

SFB 649 Discussion Paper 2011-005

Local Quantile Regression

Wolfgang Karl Härdle*
Vladimir Spokoiny**
Weining Wang*



* Humboldt-Universität zu Berlin, Germany
* Weierstrass Institute (WIAS) Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Local Quantile Regression *

Wolfgang Karl Härdle †, Vladimir Spokoiny ‡, Weining Wang §

December 13, 2010

Abstract

Conditional quantile curves provide a comprehensive picture of a response contingent on explanatory variables. Quantile regression is a technique to estimate such curves. In a flexible modeling framework, a specific form of the quantile is not a priori fixed. Indeed, the majority of applications do not per se require specific functional forms. This motivates a local parametric rather than a global fixed model fitting approach. A nonparametric smoothing estimate of the conditional quantile curve requires to consider a balance between local curvature and variance. In this paper, we analyze a method based on a local model selection technique that provides an adaptive estimate. Theoretical properties on mimicking the oracle choice are offered and applications to stock market and weather analysis are presented.

Keywords: Conditional Quantiles; Semiparametric and Nonparametric Methods; Asymmetric Laplace Distribution; Exponential Risk Bounds; Adaptive Bandwidth Selection.

JEL classification: C00, C14, J01, J31

1 Introduction

Quantile regression is gradually developing into a comprehensive approach for the statistical analysis of linear and nonlinear response models. Since the rigorous treatment of linear quantile regression by Koenker and Bassett (1978), richer models have been introduced into the literature, among them are nonparametric, semi parametric and additive

*The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged.

†Professor at Humboldt-Universität zu Berlin and Director of CASE - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany and National Central University, Graduate Institute of Statistics, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan (R.O.C.). Email:haerdle@wiwi.hu-berlin.de

‡Professor at Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Mohrenstr. 39, 10117 Berlin, Germany. Email:spokoiny@wias-berlin.de

§Research associate at Ladislaus von Bortkiewicz Chair, the Institute for Statistics and Econometrics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email:wangwein@cms.hu-berlin.de

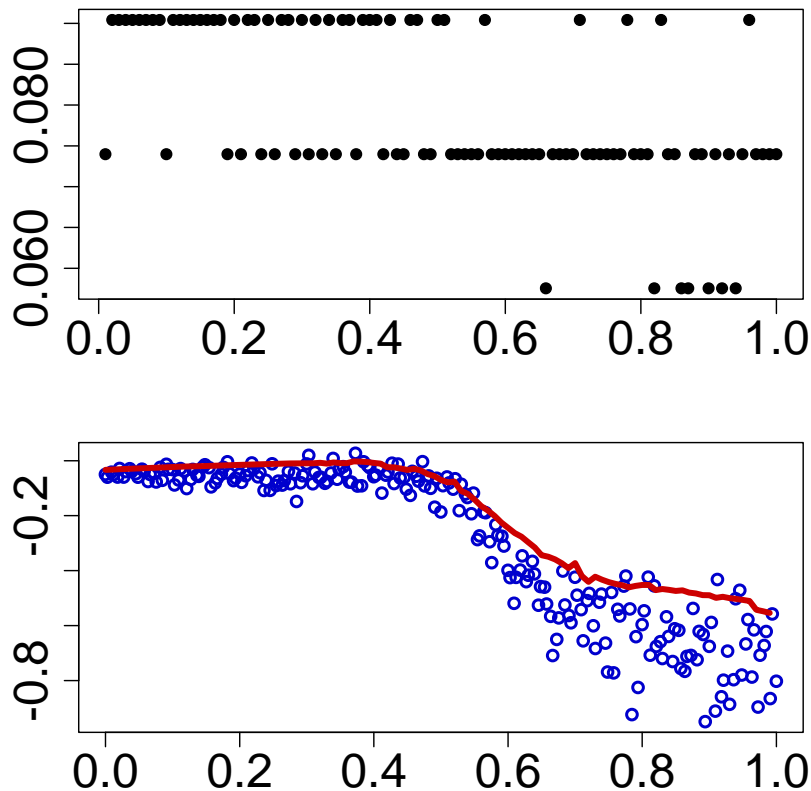


Figure 1: The bandwidth sequence (upper panel), plot of data (x variable scaled to $[0, 1]$) and the estimated 90% quantile curve (lower panel)

approaches. Quantile regression or conditional quantile estimation is a crucial element of analysis in many quantitative problems. In financial risk management, the proper definition of quantile based Value at Risk impacts asset pricing, portfolio hedging and investment evaluation, see Engle and Manganelli (2004), Cai and Wang (2008), Fitzenberger and Wilke (2006). In labor market analysis of wage distributions, education effects and earning inequalities are analysed via quantile regression. Other fields of applications of conditional quantile studies include conditional data analysis of children growth and ecology, where it accounts for the unequal variation of response variables, see James, Hastie and Sugar (2010).

In applications, the predominantly used linear form of the calibrated models was mainly determined by practical and numerical reasonings. More flexible functional forms have been put forward in the literature in only rare cases, Kong, Linton and Xia (2010) and Yu and Jones (1997).

The aim of this research is to propose an adaptive local quantile regression algorithm that is easy to implement and works for a wide class of applications. We describe in detail the adaptation technique that is based on local parametric models. The main conclusion is that the proposed algorithm is feasible and beneficial for quantile smoothing and helps in proposing alternatives to more restricted parametric models. Figure 1 presents an example of our results for analyzing the Lidar data set, see Ruppert, Wand and Carroll (2003). The presented quantile curve switches smoothness in the middle, and it is naturally reflected by the bandwidth sequence (upper panel) selected by the proposed technique. In the

presence of changing to sharper slope of the curve, the bandwidth gets smaller to attain better estimation.

This article is organized as follows. In Section 2, exponential risk bounds for conditional quantiles are established using the representation of quantiles as QMLEs of the asymmetric Laplace distribution. In Section 3, we introduce the adaptation scheme and the propagation condition to control the significance level and derive theoretical bounds for critical values. In Section 4, the ‘‘Small Modeling Bias’’ is studied. In Section 5, ‘‘Stability’’ and ‘‘Oracle’’ properties are given. In Section 6, Monte Carlo simulations are conducted to illustrate the proposed methodology. In Section 7, we apply our method on checking tail dependency among portfolio stocks; and on estimation of quantile curves for temperature risk factors.

2 Parametric Risk Bounds

Consider the MLE for the asymmetric Laplace distribution (ALD), which plays a key role in quantile estimation. An important advantage of this QMLE representation is that it and the corresponding fitted likelihood admit a closed form.

A random variable has an $ALD(\mu, \sigma, \tau)$ if its pdf is:

$$f(u|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\rho_\tau \left(\frac{u-\mu}{\sigma} \right) \right\}, \quad -\infty < u < \infty$$

where $\rho_\tau(u) = u\{\tau\mathbf{1}(u \geq 0) - (1-\tau)\mathbf{1}(u < 0)\}$ is a *loss function*. The parameter $0 < \tau < 1$ is the level of the quantile, μ a location parameter and $\sigma > 0$ a scale parameter, respectively. In the following, we restrict the analysis to the standardized situation $\mu = 0, \sigma = 1$ and abbreviate it as $ALD(\tau)$. The median, for example, is related to

$$f \left(u|0, 1, \frac{1}{2} \right) = \frac{1}{4} \exp \left(-\frac{|u|}{2} \right), \quad -\infty < u < \infty.$$

the pdf of a symmetric Laplace distribution.

Theorem 2.1 *Let Y_1, \dots, Y_n be a sample modeled by $Y_i = \theta^* + \varepsilon_i$ with i.i.d. $\varepsilon_i \sim ALD(\tau)$, and θ^* a location parameter. Let $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ denote the order statistics of $\{Y_i\}_{i=1}^n$. Then the QMLE $\tilde{\theta}$ of θ^* is*

$$\tilde{\theta} = Y_{([n\tau]+1)}$$

Proof. The log-likelihood $L(\theta)$ is

$$\begin{aligned} L(\theta) &= \log \left[\prod_{i=1}^n \tau(1-\tau) \exp \{ -\rho_\tau(Y_i - \theta) \} \right] \\ &= n \log \{ \tau(1-\tau) \} - \frac{1}{2} \sum_{i=1}^n \{ |Y_i - \theta| + (2\tau - 1)(Y_i - \theta) \} \\ &= n \log \{ \tau(1-\tau) \} + (1-\tau) \sum_{i=1}^n (\theta - Y_i) \mathbf{1}(\theta > Y_i) - \tau \sum_{i=1}^n (Y_i - \theta) \mathbf{1}(\theta \leq Y_i). \end{aligned}$$

Now one computes the derivative w.r.t. θ . Suppose that $Y_{(k)} \leq \theta < Y_{(k+1)}$, $k = 0, 1, \dots, n$, where $Y_{(0)} < Y_{(1)}$ and $Y_{(n+1)} > Y_{(n)}$ are any two real values, then

$$L'(\theta) = -(1 - \tau)k + \tau(n - k) = -k + \tau n.$$

We have $L'(\theta) = 0$, for $k = [n\tau]$.

Theorem 2.2 For any θ the log-likelihood ratio in quantile QMLE framework is:

$$\begin{aligned} L(\tilde{\theta}, \theta) &= L(\tilde{\theta}) - L(\theta) \\ &= n(\tilde{\theta} - \theta)(2\tau - 1)/2 + \frac{1}{2} \sum_{i=1}^n \left\{ |Y_i - \theta| - |Y_i - \tilde{\theta}| \right\}. \end{aligned}$$

Proof. For any two points $\tilde{\theta}$ and θ , the log-likelihood ratio is:

$$\begin{aligned} L(\tilde{\theta}, \theta) &= \frac{1}{2} \sum_{i=1}^n \left\{ |Y_i - \theta| + (2\tau - 1)(Y_i - \theta) \right\} - \left\{ |Y_i - \tilde{\theta}| + (2\tau - 1)(Y_i - \tilde{\theta}) \right\} \\ &= \frac{1}{2} \sum_{i=1}^n \left\{ |Y_i - \theta| - |Y_i - \tilde{\theta}| + (2\tau - 1)(\tilde{\theta} - \theta) \right\} \\ &= n \frac{2\tau - 1}{2} (\tilde{\theta} - \theta) + \frac{1}{2} \sum_{i=1}^n (|Y_i - \theta| - |Y_i - \tilde{\theta}|). \end{aligned} \quad (1)$$

Next, consider the Kullback-Leibler divergence $\mathcal{K}(\cdot, \cdot)$ defined for two pdf's $f(y, \theta), f(y, \theta')$ with parameters θ, θ' :

$$\mathcal{K}(\theta, \theta') = \mathbb{E}_\theta \log \frac{f(Y, \theta)}{f(Y, \theta')}.$$

Theorem 2.3 Let $f(y, \theta)$ be the density of $Y = \theta + \varepsilon$, $\varepsilon \sim \text{ALD}(\tau)$. The Kullback-Leibler divergence $\mathcal{K}(\theta, \theta')$ is:

$$\mathcal{K}(\theta, \theta') = \begin{cases} \frac{\tau}{\tau-1} + \tau(\theta - \theta') + \frac{\tau}{1-\tau} \exp \left\{ (1 - \tau)(\theta' - \theta) \right\}, & \theta \geq \theta', \\ \frac{\tau-1}{\tau} + (1 - \tau)(\theta' - \theta) + \frac{1-\tau}{\tau} \exp \left\{ -\tau(\theta' - \theta) \right\}, & \theta < \theta'. \end{cases} \quad (2)$$

Proof. The theorem may be proved by simple trivial algebraic calculations.

Remark. If θ' is close to θ , then by Taylor expansion, one sees:

$$\mathcal{K}(\theta, \theta') \approx \frac{\tau(1 - \tau)}{2} (\theta' - \theta)^2 \quad (3)$$

Define tail functions as:

$$\begin{aligned} \lambda^+(y) &= -(y/\tau)^{-1} \log \{ \tau^{-1} \text{P}(\varepsilon_i > y) \}, y \geq 0 \\ \lambda^-(y) &= -\{y/(1 - \tau)\}^{-1} \log \{ (1 - \tau)^{-1} \text{P}(\varepsilon_i < y) \}, y < 0 \end{aligned} \quad (4)$$

If we make assumptions on the error distribution (in a QMLE framework), we write $\lambda(\cdot)$ for $\lambda^{+/-}(y)$. Introduce $\lambda_0(\cdot)$ for $\lambda_0^{+/-}(y)$ in the case of true unknown tail function,

Define the rate function as $\mathfrak{M}(\mu, \theta, \theta^*) \stackrel{\text{def}}{=} -\log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\mu L(\theta, \theta^*)\}$ which we abbreviate for fixed μ as $\mathfrak{M}(\theta, \theta^*)$. $\mathbf{E}_{\lambda(\cdot), \theta^*}$ means taking expectation under the distribution $\lambda(\cdot)$ of rv. Y_i .

Assume $\lambda^+(y)$ and $\lambda^-(y) \rightarrow 0$ as $y \rightarrow +/ - \infty$, and $\lambda^{+/-}(y)$ is bounded from below by $\mu^{+/-}(z)$ and recall the definition of $\mathcal{A}(z, \theta^*) \stackrel{\text{def}}{=} \{\theta : \mathfrak{M}(\theta, \theta^*) \leq z\}$ and $\mu^+(z) > C \log z/z$ or $(\mu^-(z) < C \log z/z)$.

The bound for the local constant estimator is described in

Theorem 2.4 *Let $\{Y_i\}_{i=1}^n$ be defined as $Y_i = \theta^* + \varepsilon_i$, with ε_i as i.i.d. random noise, and $F_\varepsilon^{-1}(\tau) = 0$.*

Then we have

$$\mathbf{P}_{\lambda(\cdot), \theta^*} \{L(\tilde{\theta}, \theta^*) > \xi\} \leq C \exp(-\xi) \quad (5)$$

and for any $r > 0$

$$\mathbf{E}_{\lambda(\cdot), \theta^*} \{|L(\tilde{\theta}, \theta^*)|^r\} \leq \mathfrak{R}_r, \quad (6)$$

where \mathfrak{R}_r is defined as the moment bound for the likelihood ratio process.

Proof. See appendix.

Recall that when Y_i s do not follow ALD, quantile estimation may be done in a QMLE framework as in Theorem 2.1. W.L.O.G., assume that $\theta^* = 0$.

$$\tilde{\theta} = \arg \max_{\theta} L(\theta) = \arg \min_{\theta} \sum_{i=1}^n -\log\{\tau(1-\tau)\} + \rho_\tau(Y_i - \theta),$$

The general case ($\theta^* \neq 0$) can be reduced to this one by a simple change of variables. (4) gives $\mathbf{P}(Y > y) = \exp\{-\lambda^+(y)y/\tau\}\tau$ for $y \geq 0$, and for $y < 0$, $\mathbf{P}(Y < y) = \exp\{-\lambda^-(y)y/(1-\tau)\}(1-\tau)$, showing that the tail function identifies the cdf of Y .

The case with $\lambda^+(y) \geq \lambda_0^+ > 0$ $\{\lambda^-(y) \leq \lambda_0^- < 0\}$ corresponds to light tails, while $\lambda^+(y) \rightarrow 0$ $\{\lambda^-(y) \rightarrow 0\}$ as $|y| \rightarrow \infty$ means heavy tails of the distribution $\mathbf{P}(\cdot)$.

For example, a random variable $Y \sim \text{Pareto}(\alpha, x_0)$, $x_0 > 0$, with density function $\alpha x_m^\alpha / y^{\alpha+1}$, and $\mathbf{P}(Y > y) = (x_m/y)^\alpha$ has the tail function behavior:

$$\begin{aligned} \lim_{y \rightarrow \infty} \lambda^+(y) &= \lim_{y \rightarrow \infty} - (y/\tau)^{-1} (\alpha \log x_0 - \alpha \log y - \log \tau) \\ &= 0, \end{aligned}$$

By contrast the light tail random variable $Y \sim \exp(\lambda)$ with density function $\lambda \exp\{-\lambda y\} \mathbf{1}(y \geq 0)$:

$$\begin{aligned} \lim_{y \rightarrow \infty} \lambda^+(y) &= \lim_{y \rightarrow \infty} - \tau y^{-1} (-\lambda y - \log \tau) \\ &= \lambda \tau > 0. \end{aligned}$$

The interesting case of course is the one where $\lambda^+(y)$ is positive and monotonously decreases to 0 in $y > 0$, and $\lambda^-(y)$ is negative and monotonously increases to 0 as $y \rightarrow \infty$. Assume further that $\lambda^+(y)\{\lambda^-(y)\}$ is sufficiently regular and its first derivative $\lambda^{+'}(y)\{\lambda^{-'}(y)\}$ is uniformly continuous on \mathbb{R} .

2.1 Local Polynomial Estimation

Let $\{X_i, Y_i\}_{i=1}^n$ be independent random variables, we use the following model for construction:

$$Y_i = l(X_i) + \varepsilon_i, \quad (7)$$

the τ th quantile curve follows with $P(\varepsilon_i > 0) = \tau$.

We concentrate on $X \in \mathbb{R}^1$ in this paper. It is not difficult to extend to the d -dimensional case $X \in \mathbb{R}^d$ $d > 1$. For fixed $\tau \in (0, 1)$ and $u \in \mathbb{R}$, assume that the conditional quantile function $F_{Y|x}^{-1}(\tau) = l(x)$ can be approximated by a polynomial, in the vicinity of u :

$$l(x) \approx \theta_0(u) + \theta_1(u)(x - u) + \theta_2(u)(x - u)^2 + \dots + \theta_p(u)(x - u)^p. \quad (8)$$

Then we have the corresponding QMLE :

$$\begin{aligned} \tilde{\theta}(u) &= \{\tilde{\theta}_0(u), \tilde{\theta}_1(u), \dots, \tilde{\theta}_p(u)\}^\top \\ &\stackrel{\text{def}}{=} \operatorname{argmax}_{\theta \in \Theta} L(W, \theta) \\ &= \operatorname{argmax}_{\theta \in \Theta} \left\{ \log \tau(1 - \tau) \sum_{i=1}^n w_i - \sum_{i=1}^n \rho_\tau(Y_i - \theta^\top \psi_i) w_i \right\} \\ &= \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n \rho_\tau(Y_i - \theta^\top \psi_i) w_i, \end{aligned} \quad (9)$$

where $\psi_i = \{1, (X_i - u), (X_i - u)^2, \dots, (X_i - u)^p\}^\top$, $w_i = K\{(X_i - u)/h\}$ and $W = (w_1, \dots, w_n)^\top$, for i in $1, 2, \dots, n$.

We use Theorem 2.8 in Spokoiny (2011) to prove a result similar to (6) for local polynomial approximations (8), consider the following definitions, (from now on abbreviate $\mathbb{E}_{\lambda_0, \theta(\cdot)}$ as \mathbb{E}):

$$\begin{aligned} \zeta(\theta) &\stackrel{\text{def}}{=} L(W, \theta) - \mathbb{E} L(W, \theta) \\ \zeta(\theta, \theta^*) &\stackrel{\text{def}}{=} \zeta(\theta) - \zeta(\theta^*) \\ \mathfrak{N}(\mu, \theta, \theta^*) &\stackrel{\text{def}}{=} -\log \mathbb{E} \exp\{\mu \zeta(\theta, \theta^*)\} \\ \mathfrak{M}(\mu, \theta, \theta^*) &\stackrel{\text{def}}{=} -\log \mathbb{E} \exp\{\mu L(W, \theta, \theta^*)\} = -\log \mathbb{E} \exp[\mu \{L(W, \theta) - L(W, \theta^*)\}] \end{aligned} \quad (10)$$

The two conditions involved are (ED) and (\mathcal{L}):

1.(ED) $\nabla \zeta(\theta)$ satisfying the condition (ED):

$\exists \lambda^* > 0$ and a symmetric positive matrix V^2 s.t. $\forall \lambda, 0 < \lambda \leq \lambda^*$,

$$\log \mathbb{E} \exp\{\lambda \gamma^\top \nabla \zeta(\theta)\} \leq \nu_0 \lambda^2 \|V^2 \gamma\|^2 / 2, \quad (11)$$

for some fixed $\nu_0 \geq 1$ and $\gamma \in \mathbb{R}^{p+1}$.

Set $V^2(\theta) = \sum_{i=1}^n \psi_i \psi_i^\top p_i(\psi_i^\top \theta) w_i$, with $p_i(\cdot)$ as the density of Y_i .

2. (\mathcal{L}) For $R > 0$, there exists constant $\mathfrak{a} = \mathfrak{a}(R) > 0$ such that it holds on the set $\Theta_0(R) = \{\theta : \|V(\theta - \theta^*)\|^2 \leq R\}$:

$$-\mathbb{E} L(W, \theta, \theta^*) \geq \mathfrak{a} \|V(\theta - \theta^*)\|^2 / 2 \quad (12)$$

Restrict to a local central point θ^* . Define the matrix $V_0 = V(\theta^*)$. Assume that the density of Y_i s are uniformly bounded from below in $\Theta_0(R)$, then we have (12). The detailed proof of condition (ED) and (\mathcal{L}) is deferred to the appendix.

Given (ED) on $\Theta_0(R)$ with the matrix V_0 and $\lambda^* = N^{1/2}(\Psi)\delta^*/C(\Psi, R)$, where

$$C(\Psi, R) \stackrel{\text{def}}{=} \sup_{\gamma \in R^p} \sup_{\theta \in \Theta_0(R)} \frac{\|V(\theta)\gamma\|}{\|V(\theta^*)\gamma\|}.$$

We may apply Theorem 2.8 of Spokoiny (2011) to obtain:

Theorem 2.5 *Let $V_0 = V(\theta^*)$ and R be so that $\lambda^* \geq \mathfrak{q}R/\mu$ for $\mu = \mathfrak{a}/(2\nu_0\mathfrak{q}^3)$. We have for any $0 < r < R$:*

$$\mathbb{P}(\|V(\tilde{\theta} - \theta^*)\| > r) \leq \exp\{-\mathfrak{a}^2 r^2 / (8\nu_0\mathfrak{q}^3) + \mathfrak{Q}_0(p, \mathfrak{q})\} + \exp\{-\mathfrak{g}(R)\}$$

with $\mathfrak{Q}_0(p, \mathfrak{q}) = pC(\mathfrak{q}) + \mathfrak{q}^2 + \log(\mathfrak{q}^2 - \mathfrak{q})$. Moreover, if $R \leq \mu\lambda^*/(\nu_0\mathfrak{q})$, then for any $\mathfrak{z} > 0$ with $2\mathfrak{z}/\mathfrak{a} \leq R^2$, it holds

$$\mathbb{P}\{L(\tilde{\theta}, \theta^*) > \mathfrak{z}\} \leq \exp\{-\mathfrak{z}\mathfrak{a}/(\nu_0\mathfrak{q}^3) - \mathfrak{Q}_0(p, \mathfrak{q})\} + \exp\{-\mathfrak{g}(R)\}, \quad (13)$$

where $\mathfrak{g}(R)$ is defined as:

$$\mathfrak{g}(R) \stackrel{\text{def}}{=} \inf_{\theta \in \Theta} \sup_{\mu \in M} \{\mu R + \mathfrak{C}(\mu, \theta, \theta^*)\} \quad (14)$$

$$= \inf_{\theta \in \Theta} \mathfrak{g}(R, \theta, \theta^*) \quad (15)$$

with

$$\mathfrak{C}(\mu, \theta, \theta^*) = \mathfrak{q}^{-1}\mathfrak{M}(\mathfrak{q}\mu, \theta, \theta^*) - (p+1) \log_+\{\epsilon^{-1}\mu^*\|V(\theta - \theta^*)\|\} - \mathfrak{Q}(p, \mathfrak{q}), \quad (16)$$

where $\log\{(x)\mathbf{1}(x \geq 1)\}$.

$$\mathfrak{Q}(p, \mathfrak{q}) = -\nu_0\mathfrak{q}\epsilon^2/2 - \log(2+2p) - pC(\mathfrak{q}) - \log(1 - \mathfrak{a}^{-(p+1)}), \quad (17)$$

$$C(\mathfrak{q}) = \log \mathfrak{q} \log\{2(1 - \mathfrak{q}^{-1})\},$$

$\mathfrak{q} > 1, \mathfrak{a}, \nu_0, p, \epsilon$ are positive real constants.

To achieve a confidence set, we want that $\exp\{-\mathfrak{g}(R)\}$ in (13) is negligible for moderately large sample size or not so heavy tails. Typically, the tail function is assumed to hold $\lambda^+(\theta)\theta > C \log \theta$. We need again to bound the rate function from below by a monotonically increasing function.

Define $\tilde{\varepsilon}_i = Y_i - \psi_i^\top \theta^*$ and their distributions $P_i(\varepsilon \in A) = \mathbb{P}\{(Y_i - \psi^\top \theta^*) \in A\}$ for any Borel set A on the real line. If $Y_i = \psi_i^\top \theta^* + \varepsilon_i$ is the true model, then P_i coincides with distribution of each ε_i . For simplicity of presentation, $\lambda^+(y)\{\lambda^-(y)\}$ is assumed to be sufficient regular and its first derivative is uniform continuous in R . Therefore as in the local constant case, we have $|y\lambda_i^+(y)|' \in [0, 1]$ and $|y\lambda_i^+(y)| < 1$. Define

$$\mathfrak{M}^*(\theta, \theta^*) \stackrel{\text{def}}{=} \sup_{\mu} \mathfrak{M}(\mu, \theta, \theta^*)$$

Theorem 2.6 We can achieve, with $y > 0$, and $\mu_i = \lambda_i^+(y)/(2\tau)$

$$\mathfrak{M}^*(\theta, \theta^*) \geq \mathfrak{M}(\mu, \theta, \theta^*) \geq \sum_{i=1}^n \log\{1 + \psi_i^\top(\theta - \theta^*)\lambda^+(\theta)/\tau\} + \sum_{i=1}^n \left\{ \frac{(2\tau - 2)(\psi_i^\top\theta - \psi_i^\top\theta^*)\lambda^+(\theta)}{2\tau} \right\}.$$

Proof. See appendix.

The moment bounds for the likelihood ratio process can be obtained by the following theorem:

Theorem 2.7 Define the positive loss function $|L(\tilde{\theta}, \theta^*)|^r$ ($r > 0$). Then

$$\mathbb{E}_{\lambda(\cdot), \theta^*} |L(\tilde{\theta}, \theta^*)|^r \leq \mathfrak{R}_r,$$

where $\mathfrak{R}_r > 0$ is a constant.

Proof. See appendix.

3 Propagation and Bounds for Critical Values

In section 2, we derived the parametric risk bounds for local constant and local polynomial quantile estimates. These bounds provide us a fundament on which to carry out a local adaptive procedure. For fixed x , a sequence of ordered weights $W^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_n^{(k)})^\top$ is defined $w_i^{(k)} = K_{h_k}(x - X_i)$, where $(h_1 < h_2 < \dots < h_K)$. The weights of a local model determine a local estimation of $l(x)$ as well as the derivatives $\hat{l}^{(i)}(x)$ of $l(x)$, see (9). Note that $\hat{\theta}_k(x)$ is a $p+1$ dimensional vector, with $\hat{l}^{(i)}(x) = i!\hat{\theta}_{k,i}(x)$ and $i \in 1, \dots, p$. $\hat{\theta}_{k,i}(x)$ is the i th component of $\hat{\theta}_k(x)$.

The adaptation algorithm is described by:

- Start with $\hat{\theta}_1(x) = \tilde{\theta}_1(x)$.
- For $k \geq 2$, $\tilde{\theta}_k(x)$ is accepted and $\hat{\theta}_k(x) = \tilde{\theta}_k(x)$, if $\tilde{\theta}_{k-1}(x)$ was accepted and

$$L\{W^{(\ell)}, \tilde{\theta}_\ell(x), \tilde{\theta}_k(x)\} \leq \mathfrak{z}_\ell, \ell = 1, \dots, k-1, \quad (18)$$

where $\hat{\theta}_k(x)$ is the latest accepted estimate after the first k steps. Note that (18) is equivalent to say that $\tilde{\theta}_k(x)$ is covered in all the confidence set of $\tilde{\theta}_\ell(x)$ ($\ell < k$).

The implementation of the procedure requires to fix the set of critical values of $\mathfrak{z}_1, \dots, \mathfrak{z}_{k-1}$. The proposed approach suggests to tune these parameter by the so called propagation condition. For every step k , we require that the estimate at step k $\hat{\theta}_k(x)$ is sufficiently close to the MLE in step k : $\tilde{\theta}_k$, in the sense that:

$$\mathbb{E}_{\lambda(\cdot), \theta^*} |L\{W^{(k)}, \tilde{\theta}_k(x), \hat{\theta}_k(x)\}|^r \leq \alpha \mathfrak{R}_r \quad (19)$$

α and r are two hyper-parameters. The role of α is similar to the significance level of a test, while r denotes the power of the loss function. A larger r and a smaller α would lead to an increase in critical values. Note that (19) relies on the artificial parametric model $P_{\lambda(\cdot), \theta^*}$ not the true model $P_{\lambda_0(\cdot), \theta(\cdot)}$. A choice of critical values \mathfrak{z}_l ($l \in 1, 2, \dots, k$) can be based on the following steps:

- Consider first only \mathfrak{z}_1 and let $\mathfrak{z}_2 = \dots = \mathfrak{z}_{K-1} = \infty$, leading to the estimates $\hat{\theta}_k(x)$ for $k = 2, \dots, K$.
- The value \mathfrak{z}_1 is selected as the minimal one for which

$$\sup_{\theta^*} \mathbb{E}_{\lambda(\cdot), \theta^*} \frac{|L\{W^{(k)}, \tilde{\theta}_k(x), \hat{\theta}_k(\mathfrak{z}_1, x)\}|^r}{\mathfrak{R}_r} \leq \frac{\alpha}{K-1}, k = 2, \dots, K.$$

- Set $\mathfrak{z}_{k+1} = \dots = \mathfrak{z}_{K-1} = \infty$ and fix \mathfrak{z}_k , which leads to the set of parameters $\mathfrak{z}_1, \dots, \mathfrak{z}_k, \infty, \dots, \infty$ and the estimates $\hat{\theta}_m(\mathfrak{z}_1, \dots, \mathfrak{z}_k)$ for $m = k+1, \dots, K$. Select \mathfrak{z}_k s.t.

$$\sup_{\theta^*} \mathbb{E}_{\lambda(\cdot), \theta^*} \frac{|L\{W^{(k)}, \tilde{\theta}_m(x), \hat{\theta}_m(\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_k, x)\}|^r}{\mathfrak{R}_r} \leq \frac{k\alpha}{K-1},$$

$$m = k+1, \dots, K.$$

The theorem below assures the existence of critical values constructed via the above technique by presenting upper bounds for the critical values.

Assume that for some constants $0 < u_0 < u < 1$, it holds $0 < \|V_k^{-1}V_{k-1}^2V_k^{-1}\|_\infty \leq u < 1$, define $h_k = C_1C_2^k$, $k = 1, \dots, K$.

Theorem 3.1 *Suppose that $r > 0, \alpha > 0$. Under the assumptions of probability bounds in (13), there are a_0, a_1, a_2 , s.t. the propagation condition is fulfilled with the choice of $\mathfrak{z}_k = a_0r \log(\rho^{-1}) + a_1r \log(h_K/h_k) + a_2 \log(nh_k)$.*

Proof.

We need

$$\mathbb{E}_{\lambda(\cdot), \theta^*} \{|L(W^{(k)}, \tilde{\theta}_k(x), \hat{\theta}_k(x))|^r\} \leq \alpha \mathfrak{R}_r, \quad (20)$$

first we prove,

$$\mathbb{E}_{\lambda(\cdot), \theta^*} \{|L(W^{(k)}, \tilde{\theta}_l(x), \hat{\theta}_k(x))|^r\} \leq \alpha(3\|V_k/V_l\|\mathfrak{a}_k/\mathfrak{a}_l u_0^{-1}[\mathfrak{z}_l + \mathfrak{Q}_0\{p\}] + u_0^{-1}[\mathfrak{z}_k + \mathfrak{Q}_0\{p\}])^r, \quad (21)$$

To prove (21), we apply the following:

Theorem Spokoiny (2011) Suppose (ED) and (\mathcal{L}) on $\Theta_0(R)$, where $R \leq \mu_0\lambda^*/(\nu_0\mathfrak{q})$ for $\mu_0 = \mathfrak{a}/(2\nu_0\mathfrak{q})$. Then for any $\mathfrak{z} > 0$ on a random set of probability at least $1 - e^{-\mathfrak{z}}$ it holds for all $\theta \in \Theta_0(R)$

$$L(\theta, \theta^*) \leq -\mathfrak{a}\|V_0(\theta - \theta^*)\|^2/4 + \mu_0^{-1}\{\mathfrak{z}_k + \mathfrak{Q}_0(p)\}$$

Similarly, if

$$-\mathbb{E}_{\lambda(\cdot), \theta^*} L(W, \theta, \theta^*) \geq \mathfrak{a}_1\|V_0(\theta - \theta^*)\|^2/2 \quad (22)$$

then it holds on the same random set

$$L(\theta, \theta^*) \geq -3\mathfrak{a}_1\|V_0(\theta - \theta^*)\|^2/4 - \mu_0^{-1}\{\mathfrak{z}_k + \mathfrak{Q}_0(p)\}, \theta \in \Theta_0(R)$$

To prove (21), we have for $u > 0$,

$$\begin{aligned}
|L(W^{(k)}, \tilde{\theta}_l, \theta^*)| &\leq \max\{L(W^{(k)}, \tilde{\theta}_l, \theta^*), -L(W^{(k)}, \tilde{\theta}_l, \theta^*)\} \\
&\leq \max\{L(W^{(k)}, \tilde{\theta}_l, \theta^*), -L(W^{(k)}, \tilde{\theta}_l, \theta^*)\} + uL(W^{(l)}, \tilde{\theta}_l, \theta^*) \\
&\leq \max_{\theta} \{\max\{L(W^{(k)}, \theta, \theta^*), -L(W^{(k)}, \theta, \theta^*)\} + uL(W^{(l)}, \theta, \theta^*)\} \\
&\leq \max\{-\mathbf{a}_k \|V_k(\theta - \theta^*)\|^2 + \mu_0^{-1}[\mathfrak{z}_k + \mathfrak{Q}_0(p)], \\
&\quad 3\mathbf{a}'_k \|V_k(\theta - \theta^*)\|^2/4 + \mu_0^{-1}[\mathfrak{z}_k + \mathfrak{Q}_0(p)]\} \\
&\quad -u\mathbf{a}_l \|V_l(\theta - \theta^*)\|^2/4 + u\mu_0^{-1}[\mathfrak{z}_l + \mathfrak{Q}_0(p)]
\end{aligned}$$

Using $u > 3\|V_k/V_l\|\mathbf{a}'_k/\mathbf{a}_l$ to guarantee $|L(W^k, \tilde{\theta}_l, \theta^*)| < \infty$, the proof follows.

To prove (20), define $\{\mathfrak{B}_k : \hat{k} = k - 1\}$, by Cauchy-Schwartz inequality,

$$\begin{aligned}
\mathbb{E}_{\lambda(\cdot), \theta^*} [L\{W^{(k)}, \tilde{\theta}_k(x), \hat{\theta}_k(x)\}^r] &= \sum_{l=1}^k \mathbb{E}_{\lambda(\cdot), \theta^*} \{|L(W^{(k)}, \tilde{\theta}_k(x), \tilde{\theta}_{l-1}(x))|^{2r}\}^{1/2} \mathbb{P}(\mathfrak{B}_l)^{1/2} \\
&\leq \sum_{l=1}^k 2^{(2r-1)+} [\mathbb{E}_{\lambda(\cdot), \theta^*} |L\{W^{(k)}, \tilde{\theta}_k(x), \theta^*\}|^{2r} + \mathbb{E}_{\lambda(\cdot), \theta^*} \{|W^{(k)}, \tilde{\theta}_l(x), \theta^*\}|^{2r}] \mathbb{P}(\mathfrak{B}_l)^{1/2} \\
&\leq \sum_{l=1}^k C\{\mathfrak{R}_{2r}\} (\mathbf{a}'_k \|V_k/V_l\|/\mathbf{a}_l)^{2r} \exp\{-C_1 \mathfrak{z}_l/2\}
\end{aligned}$$

4 “Small Modeling Bias” Condition

Now we extend the condition (20) to the situation when the parametric assumption is not precisely fulfilled but the deviation from it is small in a modeling bias sense. We measure the deviation via the Kullback Leibler divergence between the nonparametric measure and the assumed parametric measure:

$$\Delta_{\lambda(\cdot), \lambda_0(\cdot)}(W^{(k)}, \theta) = \sum_{i=1}^n \mathcal{K}\{\mathbb{P}_{\theta(X_i), \lambda_0(\cdot)}, \mathbb{P}_{\theta, \lambda(\cdot)}\} \mathbf{1}\{w_i^{(k)} > 0\},$$

where $\mathbb{P}_{\theta(X_i), \lambda_0(\cdot)}, \mathbb{P}_{\theta, \lambda(\cdot)}$ correspond to the marginal distributions of Y_i with respect to $\theta(\cdot)$ and θ . Here we tacitly assume that a misspecified $\lambda_0(\cdot)$ will not induce a big deviation, see Section 6 for evidence on this assumption, so assume:

$$\Delta_{\lambda(\cdot), \lambda_0(\cdot)}(W^{(k)}, \theta) \approx \Delta_{\lambda(\cdot), \lambda(\cdot)}(W^{(k)}, \theta)$$

We define a bound by the following “small modeling bias” (SMB) condition:

$$\Delta_{\lambda(\cdot), \lambda_0(\cdot)}(W^{(k)}, \theta) \leq \Delta, \tag{23}$$

and the oracle $k^* \stackrel{\text{def}}{=} \arg \max_k \{\Delta_{\lambda(\cdot), \lambda_0(\cdot)}(W^{(k)}, \theta) \leq \Delta\}$.

Lemma 1 *Let P, P_0 , be two measures s.t. $\mathbb{E} \log(dP/dP_0) \leq \Delta < \infty$, For any random variable \mathfrak{z} , with $\mathbb{E} \mathfrak{z} < \infty$, we have $\mathbb{E} \log(1 + \mathfrak{z}) \leq \Delta + \mathbb{E} \mathfrak{z}$.*

Proof: $f(x) = xy - x \log x + x$ attains maximum at the point $x = e^y$, thus $f(x) \leq f(e^y)$, and we have $xy \leq x \log x - x + e^y$. Let $x = dP/dP_0$ and $y = \log(1 + \mathfrak{z})$,

$$\begin{aligned} \mathbb{E}_0 dP/dP_0 \log(1 + \mathfrak{z}) &= \mathbb{E} \log(1 + \mathfrak{z}) \\ &\leq \mathbb{E}_0 (dP/dP_0 \log dP/dP_0 - dP/dP_0 + 1 + \mathfrak{z}) \\ &\leq \mathbb{E} \log \frac{dP}{dP_0} + \mathbb{E}_0 \mathfrak{z} \\ &\leq \Delta + \mathbb{E}_0 \mathfrak{z} \end{aligned}$$

From the above lemma and (23), we can derive the from the propagation condition the propagation property:

$$\mathbb{E}_{\lambda_0(\cdot), \theta(\cdot)} \log\{1 + |L(\tilde{\theta}_k, \hat{\theta}_k)|^r / \mathfrak{R}_r\} \leq \Delta + \alpha, \quad (24)$$

when the SMB condition is fulfilled:

$$\Delta_{\lambda(\cdot), \lambda_0(\cdot)}(W^{(k)}, \theta) \leq \Delta, \forall k < k^*.$$

5 Stability and Oracle Property

Due to the ‘‘propagation’’ result (20), the accuracy of the sequential test is guaranteed when the SMB assumption is fulfilled. In addition, we also need to make sure that when our final estimated step \hat{k} overshoots the oracle k^* ($\hat{k} > k^*$), the estimate does not vary too much. The stability condition shows that in the case of overshooting $\hat{k} > k^*$, the estimate is accurate enough in the sense that,

$$L(W^{(k^*)}, \tilde{\theta}_{k^*}, \hat{\theta}_{\hat{k}}) \mathbf{1}\{\hat{k} > k^*\} \leq \mathfrak{z}_{k^*}.$$

The ‘‘stability’’ property naturally follows from the setup of our test.

Combination of the ‘‘propagation’’ and ‘‘stability’’ statements implies the ‘‘oracle’’ property, under the SMB condition,

$$\mathbb{E} \log\{1 + \frac{L(W^{(k^*)}, \tilde{\theta}_{k^*}(x), \theta)^r}{\mathfrak{R}_r}\} \leq \Delta + 1 \quad (25)$$

$$\mathbb{E} \log\{1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x))|^r}{\mathfrak{R}_r}\} \leq \rho + \Delta \quad (26)$$

Proof. (25) is a trivial consequence of (24) and ‘‘stability’’. We now prove (26).

$$\begin{aligned} &\mathbb{E}_{\lambda_0(\cdot), \theta(\cdot)} \log\{1 + \frac{|L\{W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x)\}|^r}{\mathfrak{R}_r}\} \\ &= \mathbb{E}_{\lambda_0(\cdot), \theta(\cdot)} \log\{1 + \frac{|L\{W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x)\}|^r}{\mathfrak{R}_r}\} \mathbf{1}(\hat{k} \leq k^*) \\ &\quad + \mathbb{E}_{\lambda_0(\cdot), \theta(\cdot)} \log\{1 + \frac{|L\{W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x)\}|^r}{\mathfrak{R}_r}\} \mathbf{1}(\hat{k} > k^*) \\ &\leq \Delta + \mathbb{E}_{\lambda_0(\cdot), \theta(\cdot)} \frac{|L\{W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x)\}|^r}{\mathfrak{R}_r} + \log(1 + \frac{|L\{W^{(k^*)}, \tilde{\theta}_{k^*}(x), \tilde{\theta}_{\hat{k}}(x)\}|^r}{\mathfrak{R}_r}) \mathbf{1}(\hat{k} > k^*) \\ &\leq \Delta + \rho + \log(1 + \mathfrak{z}_{k^*} / \mathfrak{R}_r) \end{aligned}$$

Table 1: Critical Values with different r and α

$\alpha = 0.25, r = 0.5$	6.123	2.333	0.987	3.678e-05	0.000
$\alpha = 0.5, r = 0.5$	4.616	1.578	0.357	2.472e-05	0.000
$\alpha = 0.6, r = 0.5$	3.203	0.679	0.025	0.006	7.278e-05
$\alpha = 0.25, r = 0.75$	9.127	3.288	1.031	0.126	5.675e-05
$\alpha = 0.25, r = 1$	12.75	4.280	1.224	1.095e-04	0.000

Table 2: Critical Values with Different τ

$\tau = 0.05$	6.464	2.204	0.620	3.345e-05	0.000
$\tau = 0.5$	7.997	3.089	0.986	0.300e-05	0.000
$\tau = 0.75$	9.203	3.910	1.106	0.123	7.254e-05
$\tau = 0.95$	8.589	5.452	1.904	0.334	1.203e-05

6 Monte Carlo Simulation

This section aims at illustrating the local quantile estimation technique at different levels $\tau = 0.05, 0.5, 0.75, 0.95$ and for different noise distributions. Tail functions we select from a) Laplace, b) normal and c) student $t(3)$ distribution, the errors brought by misidentification of the noise distribution are also studied. The global bandwidth selection for quantile regression follows the proposal of Yu and Jones (1998), in which they consider a rule of thumb bandwidth based on the assumption that the quantiles are parallel. We also compare with the proposals of Cai and Xu (2008), where an approach based on a nonparametric version of the Akaike information criterion (AIC) is implemented. Global bandwidth selectors are compared with the localizing technique presented here.

6.1 Critical Values

The critical values are simulated via the ‘‘propagation condition’’ (19).

Table 1 shows the critical values with several choices of α and r with $\tau = 0.75$, $m = 10000$ Monte Carlo samples, and bandwidth sequence $(8, 14, 19, 25, 30, 36, 41, 52, 63) * 0.001$. Critical values decrease when α increases, and increase when r increases. The critical values for the last 3 bandwidths is actually equal to 0.

The bandwidth sequence in Table 2 displays the values of different critical values with different τ , $\alpha = 0.25$, $r = 0.5$, $m = 10000$ Monte Carlo samples, bandwidth sequence $(8, 14, 19, 25, 30, 36, 41, 52, 63) * 0.001$, and normal noise. Critical values are roughly of the same level with respect to different τ .

Table 3 displays the critical values for three alternative bandwidth sequences, i.e. $(8, 16, 25, 36, 49, 63, 79, 99) * 0.001$, $(5, 8, 14, 19, 27, 36, 46, 58, 66) * 0.001$, $(8, 14, 19, 25, 30, 36, 41, 52, 63) * 0.001$, with $\alpha = 0.25$, $r = 0.5$, and $\tau = 0.85$. We see that different bandwidth sequence would return us different level of critical values. In applications, critical values are chosen therefore in a scenario based fashion.

The critical values in Table 4 are simulated under t -distribution with three degree of

Table 3: Critical Values with Different Bandwidth Sequences

hseq1	11.33	1.243	6.933e-05	0.000	0.000
hseq2	18.39	6.479	2.230	0.469	8.738e-05
hseq3	6.123	2.333	0.987	3.678e-05	0.000

freedom, normal distribution and ALD(0, 1,0.5). The critical values apparently increase when the distribution tails get fatter.

Table 4: Critical Values with Different Noise Distributions

N(0,1)	11.50	4.924	2.514	1.313	2.765e-05
ALD(0,1,0.5)	14.05	6.554	3.304	1.443	5.879e-05
t(3)	15.42	8.707	2.370	0.342	3.898e-05

In Table 5, critical values are shown in the same circumstances as in Table 4 except for changing to the local linear case. Since we introduce one more variable (trend) to estimate, critical values doubled or tripled compared to the local constant case. The behavior with respect to tail functions stays the same.

Table 5: Local Linear

N(0,1)	29.97	58.64	43.21	33.41	19.43	07.40
ALD(0,1,0.5)	45.28	74.51	66.43	50.42	31.42	13.50
t(3)	51.77	84.94	59.28	44.99	29.07	11.57

6.2 Comparison of Different Bandwidth Selection Techniques

We illustrate our proposal by considering $x \in [0, 1]$, and estimation of the conditional quantile function $l_\tau(x)$ at point $x = 0.5$, $\tau = 0.75$. The sample with ($n = 1000$) are simulated under three scenarios:

$$f^{[1]}(x) = \begin{cases} 0 & \text{if } x \in [0, 0.333]; \\ 8 & \text{if } x \in (0.333, 0.666]; \\ -1 & \text{if } x \in (0.666, 1] \end{cases}$$

$$f^{[2]}(x) = 2x(1 + x)$$

$$f^{[3]}(x) = \sin(k_1x) + \cos(k_2x)\mathbf{1}\{x \in (0.333, 0.666)\} + \sin(k_2x)$$

The noise distributions considered are:

- Normal distribution with 0 mean and 0.03 as variance

- ALD(0,2,0.5)
- t -distribution with 3 degree of freedom

Figure 2 presents pictures with different noise distributions and comparison of boxplots of difference between estimations and the true curves in the local constant case. Figure 3 and 4 show in the local linear case function estimations and the first derivatives as well. The adaptive method always outperforms methods with fixed bandwidth, especially in the presence of jump. Table 6 further confirms our conclusion, by errors using four methods are evaluated with 1000 samples.

Table 6: Comparison of Monte Carlo errors at $x = 0.5$ with 1000 samples

	Fixed bandw	Local Constant	Local linear	Fixed bandw (Cai)
$f^{[1]}(x)$	0.654	0.172	0.169	0.378
$f^{[2]}(x)$	0.206	0.008	0.008	0.245
$f^{[3]}(x)$	0.137	0.021	0.019	0.123

Table 7 offers an evaluation of misidentification errors with critical values simulated from ALD distribution while data are simulated from other distributions. Table 7 gives L_1 errors between \hat{l} and l and lets us conclude that misspecification of tail loss would not contaminate our results significantly.

Table 7: Comparison of Error Misspecification

	Local Constant {N(0,1)}	Local Constant {t(3)}	Local linear {N(0,1)}
$f^{[1]}(x)$	0.252	0.220	0.169
$f^{[2]}(x)$	0.070	0.016	0.043
$f^{[3]}(x)$	0.009	0.021	0.019

7 Application

In the study of financial products, it is very important to detect and understand tail dependence among underlyings such as stocks. In particular, the tail dependence structure represents the degree of dependence in the corner of the lower-left quadrant or upper-right quadrant of a bivariate distribution. Hauksson, Michel, Thomas, Ulrich and Gennady (2001) and Embrechts and Straumann (1999) provide good access to the literature on tail dependence and Value at Risk. With the adaptive quantile technique, we provide an alternative approach to study tail dependence.

Figure 5 shows the shape of conditional quantile curves from a bivariate normal random sample. The correlation is calibrated from real data as given in Figure 6, where X is standardized returns from stock “clpholdings” from Hong Kong Hangseng Index, and Y is returns from stock “cheung kong”. Figure 6 and Figure 7 show the empirical conditional quantile curves actually deviate from the one calculated from normal distributions in Figure 5, which implies non normality. The motivation of doing adaptive bandwidth selection is clear to see, from Figure 6 and Figure 7, the dependency structure change is

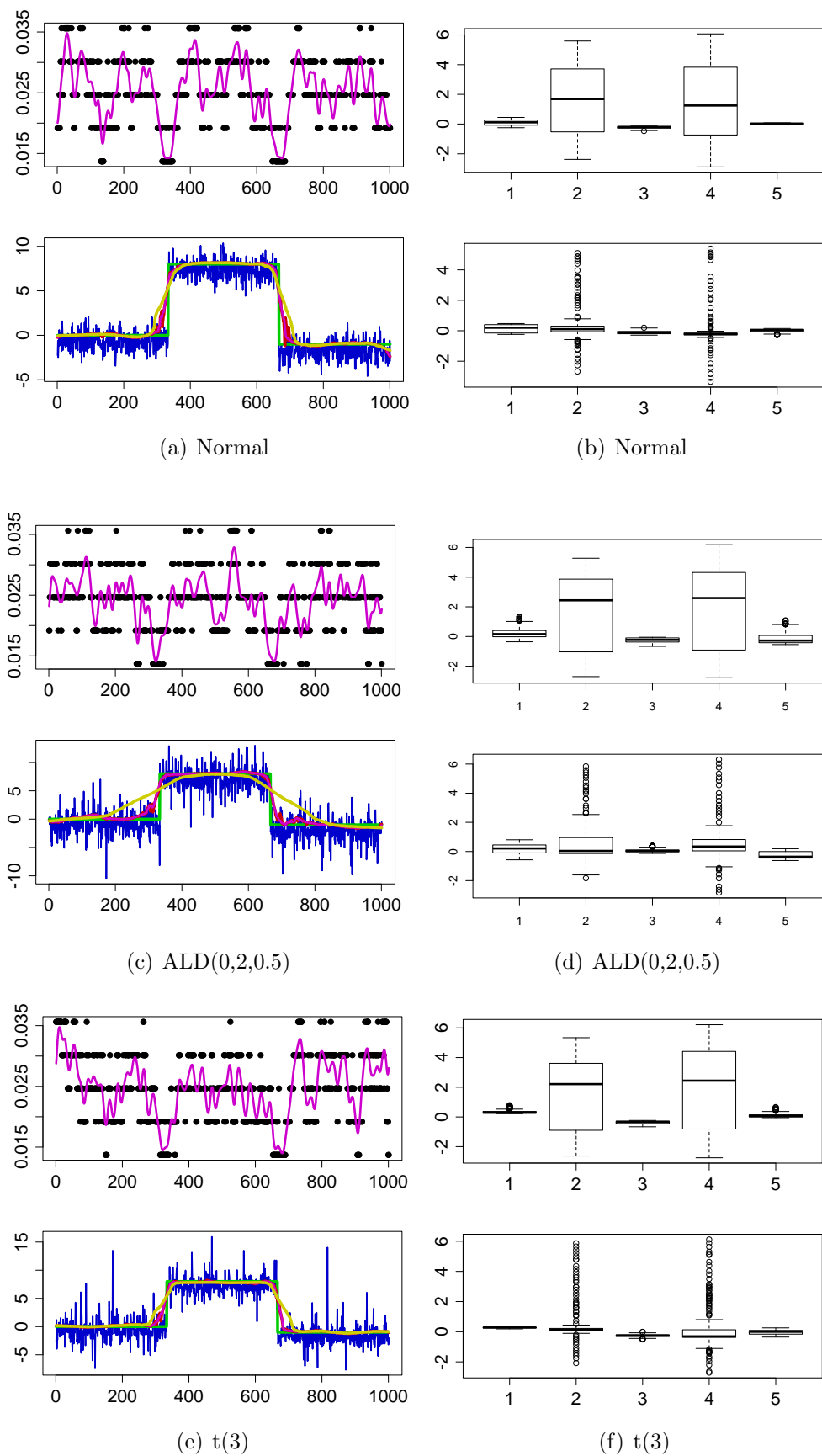


Figure 2: The bandwidth sequence (upper left panel), the data with noise (blue) , the adaptive estimation of 0.75 quantile (red), the quantile smoother with fixed optimal bandwidth = 0.06 (yellow); boxplot of block residuals fixed bandwidth (upper right), adaptive bandwidth (lower right)

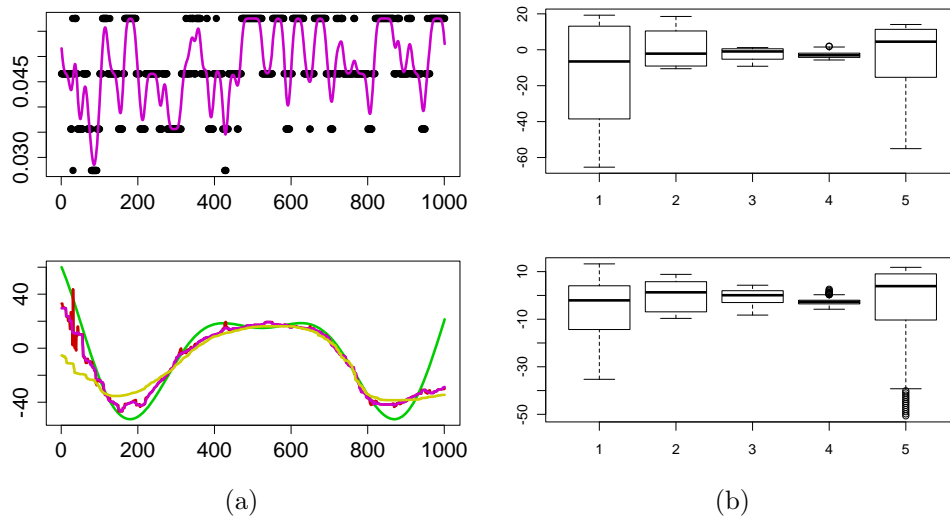


Figure 3: The bandwidth sequence (upper left panel), the data with $t(3)$ noise (blue), the adaptive estimation of 0.75 quantile (blue), the quantile smoother with fixed optimal bandwidth = 0.06 (yellow); The blocked residual (right)

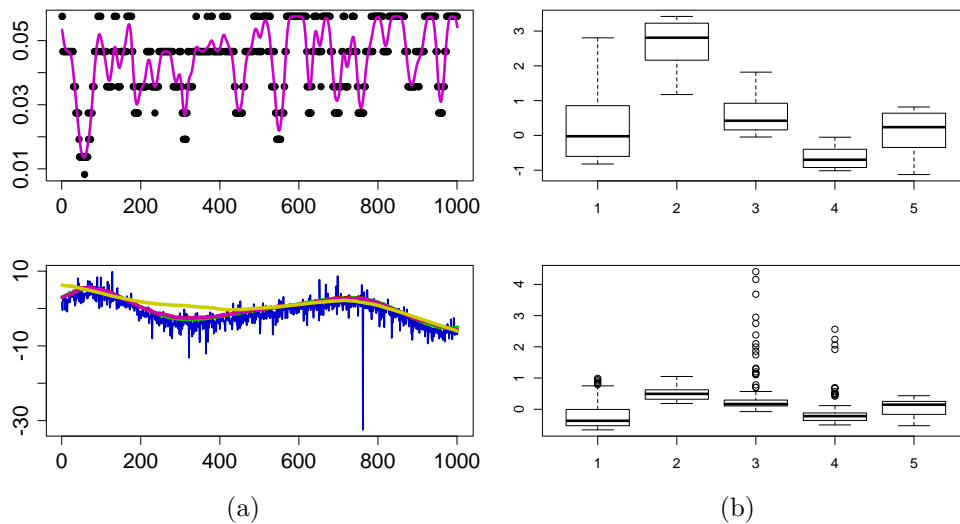


Figure 4: The bandwidth sequence (upper panel), the data with $ALD(0,2,0.5)$ noise (blue), the adaptive estimation of 0.85 quantile (blue), the quantile smoother with fixed optimal bandwidth = 0.06 (yellow)

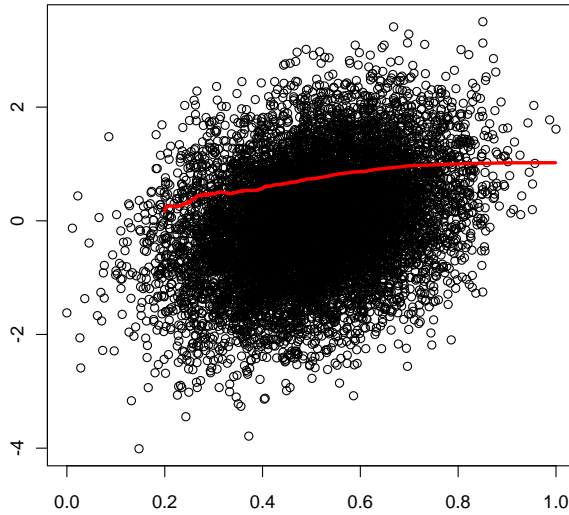


Figure 5: Plot of quantile curve for two normal random variables with $\rho = 0.3204$, $\tau = 0.75$

more obvious compared with the fixed bandwidth curve. Moreover, the flexible adaptive curve is not a consequence of overfitting since it still mostly lies in the confidence bands produced by fixed bandwidth estimation, see Härdle and Song (2010).

Figure 8 shows the first derivative curve for above example. The curve gets more volatile while x increases until a drastically change, then it turns flat.

We measure deviation from normality by accumulated L_1 distance and examine different combination of stocks from Hong Kong Hangseng Index. The results is summarized in Table 8.

Table 8: Summary of Combination of Stocks

	Chalco	Cosco pacific	Bank of China
New world devo	0.252	0.220	0.169
Sino land	0.070	0.016	0.043
Swire pacific A	0.009	0.021	0.019

Another application of quantile function estimation is in temperature data analysis, which is of key interest to price temperature derivatives. Quantile regression can provide a more flexible and complete approach to understand the temperature risk drivers.

Denote daily temperature as $T \mapsto (t, j)$, with $t = 1, \dots, \tau = 365$ days, $j = 0, \dots, J$ years. The time series decomposition for $T_{t,j}$ is given as:

$$\begin{aligned}
 X_{t,j} &= T_{t,j} - \Lambda_t \\
 X_{t,j} &= \sum_{l=1}^L \beta_l X_{t-l,j} + \sigma_t \varepsilon_{t,j} \\
 \varepsilon_{t,j} &\sim \mathbf{N}(0, 1),
 \end{aligned} \tag{27}$$

$$\hat{\varepsilon}_{t,j} = X_{365j+t} - \sum_{l=1}^L \hat{\beta}_l X_{365j+t-l} \tag{28}$$

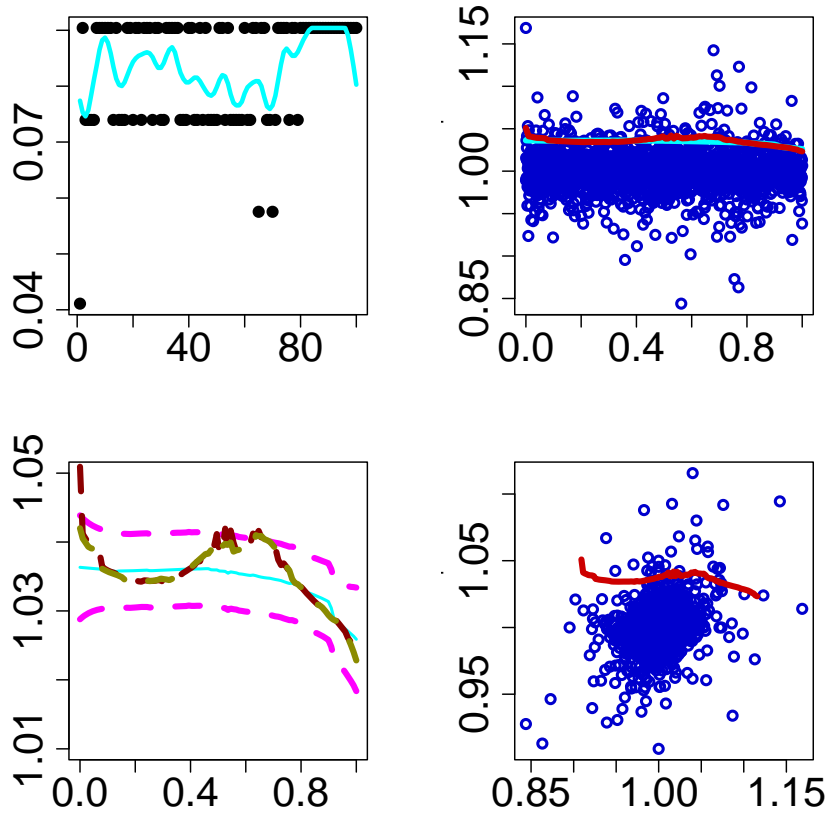


Figure 6: The bandwidth sequence with smoothed bandwidth curve(upper left panel); Scatter plot of stock returns (upper right panel) , the adaptive estimation of 0.75 quantile (red), the quantile smoother with fixed optimal bandwidth = 0.06 (cyan); fixed bandwidth curve (cyan), adaptive bandwidth curve (red dashed), confidence band (magenta) (lower left panel); adaptive bandwidth with normal scale (lower right panel)

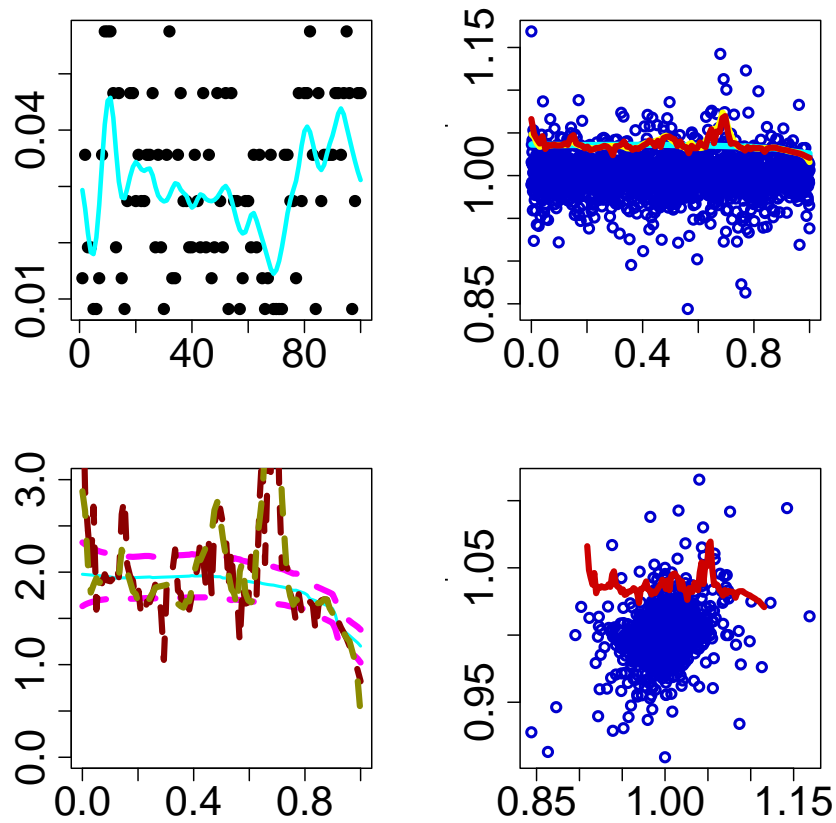


Figure 7: The bandwidth sequence with smoothed bandwidth curve (upper left panel); Scatter plot of stock returns (upper right panel), the adaptive estimation of 0.75 quantile (red), the quantile smoother with fixed optimal bandwidth = 0.06 (cyan); fixed bandwidth curve (cyan), adaptive bandwidth curve (red dashed), confidence band (magenta) (lower left panel); adaptive bandwidth with normal scale (lower right panel)

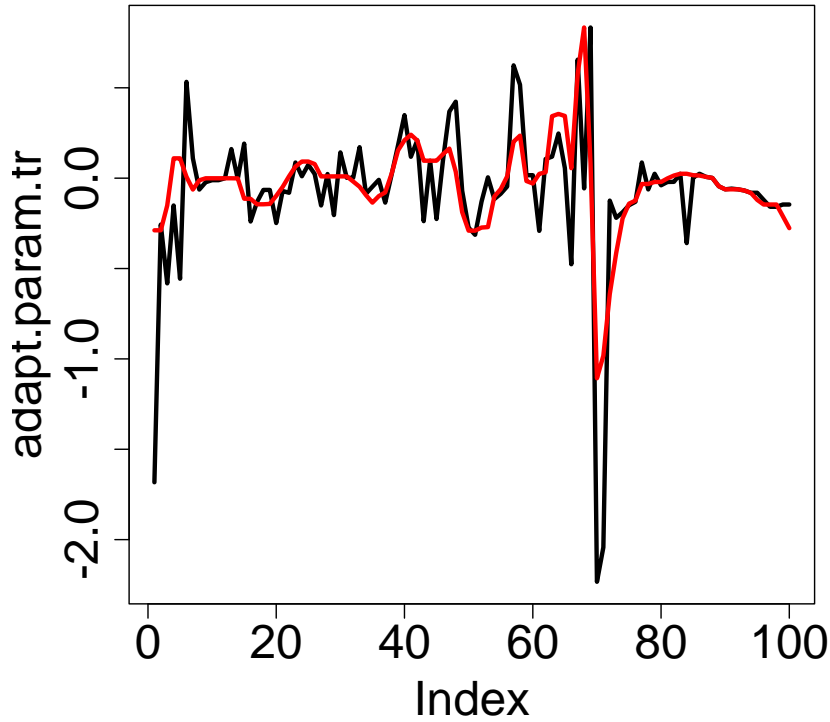


Figure 8: The adaptive trend curve (black), smoothed adaptive curve (red). $\tau = 0.75$

where $T_{t,j}$ is the temperature at day t in year j , Λ_t denotes the seasonality effect and σ_t the seasonal volatility.

We are interested specifically in the stochastic risk drivers $\varepsilon_{t,j}$, Figure 9 presents a time series plot of $\hat{\varepsilon}_{t,j}/\hat{\sigma}_t$, and the estimated 90% quantile function. By zooming in the curve, we observe a very interesting phenomena: an increase of trend of the standardized residual over years, which implies the existence of a potentially non stationarity trend possibly caused by global warming.

To further understand the risk factors, we analyze the quantile functions of $\hat{\varepsilon}_{t,j}^2$ over 12 years, and average the over every 4 years for comparison, see Figure 10 and Figure 11. The differences between Berlin and Kaoshiung are easy to see, the variance function peaks from Jan-Feb, while for Berlin the peaks come more in summer. Moreover, there is a tendency for Kaoshiung to be more volatile while this phenomenon does not appear in Berlin.

Table 9 summarizes statistics from the normality test of standardized residuals from three methods in Kaoshiung. The first method is to estimate $\hat{\sigma}$ via median curve, the second method is to estimate $\hat{\sigma}$ by $\{l_{\varepsilon,0.75} - l_{\varepsilon,0.25}\}/1.34$ (1.34 is the inter quartile range of a standard normal distribution), the third method is to get the conditional mean. The fixed mean (fourth method) and adaptive bandwidth are compared.

A general fact from Table 9 and 10 is that Berlin has more normal residuals than Kaoshiung. Our method three mean regression is always better in getting more normal residuals, and method two from adaptive quantile is compatible with method three. It means quantiles at higher or lower levels are better to explain the extreme happened in volatility function. The fix bandwidth as usual performs poorly. Therefore we conclude that our adaptive technique is useful in modeling temperature residuals.

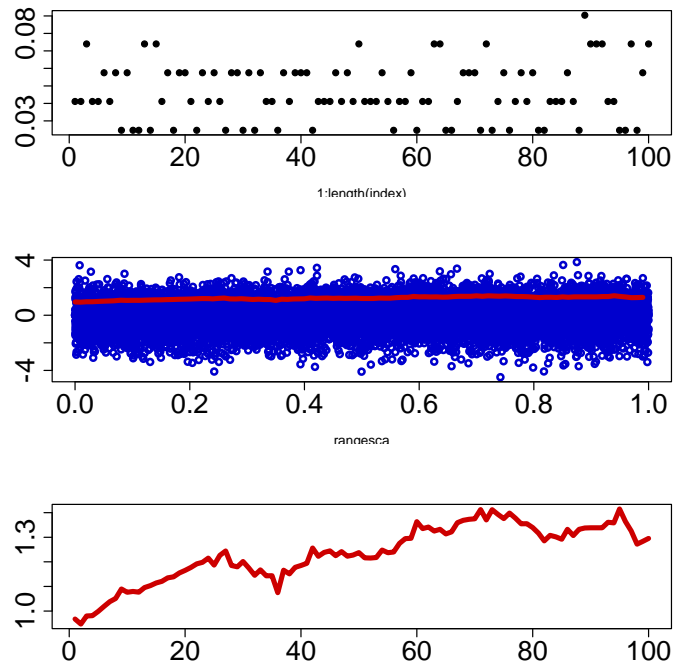


Figure 9: Plot of quantile for standardized weather residuals over 51 years at Berlin, 90% quantile .

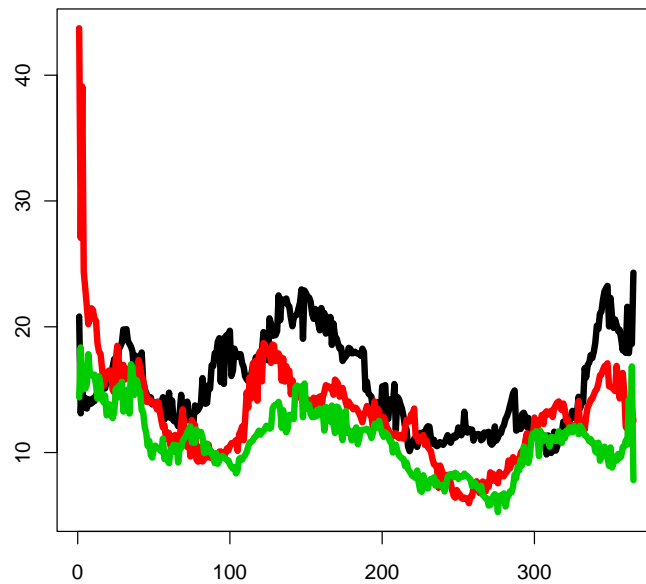


Figure 10: Estimated 90% quantile of variance functions, Berlin, average over 1995 – 1998, 1999 – 2002 (red), 2003 – 2006 (green)

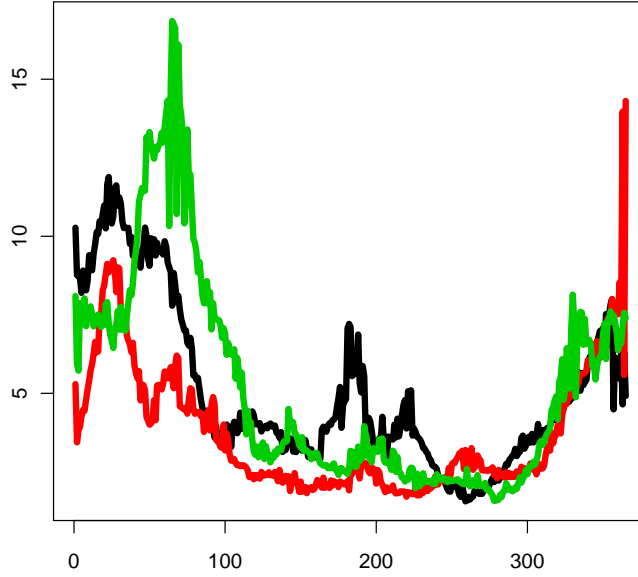


Figure 11: Estimated 90% quantile of variance functions, Kaoshiung, average over 1995 – 1998, 1999 – 2002 (red), 2003 – 2006 (green)

Table 9: P-values of Normality Tests:Berlin

	AD	JB	KS
1	0.000	0.010	0.060
2	0.062	0.000	0.020
3	0.054	0.487	0.171
4	0.009	0.000	0.002

8 Appendix

8.1 The Local Constant Case

The tail function behavior implies that $\{y\lambda^+(y)\}' \in [0, 1]$ ($\{y\lambda^-(y)\}' \in [-1, 0]$) and hence:

$$|y\lambda^{+'}(y)| = |\{y\lambda^+(y)\}' - \lambda^+(y)| < 1.$$

Let

$$\begin{aligned} m(\theta) &\stackrel{\text{def}}{=} \mathbf{E}_{\lambda(\cdot), \theta^*} \rho_\tau(|Y - \theta|), \\ q(\theta) &\stackrel{\text{def}}{=} \mathbf{P}(Y \leq \theta) - \mathbf{P}(Y > \theta) \\ q^{(1)}(\theta) &\stackrel{\text{def}}{=} (1 - \tau)\mathbf{P}(Y \leq \theta) - \tau\mathbf{P}(Y \geq \theta) \end{aligned}$$

and observe $m'(\theta) \stackrel{\text{def}}{=} (1 - \tau)\mathbf{P}(Y < \theta) - \tau\mathbf{P}(Y \geq \theta)$.

Next, define $\ell(y, \theta, 0)$ as the likelihood ratio function for one observation,

$$\ell(y, \theta, 0) = (2\tau - 1)\theta - \{|Y - \theta| - |Y|\}$$

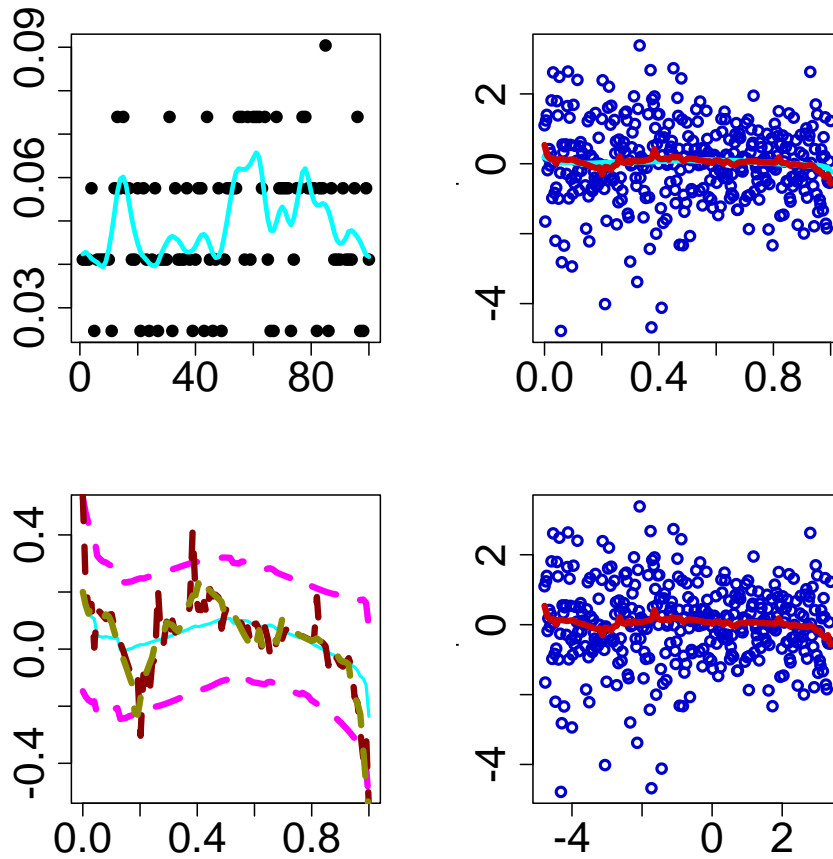


Figure 12: The bandwidth sequence with smoothed bandwidth curve (upper left panel); Scatter plot of temperature residuals with (upper right panel) , the adaptive estimation of 0.5 quantile (red) , the quantile smoother with fixed optimal bandwidth = 0.06 (cyan); fixed bandwidth curve (cyan), adaptive bandwidth curve (red dashed), confidence band (magenta) (lower left panel); adaptive bandwidth with normal scale (lower right panel)

Table 10: P-values of Normality Tests:Kaoshiung

	AD	JB	KS
1	0.000	0.000	0.000
2	1.03e-05	0.077	0.043
3	2.37e-06	0.742	0.674
4	0.000	0.021	0.019

It is also clear that $|q(\theta)| \leq 1$. For $\theta \geq 0$, it holds

$$\ell'(y, \theta, 0) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} \ell(y, \theta, 0) = \begin{cases} 2, & y \in (0, \theta), \\ 0, & \text{otherwise,} \end{cases}$$

and $\ell(y, \theta, 0) = 2\tau\theta - 2\theta$ for $y < 0$. Therefore, integration by parts yields

$$\begin{aligned} \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\mu\ell(Y, \theta, 0)\} &= - \int \exp\{\mu\ell(y, \theta, 0)\} dP(Y > y) \\ &= \exp\{\mu(2\tau\theta - 2\theta)\} + 2\mu \int_0^\theta \ell'(y, \theta, 0) \exp\{\mu\ell(y, \theta, 0)\} P(Y > y) dy \\ &= \exp\{\mu(2\tau\theta - 2\theta)\} + \mu \int_0^\theta \exp\{\mu(2\tau\theta - 2\theta + 2y)\} \tau \exp\{-\frac{1}{\tau}y\lambda^+(y)\} dy \\ &= \exp\{\mu(2\tau\theta - 2\theta)\} + 2\mu\tau \exp\{\mu(2\tau\theta - 2\theta)\} \int_0^\theta \exp[(y/\tau)\{\lambda^+(\theta) - \lambda^+(y)\}] dy, \end{aligned}$$

where we fix $\mu(\theta) = \lambda^+(\theta)/\{2\tau\}$. Monotonicity of $\lambda^+(y)$ implies

$$\begin{aligned} \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\mu(\theta)\ell(Y_1, \theta, 0)\} &\leq \exp\{\mu(2\tau\theta - 2\theta)\} (1 + 2\mu\tau\theta) \\ &= \exp\{\lambda^+(\theta)(2\tau\theta - 2\theta)/2\tau\} (1 + \lambda^+(\theta)\theta). \end{aligned}$$

Therefore, for $\theta > 0$,

$$\mathfrak{M}(\theta, 0) = -\log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\mu(\theta)\ell(Y, \theta, 0)\} \geq -\log\{1 + \theta\lambda^+(\theta)\} + \left(\frac{-2\tau + 2}{2\tau}\right)\theta\lambda^+(\theta). \quad (29)$$

For $\theta < 0$, take $\mu = \lambda^-(\theta)/2(1 - \tau)$ the monotonicity of $\lambda^-(y)$ implies

$$\mathfrak{M}(\theta, 0) = -\log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\mu(\theta)\ell(Y, \theta, 0)\} \geq -\log\{1 + \theta\lambda^-(\theta)\} + \left\{\frac{\tau}{(1 - \tau)}\right\}\theta\lambda^-(\theta)$$

(29) states the identifiability of the function, now we prove the other condition to achieve the risk bound in Golubev and Spokoiny (2009).

Define

$$h(\delta, \gamma; \theta) \stackrel{\text{def}}{=} \log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\left\{2\delta \frac{\gamma^\top \nabla \zeta_0(\theta)}{\sqrt{\gamma^\top v(\theta) \gamma}}\right\}$$

where

$$v(\theta) \stackrel{\text{def}}{=} \mathbf{E} \nabla \zeta_0(\theta)^\top \nabla \zeta_0(\theta).$$

The smoothness condition is shown as:

$$h(\delta, \gamma; \theta) \leq 2\nu_0^2\delta^2, \delta < \bar{\delta}, \nu_0 \geq 1. \quad (30)$$

To prove (30), define,

$$\zeta_0(\theta) \stackrel{\text{def}}{=} \mathbf{E}_{\lambda(\cdot), \theta^*}(-|Y - \theta| + |Y|) - (-|Y - \theta| + |Y|),$$

Then, for $\theta > 0$,

$$\begin{aligned} \nabla \zeta_0(\theta) &= \mathbf{1}(Y \leq \theta) - \mathbf{1}(Y > \theta) - q(\theta) \\ \mathbf{E}_{\lambda(\cdot), \theta^*} |\nabla \zeta_0(\theta)|^2 &= 1 - q^2(\theta) \\ \text{Var} \nabla \zeta_0 &= \text{Var} \int_0^\theta \nabla \zeta_0(u) du \leq \theta \int_0^\theta \mathbf{E}_{\lambda(\cdot), \theta^*} |\nabla \zeta_0(u)|^2 du = \theta \int_0^\theta \{1 - q^2(u)\} du \end{aligned}$$

and

$$\theta^{-2} \text{Var} \zeta_0(\theta) = \theta^{-1} \int_0^\theta \{1 - q^2(u)\} du \rightarrow 0, \theta \rightarrow \infty$$

because $q(\theta) \rightarrow 1$. Therefore,

$$\nabla \zeta_1(\theta) = \partial \zeta_1(\theta) / \partial \theta = \lambda^+(\theta) \nabla \zeta_0(\theta) + \theta \lambda^{+'}(\theta) \zeta_0(\theta) / \theta.$$

The conditions $|\nabla \zeta_0(\theta)| \leq 1$, $|\zeta_0(\theta) / \theta| \leq 1$, $|\theta \lambda^{+'}(\theta)| \leq 1$, $\lambda^+(\theta) \rightarrow 0$ and $\text{Var}\{\zeta_0(\theta) / \theta\} \rightarrow 0$ as $\theta \rightarrow \infty$, easily imply $h(\delta, \gamma; \theta) \leq 2\nu_0^2\delta^2$ for some fixed $\delta < \bar{\delta}, \nu_0 \geq 1$, thus the smoothness condition (30) is satisfied.

Moreover, if $\mathbf{E}_{\lambda(\cdot), \theta^*} |Y_1|^\gamma < \infty$ for some $\gamma > 0$, which leads to bound for the loss $\tilde{u} = |\tilde{\theta} - \theta^*|$:

Define

$$\Omega(\rho, s)_{\lambda(\cdot), \theta^*} \stackrel{\text{def}}{=} \mathbf{E}_{\lambda(\cdot), \theta^*} \sup_{\theta \in \Theta} \exp[\rho\{\mu(\theta)L(\theta, \theta^*) + s\mathfrak{M}(\theta, \theta^*)\}]$$

The conditions of Theorem 3.2 in Golubev and Spokoiny (2009) are therefore satisfied, so we have

$$\Omega(\rho, s) \leq C' |(1 - \rho)(1 - s)|^{-1/2}$$

and we can have further,

$$\mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\rho^2 n [\tilde{u} \lambda^+(\tilde{u}) - \log\{1 + \tilde{u} \lambda^+(\tilde{u})\}]\} \leq \frac{C}{(1 - \rho)}$$

with some fixed constant C', C provided that n exceeds some minimal sample size n_0 .

To prove (5), we assume that $\lambda^+(\theta)$ is bounded from below by $\mu^+(z) > 0$ in every set $\mathcal{A}(z, \theta^*) \stackrel{\text{def}}{=} \{\theta : \mathfrak{M}(\theta, \theta^*) \leq z\}$. Then we have:

$$\begin{aligned} \mathbf{P}\{L(\tilde{\theta}, \theta^*) > \mathfrak{z}\} &\leq \mathbf{P}\{L(\tilde{\theta}, \theta^*) > \mathfrak{z}, \tilde{\theta} \in \mathcal{A}(z, \theta^*)\} + \mathbf{P}\{\tilde{\theta} \notin \mathcal{A}(z, \theta^*)\} \\ &\leq \Omega(\rho, 0) \exp\{-\rho \mu^+(z) \mathfrak{z}\} + \Omega(\rho, s) \exp\{-\rho s z\} \end{aligned}$$

Set $\mu^+(z) \mathfrak{z} = sz$, then

$$\mathbf{P}\{L(\tilde{\theta}, \theta^*) > \mathfrak{z}\} \leq 2\Omega(\rho, s) \exp\{-\rho \mu_+(z) \mathfrak{z}\}$$

For the moment bound, we have:

$$\begin{aligned}
& \mathbf{E}_{\lambda(\cdot), \theta^*} |L(\tilde{\theta}, \theta^*)|^r \\
& \leq r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} P_{\theta^*, \lambda(\cdot)} \{L(\tilde{\theta}, \theta^*) > \mathfrak{z}\} d\mathfrak{z} \\
& \leq r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} P_{\theta^*, \lambda(\cdot)} \{\theta^* \notin \varepsilon(\mathfrak{z}), \tilde{\theta} \in \mathcal{A}(z(\mathfrak{z}), \theta^*)\} d\mathfrak{z} + r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} P[\tilde{\theta} \notin \mathcal{A}\{z(\mathfrak{z}), \theta^*\}] d\mathfrak{z} \\
& \leq r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} \Omega(\rho, 0) \exp[-\rho \mu^+ \{z(\mathfrak{z})\} \mathfrak{z}] d\mathfrak{z} + r \int_{\mathfrak{z} \geq 0} \mathfrak{z}^{r-1} \Omega(\rho, s) \exp\{-\rho s z(\mathfrak{z})\} d\mathfrak{z} \\
& = B_1 + B_2
\end{aligned}$$

when $\mu^+(z)\mathfrak{z} = sz$, $z(\mathfrak{z}) > \log(\mathfrak{z})$, then we have $B_1 < \infty$, $B_2 < \infty$. So we conclude that exists $0 < \tau_r < \infty$ such that $\mathbf{E}_{\lambda(\cdot), \theta^*} |L(\tilde{\theta}, \theta^*)|^r < \infty$.

8.2 Proof of risk bound

8.2.1 Proof of (ED)

We know that

$$\begin{aligned}
\nabla L(W, \theta) &= - \sum_{i=1}^n \rho'_\tau(Y_i - \theta^\top \psi_i) w_i \\
&= \sum_{i=1}^n \psi_i [\tau - \mathbf{1}\{(Y_i - \theta^\top \psi_i) < 0\}] w_i
\end{aligned}$$

Thus, we have,

$$\mathbf{E}_{\lambda(\cdot), \theta^*} \{\nabla L(W, \theta)\} - \nabla L(W, \theta) = \sum_{i=1}^n \psi_i (P\{(Y_i - \theta^\top \psi_i) < 0\} - \mathbf{1}\{(Y_i - \theta^\top \psi_i) < 0\})$$

Denote $\varepsilon_i(\theta) = P\{(Y_i - \theta^\top \psi_i) < 0\} - \mathbf{1}\{(Y_i - \theta^\top \psi_i) < 0\}$.

$$\begin{aligned}
\zeta(\theta) &= \mathbf{E}_{\lambda(\cdot), \theta^*} \{\nabla L(W, \theta)\} - \nabla L(W, \theta) \\
&= \sum_{i=1}^n \psi_i w_i \varepsilon_i(\theta)
\end{aligned}$$

If $\varepsilon_i(\theta)$ are identically distributed, $\mathbf{1}\{(Y_i - \theta^\top \psi_i) < 0\}$ is a Bernoulli random variable, and $P\{(Y_i - \theta^\top \psi_i) < 0\} = p$. Suppose $0 < \delta < \delta^* < 1$, and we have,

$$\begin{aligned}
\log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\delta \varepsilon\} &= \log[\exp\{\delta(p-1)\}p + \exp(\delta p)(1-p)] \\
&\leq p(1-p)\nu_0 \delta^2/2,
\end{aligned}$$

note that ν_0 depends on δ^* only. Therefore it holds for any $\gamma \in \mathbb{R}^p$ and $\rho > 0$ with $\rho|\psi_i^\top \gamma| \leq 1$ that,

$$\begin{aligned} \log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\rho \gamma^\top \nabla \zeta(\theta)\} &\leq \log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\left\{\rho \sum_{i=1}^n \gamma^\top w_i \psi_i \varepsilon_i(\theta)\right\} \\ &\leq \sum_{i=1}^n \log \mathbf{E}_{\lambda(\cdot), \theta^*} \exp\{\rho \gamma^\top w_i \psi_i \varepsilon_i(\theta)\} \\ &\leq \sum_{i=1}^n \rho^2 (\gamma^\top \psi_i w_i)^2 \sigma\{p_i(\theta)\}^2 \{\varepsilon_i(\theta)\} \nu_0 / 2 \\ &\leq \nu_0 \rho^2 \|V(\theta) \gamma\|^2 \end{aligned}$$

where $\sigma\{p_i(\theta)\}^2 \leq 1/4$ and

$$V^2(\theta) = \sum_{i=1}^n \sigma^2(\psi_i^\top \theta) \psi_i \psi_i^\top w_i^2.$$

Denote also $V^2 = (1/4) \sum_{i=1}^n \psi_i \psi_i^\top w_i^2$, and $V(\theta) \leq V$ for all θ . Then the condition (ED) is fulfilled with the matrix V and $\lambda^* = \delta^* N^{1/2}(\Psi)$ for $N(\Psi)$ defined as:

$$N^{-1/2}(\Psi) \stackrel{\text{def}}{=} \max_i \sup_{\gamma \in \mathbb{R}^p} \frac{\gamma^\top \psi_i w_i \sigma_i}{\|V \gamma\|}$$

8.2.2 The (\mathcal{L}) Condition

We know that

$$\frac{\partial \mathbf{E}_{\lambda(\cdot), \theta^*} L(W, \theta)}{\partial \theta} = - \sum_{i=1}^n \psi_i [\tau - \mathbf{P}\{(Y_i - \theta^\top \psi_i) < 0\}] w_i$$

and

$$\frac{\partial^2 \mathbf{E}_{\lambda(\cdot), \theta^*} L(W, \theta)}{\partial^2 \theta} = \sum_{i=1}^n \psi_i \psi_i^\top p_i(\psi_i^\top \theta) w_i \stackrel{\text{def}}{=} H^2(\theta)$$

Recall that $-\nabla \mathbf{E}_{\lambda(\cdot), \theta^*} L(W, \theta^*) = 0$. Now we take Taylor expansion of $-\mathbf{E}_{\lambda(\cdot), \theta^*} L(W, \theta, \theta^*)$, we conclude that, there is $\theta^o \in [\theta, \theta^*]$ such that

$$\begin{aligned} -\mathbf{E}_{\lambda(\cdot), \theta^*} L(W, \theta, \theta^*) &= \sum_{i=1}^n |\psi_i^\top (\theta - \theta^*)|^2 p_i(\psi_i^\top \theta^o) w_i \\ &= (\theta - \theta^*)^\top H^2(\theta^o) (\theta - \theta^*). \end{aligned}$$

Thus the condition (\mathcal{L}) is fulfilled in $H(\theta) - aV(\theta^*) \geq 0$ p.d. for $\theta \in \Theta_0(R)$.

8.2.3 Bound for the rate function

Next, for $\theta \geq \theta^*$, it is also clear that $\psi^\top \theta \geq \psi^\top \theta^*$.

$$\ell(y, \theta, \theta^*) = (2\tau - 1)(\psi^\top \theta - \psi^\top \theta^*) - \{|Y - \psi^\top \theta| - |Y - \psi^\top \theta^*|\}$$

$$\ell'(y, \theta, \theta^*) \stackrel{\text{def}}{=} \frac{\partial}{\partial y} \ell(y, \theta, \theta^*) = \begin{cases} 2, & y \in (\psi^\top \theta^*, \psi^\top \theta), \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, define $A_y = [y, \infty)$, integration by parts yields

$$\begin{aligned} \mathbb{E}_{\lambda(\cdot), \theta^*} \exp\{\mu \ell(Y, \theta, \theta^*)\} &= - \int \exp\{\mu \ell(y, \theta, \theta^*)\} dP(A_y) \\ &= \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} + \mu \int_{\psi_i^\top \theta^*}^{\psi_i^\top \theta} \ell'(y, \theta, \theta^*) \exp\{\mu \ell(y, \theta, \theta^*)\} P(A_y) dy \\ &= \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} \\ &\quad + 2\mu \int_{\psi_i^\top \theta^*}^{\psi_i^\top \theta} \exp[\mu\{(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*) + 2Y - 2\psi_i^\top \theta^*\}] P(A_y) dy \\ &= \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} \\ &\quad + 2\mu \int_{\psi_i^\top \theta^*}^{\psi_i^\top \theta} \exp[\mu\{(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*) + 2Y - 2\psi_i^\top \theta^*\}] \tau \exp\{-\lambda^+(y)y/\tau\} dy \\ &= \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} + 2\mu\tau (\exp[\mu\{(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*) - 2\psi_i^\top \theta^*\}]) \\ &\quad \int_{\psi_i^\top \theta^*}^{\psi_i^\top \theta} \exp\{y(-\lambda_i^+(y)/\tau + 2\mu)\} dy, \\ &= \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} + 2\mu\tau (\exp[\mu\{(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*) - 2\psi_i^\top \theta^*\}]) \\ &\quad \int_{\psi_i^\top \theta^*}^{\psi_i^\top \theta} \exp\{y(-\lambda_i^+(y)/\tau + 2\lambda_i^+(\theta)/2\tau)\} dy, \end{aligned}$$

where we fix $\mu(\theta) = \lambda_i^+(\theta)/\{2\tau\}$. Monotonicity of $\lambda^+(y)$ implies

$$\begin{aligned} \mathbb{E}_{\lambda(\cdot), \theta^*} \exp\{\mu(\theta)\ell(Y_1, \theta, 0)\} &\leq \{1 + 2\mu\tau\psi_i^\top(\theta - \theta^*) \exp(-2\mu\psi_i^\top \theta^*)\} \\ &\quad \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\} \\ &= \{1 + 2\mu\tau\psi_i^\top(\theta - \theta^*)\} \\ &\quad \exp\{\mu(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\}, \end{aligned}$$

Therefore, for $\theta \geq \theta^*$,

$$\mathfrak{M}^*(\theta, \theta^*) \geq \sum_{i=1}^n \log\{1 + \psi_i^\top(\theta - \theta^*)\lambda^+(\theta)/\tau\} + \sum_{i=1}^n \left\{ \frac{(2\tau - 2)(\psi_i^\top \theta - \psi_i^\top \theta^*)\lambda^+(\theta)}{2\tau} \right\}.$$

The case for $\theta \leq \theta^*$ can be proved in the same fashion.

8.2.4 Moment bounds

$$\begin{aligned}
\mathbb{E}_{\lambda(\cdot), \theta^*} |L(\tilde{\theta}, \theta^*)|^r \mathbf{1}(|L(\tilde{\theta}, \theta^*)|^r > R) &= - \int_0^\infty z^r d\mathbb{P}\{L(\tilde{\theta}, \theta^*) \geq z\} \\
&= r \int_0^\infty z^{r-1} \mathbb{P}\{L(\tilde{\theta}, \theta^*) \geq z\} dz \\
&= r \int_0^\infty z^{r-1} \mathbb{P}\{L(\tilde{\theta}, \theta^*) \geq z, \tilde{\theta} \in \Theta_0(R)\} dz \\
&\quad + r \int_0^\infty z^{r-1} \mathbb{P}\{L(\tilde{\theta}, \theta^*) \geq z, \tilde{\theta} \notin \Theta_0(R)\} dz \\
&\leq r \int_0^\infty z^{r-1} \exp\{-[\{\mu_1 z - \mathfrak{Q}_0(p)\}_+]\} dz \\
&\quad + r \int_0^\infty z^{r-1} \exp\{-\mathfrak{g}(R)\} dz \\
&\leq \mathfrak{R}_r
\end{aligned}$$

References

- Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall, *Journal of Econometrics* **147**: 120–130.
- Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models, *Journal of the American Statistical Association* **103**(484): 1595–1608.
- Embrechts, P., M. A. and Straumann, D. (1999). Correlation and dependency in risk management: Properties and pitfalls, *Risk Management: Value at Risk and Beyond*, Cambridge University Press pp. 176–223.
- Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles, *Journal of Business and Economic Statistics* **22**: 367–381.
- Fitzenberger, B. and Wilke, R. A. (2006). Using quantile regression for duration analysis, *Modern Econometric Analysis* **90**(1): 105–120.
- Golubev, Y. and Spokoiny, V. (2009). Exponential bounds for minimum contrast estimators, *Electronic Journal of Statistics, invited resubmission*.
- Härdle, W. K. and Song, S. (2010). Confidence bands in quantile regression, *Econometric Theory* **26**: 1180–1200.
- Hauksson, A. H., Michel, D., Thomas, D., Ulrich, M. and Gennady, S. (2001). Multivariate extremes, aggregation and risk estimation, *Quantitative Finance* **1**: 79–75.
- James, G. M., Hastie, T. J. and Sugar, C. A. (2010). Principal component models for sparse functional data, *Biometrika* **87**: 587–602.
- Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica* **46**(1): 33–50.

- Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model, *Econometric Theory* **26**: 159–166.
- Ruppert, D., Wand, M. and Carroll, R. (2003). *Semiparametric Regression*, Cambridge University Press.
- Spokoiny, V. (2011). Penalized risk bounds in parametric estimation, *Manuscript* .
- Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimation, *Computational Statistics and Data Analysis* **25**: 159–166.
- Yu, K. and Jones, M. C. (1998). Local linear quantile regression, *Journal of the American Statistical Association* **93**: 228–237.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin and Weining Wang, January 2011.
- 002 "A Confidence Corridor for Sparse Longitudinal Data Curves" by Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.
- 003 "Mean Volatility Regressions" by Lu Lin, Feng Li, Lixing Zhu and Wolfgang Karl Härdle, January 2011.
- 004 "A Confidence Corridor for Expectile Functions" by Esra Akdeniz Duran, Mengmeng Guo and Wolfgang Karl Härdle, January 2011.
- 005 "Local Quantile Regression" by Wolfgang Karl Härdle, Vladimir Spokoiny and Weining Wang, January 2011.

SFB 649, Ziegelstraße 13a, D-10117 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

