Bayesian Inference for L_p quantile regression models

Mauro Bernardi[1] Valeria Bignozzi[2] Lea Petrella[2]

[1]Department of Statistical Sciences, University of Padova, Italy
[2]MEMOTEF Department, Sapienza University of Rome, Italy

Workshop on
"Recent Advances in Quantile and M–quantile Regression"
University of Pisa, July 15, 2016

Quantile regression (QR)

QR

- simple, robust and distribution free tool for modelling conditional quantiles of a response variable as a function of some covariates, see Koenker and Basset (1978), Koenker (2005).
- account for skewness, fat-tails, outliers, truncated and censored data, and heteroskedasticity, that can shadow the nature of the dependence between the variable of interest and the covariates.
- appropriate when the underlying model is nonlinear, innovation terms are non-Gaussian, tail behaviour is of the primary interest (Lum and Gelfand 2012 and Koenker 2005).

L_p quantiles

- extend quantile methods, see Bignozzi et al. (2016);
- applications in statistics, economics and finance, (Valeria);
- elicitable risk measures (backtesting).

QR and alternatives...

Quantile regression (Koenker and Bassett, 1978)

- $q_\tau\left(Y \mid \mathbf{X}=\mathbf{x}\right) = \mathbf{x}'\boldsymbol{\beta}$;
- solve the minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i'\boldsymbol{\beta})}\left(y_i\right)\right)\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)$$

QR and alternatives...

Quantile regression (Koenker and Bassett, 1978)

- $q_\tau\left(Y \mid \mathbf{X}=\mathbf{x}\right)=\mathbf{x}'\boldsymbol{\beta}$;
- solve the minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i'\boldsymbol{\beta})}\left(y_i\right)\right)\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)$$

Asymmetric Least Square regression (ALS) (Newey and Powel, 1987)

- $e_\tau\left(Y \mid \mathbf{X}=\mathbf{x}\right)=\mathbf{x}'\boldsymbol{\beta}$;
- solve the minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i'\boldsymbol{\beta})}\left(y_i\right)\right)\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2$$

for $\tau \in (0, 1)$.

L_p–quantile regression models

L_p-quantiles, $\rho_{p,\tau}, p \geq 2, \tau \in (0,1)$

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i'\boldsymbol{\beta})}(y_i)\right)(y_i - \mathbf{x}_i'\boldsymbol{\beta})^p \,.$$

L_p–quantile regression models

L_p-quantiles, $\rho_{p,\tau}, p \geq 2, \tau \in (0,1)$

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i' \boldsymbol{\beta})}(y_i)\right) (y_i - \mathbf{x}_i' \boldsymbol{\beta})^p \,.$$

- introduced by Chen (1996);
- elicitable risk measures;
- L_2-quantile corresponds to the expectile, see Bellini et al. (2014);
- for $p \in (1,2)$, L_p–quantiles are MM estimators, see Lange (2010), Hunter and Lange (2000), Bernardi et al. (2016).

Bayesian Inference

- Yu and Moyeed (2001) propose to use the Asymmetric Laplace (AL) distribution to perform Bayesian quantile regression;
- Gerlach et al. (2012) propose to use the Asymmetric Gaussian (AG) distribution to perform Bayesian expectile regression;
- Robust regression models using L^p quantiles have been considered by Bernardi et al. (2016).

Bayesian Inference

- Yu and Moyeed (2001) propose to use the Asymmetric Laplace (AL) distribution to perform Bayesian quantile regression;
- Gerlach et al. (2012) propose to use the Asymmetric Gaussian (AG) distribution to perform Bayesian expectile regression;
- Robust regression models using L^p quantiles have been considered by Bernardi et al. (2016).

Goal of the paper

Perform Bayesian inference for L_p–quantile regression models.

Approach

Find a distribution function which extends the AL and AG.

Goals

Inference

- We introduce a misspecified likelihood for L_p–quantile regression models;

Goals

Inference

- We introduce a misspecified likelihood for L_p–quantile regression models;
- we provide a full Bayesian approach for inference on parameters of L_p–quantile regression models;

Goals

Inference

- We introduce a misspecified likelihood for L_p–quantile regression models;
- we provide a full Bayesian approach for inference on parameters of L_p–quantile regression models;
- we prove posterior consistency of the Bayesian procedure;

Goals

Inference

- We introduce a misspecified likelihood for L_p–quantile regression models;
- we provide a full Bayesian approach for inference on parameters of L_p–quantile regression models;
- we prove posterior consistency of the Bayesian procedure;
- we perform model selection under dirac spike and slab L_1 prior.

Goals

Inference

- We introduce a misspecified likelihood for L_p–quantile regression models;
- we provide a full Bayesian approach for inference on parameters of L_p–quantile regression models;
- we prove posterior consistency of the Bayesian procedure;
- we perform model selection under dirac spike and slab L_1 prior.
- US inflation data (DQMA, Bernardi et al. 2016).

The Skewed Exponential Power (SEP) distribution

The SEP density function is given by, see Zhu and Zinde-Walsh (2009)

$$f_{\text{SEP}}(y \mid \mu, \sigma, \alpha, p) = \begin{cases} \frac{1}{\sigma} K(p) \exp\left(-\frac{1}{p}\left|\frac{y-\mu}{2\alpha\sigma}\right|^p\right) & \text{if } y \le \mu \\ \frac{1}{\sigma} K(p) \exp\left(-\frac{1}{p}\left|\frac{y-\mu}{2(1-\alpha)\sigma}\right|^p\right) & \text{if } y > \mu \end{cases}$$

- $\alpha \in (0,1)$, $\sigma > 0$, $p > 0$;
- $K(p) = \left[2p^{\frac{1}{p}}\Gamma\left(1+\frac{1}{p}\right)\right]^{-1}$ is the normalising constant;
- the location parameter μ coincides with $Q_\alpha(Y)$;
- admits a stochastic representation.

Proposition

If $U \sim \mathcal{G}\left(1+\frac{1}{p}, \frac{1}{p}\right)$ and $Y \mid U = u \sim \mathcal{U}\left(\mu - 2\alpha\sigma u^{\frac{1}{p}}, \mu\right)$, with probability α and $Y \mid U = u \sim \mathcal{U}\left(\mu, \mu + 2(1-\alpha)\sigma u^{\frac{1}{p}}\right)$ with probability $(1-\alpha)$, then $Y \sim \mathcal{SEP}(\mu, \sigma, p, \alpha)$.

SEP (Cont'ed)

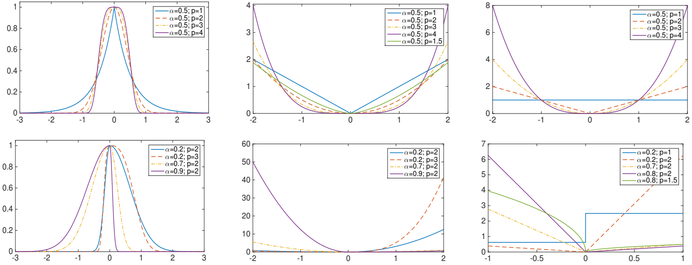

Figure : SEP, pdf, loss and score for different values of p and α.

L_p–quantiles and SEP

- to solve the minimisation problem for $\tau \in (0,1)$

$$\rho_{p,\tau}(Y) = \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{(-\infty, \mathbf{x}_i'\boldsymbol{\beta})}(y_i)\right)(y_i - \mathbf{x}_i'\boldsymbol{\beta})^p$$

is equivalent to finding the MLE for

$$\mathcal{L}(\mathbf{y} \mid \mathbf{X} = \mathbf{x}) \propto \exp\left(-\sum_{i=1}^{n} \frac{1}{p}\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{2\left(\mathbb{1}_{(\mathbf{x}_i'\boldsymbol{\beta}, \infty)}(y_i) - \alpha\right)}\right)^p\right),$$

for $\alpha = \frac{\tau^{\frac{1}{p}}}{\tau^{\frac{1}{p}} + (1-\tau)^{\frac{1}{p}}}$.

Key point

Minimising the empirical loss function for the L_p–quantiles is equivalent to maximising the likelihood function obtained from an i.i.d. data sample of the SEP distribution.

L_p–quantile regression model

The Model

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_{\alpha,p} + \varepsilon_i, \qquad \text{for all } i = 1, 2, \ldots, n$$

- $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is a random sample of size n;
- $\mathbf{x}_i = (1, x_{1,i}, x_{2,i}, \ldots, x_{q,i})'$ the associated set of $q+1$ covariates for $i = 1, 2, \ldots, n$;
- $\boldsymbol{\beta}_{\alpha,p} = \left(\beta^p_{\alpha,0}, \beta^p_{\alpha,1}, \ldots, \beta^p_{\alpha,q}\right)'$ is a vector $q+1$ unknown regression parameters;
- $\varepsilon_i \sim \mathcal{SEP}\,(0, \sigma, \alpha, p)$, i.i.d., for any $i = 1, 2, \ldots, n$

In particular,

$$y_i \mid \mathbf{x}_i \sim \mathcal{SEP}\,(\mathbf{x}_i'\boldsymbol{\beta}_{\alpha,p}, \sigma, \alpha, p)\,.$$

Prior distributions

Non–informative prior

- $\pi\left(\boldsymbol{\beta}_{p,\alpha}, \tilde{\sigma}\right) \propto 1$, where $\tilde{\sigma} = \sigma^{p}$.
- Proper joint posterior: theoretical result.

Prior distributions

Non–informative prior

- $\pi\left(\boldsymbol{\beta}_{p,\alpha}, \tilde{\sigma}\right) \propto 1$, where $\tilde{\sigma} = \sigma^p$.
- Proper joint posterior: theoretical result.

Informative priors

- $\boldsymbol{\beta}_{p,\alpha} \sim \mathcal{N}\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right)$, $\tilde{\sigma} \sim \mathcal{IG}\left(\lambda_\sigma, \eta_\sigma\right)$.
- Lasso L_1 regularised prior: multivariate approach.
- Posterior consistency: theoretical result.

Prior distributions

Non–informative prior

- $\pi\left(\boldsymbol{\beta}_{p,\alpha}, \tilde{\sigma}\right) \propto 1$, where $\tilde{\sigma} = \sigma^p$.
- Proper joint posterior: theoretical result.

Informative priors

- $\boldsymbol{\beta}_{p,\alpha} \sim \mathcal{N}\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right)$, $\tilde{\sigma} \sim \mathcal{IG}\left(\lambda_\sigma, \eta_\sigma\right)$.
- Lasso L_1 regularised prior: multivariate approach.
- Posterior consistency: theoretical result.

Dirac spike–and–slab prior

- Lasso L_1 regularised prior + dirac in zero.
- Stochastic Search Variable Selection algorithm.

Proper posterior under improper prior

Theorem

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ *be a sample of i.i.d. observations generated by the linear* L_p *quantile regression model with* $q > 1$*, under the assumption of improper diffuse prior for the regressor parameters, i.e.,* $\pi(\boldsymbol{\beta}) \propto 1$*, then*

$$0 < \int_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma)\, \pi(\boldsymbol{\beta})\, d\boldsymbol{\beta} < \infty.$$

Remark

Under improper prior the Bayesian approach to the L_p *quantile regression model generalises the Bayesian quantile regression framework of Yu and Moyeed (2001).*

Posterior consistency

Assumptions:

- Let $(Y_1, Y_2, \ldots, Y_n)$ be a sequence of independent observations of a univariate response variable and let $(X_1, X_2, \ldots, X_n)$ be one–dimensional non–random covariates.
- $\mathcal{P}_{0,i}$ denotes the true and unknown probability distribution of Y_i, with the true α–level L_p quantile of order $p \in \mathbb{N}^+$ given by

$$Q^p_\alpha \left(Y_i \mid X_i\right) = \delta_0 + \beta_0 X_i,$$

  for $\alpha \in (0, 1)$.
- Suppose that the misspecified model for

$$Y_i \sim \mathcal{SEP}\left(\cdot, \mu_i^\alpha, \sigma, p, \alpha\right),$$

  with

$$\mu_i^\alpha = \delta + \beta X_i,$$

  for $i = 1, 2, \ldots, n$.

Posterior consistency (Cont'ed)

Result: If the following condition is satisfied

$$\mathbb{E}_{\mathcal{P}_0}\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta_0, \beta_0, \sigma_0 = 1\right)}\right)\right] < +\infty,$$

then

$$\begin{aligned}\inf_{\delta,\beta} \mathbb{E}_{\mathcal{P}_0}&\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta, \beta, \sigma_0 = 1\right)}\right)\right] \\ &\geq \mathbb{E}_{\mathcal{P}_0}\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta_0, \beta_0, \sigma = 1\right)}\right)\right],\end{aligned}$$

for fixed α and p. (BQR, see Sriram et al. 2013).

Posterior consistency (Cont'ed)

Result: If the following condition is satisfied

$$\mathbb{E}_{\mathcal{P}_0}\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta_0, \beta_0, \sigma_0 = 1\right)}\right)\right] < +\infty,$$

then

$$\begin{aligned}\inf_{\delta,\beta} \mathbb{E}_{\mathcal{P}_0}&\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta, \beta, \sigma_0 = 1\right)}\right)\right] \\ &\geq \mathbb{E}_{\mathcal{P}_0}\left[\log\left(\frac{\mathcal{P}_{0,i}\left(Y_i \mid X_i\right)}{\mathcal{P}_i^{\alpha,p}\left(Y_i \mid X_i, \delta_0, \beta_0, \sigma = 1\right)}\right)\right],\end{aligned}$$

for fixed α and p. (BQR, see Sriram et al. 2013).

Proposition

The weak consistency of the posterior follows from the Schwartz (1965) theorem, since any Kullback–Leibler neighbourhood of the true density has positive probability.

The collapsed Gibbs sampler

- Choose the initial parameters value $\left(\boldsymbol{\beta}_{\alpha,p}^{(0)}, \tilde{\sigma}^{(0)}\right)$.
- Iteratively sampling $\left(\boldsymbol{\beta}_{\alpha,p}^{(k)}, \mathbf{u}^{(k)}, \tilde{\sigma}^{(k)}\right)$, for $k = 1, 2, \ldots$ from
  (i) $\tilde{\sigma}^{(k)} \sim \pi\left(\tilde{\sigma} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_{\alpha,p}^{(k-1)}\right)$, collapsed step with $\tilde{\sigma} = \sigma^p$;
  (ii) $\mathbf{u}^{(k)} \sim \pi\left(\mathbf{u} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_{\alpha,p}^{(k-1)}, \tilde{\sigma}^{(k)}\right)$;
  (iii) $\boldsymbol{\beta}_{\alpha,p}^{(k)} \sim \pi\left(\boldsymbol{\beta}_{\alpha,p} \mid \mathbf{y}, \mathbf{X}, \mathbf{u}^{(k)}, \tilde{\sigma}^{(k)}\right)$, by iteratively simulating from the complete set of full conditionals $\pi\left(\beta_{\alpha,p}^{(j)} \mid \mathbf{y}, \mathbf{X}, \mathbf{u}^{(k)}, \boldsymbol{\beta}_{\alpha,p}^{(-j)}\right)$, for $j = 0, 1, \ldots, q$.

The collapsed Gibbs sampler

- Choose the initial parameters value $\left(\boldsymbol{\beta}_{\alpha,p}^{(0)}, \tilde{\sigma}^{(0)}\right)$.
- Iteratively sampling $\left(\boldsymbol{\beta}_{\alpha,p}^{(k)}, \mathbf{u}^{(k)}, \tilde{\sigma}^{(k)}\right)$, for $k = 1, 2, \ldots$ from
  - *(i)* $\tilde{\sigma}^{(k)} \sim \pi\left(\tilde{\sigma} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_{\alpha,p}^{(k-1)}\right)$, collapsed step with $\tilde{\sigma} = \sigma^p$;
  - *(ii)* $\mathbf{u}^{(k)} \sim \pi\left(\mathbf{u} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_{\alpha,p}^{(k-1)}, \tilde{\sigma}^{(k)}\right)$;
  - *(iii)* $\boldsymbol{\beta}_{\alpha,p}^{(k)} \sim \pi\left(\boldsymbol{\beta}_{\alpha,p} \mid \mathbf{y}, \mathbf{X}, \mathbf{u}^{(k)}, \tilde{\sigma}^{(k)}\right)$, by iteratively simulating from the complete set of full conditionals $\pi\left(\beta_{\alpha,p}^{(j)} \mid \mathbf{y}, \mathbf{X}, \mathbf{u}^{(k)}, \boldsymbol{\beta}_{\alpha,p}^{(-j)}\right)$, for $j = 0, 1, \ldots, q$.

Remark

Steps (i) and (ii) above ensures draws from the conditional posterior distribution $\pi\left(\tilde{\sigma}, \mathbf{u} \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_{p,\alpha}\right)$, see, Van Dyk and Park (2008), Park and Van Dyk (2009) and Bernardi et al. (2015).

Handling sparsity and regressor selection

Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)$ be the q–dimensional vector where $\gamma_j = 1$ if the j–th covariate $\mathbf{x}_j = (x_{j,1}, x_{j,2}, \dots, x_{j,T})'$, for $j = 1, 2, \dots, q$ is included as explanatory variable in the regression model and $\gamma_j = 0$, otherwise.

$$\pi\left(\boldsymbol{\beta}_{p,\alpha} \mid \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \varrho, \varsigma\right) = \pi_0\left(\beta_{p,\alpha}^{(0)} \mid \varsigma\right) \pi_{\mathsf{SL}}\left(\boldsymbol{\beta}_{p,\alpha}^{\gamma} \mid \boldsymbol{\Sigma}\right) \prod_{j:\gamma_j=0} \pi_{\mathsf{SP}}\left(\beta_{p,\alpha}^{(j)}\right)$$

$$\pi_0\left(\beta_{p,\alpha}^{(0)} \mid \varsigma\right) = \mathcal{L}_1\left(\beta_{p,\alpha}^{(0)} \mid 0, \varsigma\right)$$

$$\pi_{\mathsf{SP}}\left(\beta_{p,\alpha}^{(j)}\right) = \delta_0\left(\beta_{p,\alpha}^{(j)}\right)$$

$$\pi_{\mathsf{SL}}\left(\boldsymbol{\beta}_{p,\alpha}^{(\gamma)} \mid \boldsymbol{\Sigma}_{\gamma}, \boldsymbol{\gamma}\right) = \mathcal{L}_r\left(\boldsymbol{\beta}_{p,\alpha}^{(\gamma)} \mid \mathbf{O}_{\gamma}, \boldsymbol{\Sigma}_{\gamma}\right),$$

where $\delta_0\left(\beta_{p,\alpha}^{(j)}\right)$ is a point mass at zero and $\mathcal{L}_r\left(\boldsymbol{\beta}_{p,\alpha}^{(\gamma)} \mid \mathbf{O}_{\gamma}, \boldsymbol{\Sigma}_{\gamma}\right)$ denotes the r–dimensional Laplace distribution with $r = \sum_{j=1}^{q} \gamma_j$.

Handling sparsity and regressor selection (Cont'ed)

- Dirac spike–and–slab L_1 prior: Hans (2009, 2010), Bernardi (2016).

Handling sparsity and regressor selection (Cont'ed)

- Dirac spike–and–slab L_1 prior: Hans (2009, 2010), Bernardi (2016).
- Alternative: DQMA approach of Bernardi et al. (2016).

Handling sparsity and regressor selection (Cont'ed)

- Dirac spike–and–slab L_1 prior: Hans (2009, 2010), Bernardi (2016).
- Alternative: DQMA approach of Bernardi et al. (2016).
- Here a multivariate slab prior accounts for L_1 shrinkage and regression correlation. Marginals are L_1.

Handling sparsity and regressor selection (Cont'ed)

- ▷ Dirac spike–and–slab L_1 prior: Hans (2009, 2010), Bernardi (2016).
- ▷ Alternative: DQMA approach of Bernardi et al. (2016).
- ▷ Here a multivariate slab prior accounts for L_1 shrinkage and regression correlation. Marginals are L_1.
- ▷ Hierarchical prior specification is completed by

$$
\begin{aligned}
\varrho &\sim \mathcal{B}e\left(\varrho \mid \alpha_\varrho, \beta_\varrho\right) \\
\varsigma &\sim \mathcal{IG}\left(\varsigma \mid \psi, \varpi\right) \\
\boldsymbol{\Sigma} \mid \boldsymbol{\gamma} &\sim \mathcal{IW}_r\left(\boldsymbol{\Sigma}_\gamma \mid c, \mathbf{C}_\gamma\right).
\end{aligned}
$$

Handling sparsity and regressor selection (Cont'ed)

- Dirac spike–and–slab L_1 prior: Hans (2009, 2010), Bernardi (2016).
- Alternative: DQMA approach of Bernardi et al. (2016).
- Here a multivariate slab prior accounts for L_1 shrinkage and regression correlation. Marginals are L_1.
- Hierarchical prior specification is completed by

$$
\begin{aligned}
\varrho &\sim \mathcal{B}e\left(\varrho \mid \alpha_\varrho, \beta_\varrho\right) \\
\varsigma &\sim \mathcal{IG}\left(\varsigma \mid \psi, \varpi\right) \\
\boldsymbol{\Sigma} \mid \boldsymbol{\gamma} &\sim \mathcal{IW}_r\left(\boldsymbol{\Sigma}_\gamma \mid c, \mathbf{C}_\gamma\right).
\end{aligned}
$$

- Stochastic Search Variable Selection (SSVS) algorithm: exploits conjugacy with data augmentation and the structure of the conditionals.

SSVS algorithm

- Choose the initial parameters value $\left(\boldsymbol{\beta}_{p,\alpha}^{(0)}, \tilde{\sigma}^{(0)} \varrho^{(0)}, \varsigma^{(0)}, \boldsymbol{\Sigma}^{(0)}\right)$.
- Iteratively sampling $\left(\boldsymbol{\beta}_{p,\alpha}^{(k)}, \tilde{\sigma}^{(k)}, \mathbf{u}^{(k)}, \boldsymbol{\gamma}^{(k)}, \varrho^{(k)}, \varsigma^{(k)}, \boldsymbol{\Sigma}^{(k)}\right)$, for $k = 1, 2, \ldots$ from
  - *(i)* $\tilde{\sigma}^{(k)}, \mathbf{u}^{(k)}, \boldsymbol{\beta}_{p,\alpha}^{(k)}$, sampled as before with the only difference the FC of $\beta_{p,\alpha}^{(j)}$ is GH;
  - *(ii)* $\gamma_j^{(k)} \sim \pi\left(\gamma_j \mid \mathbf{y}, \mathbf{X}_\gamma, \mathbf{u}^{(k)}, \boldsymbol{\beta}_{p,\alpha}^{(\gamma,k)}, \tilde{\sigma}^{(k)}, \boldsymbol{\gamma}_{-j}^{(k-1)}, \varrho^{(k-1)}, \varsigma^{(k-1)}, \boldsymbol{\Sigma}^{(k-1)}\right)$, for $j = 1, 2, \ldots, q$, which is Bernoulli with parameter $\mathbb{P}(\gamma_j = 1) = 1 - \tilde{\pi}_j$ with

$$\tilde{\pi}_j = \frac{1}{1 + \frac{\varrho}{1-\varrho} R_j}$$

$$\begin{aligned} R_j &= \int_{\underline{\beta}_j}^{\overline{\beta}_j} \pi\left(\beta_{p,\alpha}^{(j)} \mid \mathbf{y}, \mathbf{X}, \mathbf{u}, \boldsymbol{\beta}_{p,\alpha}^{(-j)}, \tilde{\sigma}, \varrho, \varsigma, \boldsymbol{\gamma}\right) d\beta_{p,\alpha}^{(j)} \\ &= F_{\beta_{p,\alpha}^{(j)}}\left(\overline{\beta}_j\right) - F_{\beta_{p,\alpha}^{(j)}}\left(\underline{\beta}_j\right), \end{aligned}$$

  - *(iii)* $\varrho^{(k)}, \varsigma^{(k)}, \boldsymbol{\Sigma}^{(k)}$, exploits conjugacy.

Applications

- Melbourne daily maximum temperature;

Applications

- ▷ Melbourne daily maximum temperature;
- ▷ New Zealand population dataset;

Applications

- Melbourne daily maximum temperature;
- New Zealand population dataset;
- Car dataset;

Applications

- Melbourne daily maximum temperature;
- New Zealand population dataset;
- Car dataset;
- Forecast VaR of large dimensional financial indexes;

Applications

- ▷ Melbourne daily maximum temperature;
- ▷ New Zealand population dataset;
- ▷ Car dataset;
- ▷ Forecast VaR of large dimensional financial indexes;
- ▷ **US inflation data (DQMA, Bernardi et al. 2016)**.

Application: US Inflation

Generalised Phillips curve

1. Relevance of covariates to predict current inflation at different quantile levels
2. Verify empirically whether predictors for high and low inflation are different or their relevance change over time.

Different quantile levels

The focus on quantiles of the predicted variable helps in discerning periods characterised by different economic implications and, in particular, those featured by low and high inflation levels.
Relevant effects:

1. inclusion probability may be different at different τ–levels;
2. those probabilities may substantially change during period of high inflation with respect to those of low inflation.

US Inflation (Cont'ed)

Time-variations

The degree of inflation pressure in the economy may have potential effects on the real side of the economy and on the overall level of output produced and may influence the business cycle amplitude and period.

Autoregressive model with exogenous of order M

We extend the ARX(p) of Stock and Watson (1999) and Koop and Korobilis (2012) to the L_p –quantile framework

$$q_\tau\left(y_t, \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \mathbf{x}'_{t-1}\boldsymbol{\beta} + \sum_{j=1}^{M} \gamma_j y_{t-j} \tag{1}$$

where $y_t = 100 \log\left(\frac{P_t}{P_{t-1}}\right)$, with P_t being a price index.

US Inflation Data

Variable	Name
FX	CAN/US, GER/US, JPN/US and US/UK
Money	M1, M2
Bank loans	bank prime loan rate
	Commercial and Industrial Loans, All Commercial Banks
	Consumer Loans, All Commercial Banks
Interest Rates	30-Year Conventional Mortgage Rate
	Bank Prime Loan Rate
	3-Month Treasury Bill
	10-Year Treasury Constant Maturity Rate
	Long-Term Government Bond Yields: 10-year
	Moody's Seasoned Aaa Corporate Bond Yield
	Effective Federal Funds Rate
	1–Month Certificate of Deposit
Employment	Civilian Unemployment Rate
	Number of Civilians Unemployed for 15 Weeks & Over
	Number of Civilians Unemployed – Less than 5 Weeks
	Civilian Labor Force Participation Rate
	All Employees: Total nonfarm
	All Employees: Total Private Industries
Expectations	University of Michigan: Inflation Expectation
	University of Michigan: Consumer Sentiment
Real estate	Housing Starts Total: New Privately Owned Units Started
	New One Family Houses Sold: United States
	Real Estate Loans at All Commercial Banks
Production	Index of Aggregate Weekly Hours
	Industrial Production Index
	ISM Manufactoring: PMI Composite Index
	ISM Manufacturing: Supplier Deliveries Index
	Real personal consumption expenditures
	Capacity of Utilization: Total Industry
	Motor Vehicle Assemblies: Total motor vehicle assemblies
Oil	Spot Oil Price: West Texas Intermediate
Finance	SP500 index

US Inflation (Cont'ed)

Predictor	UNRATE	MORTG	MPRIME	CDM1	FEDFUNDS	UNEMPL	OILPRICE	INFEXP
$\tau = 0.10$	$\perp\triangle\square$	$\perp\triangle\square$	–	–	–	$\perp\triangle\square\bigstar$	–	–
$\tau = 0.25$	–	–	–	–	–	$\perp\triangle\square\bigstar$	–	$\perp\triangle\square\bigstar$
$\tau = 0.50$	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	–	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	–	$\perp\triangle\square\bigstar$
$\tau = 0.75$	–	–	–	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$	$\perp\triangle\square\bigstar$
$\tau = 0.90$	–	–	–	$\perp\triangle\square$	$\perp\triangle\square$	–	$\perp\triangle\square\bigstar$	–

Symbol	Legend
$\perp$	$p = 2$
$\triangle$	$p = 3$
$\square$	$p = 4$
$\bigstar$	$p = 5$

Discussion and conclusion

Main contributions

- L_p quantile regression model generalises QR and Expectile regression;
- Bayesian inference and model selection using Spike–and–Slab LASSO prior;
- optimality of the SEP likelihood (posterior consistency);
- several applications in statistics, economics and finance.

Further research directions

- for $p = 1, 2, \ldots,$ we have a sequence of (conditional) quantile measures;
- relevant regressors depend on $\tau \in (0, 1)$ as well as on $p = 1, 2, \ldots;$
- possible solution: $p \sim \mathcal{DP}(\alpha_0, G)$, where $G \sim \mathcal{P}(\lambda)$.

Thank you for your kind attention!

References I

- Aigner, D. J., Amemiya, T. and Poirier, D. J. (1976). On the estimation of production frontiers: Maximum Likelihood Estimation of the parameters of a discontinuous density function. *Internat. Econom. Rev.*, **17**, 377-396.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, **9**(3), 203–228.
- Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quant. Finance*, **15**(5), 725-733.
- Bernardi, M., Gayraud, G. and Petrella, L. (2015). Bayesian tail risk interdependence using quantile regression.*Bayesian Anal.*, **10**(3), 553-603.
- Bernardi, M., Gayraud, G. and Petrella, L. (2015). Posterior rate for marginal and conditional quantiles under asymmetric Laplace distribution misspecification.*Working Paper*.

References II

- Bernardi M, Casarin R. Maillet, B. and Petrella L., (2016). Dynamic Model Averaging for Bayesian Quantile Regression. *arXiv:1602.00856*.
- Bernardi M., (2016). Quantile model selection using dirac spike and slab L_1 prior. *Working Paper*.
- Bernardi M., Bottone M. and Petrella L., (2016). Bayesian Robust Quantile Regression. *arXiv:1605.05602*.
- Chen, Z. (2001). Conditional L_p-quantiles and their application to the testing of symmetry in non-parametric regression. *Statist. Probab. Lett*, **29**, 107–115.
- Delbaen, F., Bellini, F., Bignozzi, V., Ziegel, J. F. (2015). On convex risk measures with the CxLS property. *Finance Stoch.*, forthcoming.

References III

- Gerlach, R. H., Chen W. S. and Lin, L. (2012). Bayesian semi-parametric expected shortfall forecasting in financial markets. *Business Analytics Working Paper Series*.
- Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.*, **106**(494), 746–762.
- Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.
- Koenker, P., (2005). *Quantile Regression*. Cambridge University Press, Cambridge.
- Koenker, R. and Basset, G. (1978). Regression Quantiles. *Econometrica*, **46**, 33-50.
- Stock, J. and Watson, M., (1999). Forecasting inflation. *Journal of Monetary Economics*, **44**, pp. 293–335.

References IV

- Stock, J. and Watson, M., (2007). Why has U.S. inflation become harder to forecast? *Journal of Monetary Credit and Banking*, **39**, pp. 3–33.
- Yu, K. and Moyeed, R. A. (2001). Bayesian Quantile Regression. *Statist. Probab. Lett*, **54**, 437–447.
- Ziegel, J. (2014). Coherence and elicitability. *Math. Finance*, forthcoming.
- Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *J. Econometrics*, **148**(1), 86–99.