Quantile Autoregression

Roger Koenker

University of Illinois, Urbana-Champaign

University of Copenhagen 18-20 May 2016

Based on joint work with Zhijie Xiao, Boston College.

Outline

Introduction

In classical regression and autoregression models

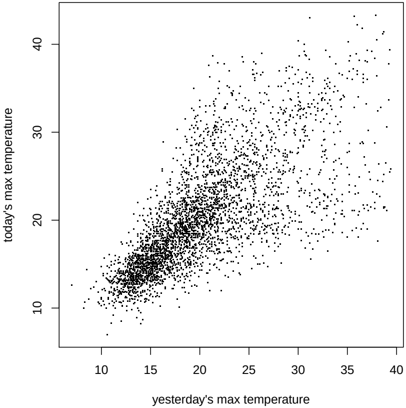
$$
\begin{aligned}
y_i &= h(x_i, \theta) + u_i, \\
y_t &= \alpha y_{t-1} + u_t
\end{aligned}
$$

conditioning covariates influence only the location of the conditional distribution of the response:

$$\text{Response} = \text{Signal} + \text{IID Noise.}$$

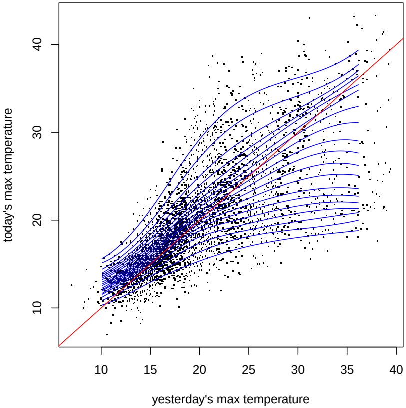
But why should noise always be so well-behaved?

A Motivating Example



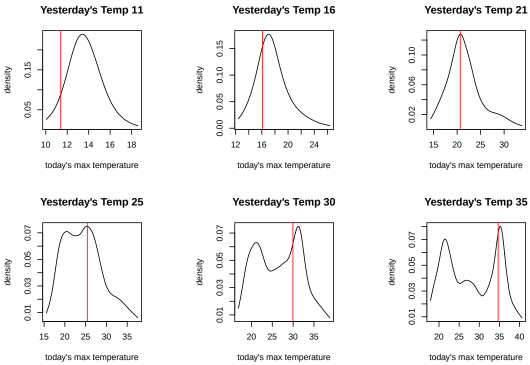
Daily Temperature in Melbourne: An AR(1) Scatterplot

Estimated Conditional Quantiles of Daily Temperature



Daily Temperature in Melbourne: A Nonlinear QAR(1) Model

Conditional Densities of Melbourne Daily Temperature



Location, scale and shape all change with y_{t-1}.
When today is hot, tomorrow's temperature is bimodal!

Linear AR(1) and QAR(1) Models

The classical linear AR(1) model

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

with iid errors, $u_t : t = 1, \cdots, T$, implies

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

and conditional quantile functions are all parallel:

$$Q_{y_t}(\tau | \mathcal{F}_{t-1}) = \alpha_0(\tau) + \alpha_1 y_{t-1}$$

with $\alpha_0(\tau) = F_u^{-1}(\tau)$ just the quantile function of the u_t's.
But isn't this rather boring? What if we let α_1 depend on τ too?

A Random Coefficient Interpretation

If the conditional quantiles of the response satisfy:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau) y_{t-1}$$

then we can generate responses from the model by replacing τ by uniform random variables:

$$y_t = \alpha_0(u_t) + \alpha_1(u_t) y_{t-1} \quad u_t \sim \text{iid } U[0,1].$$

This is a very special form of random coefficient autoregressive (RCAR) model with comonotonic coefficients.

On Comonotonicity

Definition: Two random variables $X, Y : \Omega \to \mathbb{R}$ are comonotonic if there exists a third random variable $Z : \Omega \to \mathbb{R}$ and increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$.

- If X and Y are comonotonic they have rank correlation one.
- From our point of view the crucial property of comonotonic random variables is the behavior of quantile functions of their sums, X, Y comonotonic implies:

$$F_{X+Y}^{-1}(\tau) = F_X^{-1}(\tau) + F_Y^{-1}(\tau)$$

- X and Y are driven by the same random (uniform) variable.

The QAR(p) Model

Consider a p-th order QAR process,

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau)y_{t-1} + \ldots + \alpha_p(\tau)y_{t-p}$$

Equivalently, we have random coefficient model,

$$\begin{aligned} y_t &= \alpha_0(u_t) + \alpha_1(u_t)y_{t-1} + \cdots + \alpha_p(u_t)y_{t-p} \\ &\equiv x_t^\top \alpha(u_t). \end{aligned}$$

Now, all $p+1$ random coefficients are comonotonic, functionally dependent on the same uniform random variable.

Vector QAR(1) representation of the QAR(p) Model

$$Y_t = \mu + A_t Y_{t-1} + V_t$$

where

$$\mu = \begin{bmatrix} \mu_0 \\ 0_{p-1} \end{bmatrix}, \; A_t = \begin{bmatrix} a_t & \alpha_p(u_t) \\ I_{p-1} & 0_{p-1} \end{bmatrix}, \; V_t = \begin{bmatrix} \nu_t \\ 0_{p-1} \end{bmatrix}$$

$$a_t = [\alpha_1(u_t), \ldots, \alpha_{p-1}(u_t)],$$

$$Y_t = [y_t, \cdots, y_{t-p+1}]^\top,$$

$$\nu_t = \alpha_0(u_t) - \mu_0.$$

It all looks rather complex and multivariate, but it is really still nicely univariate and very tractable.

Slouching Toward Asymptopia

We maintain the following regularity conditions:

A.1 $\{v_t\}$ are iid with mean 0 and variance $\sigma^2 < \infty$. The CDF of v_t, F, has a continuous density f with $f(v) > 0$ on $\mathcal{V} = \{v : 0 < F(v) < 1\}$.

A.2 Eigenvalues of $\Omega_A = E(A_t \otimes A_t)$ have moduli less than unity.

A.3 Denote the conditional CDF $\Pr[y_t < y|\mathcal{F}_{t-1}]$ as $F_{t-1}(y)$ and its derivative as $f_{t-1}(y)$, f_{t-1} is uniformly integrable on $\mathcal{V}$.

Stationarity

Theorem 1: Under assumptions A.1 and A.2, the QAR(p) process y_t is covariance stationary and satisfies a central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t - \mu_y) \Rightarrow N\left(0, \omega_y^2\right),$$

with

$$\begin{aligned}
\mu_y &= \frac{\mu_0}{1 - \sum_{j=1}^{p} \mu_p}, \\
\mu_j &= E(\alpha_j(u_t)), \quad j = 0, ..., p, \\
\omega_y^2 &= \lim \frac{1}{n} E[\sum_{t=1}^{n} (y_t - \mu_y)]^2.
\end{aligned}$$

Example: The QAR(1) Model

For the QAR(1) model,

$$Q_{y_t}(\tau | y_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau) y_{t-1},$$

or with u_t iid $U[0,1]$.

$$y_t = \alpha_0(u_t) + \alpha_1(u_t) y_{t-1},$$

if $\omega^2 = E(\alpha_1^2(u_t)) < 1$, then y_t is covariance stationary and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (y_t - \mu_y) \Rightarrow N\left(0, \omega_y^2\right),$$

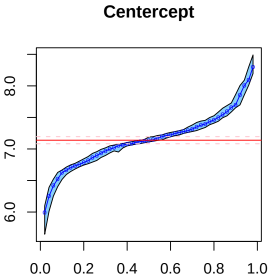
where $\mu_0 = E\alpha_0(u_t)$, $\mu_1 = E(\alpha_1(u_t)$, $\sigma^2 = V(\alpha_0(u_t))$, and

$$\mu_y = \frac{\mu_0}{(1 - \mu_1)}, \quad \omega_y^2 = \frac{(1 + \mu_1)\sigma^2}{(1 - \mu_1)(1 - \omega^2)},$$

Qualitative Behavior of QAR(p) Processes

- The model can exhibit unit-root-like tendencies, even temporarily explosive behavior, but episodes of mean reversion are sufficient to insure stationarity.
- Under certain conditions,the QAR(p) process is a semi-strong ARCH(p) process in the sense of Drost and Nijman (1993).
- The impulse response of y_{t+s} to a shock u_t is stochastic but converges (to zero) in mean square as $s \to \infty$.

Estimated QAR(1) v. AR(1) Models of U.S. Interest Rates



Data: Seasonally adjusted monthly: April, 1971 to June, 2002.
Do 3-month T-bills really have a unit root?

Estimation of the QAR model

Estimation of the QAR models involves solving,

$$\hat{\alpha}(\tau) = \text{argmin}_{\alpha} \sum_{t=1}^{n} \rho_\tau(y_t - x_t^\top \alpha),$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$, the $\surd$-function.
Fitted conditional quantile functions of y_t, are given by,

$$\hat{Q}_t(\tau | x_t) = x_t^\top \hat{\alpha}(\tau),$$

and conditional densities by the difference quotients,

$$\hat{f}_t(\tau | x_{t-1}) = \frac{2h}{\hat{Q}_t(\tau + h | x_{t-1}) - \hat{Q}_t(\tau - h | x_{t-1})},$$

The QAR Process

Theorem 2: Under our regularity conditions,

$$\sqrt{n}\Omega^{-1/2}(\hat{\alpha}(\tau) - \alpha(\tau)) \Rightarrow B_{p+1}(\tau),$$

a $(p+1)$-dimensional standard Brownian Bridge, with

$$\begin{aligned}
\Omega &= \Omega_1^{-1}\Omega_0\Omega_1^{-1}. \\
\Omega_0 &= E(x_t x_t^\top) = \lim n^{-1}\sum_{t=1}^{n} x_t x_t^\top, \\
\Omega_1 &= \lim n^{-1}\sum_{t=1}^{n} f_{t-1}(F_{t-1}^{-1}(\tau))x_t x_t^\top.
\end{aligned}$$

Inference for QAR models

For fixed $\tau = \tau_0$ we can test the hypothesis:

$$H_0 : \quad R\alpha(\tau) = r$$

using the Wald statistic,

$$W_n(\tau) = \frac{n(R\hat{\alpha}(\tau) - r)^\top [R\hat{\Omega}_1^{-1}\hat{\Omega}_0\hat{\Omega}_1^{-1}R^\top]^{-1}(R\hat{\alpha}(\tau) - r)}{\tau(1-\tau)}$$

This approach can be extended to testing on general index sets $\tau \in \mathcal{T}$ with the corresponding Wald process.

Asymptotic Inference

Theorem: Under H_0, $W_n(\tau) \Rightarrow Q_m^2(\tau)$, where $Q_m(\tau)$ is a Bessel process of order $m = \text{rank}(R)$. For fixed τ, $Q_m^2(\tau) \sim \chi_m^2$.

- Kolmogorov-Smirov or Cramer-von-Mises statistics based on $W_n(\tau)$ can be used to implement the tests.
- For known R and r this leads to a very nice theory – estimated R and/or r testing raises new questions.
- The situation is quite analogous to goodness-of-fit testing with estimated parameters.

Example: Unit Root Testing

Consider the augmented Dickey-Fuller model

$$y_t = \delta_0 + \delta_1 y_{t-1} + \sum_{j=2}^{p} \delta_j \Delta y_{t-j} + u_t.$$

We would like to test this constant coefficients version of the model against the more general QAR(p) version:

$$Q_{y_t}(\tau|x_t) = \delta_0(\tau) + \delta_1(\tau) y_{t-1} + \sum_{j=2}^{p} \delta_j(\tau) \Delta y_{t-j}$$

The hypothesis: $H_0 : \delta_1(\tau) = \bar{\delta}_1 = 1$, for $\tau \in \mathcal{T} = [\tau_0, 1 - \tau_0]$, is considered in Koenker and Xiao (JASA, 2004).

Example: Two Tests

- When $\bar{\delta}_1 < 1$ is known we have the candidate process,

$$V_n(\tau) = \sqrt{n}(\hat{\delta}_1(\tau) - \bar{\delta}_1)/\hat{\omega}_{11}.$$

where $\hat{\omega}_{11}^2$ is the appropriate element from $\hat{\Omega}_1^{-1}\hat{\Omega}_0\hat{\Omega}_1^{-1}$. Fluctuations in $V_n(\tau)$ can be evaluated with the Kolmogorov-Smirnov statistic,

$$\sup_{\tau \in \mathcal{T}} V_n(\tau) \Rightarrow \sup_{\tau \in \mathcal{T}} B(\tau).$$

- When $\bar{\delta}_1$ is unknown we may replace it with an estimate, but this disrupts the convenient asymptotic behavior. Now,

$$\hat{V}_n(\tau) = \sqrt{n}((\hat{\delta}_1(\tau) - \bar{\delta}_1) - (\hat{\delta}_1 - \bar{\delta}_1))/\hat{\omega}_{11}$$

Martingale Transformation of $\hat{V}_n(\tau)$

Khmaladze (1981) suggested a general approach to the transformation of parametric empirical processes like $\hat{V}_n(\tau)$:

$$\widetilde{V}_n(\tau) \;=\; \hat{V}_n(\tau) - \int_0^{\tau} \left[\dot{g}_n(s)^{\top} C_n^{-1}(s) \int_s^1 \dot{g}_n(r) d\hat{V}_n(r) \right] ds$$

where $\dot{g}_n(s)$ and $C_n(s)$ are estimators of

$$\dot{g}(r) \;=\; (1, (\dot{f}/f)(F^{-1}(r)))^{\top};\; C(s) = \int_s^1 \dot{g}(r)\dot{g}(r)^{\top} dr.$$

This is a generalization of the classical Doob-Meyer decomposition.

Restoration of the ADF property

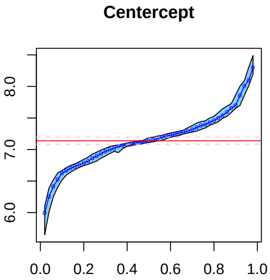
Theorem Under H_0, $\tilde{V}_n(\tau) \Rightarrow W(\tau)$ and therefore

$$\sup_{\tau \in \mathcal{T}} \|\tilde{V}_n(\tau)\| \Rightarrow \sup_{\tau \in \mathcal{T}} \|W(\tau)\|,$$

with $W(r)$ a standard Brownian motion.

- The martingale transformation of Khmaladze annihilates the contribution of the estimated parameters to the asymptotic behavior of the $\hat{V}_n(\tau)$ process, thereby restoring the asymptotically distribution free (ADF) character of the test.

Three Month T-Bills Again



A test of the “location-shift” hypothesis yields a test statistic of 2.76 which has a p-value of roughly 0.01, contradicting the conclusion of the conventional Dickey-Fuller test.

QAR Models for Longitudinal Data

- In estimating growth curves it is often valuable to condition not only on age, but also on prior growth and possibly on other covariates.
- Autoregressive models are natural, but complicated due to the irregular spacing of typical longitudinal measurements.
- Finnish Height Data: $\{Y_i(t_{i,j}) : \ j = 1, \dots, J_i, \ i = 1, \dots, n.\}$
- Partially Linear Model [Pere, Wei, Koenker, and He (2006)]:

$$\begin{aligned} Q_{Y_i(t_{i,j})}(\tau \quad | \quad & t_{i,j}, Y_i(t_{i,j-1}), x_i) = g_\tau(t_{i,j}) \\ + \quad & [\alpha(\tau) + \beta(\tau)(t_{i,j} - t_{i,j-1})]Y_i(t_{i,j-1}) + x_i^\top \gamma(\tau). \end{aligned}$$

Parametric Components of the Conditional Growth Model

τ	Boys			Girls		
	$\hat{\alpha}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\gamma}(\tau)$	$\hat{\alpha}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\gamma}(\tau)$
0.03	0.845 (0.020)	0.147 (0.011)	0.024 (0.011)	0.809 (0.024)	0.135 (0.011)	0.042 (0.010)
0.1	0.787 (0.020)	0.159 (0.007)	0.036 (0.007)	0.757 (0.022)	0.153 (0.007)	0.054 (0.009)
0.25	0.725 (0.019)	0.170 (0.006)	0.051 (0.009)	0.685 (0.021)	0.163 (0.006)	0.061 (0.008)
0.5	0.635 (0.025)	0.173 (0.009)	0.060 (0.013)	0.612 (0.027)	0.175 (0.008)	0.070 (0.009)
0.75	0.483 (0.029)	0.187 (0.009)	0.063 (0.017)	0.457 (0.027)	0.183 (0.012)	0.094 (0.015)
0.9	0.422 (0.024)	0.213 (0.016)	0.070 (0.017)	0.411 (0.030)	0.201 (0.015)	0.100 (0.018)
0.97	0.383 (0.024)	0.214 (0.016)	0.077 (0.018)	0.400 (0.038)	0.232 (0.024)	0.086 (0.027)

Estimates of the QAR(1) parameters, $\alpha(\tau)$ and $\beta(\tau)$ and the mid-parental height effect, $\gamma(\tau)$, for Finnish children ages 0 to 2 years.

Forecasting with QAR Models

Given an estimated QAR model,

$$\hat{Q}_{y_t}(\tau|\mathcal{F}_{t-1}) = x_t^\top \hat{\alpha}(\tau)$$

based on data: $y_t : \ t = 1, 2, \cdots, T$, we can forecast

$$\hat{y}_{T+s} = \tilde{x}_{T+s}^\top \hat{\alpha}(U_s), \ s = 1, \cdots, S,$$

where $\tilde{x}_{T+s} = [1, \tilde{y}_{T+s-1}, \cdots, \tilde{y}_{T+s-p}]^\top$, $U_s \sim U[0, 1]$, and

$$\tilde{y}_t = \begin{cases} y_t & \text{if} \quad t \leqslant T, \\ \hat{y}_t & \text{if} \quad t > T. \end{cases}$$

Conditional density forecasts can be made based on an ensemble of such forecast paths.

Linear QAR Models May Pose Statistical Health Risks

- Lines with distinct slopes eventually intersect. [Euclid: P5]
- Quantile functions, $Q_Y(\tau|x)$ should be monotone in τ for all x, intersections imply point masses – or even worse.
- What is to be done?
  - Constrained QAR: Quantiles can be estimated simultaneously subject to linear inequality restrictions.
  - Nonlinear QAR: Abandon linearity in the lagged y_t's, as in the Melbourne temperature example, both parametric and nonparametric options are available.

Nonlinear QAR Models via Copulas

An interesting class of stationary, Markovian models can be expressed in terms of their copula functions:

$$G(y_t, y_{t-1}, \cdots, y_{y-p}) = C(F(y_t), F(y_{t-1}), \cdots, F(y_{y-p}))$$

where G is the joint df and F the common marginal df.

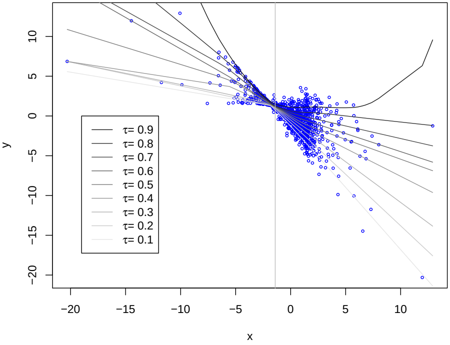
- Differentiating, $C(u, v)$, with respect to u, gives the conditional df,

$$H(y_t | y_{t-1}) = \frac{\partial}{\partial u} C(u, v)|_{(u=F(y_t), v=F(y_{t-1}))}$$

- Inverting we have the conditional quantile functions,

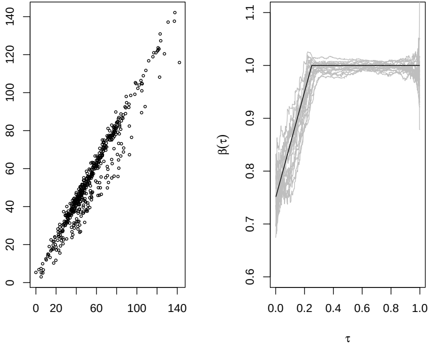
$$Q_{y_t}(\tau | y_{t-1}) = h(y_{t-1}, \theta(\tau))$$

Example 1 (Fan and Fan)



Model: $Q_{y_t}(\tau|y_{t-1}) = -(1.7 - 1.8\tau)y_{t-1} + \Phi^{-1}(\tau)$.

Example 2 (Near Unit Root)



Model: $Q_{y_t}(\tau|y_{t-1}) = 2 + \min\{\frac{3}{4} + \tau, 1\} y_{t-1} + 3\Phi^{-1}(\tau)$.

Conclusions

- QAR models are an attempt to expand the scope of classical linear time-series models permitting lagged covariates to influence scale and shape as well as location of conditional densities.
- Efficient estimation via familiar linear programming methods.
- Random coefficient interpretation nests many conventional models including ARCH.
- Wald-type inference is feasible for a large class of hypotheses; rank based inference is also an attractive option.
- Forecasting conditional densities is potentially valuable.
- Many new and challenging open problems. . . .