A Confidence Corridor for Expectiles

Esra Akdeniz Duran
Mengmeng Guo
Wolfgang Karl Härdle

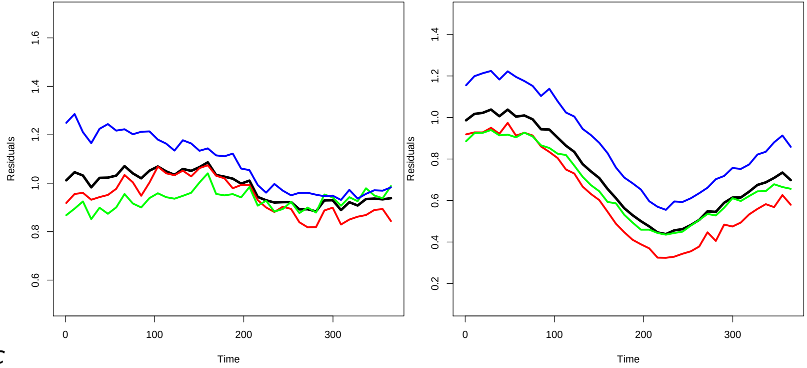
Ladislaus von Bortkiewicz
Chair of Statistics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Expectile Regression

- Standard regression: average behaviour of Y given x
- QR: bigger picture of conditional response
- ER: alternative to QR

Example

- Financial Market
  - VaR (Value at Risk), Kuan et al.(2009)
  - Expected shortfall, Taylor(2008)
- Demographic Research
  - Smooth frontier curve construction, Schnabel and Eilers(2009)
- Heteroscedasticity and/or conditional symmetry test, Newey and Powell(1987)



CC

Figure 1: Expectiles ($\tau = 0.9$) for Berlin (left) and Taipei (right) Temperature Residuals from 1948-2007: Average Expectile from 1948-2007, Average Expectile from 1948-1967, Average Expectile from 1968-1987, Average Expectile from 1988-2007

Questions

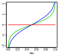
- Stochastic fluctuation of ER
- Confidence Corridors for ER
- Functional form tests

Outline

1. Motivation ✓
2. Quantiles and Expectiles
3. Confidence Corridor for Expectiles
4. Simulation
5. Application

Notations

$\{(X_i, Y_i)\}_{i=1}^n$ i.i.d. random variables
$f(x, y)$ joint pdf ,
$F(x, y)$ joint cdf,
$f(y|x)$ conditional pdfs ,
$F(y|x)$ conditional cdfs



Quantile Regression

$$Y = l(x) + \varepsilon, \qquad \text{with} \qquad F^{-1}_{Y|x}(\tau) = 0$$

- $l(x)$: QR

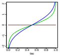
$$l(x) = \arg\min_{\theta} \mathsf{E}\{\rho_\tau(Y - \theta)|X = x\}$$

with a "check function"

$$\rho_\tau(u) = u|\tau - \mathbf{I}\{u \in (-\infty, 0)\}| \qquad \tau \in (0, 1)$$

- $l_n(x)$: estimated QR

$$l_n(x) = \arg\min_{\theta} n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_i - \theta) K_h(x - X_i)$$



Expectile Regression

Now

$$\rho_\tau(u) = u^2 |\tau - \mathbf{I}\{u \in (-\infty, 0)\}| \qquad \tau \in (0,1)$$

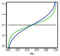
- $v(x)$: ER

$$v(x) = \arg\min_\theta \mathsf{E}\{\rho_\tau(Y - \theta)|X = x\}$$

- $v_n(x)$: estimated ER

$$v_n(x) = \arg\min_\theta n^{-1} \sum_{i=1}^{n} \rho_\tau(Y_i - \theta) K_h(x - X_i)$$

- $K_h(u) = \frac{1}{h} K(\frac{u}{h})$ with bandwidth h



Quantile Regression

τ quantile curve $l(x)$ satisfies

$$F\{l(x)|x\} = \int_{-\infty}^{l(x)} dF(Y|x) = \tau$$

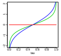
$$l(x) = F^{-1}_{Y|x}(\tau)$$

τ quantile curve estimator:

$$l_n(x) = \hat{F}^{-1}_{Y|x}(\tau)$$

where nonparametric estimation of F(y|x) is

$$\hat{F}(y|x) = \frac{\sum_{i=1}^{n} K_h(x - X_i)\,\mathbf{I}(Y_i < y)}{\sum_{i=1}^{n} K_h(x - X_i)}$$



Expectile Regression

τ expectile curve $v(x)$ satisfies

$$G_{Y|x}(v) = \frac{\int_{-\infty}^{v(x)} |Y - v(x)| \, dF(Y|x)}{\int_{-\infty}^{\infty} |Y - v(x)| \, dF(Y|x)} = \tau$$

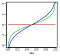
$$v(x) = G_{Y|x}^{-1}(\tau)$$

τ expectile curve estimator:

$$v_n(x) = \hat{G}_{Y|x}^{-1}(\tau)$$

where the nonparametric estimation of $G_{Y|x}(v)$ is

$$\hat{G}_{Y|x}(v) = \frac{\sum_{i=1}^{n} K_h(x - X_i) \, \mathbf{I}(Y_i < y)|y - v|}{\sum_{i=1}^{n} K_h(x - X_i)|y - v|}$$



Expectile Regression

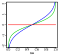
$v(x)$ and $v_n(x)$ can be treated as a zero (w.r.t. θ) of the functions respectively

$$H(\theta, x) = G_{Y|x}(\theta) - \tau \qquad (1)$$

$$H_n(\theta, x) = \hat{G}_{Y|x}(\theta) - \tau = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)\psi(Y_i - \theta)}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)|Y_i - \theta|} \qquad (2)$$

where

$$\begin{aligned} \psi(u) &= (\tau\, \mathbf{I}\{u \in (0, \infty)\} - (1-\tau)\, \mathbf{I}\{u \in (-\infty, 0)\})|u| \\ &= (\mathbf{I}\{u \in (-\infty, 0)\} - \tau)|u|, \end{aligned}$$



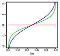
Expectile-Quantile Correspondence

Fixed x, define $w(\tau)$ such that $v_{w(\tau)}(x) = l(x)$ then $w(\tau)$ is related to $l(x)$ via

$$w(\tau) = \frac{\tau l(x) - \int_{-\infty}^{l(x)} y dF(y|x)}{2\,\mathsf{E}(Y|x) - 2\int_{-\infty}^{l(x)} y dF(y|x) - (1-2\tau)l(x)} \qquad (3)$$

For example, $Y \sim U(0,1)$, then $w(\tau) = \tau^2/(2\tau^2 - 2\tau + 1)$

Expectile corresponds to quantile with transformation w.



Expectile and Quantile Curves

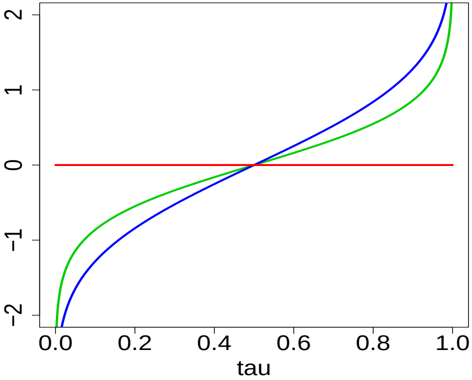


Figure 2: Expectile (green) and Quantile (blue) for $N(0,1)$

Expectiles versus Quantiles

- ER: global dependence of the distribution, Koenker(2005)
- ER: easier to calculate, Efron(1991)
- One to one mapping from expectiles to quantiles, Jones(1994)
- ER: more sensitive to outliers, Schnabel and Eiler(2009)

Assumptions

(A1) $K(\cdot)$ is positive, symmetric, has support $[-A, A]$ and Lipschitz ctsly differentiable;

(A2) $(nh)^{-1/2}(\log n)^{3/2} \to 0$, $(n \log n)^{1/2} h^{5/2} \to 0$, $(nh^3)^{-1}(\log n)^2 \leqslant M$;

(A3) $h^{-3}(\log n) \int_{|y|>a_n} f_Y(y) dy = \mathcal{O}(1)$, $\{a_n\}_{n=1}^{\infty}$ a sequence of constants tending to infinity as $n \to \infty$;

(A4) $\inf_{x \in J} |p(x)| \geqslant p_0 > 0$, where $p(x) = \partial \, \mathsf{E}\{\psi(Y - \theta)|x\}/\partial\theta|_{\theta=v(x)}$;

Assumptions

(A5) ER $v(x)$ is Lipschitz twice ctsly differentiable, for all $x \in J$;

(A6) $0 < m_1 \leqslant f_X(x) \leqslant M_1 < \infty, \; x \in J.$, and the conditional density $f(\cdot|y), \; y \in \mathbb{R}$, is uniform locally Lipschitz continuous of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leqslant 1$.

Normality

Under regularity assumptions, we have

Theorem

$$\sqrt{nh}\{v_n(x) - v(x)\} \xrightarrow{\mathcal{L}} N\{0, V(x)\} \tag{4}$$

with

$$V(x) = \lambda(K)\sigma^2(x)/\{f_X(x)p(x)^2\}$$

where

$$\lambda(K) = \int_{-A}^{A} K^2(u)du$$

$$\sigma^2(x) = \mathsf{E}[\psi^2\{Y - v(x)\}|x]$$

$$p(x) = \partial\,\mathsf{E}\{\psi(Y - \theta)|x\}/\partial\theta|_{\theta=v(x)}$$

Uniform Convergence

Theorem

Let $H(\theta, x)$ and $H_n(\theta, x)$ be given by (1) and (2). For some constant A^* not depending on n, we have a.s.

$$\sup_{\theta \in I} \sup_{x \in J} |H_n(\theta, x) - H(\theta, x)| \leq A^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{5}$$

Thus also:

$$\sup_{x \in J} |v_n(x) - v(x)| \leq B^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{6}$$

Theorem

$$P\Big((2\delta \log n)^{1/2}\Big[\sup_{x\in J} r(x)|\{v_n(x)-v(x)\}|/\lambda(K)^{1/2}-d_n\Big]<z\Big)$$

$$\longrightarrow \exp\{-2\exp(-z)\}, \qquad \textit{as} \qquad n\to\infty.$$

$$r(x)=(nh)^{-\frac{1}{2}}p(x)\{f_X(x)/\sigma^2(x)\}^{\frac{1}{2}}$$

and δ, $\lambda(K)$, d_n are suitable scaling parameters.

Uniform Confidence Bands

Theorem

An approximate $(1-\alpha)\times 100\%$ *confidence bands for* $v(x)$ *is*

$$v_n(x) \pm (nh)^{-1/2}\{\hat{\sigma}^2(x)\lambda(K)/\hat{f}_X(x)\}^{1/2} * \hat{p}^{-1}(x)\{d_n + c(\alpha)(2\delta \log n)^{-1/2}\} \quad (7)$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$

$\hat{f}_X(x)$, $\hat{\sigma}^2(x)$ *and* $\hat{p}(x)$ *are consistent estimates for* $f_X(x)$, $\sigma^2(x)$ *and* $p(x)$.

Simulation

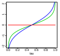
Simulate $\{(X_i, Y_i)\}_{i=1}^{n}$ with $n = 500$, and $X \sim$U[0,3]

$$Y = 1.5X^2 + 4 + \cos(3X) + \varepsilon$$

where $\varepsilon \sim \mathrm{N}(0, 1)$.
The theoretical ER (fixed τ) is

$$v(x) = 1.5x^2 + 4 + \cos(3x) + v_N(\tau)$$

where $v_N(\tau)$ is the τ-quantile of the standard Normal distribution.



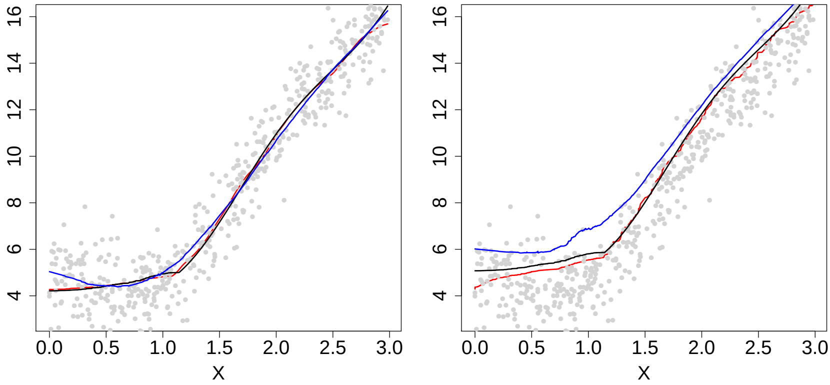


Figure 3: $\tau = 0.5$(left) and $\tau = 0.9$(right) Estimated Quantile and Expectile Plot. Quantile Curve, Theoretical Expectile Curve, Estimated Expectile Curve

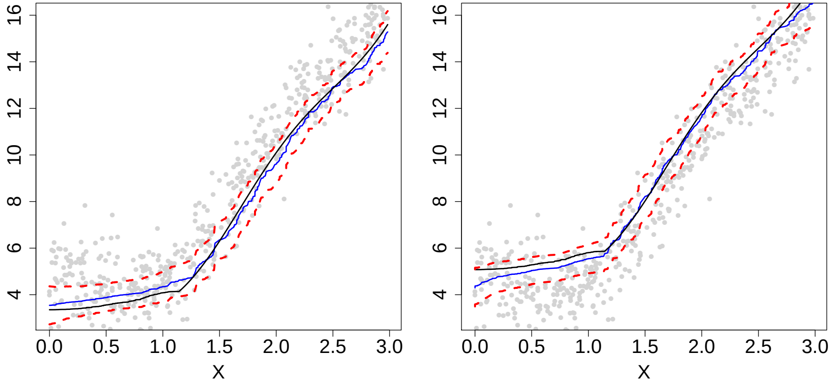


Figure 4: Uniform Confidence Bands for Expectile Curve $\tau = 0.1$ and $\tau = 0.9$. Theoretical Expectile Curve, Estimated Expectile Curve, 5% – 95% Uniform Confidence Bands

Temperature Residuals

The time series decomposition is:

$$
\begin{aligned}
X_{365j+t} &= T_{t,j} - \Lambda_t \\
X_{365j+t} &= \sum_{l=1}^{L} \beta_{lj} X_{365j+t-l} + \varepsilon_{t,j} \qquad (8)
\end{aligned}
$$

where $T_{t,j}$ is the temperature at day t in year j, Λ_t denotes the seasonality effect, $t = 1, \cdots, \tau = 365$ days and $j = 0, \cdots, J$ years.

Stations in Taiwan

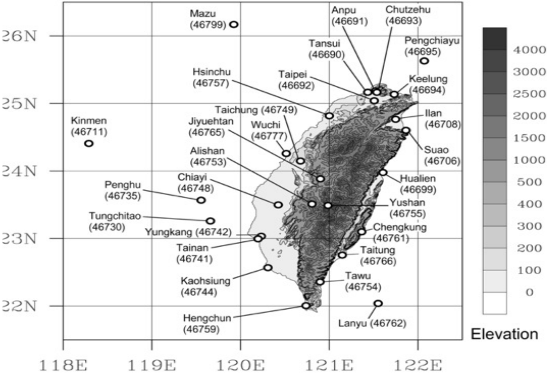


Figure 5: The locations of CWB (Central Weather Bureau) weather stations.

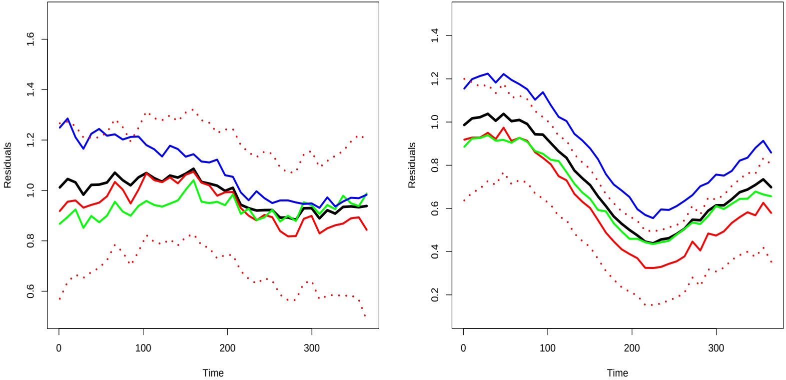


Figure 6: 0.9-expectile curves for Berlin (left) and Taipei (right) temperature residuals from 1948-2007 with the 5% – 95% confidence bands for the first 20 years expectiles

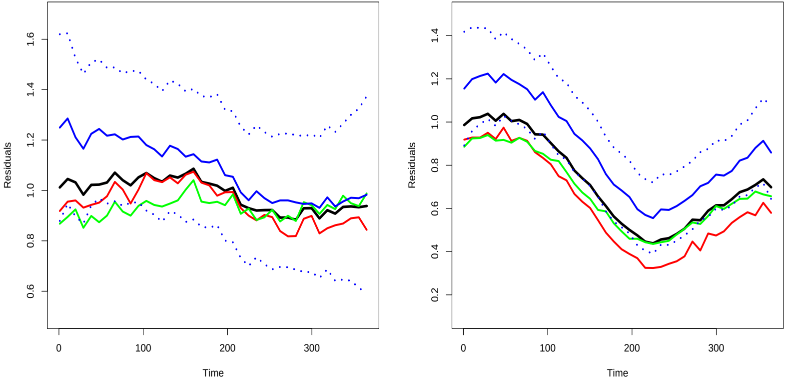


Figure 7: 0.9-expectile curves for Berlin (left) and Taipei (right) temperature residuals from 1948-2007 with the 5% − 95% confidence bands for the latest 20 years expectiles

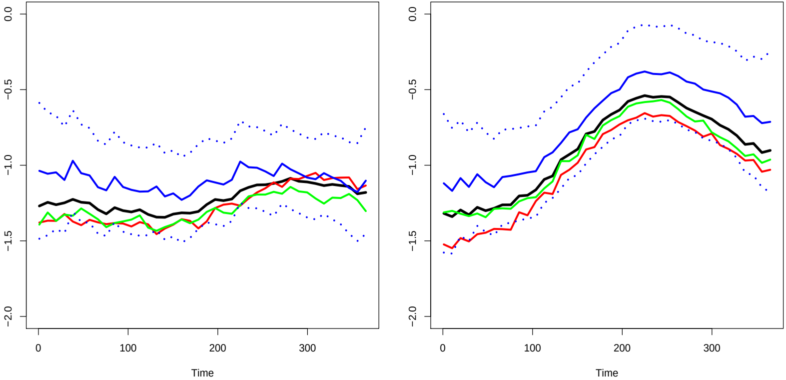


Figure 8: 0.01-expectile curves for Berlin (left) and Taipei (right) temperature residuals from 1948-2007 with the 5% – 95% confidence bands for the first 20 years expectiles

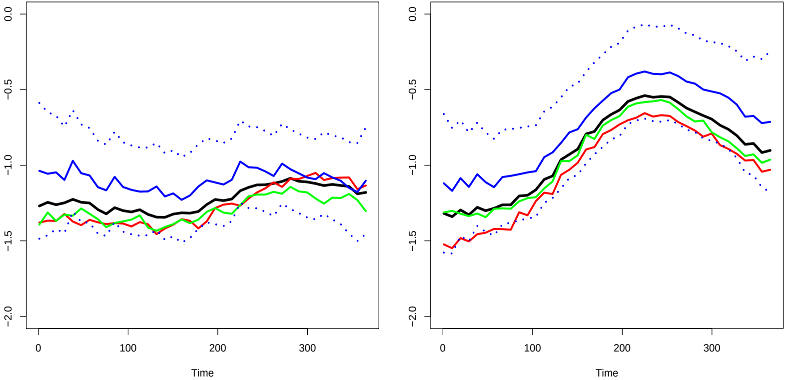


Figure 9: 0.01-expectile curves for Berlin (left) and Taipei (right) temperature residuals from 1948-2007 with the 5% – 95% confidence bands for the latest 20 years expectiles

Conclusion

- Expectiles capture the tail behaviours of the distribution.
- Expectiles can be calculated at very high and very low percentages.
- The temperature risk drivers of Berlin and Taipei are different.

Reference

W. Härdle
Asymptotic Maximal Deviation of M-Smoothers
Journal of Multivariate Analysis, 29:163-179, 1989.

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

W. Härdle and S. Song
Confidence Bands in Quantile Regression
Econometric Theory, 26:1-22, 2010.

C. M. Kuan and Y. H. Yeh and Y. C. Hsu
Assesing Value at Risk with CARE–the Conditional Autoregressive Expectile models
Journal of Econometrics, 150:261-270, 2009.

W. K. Newey and J. L. Powell
Asymmetric Least Squares Estimation and Testing
Econometrica, 55:819-847, 1987.

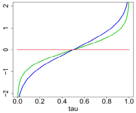
S.Schnabel and P.Eilers
An Analysis of Life Expectancy and Economic Production Using Expectile Frontier Zones
Demographic Research, 21(5):109-134, 2009.

A Confidence Corridor for Expectiles

Esra Akdeniz Duran

Mengmeng Guo

Wolfgang Karl Härdle



Ladislaus von Bortkiewicz
Chair of Statistics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de