)d\gamma$.

We deal with the computations (i)-(ii) as follows:

(i) As $\overline{q}_\gamma(u)=\int_\alpha^1 q_\gamma(u)d\gamma$ we first have to compute quantiles $q_\gamma(u)$ for $\gamma\in(0,1)$. Since the density is continuous we have $F_u(q_\gamma(u))=\gamma$, where F_u is the distribution function corresponding to (8). Without knowing the integrating constant in (10) we approximate the distribution function as follows. Let: $\tilde{F}_u(\overline{u}_i)=\sum_{s=1}^{i}f(\overline{u}_s),i=1,\cdots,N$, where $\mathcal{G}=\{\overline{u}_i,i=1,\cdots,N\}$ is a fixed grid symmetrically placed around zero, N is the number of grid points, and $\overline{u}_{[N/2]}=0$. Here, $[w]$ denotes the integer part of number w. Next we normalize $F_u(\overline{u}_i)=\mathcal{K}^{-1}\tilde{F}_u(\overline{u}_i)$, where $\mathcal{K}=\sum_{i=1}^{N}f(\overline{u}_i)=F_u(\overline{u}_N)$. We check whether $F_u(\overline{u}_i)<\varepsilon$ where $\varepsilon>0$ is a small constant, otherwise we extend the grid in both directions. The right direction is checked using $f_u(\overline{u}_N)/\max\limits_{i=1,\cdots,[N/2]}f(\overline{u}_i)<\varepsilon$.[5]

[4]Such priors have the form $p(\sigma)\propto\sigma^{-(\overline{K}+1)}\exp\left(-\overline{A}\sigma^{-1}\right)$, $\overline{A},\overline{K}\geq 0$, and they facilitate analytical integration with respect to σ. They are considered sufficiently flexible in the statistical literature. If a non-conjugate prior is used, our methods still go through at the expense of having σ as an additional parameter,

[5]We set $\varepsilon=10^{-12}$ in all computations.

(ii) The integral $\overline{q}_\gamma(u) = \int_\alpha^1 q_\gamma(u)d\gamma$ is computed using quadrature. As α is close to one, 12-point Gaussian quadrature has been found quite accurate.[6]

These computations provide us with $\overline{\varrho}_\alpha(u)$ for $u \in \mathcal{G}$. For $u \notin \mathcal{G}$ we use interpolation. If the grid is fine and extended[7] the error of approximation is trivial. Interpolation allows to compute (9) and (14) quickly and efficiently provided we sort $\{u_t(\boldsymbol{\beta}), t = 1, \cdots, T\}$ in advance. This procedure has a light set up cost in terms of CPU time but afterwards the posterior can be evaluated quickly for a large number of different values of $\boldsymbol{\beta}$ and σ as required by SMC procedures for Bayesian statistical inference. It is known that using linear programming for large n, "the linear program could become large-scaled when n is large and decomposition algorithms may be needed" (RRM, p.150).

3 Numerical techniques

The purpose of this section is to propose an approach that can be used to obtain a sample $\{\boldsymbol{\beta}^{(s)}, s = 1, \ldots, S\}$ (for given α) that converges in distribution to (14). Once we have such a sample, marginal posterior densities and moments of parameters or functions of interest can be computed easily from the available sample. For example, posterior expectations can be computed as:

$$\bar{\boldsymbol{\beta}} \equiv \mathcal{E}(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \int \boldsymbol{\beta} p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) d\boldsymbol{\beta} \simeq S^{-1} \sum_{s=1}^{S} \boldsymbol{\beta}^{(s)}. \tag{16}$$

The posterior covariance matrix can be approximated by:

$$\mathcal{E}[(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})'|\mathbf{y}, \mathbf{X}] \simeq S^{-1} \sum_{s=1}^{S} (\boldsymbol{\beta}^{(s)} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(s)} - \bar{\boldsymbol{\beta}})'. \tag{17}$$

The approximation can be made arbitrarily good by increase the number of simulations, S. A particularly attractive simulation-based approach to realize such a sample is to use the Particle Filtering (PF) approach due to Chopin (2002) who proposed a sequential PF for static models.[8] This is also known as Sequential Monte Carlo (SMC). Given a target posterior $p(\theta|Y) := p(\theta|Y_{1:T})$ a particle system is a sequence $\{\theta_j, w_j\}$ such that $E(h(\theta)|Y) := \int h(\theta)p(\theta|Y)d\theta \simeq \lim_{J\to\infty} \frac{\sum_{j=1}^J w_j h(\theta_j)}{\sum_{j=1}^J w_j}$, almost surely, for any measurable function h, provided the expectation exists. We consider the sequence of posterior distributions $p_t := p(\theta|Y_t)$. The PF algorithm is as follows. For "cycles" $r = 1, ..., R$ and $m = 1, ..., M$ do:

Step 1. Reweight: update the weights $w_j \leftarrow w_j \frac{p_{t+1}(\theta_j)}{p_t(\theta_j)}, j = 1, ..., J$.

Step 2: Resampling: resample $\{\theta_j, w_j\}_{j=1}^H \to \{\theta_j^r, 1\}_{j=1}^J$.

Step 3. Move: draw $\theta_j^m \sim \mathcal{K}_{t+1}(\theta_j^r), j = 1, ..., J$, where $\mathcal{K}_{t+1}$ is any transition kernel whose stationary distribution is p_{t+1}.

Step 4. Loop: $t \leftarrow t + 1, \{\theta_j, w_j\}_{j=1}^J \leftarrow \{\theta_j^m, 1\}_{j=1}^J$ and return to Step 1.

Chopin (2002) recommends the independence Metropolis algorithm to select the kernel, which requires a source distribution. A possible choice, if we sampled from p_n $(n < T)$, with respect to p_{n+s} is $\mathcal{N}(\hat{E}_{n+s}, \hat{V}_{n+s})$ where

$$\hat{E}_{n+s} = \frac{\sum_{j=1}^J w_j \theta_j}{\sum_{j=1}^J w_j}, \; \hat{V}_{n+s} = \frac{\sum_{j=1}^J w_j \left(\theta_j - E_{n+p}\right)\left(\theta_j - E_{n+p}\right)'}{\sum_{j=1}^J w_j}.$$

The strategy can be parallelized easily. If R processors are available, we can partition the particle system into R subsets, say S_p, $p = 1, ..., R$, and implement computations for particles of S_p in processor p. The algorithm can deal with new data at a nearly geometric rate and therefore the frequency of exchanging information between processors (after reweighting) decreases at a rate exponential to n, which is highly efficient.

[6] Compared to 24- and 40-point Gauss-Kronrod rules as implemented in the IMSL Mathematical and Statistical Library. The computations were checked using a high-accuracy Simpson's rule with tolerance 10^{-7}.

[7] We have experimented with N =100, 1,000, 5,000 and 10,000 points. We have found that 1,000 points are adequate but in all computations we use N=10,000.

[8] The parallel nature of SMC implies that we can compute a number of posteriors corresponding to different values of α, in parallel.

Resampling according to $\theta_j^m \sim K_t(\theta_j^r, .)$ reduces particle degeneracy (Gilks and Berzuini, 2001) since identical replicates of a single particle are replaced by new ones without altering the stationary distribution. For this application using $J = 2^{12}$ particles gave a mean squared error in posterior means of 10^{-5} over 100 runs. Moreover, $R = M = 10^4$ worked acceptably well in our application.

Chopin (2004) introduces a variation of SMC in which the observation dates at which each cycle terminates (say $t_1, ..., t_L$) and the parameters involved in specifying the Metropolis updates (say $\lambda_1, ..., \lambda_L$) are specified. Therefore, $0 = t_0 < t_1 < ... < t_L = T$ and we have the following scheme (we rely heavily on Durham and Geweke, 2013).

Step 1. Initialize $l = 0$ and $\theta_{jn}^{(l)} \sim p(\theta)$, $j \in \mathcal{J}, n \in \mathcal{N}$.

Step 2. For $l = 1, ..., L$:

(a) Correction phase:

(i) $w_{jn}(t_{l-1}) = 1$, $j \in \mathcal{J}, n \in \mathcal{N}$.

(ii) For $s = t_{l-1} + 1, ..., t_l$:

$$w_{jn}(s) = w_{jn}(s-1)p(y_s|y_{1:s-1}, \theta_{jn}^{(l-1)}),\ j \in \mathcal{J}, n \in \mathcal{N}.$$

(iii) $w_{jn}^{(l-1)} := w_{jn}(t_l)$, $j \in \mathcal{J}, n \in \mathcal{N}$.

(b) Selection phase, applied independently to each group $j \in \mathcal{J}$: Using multinomial or residual sampling based on $\left\{w_{jn}^{(l)}, n \in \mathcal{N}\right\}$, select

$$\{\theta_{jn}^{(l,0)}, n \in N\}$$

from $\{\theta_{jn}^{(l-1)}, n \in N\}$.

(c) Mutation phase, applied independently across $j \in \mathcal{J}, n \in \mathcal{N}$:

$$\theta_{jn}^{(l)} \sim p(\theta|y_{1:t}, \theta_{jn}^{(0)}, \lambda_l), \tag{18}$$

where the drawings are independent and the p.d.f above satisfies the invariance condition:

$$\int_{\Theta} p(\theta|y_{1:t_l}, \theta^*, \lambda_l)p(\theta^*|y_{1:t_l})d\nu(\theta^*) = p(\theta|y_{1:t_l}). \tag{19}$$

Step 3. $\theta_{jn} := \theta_{jn}^{(l)}$, $j \in \mathcal{J}, n \in \mathcal{N}$.

At the end of every cycle, the particles $\theta_{jn}^{(l)}$ have the same distribution $p(\theta|y_{1:t_l})$. The amount of dependence within each group depends upon the success of the Mutation phase which avoids degeneracy.

4 Empirical applications

4.1 Style Classification

For the sake of comparison we use the same data as in RRM. It is a case study taken from the "Style Classification with Quantile Regression" documentation in Portfolio Safeguard (American Optimal Decisions, Inc., 2011, cited in RRM) and deals with the negative return of the Fidelity Magellan Fund as predicted by the explanatory variables: Russell 1000 Growth Index (x_1, RLG), Russell 1000 Value Index (x_2, RLV), Russell Value Index (x_3, RUJ), and Russell 2000 Growth Index (x_4, RUO). The indices classify the style of the fund. There are $n = 1,264$ total observations available. We start by considering a linear model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + u_t \tag{20}$$

and compare the obtained regression coefficients for least-squares, quantile, and super-quantile regression when α is 0.75 and 0.90, as shown in Table 1. We use $L = 100$ and alternatively $L = 200$ and $L = 300$ for sensitivity purposes. An endogenous

selection of L is possible along the lines of Durham and Geweke (2013). Since the algorithm is robust with $L = 100$ we did not implement this and we leave it for future research.

Table 1. Empirical results

		Reported results from RRM				Bayesian	
	OLS	QR ($\alpha = 0.75$)	SQR ($\alpha = 0.75$)	QR ($\alpha = 0.90$)	SQR ($\alpha = 0.90$)	SQR ($\alpha = 0.75$)	SQR ($\alpha = 0.90$)
β_0	0.0010	0.0045	0.0095	0.0089	0.0138	0.0094 (0.0012)	0.0091 (0.014)
β_1	-0.5089	-0.5438	-0.5036	-0.5177	-0.4837	-0.5037 (0.015)	-0.5203 (0.032)
β_2	-0.5180	-0.4518	-0.4723	-0.4602	-0.4912	-0.4722 (0.015)	-0.4701 (0.025)
β_3	0.0484	0.0159	0.0192	0.0156	0.0223	0.0193 (0.0014)	0.0155 (0.014)
β_4	0.0061	0.0173	0.0009	-0.0001	-0.0019	0.0009 (0.0004)	-0.0025 (0.001)

Notes: QR stands for Quantile Regression. SQR stands for Super-quantile Regression. Numbers in parentheses are posterior standard deviations.

The results from Bayesian SQR are close to the results reported by RRM in their study. However, they did not report standard errors, a task that is simple in the Bayesian implementation of the problem -although, as we have mentioned, they did mention the possibility of using the bootstrap.[9] Moreover, we avoid large-scale linear programming problems which for moderate sample sizes may require specialized decomposition algorithms for their implementation. The marginal posterior densities of the slope parameters are reported in Figure 1a through 1d.

[9]The following argument may be relevant: "In this sense, the bootstrap distribution represents an (approximate) nonparametric, noninformative posterior distribution for our parameter. But this bootstrap distribution is obtained painlessly — without having to formally specify a prior and without having to sample from the posterior distribution. Hence we might think of the bootstrap distribution as a "poor man's" Bayes posterior. By perturbing the data, the bootstrap approximates the Bayesian effect of perturbing the parameters, and is typically much simpler to carry out." (Hastie, Tibshirani, and Friedman, p.272).

Figure 1a.

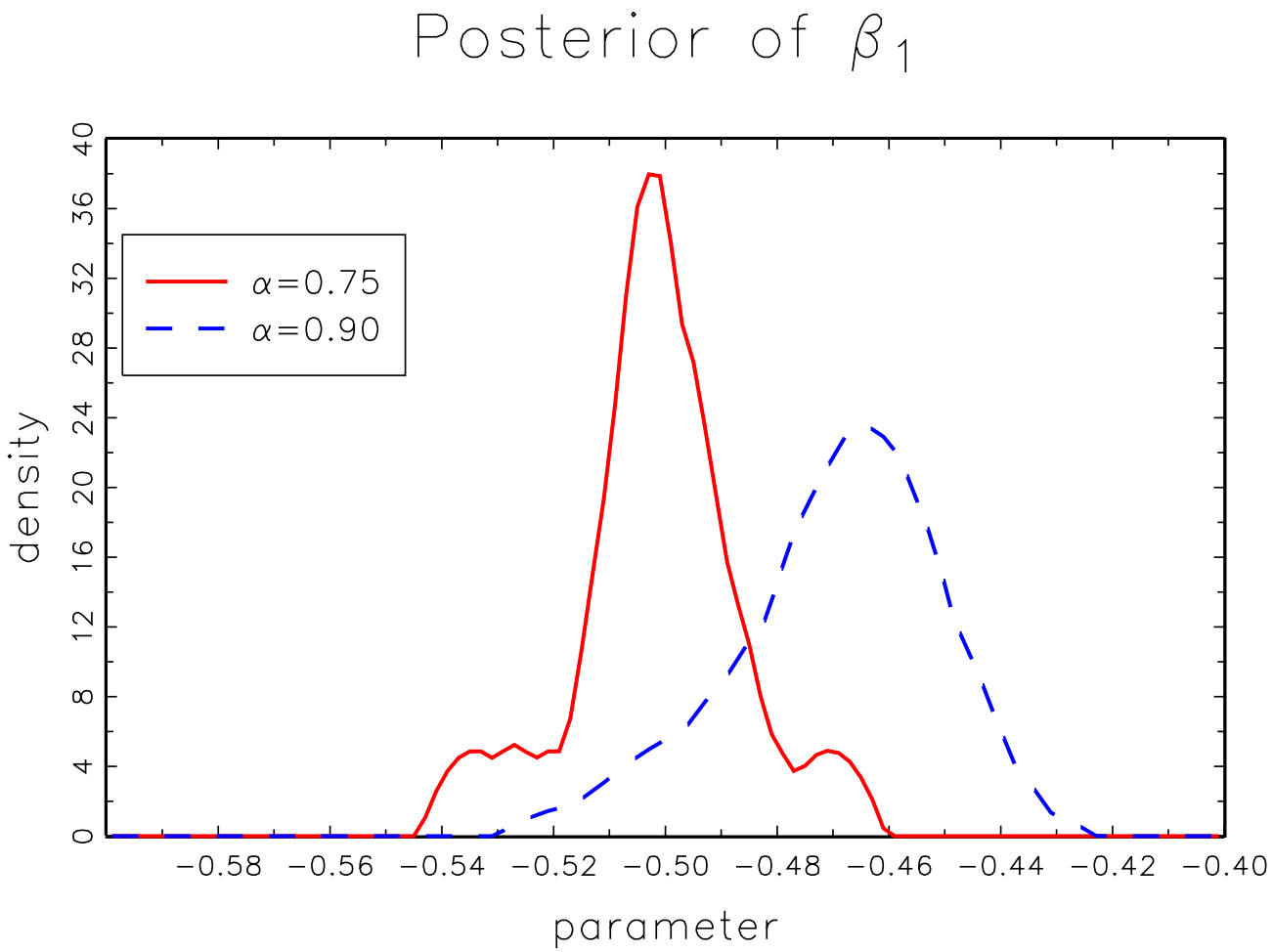

Posterior of β_1
density
parameter
α=0.75
α=0.90
0
4
8
12
16
20
24
28
32
36
40
−0.58
−0.56
−0.54
−0.52
−0.50
−0.48
−0.46
−0.44
−0.42
−0.40

Figure 1b.

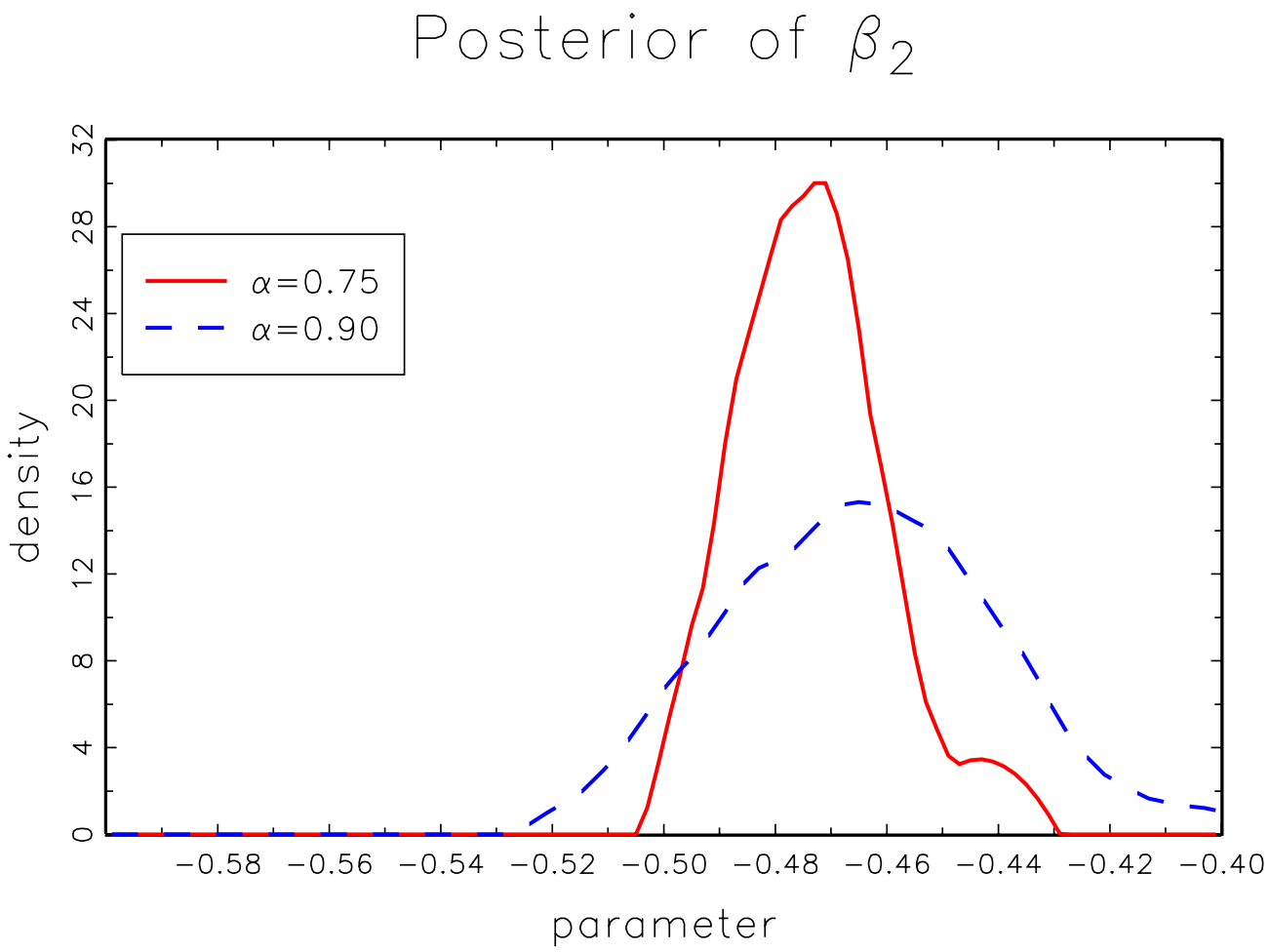

Posterior of β_2
density
parameter
α=0.75
α=0.90
0
4
8
12
16
20
24
28
32
−0.58
−0.56
−0.54
−0.52
−0.50
−0.48
−0.46
−0.44
−0.42
−0.40

Figure 1c.

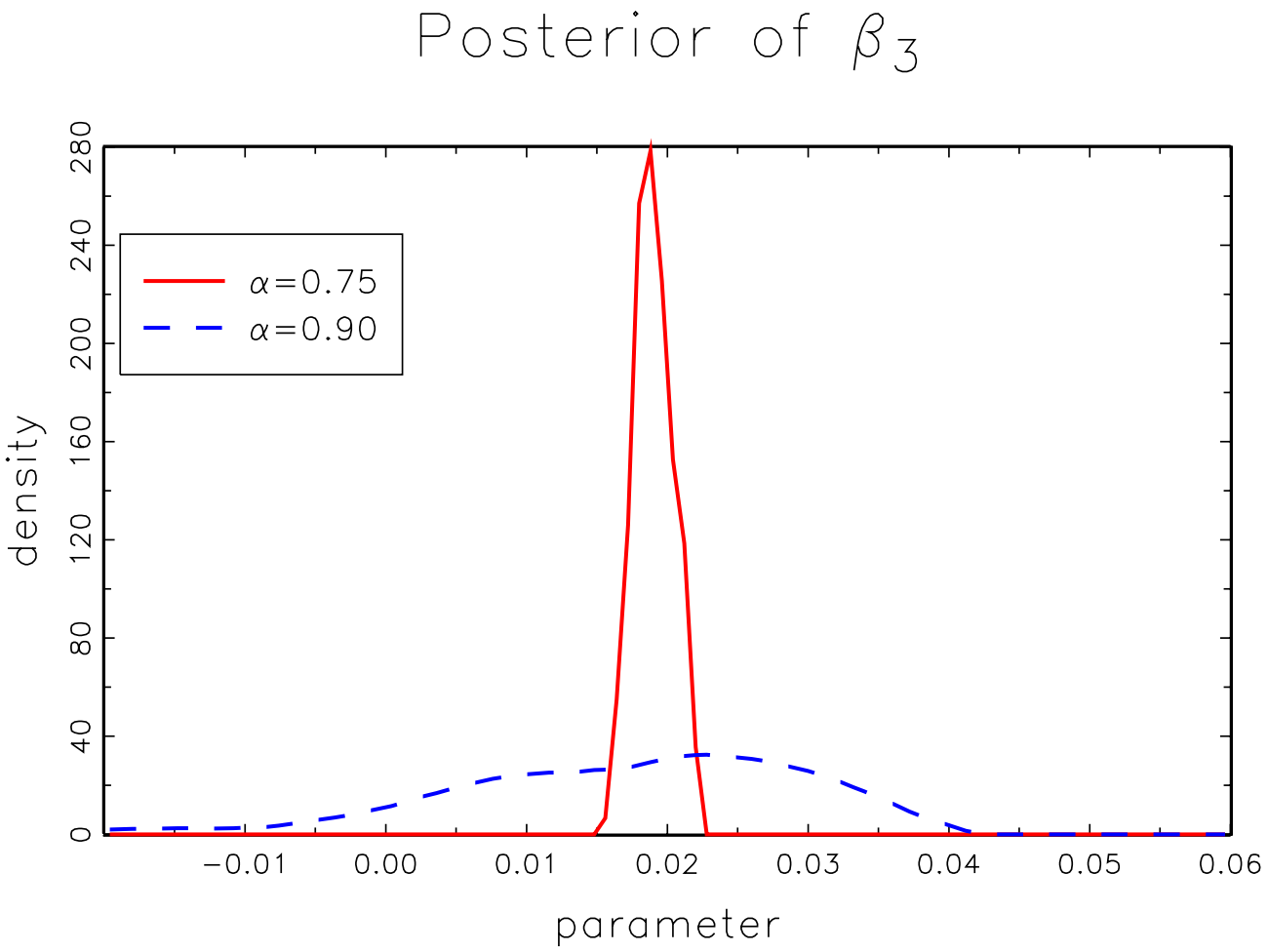
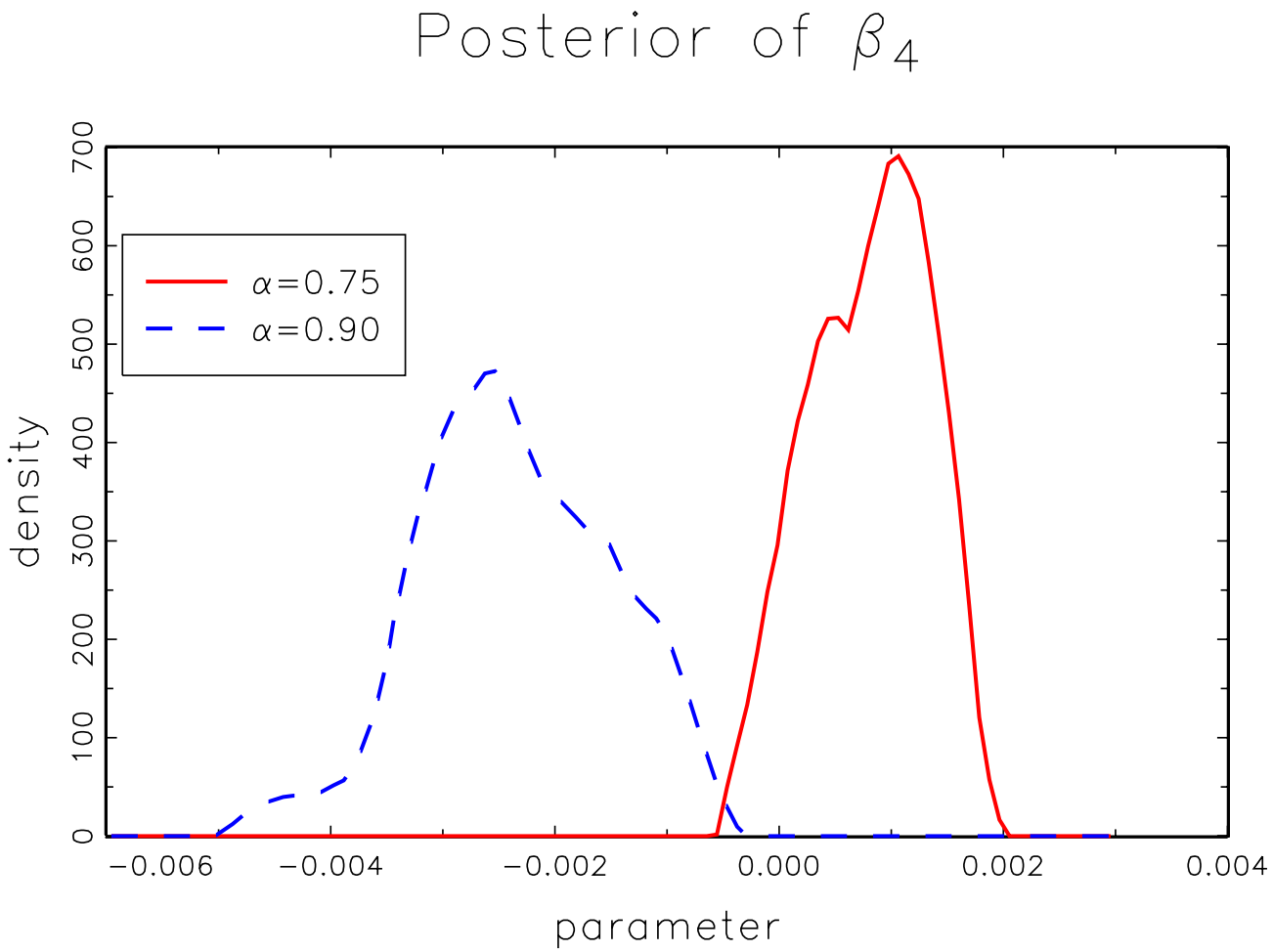

Posterior of β_3
density
parameter
α=0.75
α=0.90
0
40
80
120
160
200
240
280
−0.01
0.00
0.01
0.02
0.03
0.04
0.05
0.06

Figure 1d.



Using the same **X** data and the estimated posterior means of the parameters, we generate **y** using a Student-t distribution with 5 degrees of freedom for the errors. Our running times are reported below.

Table 2. Running times

Sample size, $n \rightarrow$	100	200	300	400	500	1,000	1,500	2,000
Running time (hrs)	0.006	0.01	0.03	0.06	0.12	0.28	0.60	1.35

Notes: Running times are based on 10,000 iterations of the PF algorithm using 10^{12} particles. We use the High Performance Computing end at Lancaster University.The combined facility offers 3000 CPU cores, 11TB of memory, 32TB of high performance filestore and 1PB of medium performance filestore. The service combines what was the previously separately supported services of local high performance computing (HPC) users and the local Particle Physics research group (GridPP). The cluster operating system is Scientific Linux, with job submission handled by Sun Grid Engine (SGE). The service supports a wide variety of third-party software including numerical packages, libraries and C and Fortran compilers.

4.2 Energy prices: Rockets and feathers

It is often the case that oil companies adjust the retail gasoline prices more quickly to cost increases than to cost decreases, creating an asymmetric adjustment path towards the long-run equilibrium, known as the "*rockets and feathers*" hypothesis. This perceived asymmetry in retail price adjustment to changes in crude oil prices is commonly attributed to "gouging" engaged by vertically integrated firms in an effort to increase retail profits, which in turn is made possible by their market power (Borenstein, et al, 1997, Deltas, 2008).

Hence many important papers have tried to investigate the existence of price asymmetry across countries or even regions with conflicting results. Bacon (1991), for example, provided sufficient evidence in favor of the price asymmetry debate in the UK. Similarly, Borenstein et al. (1997) argued that retail prices in the UK over the period 1986 to 1992 responded more quickly to crude oil price increases than to decreases. Galeotti et al. (2003) studied trends in Germany, France, UK, Italy and Spain from January 1985 to June 2000 and concluded that "*rockets and feathers*" appear to dominate the price adjustment mechanism of gasoline markets in many European countries. Deltas (2008) reported that U.S. states with high average retail-wholesale margins experienced a slower adjustment and a more asymmetric response in retail prices. Lewis and Noel (2011) use panel data from U.S. cities with and without Edgeworth price cycles, and demonstrate that prices in markets without cycles respond much more slowly to wholesale cost fluctuations than in cities with cycles. Greenwood-Nimmo and Shin (2013), in contrast, did not support the existence of "*rockets and feathers*" in the UK gasoline market. Similarly, Lamotte et al (2013) argued that the retail price of gasoline in the U.S and France respectively did not respond contemporaneously to crude oil shocks.

Our empirical analysis is based on a panel data set of 9,888 monthly observations spanning the period from January 1994 to February 2011. The sample includes 48 contiguous US states except for Maine and Connecticut where no data is available. Retail and wholesale (rack) motor gasoline prices before taxes and duties are obtained from the EIA of the U.S. Department of Energy. Spot prices of conventional gasoline (measured in dollars per gallon) traded in New York Harbor are taken also from the EIA. The price of bulk unleaded gasoline delivered to New York Harbor is considered a good proxy for the input cost since, on average, 96% of the wholesale price is represented by the cost of gasoline at the hub.[10] Crude oil price measured in dollars per barrel accounts for the Cushing, OK WTI Spot Price FOB extracted also from EIA.

The model is an extension of Greenwood-Nimmo and Shin (2013) to a panel data framework. Specifically, the model is as follows. Suppose we have an asymmetric long-run relationship of the form:

$$y_{it} = \mu_i + \boldsymbol{\chi}_{it}^{+\prime}\boldsymbol{\alpha}^{+} + \boldsymbol{\chi}_{it}^{-\prime}\boldsymbol{\alpha}^{-} + u_{it}, \ \ \forall i = 1, ..., n', \ t = 1, ..., T \tag{21}$$

where μ_i represents fixed effects, $\mathbf{x}_{it}$ is a $k\times 1$ vector of original regressors, and $\boldsymbol{\chi}_{it}^{+} = \sum_{j=1}^{t} \max(\Delta\mathbf{x}_{ij}, \mathbf{0})$, $\boldsymbol{\chi}_{it}^{-} = \sum_{j=1}^{t} \min(\Delta\mathbf{x}_{ij}, \mathbf{0})$ are partial sum processes representing positive and negative changes respectively. Moreover $\Delta\boldsymbol{\chi}_{it} = \mathbf{v}_{it}$ so that $\boldsymbol{\chi}_{it} \equiv \boldsymbol{\chi}_{i0} + \boldsymbol{\chi}_{it}^{+} + \boldsymbol{\chi}_{it}^{-}$ follows a unit root process. The dependent variable, y_{it}, is the the retail price of unleaded petrol, derv, kerosene, or gasoline, both pre tax and duty.

We have the following conditional Panel Vector-Error-Correction Model (PVECM) also known as nonlinear ARDL (autoregressive distributed lag model, Pesaran and Shin, 1998 and Pesaran et al, 2001):

$$\begin{aligned}\Delta y_{it} = \rho y_{i,t-1} + \boldsymbol{\delta}^{+\prime}\boldsymbol{\chi}_{i,t-1}^{+} + \boldsymbol{\delta}^{-\prime}\boldsymbol{\chi}_{i,t-1}^{-} + \textstyle\sum_{j=1}^{p-1}\gamma_j \Delta y_{i,t-j} + \\ \textstyle\sum_{j=0}^{q-1}\left(\boldsymbol{\pi}_j^{+\prime}\Delta\boldsymbol{\chi}_{i,t-j}^{+} + \boldsymbol{\pi}_j^{-\prime}\Delta\boldsymbol{\chi}_{i,t-j}^{-}\right) + e_{it}, \ i = 1, ..., n', \ t = 1, ..., T.\end{aligned} \tag{22}$$

The nonlinearity refers to the asymmetry of both the long-run relationship and the short-run error correction mechanism. A stable long-run relationship requires $\rho = 0, \boldsymbol{\delta}^{+} = \boldsymbol{\delta}^{-} = \mathbf{0}$. Short-run symmetry requires the following parameter restrictions:

$$\boldsymbol{\pi}_j^{+} = \boldsymbol{\pi}_j^{-}, \forall j = 0, ..., q-1,$$

or alternatively:

$$\textstyle\sum_{j=0}^{q-1}\boldsymbol{\pi}_j^{+} = \sum_{j=0}^{q-1}\boldsymbol{\pi}_j^{-}. \tag{23}$$

We can include a constant in (22) assuming there are deterministic trends in (21) which is a reasonable assumption. If we stack all time and state observations in (22) we obtain:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{24}$$

[10]The author is grateful to Michael Polemis for providing and explaining the data.

where $\mathbf{y} = [\Delta y_{it}, i = 1, ..., n, t = 1, ..., T]$, $\mathbf{X}$ denotes all right-hand-regressors in (22) and $\boldsymbol{\beta}$ denotes the associated vector of coefficients. Our empirical results are summarized in Table 3.

Table 3. Empirical results

	unleaded	derv	kerosene	gasoline
ρ	-0.34	-0.32	-0.30	-0.30
	(0.22)	(0.20)	(0.27)	(0.30)
β^+	0.71	0.79	0.85	0.92
	(0.044)	(0.035)	(0.051)	(0.048)
β^-	0.69	0.73	0.77	0.89
	(0.039)	(0.030)	(0.018)	(0.014)
π_0^+	0.41	0.45	0.47	0.51
	(0.025)	(0.032)	(0.018)	(0.025)
π_0^-	0.38	0.20	0.37	0.41
	(0.021)	(0.013)	(0.018)	(0.016)
$\sum_{j=1}^{q-1} \pi_j^+$	0.79	0.44	0.80	0.95
	(0.033)	(0.040)	(0.035)	(0.041)
$\sum_{j=1}^{q-1} \pi_j^-$	0.63	0.20	0.45	0.32
	(0.038)	(0.014)	(0.012)	(0.011)
Bayes factors (BF) favoring symmetry				
$\beta^+ = \beta^-$	0.98	1.34	12.22	41.50
$\pi_0^+ = \pi_0^-$	1.22	17.44	21.36	38.12
$\sum_{j=1}^{q-1} \pi_j^+ = \sum_{j=1}^{q-1} \pi_j^-$	1.17	19.20	25.43	44.57

Notes: Posterior standard deviations appear in parentheses. Reported are posterior means of the parameters. We set the number of lags to $q = 4$ using Akaike's information criterion after a preliminary least squares examination of the data. All computations are based on 20,000 iterations of the particle filter algorithm using 10^6 particles. Bayes factors (BF) are computed using the Particle Filter approximation in section 2 which delivers automatically the marginal or integrated likelihood of the model, viz. if the data are denoted by D then $p(D) = \int_\Theta \mathscr{L}(\theta; D)p(\theta)d\theta$, where $\mathscr{L}(\theta; D)$ is the likelihood and $p(\theta)$ is the prior.

Our interest, naturally, centers on the symmetry hypothesis which refutes the rockets-feathers hypothesis. The best way to discriminate between the two hypotheses is, probably, to examine further the parameter

$$\delta = \sum_{j=1}^{q-1} \pi_j^+ - \sum_{j=1}^{q-1} \pi_j^-, \tag{25}$$

which is the total effect from a price increase and price decrease, as it relies on the short-run dynamic multipliers. The impact multipliers are simply π_0^+, π_0^-. To show some light on the issues, we compare marginal posterior densities of δ in (25) with i) a normal approximation (which relies on an explicit assumption of normality of the error terms of (22), and ii) a bootstrap approximation using the D^ν linear program proposed by RRM. For the bootstrap implementation, we use 10,000 replications of a parametric bootstrap using re-centering of the residuals (Horowitz, 2001). The relevant densities are reported in Figures 2a through 2d, and correspond to the type of the dependent variable (unleaded, derv, kerosene and gasoline).

Figure 2a.

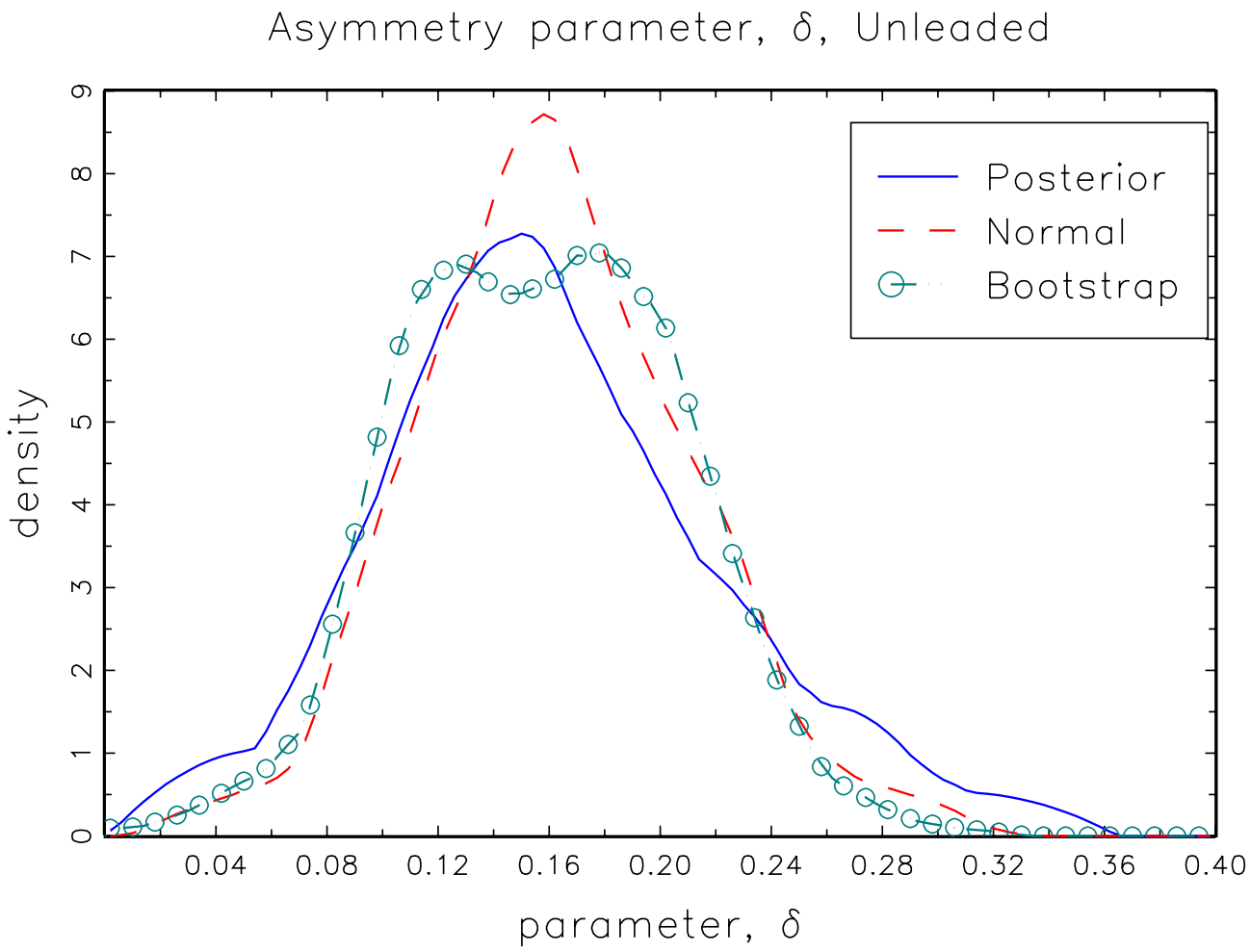

Asymmetry parameter, δ, Unleaded
density
parameter, δ
Posterior
Normal
Bootstrap
0
1
2
3
4
5
6
7
8
9
0.04
0.08
0.12
0.16
0.20
0.24
0.28
0.32
0.36
0.40

Figure 2b.

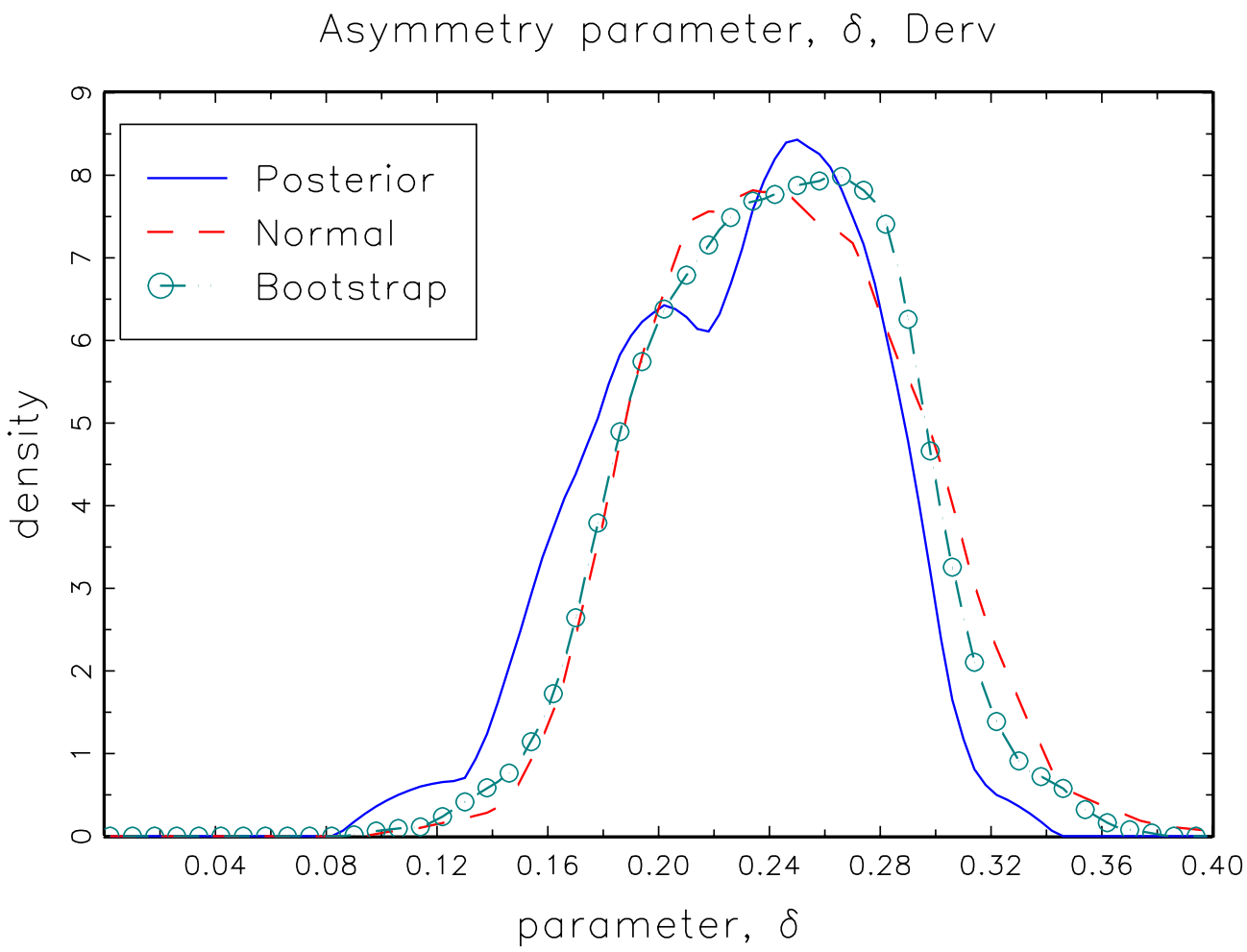

Asymmetry parameter, δ, Derv
Posterior
Normal
Bootstrap
density
0
1
2
3
4
5
6
7
8
9
0.04
0.08
0.12
0.16
0.20
0.24
0.28
0.32
0.36
0.40
parameter, δ

Figure 2c.

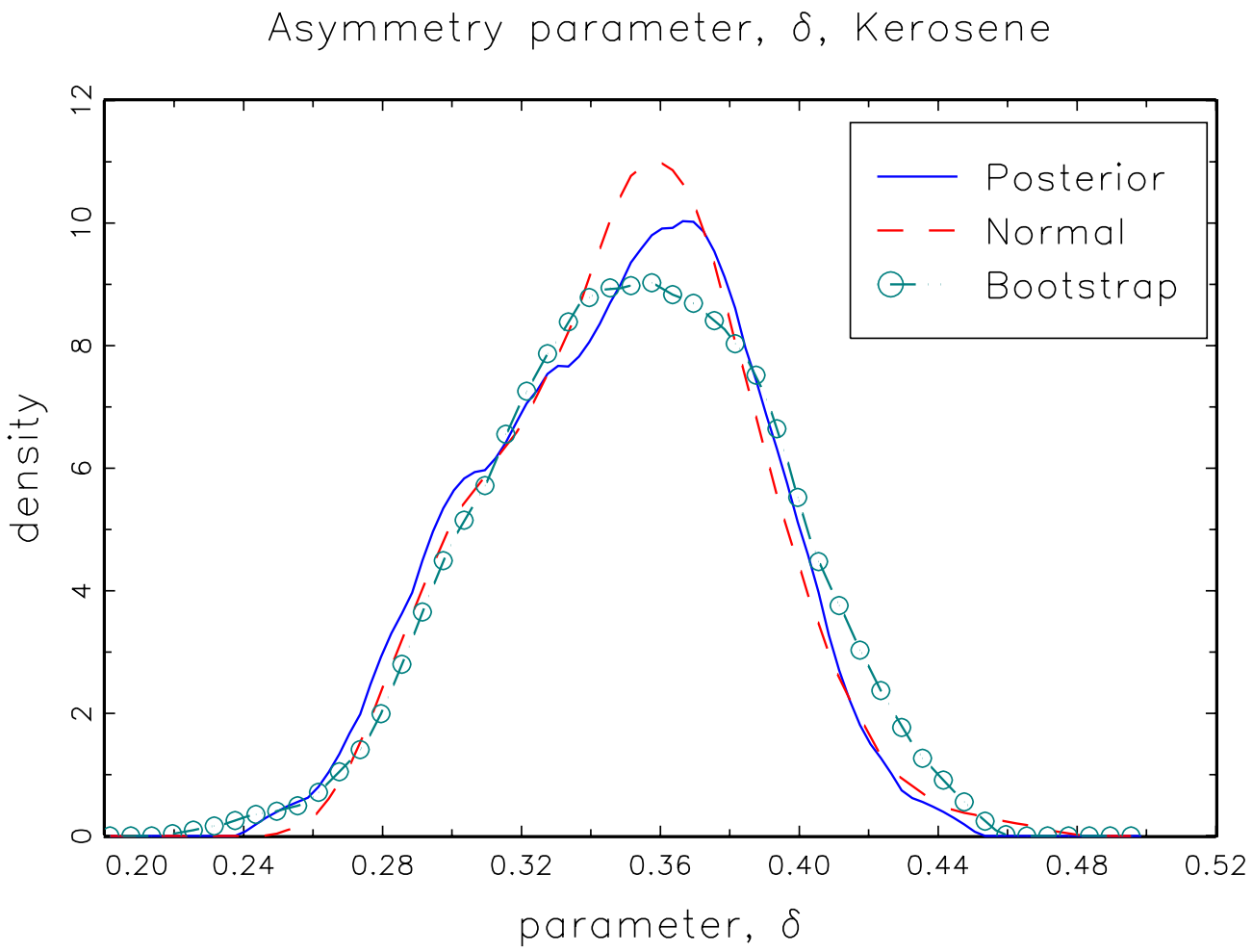
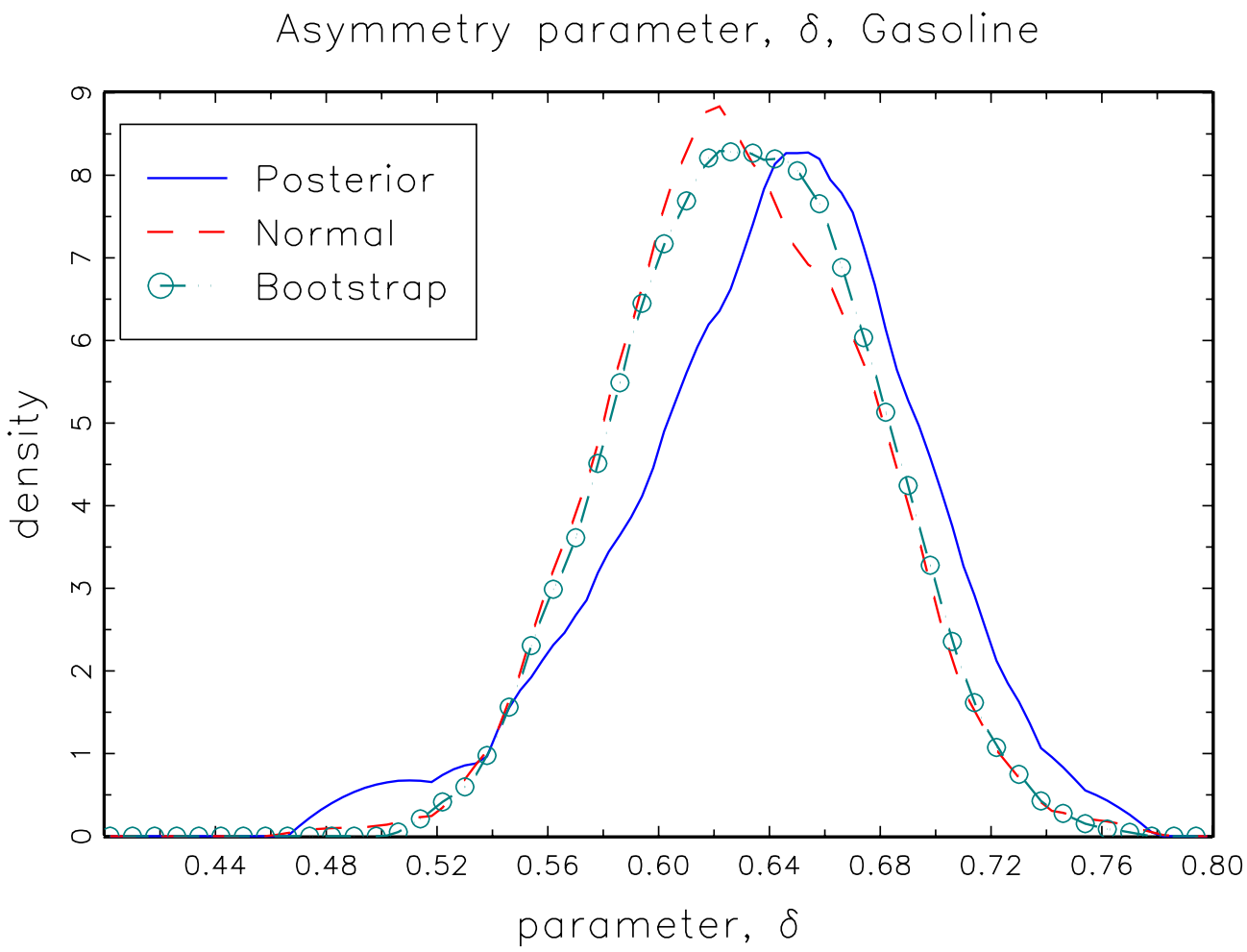

Asymmetry parameter, δ, Kerosene
density
parameter, δ
Posterior
Normal
Bootstrap
0
2
4
6
8
10
12
0.20
0.24
0.28
0.32
0.36
0.40
0.44
0.48
0.52

Figure 2d.



From the evidence in Figures 2a-2d the marginal posterior densities of asymmetry parameter, δ, have negligible probability mass around zero and they are concentrated in positive values. This is substantive evidence in favor of the "rockets and feathers" hypothesis.

5 Robustness

As our interpretation of CV@R regression rests upon a particular assumption concerning the error terms, it is reasonable to ask whether alternative error generating mechanisms affect the Bayesian CV@R regression in important ways. This effect, if one exists to begin with, should be more pronounced in relatively small samples. Under broad regularity conditions, posterior distributions as well as the bootstrap converge to the same asymptotic normal distribution. As the data and model in (22) represent a realistic scenario, we keep the same $\mathbf{X}$ and, for different sample sizes n, we assume alternative generating mechanisms for the errors in $\mathbf{e}$. The true values of the parameters, $\boldsymbol{\beta}$, are set to their posterior means, estimated and reported in the previous section. We consider the following alternative distributions for the error term: i) A normal distribution. ii) A Student-t distribution with 1,3,5 and 10 degrees of freedom. iii) A Laplace distribution, in which the maximum likelihood estimator coincides with the least absolute deviations (LAD) regression estimator. iv) Two- and three-component mixtures of normals. The two-component mixture is, in obvious notation, $\frac{1}{2}\times\mathcal{N}(-1,0.1^2)+\frac{1}{2}\times\mathcal{N}(1,0.3^2)$. The three component mixture is $0.25\times\mathcal{N}(-1,0.1^2)+0.10\times\mathcal{N}(0,0.3^2)+0.65\times\mathcal{N}(2,0.5^2)$. In all cases, the errors are re-scaled using the same parameter, the standard deviation of the

dependent variable in (22) or (24).

The sample size is set to n =25, 50, and 100. The Bayes CV@R estimators are applied to 10,000 different data sets (in which the true parameters $\boldsymbol{\beta}$ and the regressors $\mathbf{X}$ are kept fixed). For each replication, the Bayes CV@R estimator relies on 15,000 iteration the first 5,000 of which are discarded to mitigate the possible impact of start-up effects. It is based on the Particle Filter using 10^{12} particles.

The simulation results in terms of root mean squared error (RMSE) are reported in Table 4.

Table 4. RMSE results for CV@R

	$n = 25$	$n = 50$	$n = 100$
normality	0.031	0.014	0.0029
Student-t, df=1	0.028	0.012	0.0025
Student-t, df=3	0.029	0.015	0.00027
Student-t, df=5	0.031	0.014	0.0028
Student-t, df=10	0.031	0.014	0.0029
Laplace	0.025	0.012	0.0017
mixture, two normals	0.027	0.016	0.0021
mixture, three normals	0.026	0.012	0.0020

Notes: "df" stands for degrees of freedom of the Student-t distribution. The error terms are scaled using the same parameter, the standard deviation of the dependent variable in (22) or (24). The Bayes CV@R estimators are applied to 10,000 different data sets (in which the true parameters $\boldsymbol{\beta}$ and the regressors $\mathbf{X}$ are kept fixed). For each replication, the Bayes CV@R estimator relies on 5,000 iteration the first 1,000 of which are discarded to mitigate the possible impact of start-up effects. It is based on the Particle Filter using 10^6 particles.

From the evidence in Table 4, it turns out that Bayesian CV@R regression has a small bias in samples of size near 25 but the bias largely disappears in samples of size 50 and 100. Moreover, the behavior is similar across different distributions of the error term, although somewhat better for the Cauchy distribution (Student-t with one degree of freedom) for n =25 and 50.

Concluding remarks

In this paper we have developed a Bayesian approach to the problem of super-quantile or CV@R Regression recently introduced by Rockafellar, Royset and Miranda (2014, RRM). Although the concept is very appealing, computations in RRM turn out to be non-trivial for moderate sample sizes in regression. Computations rely mainly on large linear programming problems. The Bayesian interpretation of super-quantile regression does not require the use of large-scale linear programs. In fact, our posterior can be explored using a particle filtering approach (Chopin, 2002) for static models. In an empirical application provided by RRM we obtain similar results and, in addition, we are able to compute marginal posterior densities and posterior standard deviations. In RRM standard errors are not provided as the asymptotic theory of super-quantile regression is not well developed. In future research, it would be worthwhile to examine the application of super-quantile or CV@R Regression in the context of Seemingly Unrelated Regressions Equations and related models like the Vector Autoregression. The concept of super-quantile tracking would be particularly suitable in this context along with applications in forecasting using the concept of LASSO priors (Tibshirani, 1996, Park and Casella, 2008).

References

Bacon, Robert W, Rockets and Feathers: The Asymmetric Speed of Adjustment of UK Retail Gasoline Prices to Cost Changes (1991). Energy Economics 13; 211– 218.

Borenstein, S, Cameron, C. A. and R. Gilbert (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? Quarterly Journal of Economics 112; 305-339.

Cai Z., Wang X. (2008). Nonparametric estimation of conditional var and expected shortfall. Journal of Econometrics 147(1):120–130.

Chernozhukov V., Umantsev, L. (2001). Conditional value-at-risk: Aspects of modeling and estimation. Empirical Economics 26(1):271–292.

Chopin, N. (2002). A sequential particle filter method for static models, Biometrika 89 (3) 539-551.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. Annals of Statistics 32, 2385-2411.

Chun, S.Y., A. Shapiro and S. Uryasev (2012). Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics. Operations Research 60 (4), 739-756.

Deltas, G. (2008). Retail Gasoline Price Dynamics and Local Market Power. Journal of Industrial Economics 56; 613-628.

Durham, G., and J. Geweke (2013). Adaptive sequential posterior simulators for massively parallel computing environments. Working paper.

Engle R.F., Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22:367–381.

Fermanian J.D., Scaillet O. (2005) Sensitivity analysis of var and expected shortfall for portfolios under netting agreements. Journal of Banking and Finance 29: 927–958.

Gilks, W.R., and C. Berzuini (2001). Following a moving target: Monte Carlo inference for dynamic Bayesian models, Journal of the Royal Statistical Society B 63, 127-146.

Greenwood-Nimmo, M., and Y. Shin (2013).Taxation and the Asymmetric Adjustment of Selected Retail Energy Prices in the UK. Economics Letters 121; 411-416.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The elements of statistical learning: Data Mining, Inference, and Prediction. Springer, Berlin.

Horowitz, J.L. (2001). The bootstrap in econometrics, chapter 52 of the Handbook of Econometrics, Volume 5, Edited by J.J. Heckman and E. Leamer, Elsevier Science, pp. 3159-3228.

Koenker, R. and Bassett, G. W. (1978). Regression Quantiles. Econometrica, 46: 33–50. 533

Kozumi, H. and Kobayashi, G. (2009). Gibbs Sampling Methods for Bayesian Quantile Regression. Technical report, Graduate School of Business Administration, Kobe University.

Lamotte, O., Porcher, T., Schalck, C., and Silvestre, S. (2013). Asymmetric Gasoline Price Responses in France. Applied Economics Letters 20; 457-461.

Leorato S., Peracchi F., Tanase, A.V. (2012). Asymptotically efficient estimation of the conditional expected shortfall. Computational Statistics and Data Analysis 56(4): 768–784.

Park, T., and G. Casella (2008). The Bayesian LASSO. Journal of the American Statistical Association 103, 681-686.

Peracchi F., Tanase A.V. (2008). On estimating the conditional expected shortfall. Applied Stochastic Models in Buiness and Industry 24: 471–493.

Pesaran, M.H., Shin, Y., 1998. An autoregressive distributed lag modelling approach to cointegration analysis. In: Strom, S. (Ed.), Econometrics and Economic Theory: The Ragnar Frisch Centennial Symposium. Cambridge University Press, Cambridge.

Pesaran, M.H., Shin, Y., Smith, R.J. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics 16, 289–326.

Rockafellar R.T., Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26(7): 1443–1471.

Rockafellar R.T., Uryasev S., Zabarankin M. (2008) Risk tuning with generalized linear regression. Mathematics of Operations Research 33(3): 712–729.

Rockafellar, R.T. (2007). Coherent Approaches to Risk in Optimization Under Uncertainty, Tutorials in Operations Research INFORMS 2007, 38-61.

Rockafellar, R.T., J.O. Royset and S.I. Miranda (2014). super-quantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk, European Journal of Operational Research 234, 140–154.

Rockafellar, R. T. and S. P. Uryasev (2000). Optimization of conditional value-at-risk. Journal of Risk 2: 21–42.

Rockafellar, R. T. and S. P. Uryasev (2002). Conditional value-at-risk for general loss distributions. Journal of Banking and Finance 26:1443–1471.

Scaillet O. (2004) Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14(1): 115–129.

Scaillet O. (2005) Nonparametric estimation of conditional expected shortfall. Insurance and Risk Management Journal 72: 639–660.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, Ser. B, 58, 267–288

Tsionas, E. G. (2003). Bayesian Quantile Inference. Journal of Statistical Computation and Simulation, 73: 659–674.

Wozabal D., Wozabal, N. (2009) Asymptotic consistency of risk functionals. Journal of Nonparametric Statistics 21(8):977–990.

Appendix

Lemma 1. The posterior in (15) is finitely integrable.

Proof. The posterior in (15) is:

$$p(\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\alpha) \propto \left\{\sum_{t=1}^{n} \overline{\varrho}_{\alpha}(u_t(\boldsymbol{\beta}))\right\}^{-n/2}.$$

Suppose $\overline{\varrho}_{\alpha}(u_T(\boldsymbol{\beta})) = \min_{t=1,\cdots,n} \overline{\varrho}_{\alpha}(u_t(\boldsymbol{\beta}))$. Then $\{\sum_{t=1}^{n} \overline{\varrho}_{\alpha}(u_t(\boldsymbol{\beta}))\}^{-n} \leq \{n\overline{\varrho}_{\alpha}(u_T(\boldsymbol{\beta}))\}^{-n/2}$. By the mean value theorem for integrals, (9) and (2) we have $\overline{\varrho}_{\alpha}(u_T(\boldsymbol{\beta})) = q_{\gamma^*}(u_T(\boldsymbol{\beta}))$, for some $\gamma^* \in [\alpha, 1)$. Existence of $\int_{\Re^K} \{q_{\gamma^*}(u_T(\boldsymbol{\beta}))\}^{-n/2} d\boldsymbol{\beta}$ follows from the fact that the posterior distribution of quantile regression models is finitely integrable with respect to $\boldsymbol{\beta}$ as shown in Appendix C of Tsionas (2003).

Lemma 2. Any proper prior for $\boldsymbol{\beta}$ leads to a proper posterior.

Proof. Assume the prior of $\boldsymbol{\beta}$ is proportional to $p(\boldsymbol{\beta})$. From (15) we have:

$$\int \left\{\sum_{t=1}^{n} \overline{\varrho}_{\alpha}(u_t(\boldsymbol{\beta}))\right\}^{-n/2} p(\boldsymbol{\beta})d\boldsymbol{\beta} < c \cdot \int p(\boldsymbol{\beta})d\boldsymbol{\beta} < \infty,$$

for some positive and finite constant c as $\sum_{t=1}^{n} \overline{\varrho}_{\alpha}(u_t(\boldsymbol{\beta}))$ is strictly positive and, therefore, bounded below by c, in any given data set.