Bayesian inference for quantile autoregression and M-quantile regression

Keming Yu

Brunel University, UK

BNPW01, August, 2007, at Isaac Newton

- Quantile autoregression (QAR)?

- Why Bayes? and How?

- M-quantile regression?

- Why Bayes? and How?

1. Dynamic quantile regression.

- A limited number of dynamic quantiles models in the literature.

- CaViaR [Engle, Manganelli (2004)];

  QAR [Koenker, Xiao (2006)]

- CaViaR:

  $$Q_t(u) = \beta_0(u) + \beta_1(u) Q_{t-1}(u) + \beta_2(u)|y_{t-1}|.$$

- QAR: From classical AR(1):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t,$$

with iid errors: ϵ_t, $t = 1, ..., T$, implies

$$E(y_t|.) = \alpha_0 + \alpha_1 y_{t-1}$$

and conditional quantile functions are all parallel:

$$Q_t(u|.) = \alpha_0 + \alpha_1(u) y_{t-1}.$$

But the noise term ϵ_t may not be well–behaved, so let α_0 depend on u or a random coefficient interpretation

$$Q_t(u|.) = \alpha_0(u) + \alpha_1(u) y_{t-1}.$$

Then we generate responses from the model by

$$y_t = \alpha_0(u) + \alpha_1(u) y_{t-1}, \;\; u \sim iidU[0, 1].$$

- an important issue with QAR(p) is that, if the function $\beta_t(u) = \alpha_0(u) + \alpha_1(u)y_{t-1} + ... + \alpha_p(u)y_{t-p}$ is strictly increasing in u, then $\beta_t(u)$ is the conditional uth quantile of y_t.

- when $p = 0$, $y_t = \theta_0(u_t)$, then $\theta_0(.)$ is the quantile function of y_t only when it is monotonically increasing. When the condition is violated, the model is not necessarily identifiable. For example,
$y_t = |u_t - 0.5|$ and
$z_t = u_t - 0.5\, I(u_t > 0.5)$
have identically the same distributions, but have very different $\theta_0(.)$.

- Under existing theory, the coefficient function $\alpha(\tau)$ is estimated also based on minimisation of

$$\sum_t \rho_\tau(y_t - \alpha_0 - \alpha_1 y_{t-1} - \ldots - \alpha_p y_{t-p}),$$

where the *check function*

$$\rho_\tau(u) = u\,(\tau - I(u < 0)). \qquad (1)$$

- But the fitting method or algorithm doesn't guarantee the monotonicity, nor the identifiability.

- Some interesting discussion about this model and its extension can be found from Fan and Fan (2006), Keith (2006), Hallin and Werker (2006), Hafner and Linton (2006) and Robinson (2006)

2. Bayesian inference QAR

- Quantile regression:

- The traditional regression model can be viewed as a summary of all the quantile effects; that is, $\int Q(\theta|x)d\theta = E(y|x)$, where $Q(\theta|x)$ stands for the quantile regression between y and x. Under this interpretation, traditional analysis loses information due to its aggregation of possibly disparate quantile effects.

- Quantile regression, including expectile regression and M-quantile regression, is concerned with mistakes due to summarizing potentially disparate quantile effects into a single, potentially misleading, representation of the way y and x are related.

- Given regression model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \epsilon$, the τth *quantile* $(0 < \tau < 1)$ is defined as any solution, $\widehat{\boldsymbol{\beta}}(\tau)$, to the *quantile regression* minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_t \rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}). \qquad (2)$$

- It can be easily shown that the minimization of the loss function (1) is exactly equivalent to the maximization of a likelihood function formed by combining independently distributed asymmetric Laplace densities given by

$$f_\tau(u) = \tau\,(1-\tau)\,\exp\{-\rho_\tau(u)\}, \qquad (3)$$

  for $0 < \tau < 1$, where $\rho_p(u)$ as defined in (1).

  One could also incorporate location and scale parameters μ and σ, respectively, in the density in (3) to obtain

$$f_\tau(u;\mu,\sigma) = \frac{\tau\,(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{u-\mu}{\sigma}\right)\right\}.$$

Given the observations, $\boldsymbol{y} = (y_1, \ldots, y_n)$, the posterior distribution of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\beta})\, \pi(\boldsymbol{\beta}), \tag{4}$$

where $\pi(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$ and $L(\boldsymbol{y}|\boldsymbol{\beta})$ is the likelihood function written as

$$L(\boldsymbol{y}|\boldsymbol{\beta}) = \tau^n (1-\tau)^n \exp\left\{ -\sum_i \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\} \tag{5}$$

using (3) with a location parameter $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta}$.

- Bayesian nonparametric quantile regression:

- based on nonparametric scale mixture of a class of probability densities $f_\tau(\sigma)$ whose τth quantiles are zero.

- using of a nonparametric prior typical Dirichlet process (DP) prior for the mixing distribution.

- This will result in the nonparametric density

  $$g_\tau(\epsilon; G) = \int f_\tau(\epsilon; \sigma)\, dG(\sigma), \quad G \sim DP(\alpha G_0)$$

  whose τth quantile is also zero, and a hierarchical model for nonparametric quantile regression

  $$\begin{aligned} Y_i|\sigma_i &\sim g_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}; \sigma_i), \ \ i = 1, ...n, \\ \sigma_i|G &\sim G, \ \ i = 1, ..., n \\ G|\alpha, d &\sim DP(\alpha G_0), \end{aligned}$$

  where $DP(\alpha G_0)$ stands for the DP with precision parameter α and base distribution G_0. Usually, one places a gamma prior on α and a inverse gamma prior for G_0 with mean $d/(c-1)$ and gamma prior for d too.

- ALD defined in (3) could be a natural choice for $f_\tau(\epsilon;\sigma)$.
  Kottas and Krnjajić (2007) suggested a class of uniform densities:

  $$f_\tau(x) = \frac{\tau}{\sigma_1} I(-\sigma_1 < x < 0) + \frac{(1-\tau)}{\sigma_2} I(0 \leq x < \sigma_2).$$

  We may suggest a non-uniform class of family distribution densities instead:

  $$f_\tau(\epsilon|\sigma_1,\sigma_2) = \begin{cases} \frac{\tau}{\sigma_1}\exp(\frac{1}{\sigma_1}\epsilon), & \epsilon < 0 \\ \frac{1-\tau}{\sigma_2}\exp(-\frac{1}{\sigma_2}x), & \epsilon \geq 0, \end{cases} \qquad (6)$$

  where parameters $\sigma_1 > 0, \sigma_2 > 0$.

- : Bayesian QAR:

- Recall QAR

$$y_t = \alpha_0(\tau) + \alpha_1(\tau)y_{t-1} + ... + \alpha_p(\tau)y_{t-p}.$$

- We aim at inference $\alpha_i(\tau)$ to make $\alpha_0(\tau) + \alpha_1(\tau)y_{t-1} + ... + \alpha_p(\tau)y_{t-p}$ is a (conditional) increasing function of τ.

- Given the observations $\{y_t\}_{t=1}^{n}$, suppose all $y_t \geq 0$, otherwise, use $y_t - a$ instead, where a is the lowest boundary of the sample set.

- Suppose we want to estimate a set of quantiles such as those based on
  $\tau = ($
  $0.03, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.97)$.
  In general, we define a fine grid of τ-values in the internal (0, 1):
  $0 < \tau_1 < \tau_2 < ...\tau_K < 1$,
  then derive strictly increasing unknown coefficient $\alpha_i(\tau)$ $(i = 0, 1, ..., p)$ under Bayesian setting.

- There may be different ways to achieve this. For example, for $j = 0, 1, ..., p$, let $\boldsymbol{\beta}_j = (\alpha_0, \alpha_j(\tau_1), ..., \alpha_j(\tau_K))^T$. Given the initial prior $pr^I(\boldsymbol{\beta}_j)$ for $\boldsymbol{\beta}_j$, let us implement a truncated prior for $\boldsymbol{\beta}_j$ to be $pr(\boldsymbol{\beta}_j) \propto pr^I(\boldsymbol{\beta}_j) \prod_{i=2}^{K} I(\alpha_j(\tau_i) > \alpha(\tau_{i-1}))$.

- For example, if $pr^I(\boldsymbol{\beta}_j)$ is a Gaussian distribution, then we could adopt the method of Robert (1995) or that of Damien and Walker (2001) and run an extra (short) single move Gibbs sampler in each MCMC iteration.

3. M-quantile regression

- Include Expectile regression:

- Basic motivation behind expectile regression (Newey & Powell, 1987; Efron, 1991) and M-quantile regression (Breckling & Chambers, 1988; Chambers and Tzavidis, 2006) is to characterise the relationship between a response variable and explanatory variables when the behaviour of "non-average" individuals is of interest

- applications include measurement of production performance in economics and small area estimation in social sciences.

- Expectile regression $y_t = \mu(\boldsymbol{x}_t) + \epsilon_t$ can be fitted by minimizing an asymmetrically weighted least-squares criterion

$$R(\mu) = \sum_{i=1}^{n} \tau(y_i - \mu)_+^2 + (1-\tau)(y_i - \mu)_-^2,$$

where $0 < \tau < 1$ and ψ_+ and ψ_- denote the positive and negative parts of ψ.

- An expectile regression can be derived while the observations are assumed to be independently drawn from an asymmetric normal distribution
$f_\tau(u) = \frac{\sqrt{1-\tau}+\sqrt{\tau}}{4\sigma\sqrt{\pi\tau(1-\tau)}} \exp(-\sigma^{-2}|\tau - I(y \leq \mu)|(y-\mu)^2)$.

- so that parametric/nonparametric Bayesian expectile regression can be started from here.

- including inference expectile order τ which is useful in practice.

- Now extend the asymmetrical least-squares to get M-quantile regression.

- M-quantile regression: link to M-estimation.

- given an influence function such as Huber Proposal 2, $\psi(u) = uI(|u| \leq c) + c\, sgan(u)$, provided c is bounded away from zero,

- $\psi_q(u) =$

  $2\psi(s^{-1}u)\{q\, I(u > 0) + (1 - q)\, I(u \leq 0),\}$

  where $0 < q < 1$ and s is a suitable robust estimate of scale, e.g. the MAD estimates $s = median\{|u|\}/0.6745$,

- M-quantile regression is the solution of

$$\sum_{i=1}^{n} \psi_q(r_{iq\psi})\boldsymbol{x}_i = 0,$$

  where $r_{iq\psi} = y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\psi(q)$,

  and q is called M-quantile coefficient.

- A linear M-quantile regression model is one where we assumed $Q_q(\boldsymbol{x}; \psi) = \boldsymbol{x}^T \boldsymbol{\beta}_\psi(q)$. That is, we allow a different set of regression parameters for each value of q.

- M-quantile regression coefficient is an interesting and important parameter and can be used to characterise the area specification in small area estimation, the conditional profit of gene in gene expression and group specification of other applications.

- Linear interpolation which constructs estimate from a discrete set of known estimated points is used in classical estimation of M-quantile coefficient.

- The method is not very precise and the interpolant is not differentiable at the point x.

- Bayesian M-quantile regression:

- Following Huber (1981) and Stone (2005), let $\rho(.)$ be a (necessarily continuous) concave function on $(-\infty, \infty)$,the M-regression can be formulated via the log-likelihood function for robust estimation of $\mu(\boldsymbol{x})$ in the classical sense
  $l(\mu) = \sum_{i=1}^{n} \rho(y_i - \mu(\boldsymbol{x}_i))$.
  Usually, $\rho'(u) = \phi(u)$ is the influence function.
  In that context, Huber developed the asymptotic minimum variance role played by the $\rho-$function given by $\rho(u) = u^2/2$ for $|u| \leq c$ and $\rho(u) = c|u| - c^2/2$ for $u| > c$, where c is a positive constant.

- Therefore, corresponding to
$\psi_q(u) = 2\psi(s^{-1}u)\{qI(u > 0) + (1-q)I(u \leq 0)\}$,
we have

$$\rho_q(u) = 2\,\rho(\frac{u}{s})(q\,I(u > 0) + (1-q)\,I(u \leq 0)),$$

then the M-quantile regression can be formulated via the log-likelihood function for estimation of $\mu(\boldsymbol{x})$

$$l_q(\mu) = \sum_{i=1}^{n} \rho_q(y_i - \mu(\boldsymbol{x}_i)).$$

When $q = 0.5$, M-quantile reduces to M-estimation.

- To include inference for M-quantile coefficient q, consider a probability density function

$$f(x) = B(q)\,\text{exp}(-\rho_q(x)),$$

where constant $B(q)$, which is usually a function of q, satisfies

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

- For example, when the influence function $\phi(u)$ is Huber proposal 2
$\psi(u) = u\, I(|u| \leq c) + c\, sgn(u)$,

$$\frac{1}{B} = \frac{s}{2\sqrt{q}}(\gamma(\frac{1}{2}, 2c\sqrt{q}) - \gamma(\frac{1}{2}, c\sqrt{q}))$$

$$+\frac{s}{2\sqrt{1-q}}(\gamma(\frac{1}{2},\ 2c\sqrt{1-q}) - \gamma(\frac{1}{2}, c\sqrt{1-q}))$$

$$+\frac{s}{2cq}\exp(-2qc^2)+\frac{s}{2c(1-q)}\exp(-2(1-q)c^2).$$

Where

$$\gamma(\frac{1}{2}, A) = \int_0^A t^{-1/2} e^{-t}\, dt,$$

is the lower incomplete gamma function.

- Bayesian inference regression parameter $\boldsymbol{\beta}$ along with M-quantile coefficient q can be carried out by setting posterior as

$$\pi(\boldsymbol{\beta}, q) \propto \pi(\boldsymbol{\beta})\pi(q)B^n(q)\,\mathrm{exp}(-\sum_{i=1}^{n}\rho_q(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})).$$

4. Bayesian hypothesis test

Unlike classical mean regression, almost all quantile regression models, including QAR and M-quantile regression based tests are based on asymptotic theory (typical asymptotic normality and asymptotic chi-squared: these are large-sample based results under some regular conditions.) Nowadays methods developed and developing for small sample size inference seem particularly germane.

Once the Bayesian inference for the posterior distributions of some important unknown quantities of these models are in place, we can then easily move to Bayesian hypothesis test based on the Bayes factor.

$$B_{12} = \frac{\pi(M_1|\boldsymbol{y})/\pi(M_2|\boldsymbol{y})}{\pi(M_1)/\pi(M_2)}, \tag{7}$$

i.e., the ratio of posterior odds of the condition M_1 to prior odds of M_1.

SUMMARY: Benefits from Bayesina inference on these models:

- dealing with uncertainty;
- Bayesian rather than Frequentist approach;
- dealing with model identifiability;
- dealing with model choice and tests.

Thank you!

References

Angrist, J., Chernozhukov, V. and Fernndez-Val, I. (2006), Quantile regression under misspecification, with an application to the U.S. wage structure,*Econometrica,* 74, 539–563.

Buchinsky, M. (1998), Recent advances in quantile regression models, *J. Human Res.*, 33, 88–126.

Chambers, R. and Tzavidis (2006), M-quantile models for small area estimation, *Biometrika,* 93, 255–268.

Chib, S. and Greenberg, E. (2007), *Hierarchical Bayes Modeling*, Palgrave Dictionary in Economics (2nd edition), in press.

Chib, S., Jeliazkov, I. (2001), Marginal likelihood from MetropolisHastings output, *Journal of the American Statistical Association*, 96, 270–281.

Damien, P. and Walker,S.G., Sampling Truncated Normal, Beta, and Gamma Densities, *J. Computational and Graphical Statistics,* 10, 206–215.

David B. Dunson, M. Watson, Jack A. Taylor (2003) Bayesian Latent Variable Models for Median Regression on Multiple Outcomes, *Biometrics*, 59, 296–304.

Efron, B. (1991), Regression percentiles using asymmetric squared error loss, *Statistica Sinica*, 1, 93–125.

Fan, J. and Fan, Y. (2006), comments on Koenker and Xiao (2006), *J. Amer. Statist. Assoc.*, 101, 991–994.

Fernndeza, C., Koop, G. and and Steel, M. (2000), A Bayesian analysis of multiple-output production frontiers, *Journal of Econometrics,* 98, 47–79.

Geweke, J.F. (1989), Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica,*

57, 1317–1340.
Gill, J. (200), *Bayesian Methods for the Social and Behavioral Sciences,* Chapman & Hall/CRC.
Handcock, M.S. and Morris, M. (1999), *Relative Distribution Methods in the Social Sciences,* Springer-Verlag, New York.
Horowitz, J.L. and Lee, S. (2007), Nonparametric instrumental variables estimation of a quantile regression model, *Econometrica*, forthcoming.
Huber, P. J. (1981), *Robust Statistics* (Wiley, New York).
Koenker, R. and Bassett, G.S. (1978), Regression quantiles, *Econometrica*, 46, 33–50.
Kass, R.E. and Raftery, A.E. (1995), Bayes factors, *Journal of the American Statistical Association*, 90, 773–795.
Koenker, R., and Xiao, Z. (2006), Quantile autoregression, *J. Amer. Statist. Assoc.*, 101, 980–990.
Koenker R, Xiao Z (2002) Inference on the quantile regression process, *Econometrica,* 70, 1583–1612.
Koop, G. (2003), *Bayesian Econometrics,* John Wiley and Sons. Kokic, P., Chambers, R., Breckling, J. & Beare S. (1997). A measure of production performance, *Journal of Business and Economic Statistics,* 15, 445–451.
Kottas, A., and Krnjajic, M. (2005), Bayesian nonparametric modeling in quantile regression, *Technical report*.
Newey, W.K. & Powell, J.L. (1987). Asymmetric least squares estimation and testing, *Econometrica*, 55, 819–847.
Robert, C.P. (1995), Simulation of truncated normal

variables, *Statistics and Computing*, 5, 121–125.
Stone, C.J. (2005), Nonparametric M-regression with free knot splines, *J. Stat. Plan. Inference*, 130, 183–206.
Tsionas, E.G. (2003), Bayesian quantile inference, *Journal of Statistical Computation and Simulation,* 73, 659–674.
Wasserman, L. (2000), Bayesian model selection and model averaging, *Journal of Mathematical Phycology*, 44 , 92–107.
Yu, K. and Stander, J. (2007), Bayesian Analysis of a Tobit Quantile Regression Model, *J. Econometrics,* 137, 260–276.
Yu, K., Van Kerm, P. and Zhang, J. (2005), Bayesian Quantile Regression: An Application to the Wage Distribution in 1990s , *Sankhya*, 67, 1–19.
Yu, K. and Moyeed, R.A. (2001), Bayesian Quantile Regression, *Statistics and Probability Letters*, 54, 437–447.
Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics,*, John Wiley and Sons, Inc., New York.