Functional Principle Component Analysis for Generalized Quantile Regression

Mengmeng Guo　　Lan Zhou
Wolfgang Karl Härdle　　Jianhua Huang

Ladislaus von Bortkiewicz Chair
of Statistics Humboldt-Universität
zu Berlin
Department of Statistics Texas
A&M Univeristy
lvb.wiwi.hu-berlin.de
www.stat.tamu.edu

Generalized Quantiles

- Quantiles and Expectiles are “generalized quantiles", Jones (1994).
- Capture the tail behaviour of conditional distributions.
- Applications in finance, weather, demography, $\cdots$
- Some applications involve MANY GQR curves.

Estimation Method

- Kernel Smoothing
  - Quantile: Fan et.al (1994)
  - Expectile: Zhang (1994)
- Penalized Spline Smoothing
  - Quantile: Koenker et.al (1994)
  - Expectile: Schnabel and Eilers (2009)

Both can be estimated by least asymmetric weighted squares (LAWS)

Method

- Traditional: estimate individually
- Directly: estimate all parameters together

- not all information applied
- too many parameters, curse of dimensionality

Functional Principle Component Analysis(FPCA)

- a common tool to capture random curves
- dimension reduction
- meaningful interpretation of principle components
- apply FPCA and LAWS to estimate GQR curves

Figure 1: Estimated 95% expectile curves for the volatility of temperature of 30 cities in Germany from 1995-2007.

Outline

1. Motivation ✓
2. Generalized Quantile Estimation
3. FPCA for Generalized Quantile Regression Curve
4. Simulation
5. Application
6. Conclusion

Quantile and Expectile

Quantile

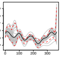
$$F_Y(l) = \int_{-\infty}^{l} dF(y) = \tau$$

$$l = F_Y^{-1}(\tau)$$

Expectile

$$G_Y(l) = \frac{\int_{-\infty}^{l} |Y - l| \, dF(y)}{\int_{-\infty}^{\infty} |Y - l| \, dF(y)} = \tau$$

$$l = G_Y^{-1}(\tau)$$



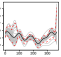
Loss Function

- Square loss $L(Y, \theta) = (Y - \theta)^2$
- Absolute value loss $L(Y, \theta) = |Y - \theta|$

Asymmetric loss function for generalized quantiles:

$$\rho_\tau(u) = |\mathbf{I}(u \leq 0) - \tau||u|^\alpha, \qquad \tau \in (0, 1) \tag{1}$$

with $\alpha \in \{1, 2\}$ and $u = Y - \theta$.



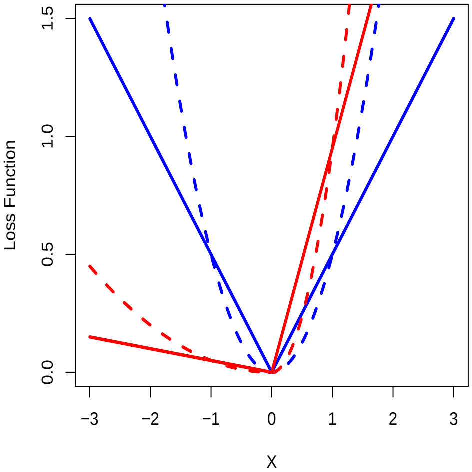


Figure 2: Loss functions for $\tau = 0.9$(red); $\tau = 0.5$(blue); $\alpha = 1$ (slide line); $\alpha = 2$ (dashed line).

Minimum contrast approach:

$$
\begin{aligned}
l_\tau &= \arg\min_\theta \; \mathsf{E}\{\rho_\tau(Y-\theta)\} \\
&= \arg\min_\theta \; (1-\tau)\int_{-\infty}^{\theta} |y-\theta|^\alpha dF(y) + \tau \int_{\theta}^{\infty} |y-\theta|^\alpha dF(y)
\end{aligned}
$$

generalized quantile regression curve:

$$
\begin{aligned}
l_t &= \arg\min_\theta \; \mathsf{E}\{\rho_\tau(Y-\theta)|X=t\} \\
&= \arg\min_\theta \; (1-\tau)\int_{-\infty}^{\theta} |y-\theta|^\alpha dF(y|t) + \tau \int_{\theta}^{\infty} |y-\theta|^\alpha dF(y|t)
\end{aligned}
$$

Single Curve Estimation

$$Y_t = l(t) + \varepsilon_t \tag{2}$$

approximate $l(\cdot)$ by a spline basis:

$$l(t) = b(t)^\top \theta_\mu \tag{3}$$

where $b(t) = \{b_1(t), \cdots, b_q(t)\}^\top$ is a vector of basis functions and θ_μ is a vector with dimension q.

Estimation

Employ a roughness penalty:

$$S(\theta_\mu) = \sum_{t=1}^{T} w_t \{Y_t - b(t)^\top \theta_\mu\}^2 + \lambda \{\theta_\mu^\top \int \ddot{b}(t)\ddot{b}(t)^\top dt \; \theta_\mu\} \quad (4)$$

where $Y = (Y_1, Y_2, \cdots, Y_T)^\top$, $\ddot{b}(t) = \frac{\partial b^2(t)}{\partial t^2}$ and w_t is the t-th element of the weight matrix defined in (6).

Weight

For expectile:

$$w_t = \begin{cases} \tau & \text{if} \quad Y_t > l(t) \\ 1 - \tau & \text{if} \quad Y_t \leq l(t) \end{cases} \tag{5}$$

and for quantile:

$$w_t = \begin{cases} \frac{\tau}{|Y_t - l(t)|} & \text{if} \quad Y_t > l(t) \\ \frac{1-\tau}{|Y_t - l(t)|} & \text{if} \quad Y_t < l(t) \\ \frac{1-\tau}{|Y_t - l(t)| + \delta} & \text{if} \quad Y_t = l(t) \end{cases} \tag{6}$$

Therefore, quantiles can be calculated by LAWS.

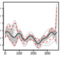
Estimation

The generalized quantile curve:

$$
\begin{aligned}
\widehat{\theta}_\mu &= \arg\min_{\theta_\mu} S(\theta_\mu) \\
&= \{B^\top W B + \lambda \int \ddot{b}(t)\ddot{b}(t)^\top dt\}^{-1}(B^\top W Y)
\end{aligned}
$$

$B = \{b(t)\}$ is the spline basis matrix with dimension $T \times q$, and W defined in (6):

$$\widehat{l}(t) = b(t)\widehat{\theta}_\mu \tag{7}$$



Mixed effect Model

Observe $i = 1, \cdots, N$ individual curves:

$$l_i(t) = \mu(t) + v_i(t) \tag{8}$$

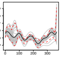
- $\mu(t)$ mean function,
- $v_i(t)$ departure from $\mu(t)$.

Approximate via

$$l_{ij} = l(t_{ij}) = b(t_{ij})^\top \theta_\mu + b(t_{ij})^\top \gamma_{ij} \tag{9}$$

where $i = 1, \cdots, N$ and $j = 1, \cdots, T_i$.

- Too many parameters to estimate.
- Very volatile when sparse data exists, James et.al (2000).



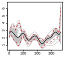
Reduced Model

$$l_i(t) = \mu(t) + \sum_{k=1}^{K} f_k(t)^\top \alpha_{ik} \tag{10}$$

- $\mu(t)$ the overall mean
- f_k *k-th* principle component functions with $f(t) = \{f_1(t), \cdots, f_K(t)\}^\top$
- $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{iK})$ principle component scores.

Represent μ and f by a basis of spline functions:

$$\begin{aligned} \mu(t) &= b(t)^\top \theta_\mu \\ f(t)^\top &= b(t)^\top \Theta_f \end{aligned}$$



$$l_{ij} = l(t_{ij}) = b(t_{ij})^\top \theta_\mu + b(t_{ij})^\top \Theta_f \alpha_i \tag{11}$$

Further, define $L_i = \{l_i(t_1), \cdots, l_i(T_i)\}^\top$,
$B_i = \{b(t_1), \cdots, b(T_i)\}^\top$ is the basis matrix with dimension $T_i \times q$.
The GQR curves in a matrix form:

$$L_i = B_i \theta_\mu + B_i \Theta_f \alpha_i \tag{12}$$

Then the model is transferred as

$$Y_i = L_i + \varepsilon_i = B_i \theta_\mu + B_i \Theta_f \alpha_i + \varepsilon_i \tag{13}$$

Constraints

$$\begin{aligned} \Theta_f^\top \Theta_f &= \mathrm{I}_K \\ \int b(t)^\top b(t) dt &= \mathrm{I}_q \end{aligned}$$

which satisfies the usual orthogonality requirements of the principle component curves:

$$\int f(t) f(t)^\top dt = \Theta_f^\top \int b(t)^\top b(t) dt \; \Theta_f = \mathrm{I}_K$$

Empirical Loss Function

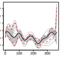
$$S = \sum_{i=1}^{N} \sum_{j=1}^{T_i} w_{ij} \{ Y_{ij} - b(t_j)^\top \theta_\mu - b(t_j)^\top \Theta_f \alpha_i \}^2 \tag{14}$$

where

$$w_{ij} = \begin{cases} \tau & \text{if} \quad Y_{ij} > l_{ij} \\ 1 - \tau & \text{if} \quad Y_{ij} \leq l_{ij} \end{cases} \tag{15}$$

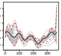
A roughness penalty applied to both mean and departure curve:

$$\begin{aligned} M_\mu &= \theta_\mu^\top \int \ddot{b}(t) \ddot{b}(t)^\top dt \; \theta_\mu \\ M_f &= \sum_{k=1}^{K} \theta_{kf}^\top \int \ddot{b}(t) \ddot{b}(t)^\top dt \; \theta_{kf} \end{aligned}$$



Asymmetric Least Square

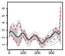
$$
\begin{aligned}
S^* &= S + \lambda_\mu M_\mu + \lambda_f M_f \\
&= \sum_{i=1}^{N} (Y_i - B_i\theta_\mu - B_i\Theta_f\alpha_i)^\top W_i (Y_i - B_i\theta_\mu - B_i\Theta_f\alpha_i) \\
&\quad + \lambda_\mu \{\theta_\mu^\top \int \ddot{b}(t)\ddot{b}(t)^\top dt \; \theta_\mu\} \\
&\quad + \lambda_f \{\sum_{k=1}^{K} \theta_{kf}^\top \int \ddot{b}(t)\ddot{b}(t)^\top dt \; \theta_{kf}\} \qquad (16)
\end{aligned}
$$



Solutions

Minimizing (16):

$$
\begin{aligned}
\widehat{\theta}_\mu &= \left\{ \sum_{i=1}^{N} B_i^\top W_i B_i + \lambda_\mu \int \ddot{b}(t)\ddot{b}(t)^\top dt \right\}^{-1} \\
&\qquad \left\{ \sum_{i=1}^{N} B_i^\top W_i (Y_i - B_i \widehat{\Theta}_f \widehat{\alpha}_i) \right\} \\
\widehat{\theta}_{jf} &= \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{ij}^2 B_i^\top W_i B_i + \lambda_f \int \ddot{b}(t)\ddot{b}(t)^\top dt \right\}^{-1} \\
&\qquad \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{ij} B_i^\top W_i (Y_i - B_i \widehat{\theta}_\mu - B_i Q_{ij}) \right\} \qquad (17)
\end{aligned}
$$

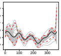


$$\widehat{\alpha}_i = \left\{ \widehat{\Theta}_f^\top B_i^\top W_i B_i \widehat{\Theta}_f \right\}^{-1} \left\{ \widehat{\Theta}_f^\top B_i^\top W_i (Y_i - B_i \widehat{\theta}_\mu) \right\} \quad (18)$$

Where

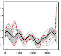
$$Q_{ij} = \sum_{k \neq j} \hat{\theta}_{kf} \hat{\alpha}_{ik}$$

and $i = 1, \cdots, N$, $j = 1, \cdots, K$.



Initial Values

1. Estimate N single curves $\widehat{l}_i$ individually.
2. Run linear regression to get $\widehat{\theta}_{\mu 0}$: $\quad \widehat{l}_i = B_i \theta_\mu + \varepsilon_i$
3. Calculate $\widetilde{l}_{i0} = \widehat{l}_i - B_i \widehat{\theta}_{\mu 0}$, and run the linear regression to get $\widehat{\Gamma}_{i0} = \Theta_{f0} \alpha_{i0}$, and $\widehat{\Gamma}_0 = (\widehat{\Gamma}_{10}, \cdots, \widehat{\Gamma}_{N0})$.
$$\widetilde{l}_{i0} = B_i \Gamma_i + \varepsilon_i$$
4. Apply SVD to decompose $\widehat{\Gamma}_{i0}$:
$$\widehat{\Gamma}_{i0} = UDV^\top = \Theta_{f0} \alpha_{i0}$$
5. Choose the first K principle components from U as $\widehat{\Theta}_{f0}$, and do regression based on $\widehat{\Theta}_{f0}$ to get $\widehat{\alpha}_{i0}$:
$$\widehat{\Gamma}_{i0} = \widehat{\Theta}_{f0} (\alpha_{i1}, \cdots, \alpha_{iK}) + \varepsilon_i \tag{19}$$



Update Procedure

1. Plug $\widehat{\Theta}_{f0}$ and $\widehat{\alpha}_{i0}$ into (17) to update θ_μ, and get $\widehat{\theta}_{\mu 1}$.
2. Plugging $\hat{\theta}_{\mu 1}$ and $\widehat{\alpha}_{i0}$ into the second equation of (17) gives $\widehat{\Theta}_{f1}$.
3. Given $\widehat{\theta}_{\mu 1}$ and $\widehat{\Theta}_{f1}$, estimate $\widehat{\alpha}_i$.
4. Recalculate the weight matrix:

$$w_{ij}^{'} = \begin{cases} \tau & \text{if} \quad Y_{ij} > \widehat{l}_{ij} \\ 1-\tau & \text{if} \quad Y_{ij} \leq \widehat{l}_{ij} \end{cases}$$

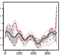
   where $\widehat{l}_{ij}$ is the *jth* element in $\widehat{l}_i = B_i \widehat{\theta}_{\mu 1} + B_i \widehat{\Theta}_{f1} \widehat{\alpha}_i$
5. Repeat step (1) to (4) until the solutions converge.

Auxiliary Parameters

- Number of knots is not crucial.
- Use 5-fold cross validation (CV) to choose the number of components and the penalty parameters.

$$CV(m) = \frac{1}{5} \sum_{i=N-(m-1)\times 5}^{N-m\times 5} \sum_{j=1}^{T_i} \widehat{w}_{ij} |Y_{ij} - \widehat{l}_{ij}|^2 \qquad (20)$$

where $m = 1, 2, \cdots, [N/5]$.



Simulation

$$Y_{it} = \mu(t) + f_1(t)\alpha_{1i} + f_2(t)\alpha_{2i} + \varepsilon_{it} \tag{21}$$

with $i = 1, \cdots, N$ and $t = 1, \cdots, T$.

The mean curve and principal component functions:

$$\begin{aligned}
\mu(t) &= 1 + t + \exp\{-(t-0.6)^2/0.05\} \\
f_1(t) &= \sin(2\pi x)/\sqrt{0.5} \\
f_2(t) &= \cos(2\pi x)/\sqrt{0.5}
\end{aligned}$$

where $\alpha_{1i} \sim \mathrm{N}(0, 36)$, $\alpha_{2i} \sim \mathrm{N}(0, 9)$.

Scenarios

- $\varepsilon_{it} \sim N(0, 0.5)$
- $\varepsilon_{it} \sim N(0, \mu(t) \times 0.5)$
- $\varepsilon_{it} \sim t(5)$

- small sample: $N = 20$, $T = 100$
- large sample: $N = 40$, $T = 150$

Theoretical τ quantile and expectile for individual i:

$$l_{it} = \mu(t) + f_1(t)\alpha_{1i} + f_2(t)\alpha_{2i} + e_\tau$$

where e_τ represents the corresponding theoretical τ-*th* quantile and expectile of the distribution of ε_{it}.

Estimators

- The individual curve:

$$
\begin{aligned}
l_i &= \mu + \sum_{k=1}^{K} f_k \alpha_{ik} \\
\widehat{l}_{i,fp} &= B_i \widehat{\theta}_\mu + B_i \widehat{\Theta}_f \widehat{\alpha}_i \\
\widehat{l}_{i,in} &: \text{Single curve, see (7)}
\end{aligned}
$$

- The mean curve:

$$
\begin{aligned}
m &= \mu(t) + e_\tau \\
m_{fp} &= \frac{1}{N} \sum_{i=1}^{N} B_i \widehat{\theta}_\mu \\
m_{in} &= \frac{1}{N} \sum_{i=1}^{N} \widehat{l}_{i,in}
\end{aligned} \tag{22}
$$

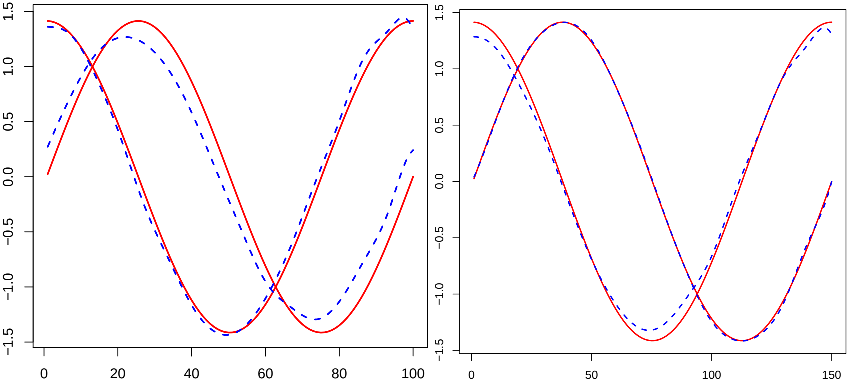


Figure 3: The estimated principle components (dashed blue) compared with the true ones (solid red) for the 95% expectile with the error term normally distributed. The left part is for $N = 20$, $T = 100$. The right one is for $N = 40$, $T = 150$.

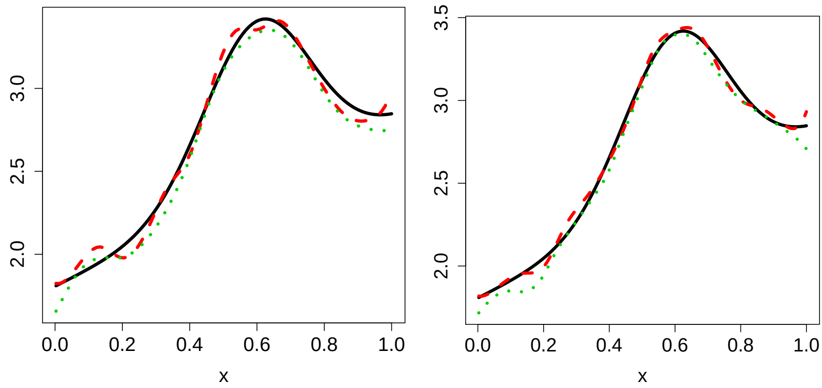


Figure 4: The estimated mean curve compared with the true mean for the 95% expectile with the error term normally distributed. The left part is for $N = 20$, $T = 100$. The right one is for $N = 40$, $T = 150$.

	Individual		*Mean*	
Sample Size	*FPCA*	*Single*	*FPCA*	*Single*
$N = 20$, $T = 100$	0.0469	0.0816	0.0072	0.0093
$N = 40$, $T = 150$	0.0208	0.0709	0.0028	0.0063
$N = 20$, $T = 100$	0.1571	0.2957	0.0272	0.0377
$N = 40$, $T = 150$	0.1002	0.2197	0.0118	0.0172
$N = 20$, $T = 100$	0.2859	0.5194	0.0454	0.0556
$N = 40$, $T = 150$	0.1531	0.4087	0.0181	0.0242

Table 1: The mean squared errors (MSE) of the FPCA and the single curve estimation for expectile curves with error term is normally distributed with mean 0 and variance 0.5 (Top Pattern), with variance $\mu(t) \times 0.5$ (Middle Pattern) and $t(5)$ distribution (Bottom Pattern).

Data

The temperature in 30 cities in Germany in 2006.

Figure 5: Maps of the 30 cities of Germany.

- The temperature T_{it} on day t for city i:

$$T_{it} = X_{it} + \Lambda_{it}$$

- The seasonal effect Λ_{it}:

$$\Lambda_{it} = a_i + b_i t + \sum_{m=1}^{M} c_{im} \cos\{\frac{2\pi(t - d_{im})}{l \cdot 365}\}$$

- X_{it} follows an $AR(p)$ process:

$$\begin{aligned} X_{it} &= \sum_{j=1}^{p_i} \beta_{ij} X_{i,t-j} + \varepsilon_{it} \qquad (23) \\ \widehat{\varepsilon}_{it} &= X_{it} - \sum_{j=1}^{p_i} \hat{\beta}_{ij} X_{i,t-j} \end{aligned}$$

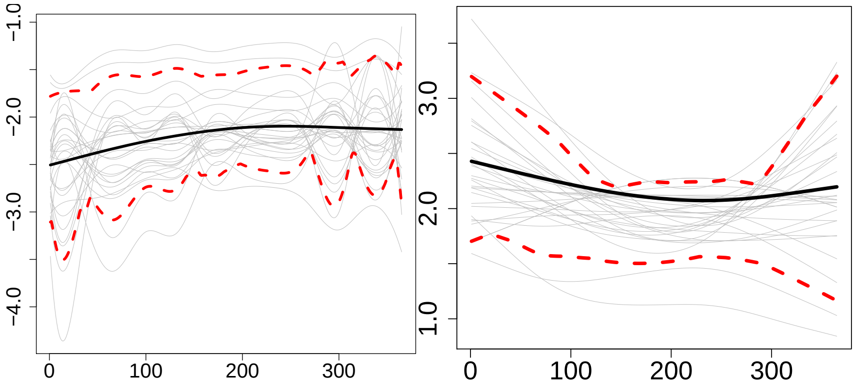


Figure 6: 5% (left) and 95% (right) estimated expectile curves of the temperature variations for 30 cities in Germany in 2006. The thick black line represents the mean curve of the expectiles.

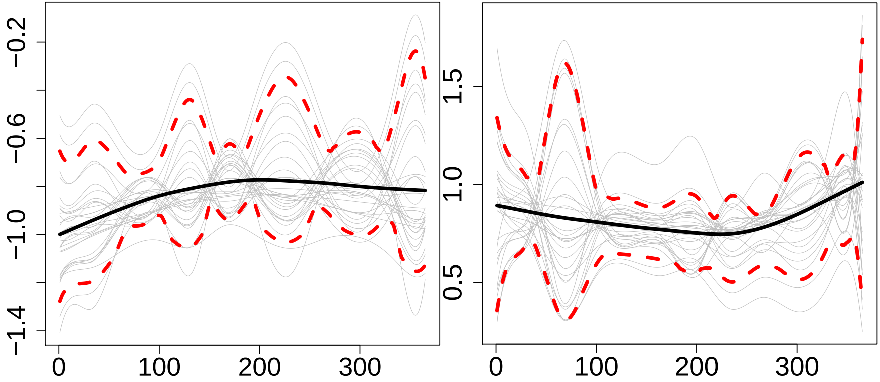


Figure 7: 25% (left) and 75% (right) estimated expectile curves of the temperature variations for 30 cities in Germany in 2006. The thick black line represents the mean curve of the expectiles.

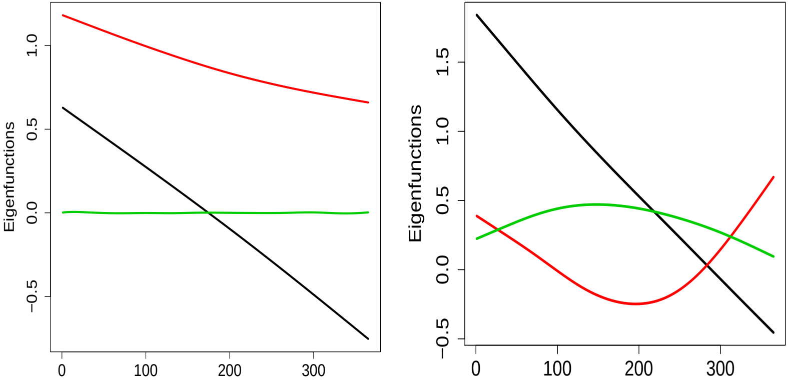


Figure 8: The estimated three eigenfunctions for 05% (left) and 95% (right) expectile curves of the temperature variation. The black one is the first eigenfunction, the red one is the second and the green one represents the third eigenfunction.

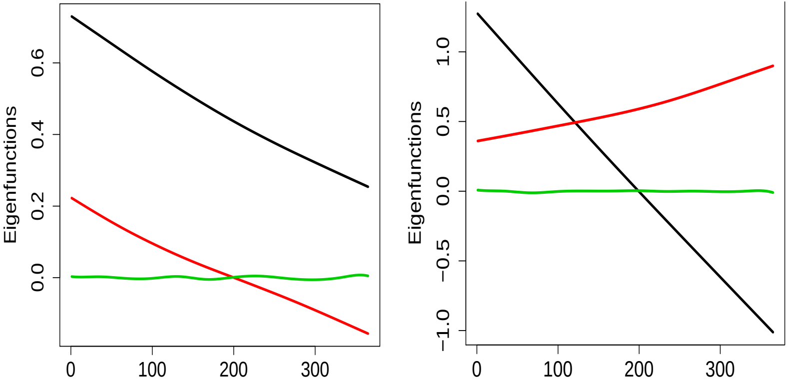


Figure 9: The estimated three eigenfunctions for 25% (left) and 75% (right) expectile curves of the temperature variation. The black one is the first eigenfunction, the red one is the second and the green one represents the third eigenfunction.

Conclusion

- Provides a novel way to estimate several generalized quantile curves simultaneously.
- Outperforms the single curve estimation, especially when the data is very volatile.
- Overcomes the data sparsity.

Reference

J. Fan and T. C. Hu and Y. K. Troung
Robust nonparametric function estimation
Scandinavian Journal of Statistics, 21:433-446, 1994.

G. James and T. Hastie and C. Sugar
Principle Component Models for Sparse Functional Data
Biometrika, 87:587-602, 2000.

M. Jones
Expectiles and M-quantiles are Quantiles
Statistics & Probability Letters, 20:149-153, 1994.

R. Koenker and P. Ng and S. Portnoy
Quantile Smoothing Splines
Biometrika, 81(4):673-680, 1994.

S.K. Schnabel and P.H.C. Eilers
Optimal Expectile Smoothing
Computational Statistics and Data Analysis, 53:4168-4177, 2009.

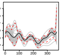
B. Zhang
Nonparametric Expectile Regression
Nonparametric Statistics, 3:255-275, 1994

W. Härdle and B. López Cabrera and O. Okhrin and W. Wang
Localising Temperature Risk
SFB discussion paper, 2011-001, Journal of Econometrics, submitted.

W. Härdle and V. Spokoiny and W. Wang
Localising Quantile Regression
SFB discussion paper, 2011-005, Scandinavian Journal of Statistics, submitted.

M. Guo and W. Härdle
Confidence Bands for Expectile Functions
SFB discussion paper, 2011-004, Asta, invited resubmission.

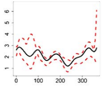
L. Zhou and J. Huang and R. Carroll
Joint Modelling of Paired Sparse Functional Data Using Principle Components
Biometrika, 95(3):601-619, 2008.



Functional Principle Component Analysis for Generalized Quantile Regression

Mengmeng Guo Lan Zhou
Wolfgang Karl Härdle Jianhua Huang

Ladislaus von Bortkiewicz Chair
of Statistics Humboldt-Universität
zu Berlin
Department of Statistics Texas
A&M Univeristy
lvb.wiwi.hu-berlin.de
www.stat.tamu.edu



30 cities in Germany

Aachen, Ausburg, Berlin, Bremen, Dresden, Dusseldorf, Emden, Essen, Fehmarn, Fichtelberg, Frankfurt, Greifswald, Hamburg, Hannover, Helgoland, Karlsruhe, Kempten, Konstanz, Leipzig, Magdeburg, Munchen, Munster, Nurburg, Rostock, Saarbrucken, Schieswig, Schwerin, Straubing, Stuttgart, Trier.