Estimating Conditional Mean and Difference Between Conditional Mean and Conditional Median

Liang Peng
Department of Risk Management and Insurance
Georgia State University
and Qiwei Yao
Department of Statistics, London School of Economics

Abstract

When a conditional distribution has an infinite variance, commonly employed kernel smoothing estimators such as local polynomial estimator for the conditional mean have a nonnormal limit, which complicates interval estimation since one has to employ different methods for the cases of finite variance and infinite variance. By estimating the middle part nonparametrically and the tail parts parametrically based on extreme value theory, this paper proposes a new estimator for the conditional mean, which results in a normal limit regardless of whether the conditional distribution has a finite variance or an infinite variance. Hence a naive bootstrap method could be employed to construct a unified interval regardless of tail heaviness. Similar result holds for estimating the difference between conditional mean and conditional median, which is a useful quantity in exploring data.

Key words and phrases: Conditional mean, heavy tail, normal limit

1 Introduction

Mean and median are two important location parameters in data exploratory analysis and the difference between mean and median gives a good indication of data skewness. When the underlying distribution has a finite variance, the sample mean has a normal limit. However, when the underlying distribution has heavy tails with a finite mean, but an infinite variance, the sample mean has a stable law limit. Therefore, in order to give an interval for a mean, one has to know whether the underlying distribution has a finite variance or an infinite variance, which can be done when some additional assumptions on the underlying distribution are imposed such as heavy tails. Under this setting, when the tail index is estimated to be less than two, a subsample bootstrap method could be employed to construct an interval for the mean; see Hall and Jing (1998).

Instead of using different methods to construct a confidence interval for the mean of a heavy tailed distribution, Peng (2001) proposed to estimate the middle part nonparametrically and tail parts parametrically based on extreme value theory, which results in an estimator always with a normal limit. Hence one could simply employ a bootstrap method or develop an empirical likelihood method to construct an interval without separating the cases of finite variance and infinite variance; see Peng (2004). This idea has been applied to estimating expected shortfall in risk management by Necir and Meraghni (2009). This paper aims to further extend this idea in Peng (2001) to estimating a conditional mean and the difference between conditional mean and conditional median, which are useful quantities in exploring data.

Suppose that $\{(X_i, Y_i)^T\}$ is a sequence of independent and identically distributed random vectors and the conditional distribution function $F(y|x) = P(Y_i \le y|X_i = x)$ satisfies

$$\begin{cases} \lim_{t\to\infty} \frac{1-F(ty|x)}{1-F(t|x)} = y^{-\alpha(x)}, \quad y > 0 \\ \lim_{t\to\infty} \frac{1-F(t|x)}{1-F(t|x)+F(-t|x)} = p(x) \in [0,1], \end{cases} \tag{1}$$

where $\alpha(x) > 1$. Like mean and median, the conditional mean $E(Y|X = x)$ is of importance in many applications, which includes the random design regression model as a special case:

$$Y_i = m(X_i) + \epsilon_i, \tag{2}$$

where $\epsilon_i's$ are independent and identically distributed random variables with zero mean and satisfy

$$\begin{cases} \lim_{t\to\infty} \frac{P(\epsilon_i > ty)}{P(\epsilon_i > t)} = y^{-\beta}, \quad y > 0 \\ \lim_{t\to\infty} \frac{P(\epsilon_i > t)}{P(|\epsilon_i| > t)} = p \in [0,1], \end{cases} \tag{3}$$

for some $\beta > 1$.

Under model (2) and conditions (3), model (1) holds with $\alpha(x) \equiv \beta$, and it is well-known that a local smoothing estimator for $m(x)$ has a normal limit and a nonnormal limit for $\beta > 2$ and $\beta < 2$, respectively, which makes interval estimation nontrivial at all. However, when ϵ_i has a median zero, i.e., $m(x)$ is a conditional median, Hall, Peng and Yao (2002) showed that the least absolute deviations estimator has a normal limit for any $\beta > 1$, and so a bootstrap method can be employed to construct an interval without knowing whether β is larger than 2 or less than 2.

In this paper, we seek new estimators for the conditional mean, and the difference between conditional mean and conditional median under conditions (1), which should always have a normal limit for any $\alpha(x) > 1$. Therefore a bootstrap method can be employed to construct confidence intervals straightforwardly.

We organize this paper as follows. Section 2 presents the new method and asymptotic results. A simulation study is given in Section 3. All proofs are put in Section 4.

2 Main Results

First we propose a new estimator for the conditional mean $E(Y_i|X_i = x)$, which gives a normal limit regardless of whether the conditional distribution has a finite variance or an infinite variance.

Suppose our observations $\{(X_i, Y_i)^T\}_{i=1}^n$ are independent and identically distributed random vectors with distribution function $F(x, y)$ and the conditional distribution $F(y|x)$ of Y_i given $X_i = x$ satisfies (1). For a given $h = h(n) > 0$, define $N = \sum_{i=1}^n I(|X_i - x| \le h)$, let $\{(\bar{X}_j, \bar{Y}_j)\}_{j=1}^N$ denote those data pairs $\{(X_i, Y_i)\}_{i=1}^n$ such that $|X_i - x| \le h$, and let $\bar{Y}_{N,1} \le \cdots \le \bar{Y}_{N,N}$ denote the order statistics of $\bar{Y}_1, \cdots, \bar{Y}_N$. Obviously, when $h \to 0$ and $hn \to \infty$, we have $N/(nh) \xrightarrow{p} f_1(x)$, where f_1 denotes the density of X_i. Therefore we write $N_0 = [nh]$ and say $N_0 \to \infty$ instead of $N \xrightarrow{p} \infty$.

Similar to Peng (2001), we write

$$\begin{array}{rl} & E(Y_i|X_i = x) = \int_{-\infty}^{\infty} y\, dF(y|x) = \int_0^1 F^-(y|x)\, dy \\ = & \int_0^{k/N} F^-(y|x)\, dy + \int_{k/N}^{1-k/N} F^-(y|x)\, dy + \int_{1-k/N}^1 F^-(y|x)\, dy \\ := & m_1(x) + m_2(x) + m_3(x), \end{array} \tag{4}$$

where $F^-(y|x)$ denotes the generalized inverse of the conditional distribution $F(y|x)$, and $k = k(N_0) \to \infty$ and $k/N_0 \to 0$ as $N_0 \to \infty$. Based on (4) we propose to estimate the first and third terms by a parametric approximation for $F(y|x)$ via extreme value theory and to estimate the second term nonparametrically. More specifically, when $F^-(y|x) \sim c_1 y^{-1/\alpha_1}$ and $F^-(1-y|x) \sim c_2 y^{-1/\alpha_2}$ as $y \to 0$, the tail indices α_1 and α_2 can be estimated by the well-known Hill estimator (Hill (1975))

$$\hat{\alpha}_1 = \{\frac{1}{k}\sum_{i=1}^{k} \log^+(-\bar{Y}_{N,i}) - \log^+(-\bar{Y}_{N,k})\}^{-1}$$

and

$$\hat{\alpha}_2 = \{\frac{1}{k}\sum_{i=1}^{k} \log^+(\bar{Y}_{N,N-i+1}) - \log^+(\bar{Y}_{N,N-k+1})\}^{-1}$$

with $\log^+ x = \log(x \vee 1)$. In our simulation, we set $\hat{\alpha}_1 = 0$ when all $\bar{Y}_{N,i} > -1$ for $1 \le i \le k$, and $\hat{\alpha}_2 = 0$ when all $\bar{Y}_{N,i} < 1$ for $N-k+1 \le i \le N$. Note that as $N_0 \to \infty$

$$\frac{m_1(x)}{\frac{k}{N}F^-(k/N|x)} \xrightarrow{p} \int_0^1 y^{-1/\alpha(x)}\, dx = \frac{\alpha(x)}{\alpha(x)-1}$$

and

$$\frac{m_3(x)}{\frac{k}{N}F^-(1-k/N|x)} \xrightarrow{p} \int_0^1 y^{-1/\alpha(x)}\, dx = \frac{\alpha(x)}{\alpha(x)-1}.$$

Therefore the three terms in (4) can be estimated separately by

$$\hat{m}_1(x) = \frac{k}{N}\bar{Y}_{N,k}\frac{\hat{\alpha}_1}{\hat{\alpha}_1 - 1}, \quad \hat{m}_2(x) = \frac{1}{N}\sum_{i=k+1}^{N-k}\bar{Y}_{N,i}, \quad \hat{m}_3(x) = \frac{k}{N}\bar{Y}_{N,N-k+1}\frac{\hat{\alpha}_2}{\hat{\alpha}_2 - 1},$$

which leads to our new estimator for the conditional mean $m(x) = E(Y_i|X_i = x)$ as $\hat{m}(x) = \hat{m}_1(x) + \hat{m}_2(x) + \hat{m}_3(x)$. Note that one could also use other tail index estimators instead of the Hill's estimator such that the one in Dierckx, Goegebeur and Guillou (2014).

Like the study of extreme value statistics, in order to derive the asymptotic limits for $\hat{m}_1(x)$ and $\hat{m}_3(x)$, one needs to specify an approximate rate in (1), which is generally called a second order condition in extreme value theory; see De Haan and Ferreira (2006). Here we simply assume that there exist positive smoothing functions $d(x), c_1(x), c_2(x), \alpha(x) > 1, \beta(x)$ such that for y large enough

$$|1 - F(y|x) - c_1(x)y^{-\alpha(x)}| + |F(-y|x) - c_2(x)y^{-\alpha(x)}| \le d(x)y^{-\alpha(x)-\beta(x)} \quad (5)$$

uniformly in $|x - x_0| \le h$. Note that $\beta(x)$ is slightly smaller than the so-called second order parameter in extreme value theory, which can be seen from the inequality for a second order regular variation in De Haan and Ferreira (2006). Furthermore we assume the following regularity conditions:

- A1) the marginal density f_1 of X_i is positive and continuous at x_0;
- A2) functions $c_1(x), c_2(x)$ and $\alpha(x)$ have a continuous second order derivative at x_0, and functions $d(x)$ and $\beta(x)$ have a continuous first order derivative at x_0;
- A3) the conditional mean function $m(x) = \int_{-\infty}^0 F(y|x)\,dy + \int_0^\infty (1 - F(y|x))\,dy$ has a continuous second order derivative at x_0.

To show that the new estimator always has a normal limit, we rely on the following approximations.

Let $H(y)$ denote the distribution function $\bar{Y}_i$ with $x = x_0$, i.e., the conditional distribution of Y_i given $|X_i - x_0| \le h$. Put $U_i = H(\bar{Y}_i)$ for $i = 1, \cdots, N$, and so $U_1, \cdots, U_N$ are i.i.d. random variables with uniform distribution on $(0, 1)$. Let $U_{N,1} \le \cdots \le U_{N,N}$ denote the order statistics of $U_1, \cdots, U_N$. Define $G_N(v) = \frac{1}{N}\sum_{i=1}^N \mathbf{1}(U_i \le v)$, $\alpha_N(v) = \sqrt{N}\{G_N(v) - v\}$, $Q_N(0) = U_{N,1}$, $Q_N(s) = U_{N,i}$ if $\frac{i-1}{N} < s \le \frac{i}{N}$, and $\beta_N(s) = \sqrt{N}\{Q_N(s) - s\}$. Then it follows from Csörgő, Csörgő, Horváth and Mason (1986) that there exists a sequence of Brownian bridges $\{B_N(u)\}$ such that for any $\nu \in [0, 1/4)$ and $\lambda > 0$

$$\begin{cases} \sup_{U_{N,1} \le u \le N_{N,N}} \frac{u^\nu |\alpha_N(u) - B_N(u)|}{u^{1/2-\nu}(1-u)^{1/2-\nu}} = O_p(1) \\ \sup_{\lambda/N \le s \le 1-\lambda/N} \frac{N^\nu |\beta_N(s) + B_N(s)|}{s^{1/2-\nu}(1-s)^{1/2-\nu}} = O_p(1). \end{cases} \quad (6)$$

Theorem 1. *Suppose (5) and Conditions A1)–A3) hold. Put $N_0 = [nh]$, $\alpha_0 = \alpha(x_0)$, $\beta_0 = \beta(x_0)$, and further assume that as $n \to \infty$*

$$\begin{cases} k \to \infty, \quad k/N_0 = o(1), \quad \sqrt{k}h^2(\log N_0)^2 = o(1), \\ k = o(N_0^{\frac{2\beta_0}{\alpha_0 + 2\beta_0}}), \quad \frac{\sqrt{N_0}}{\sigma(k/N_0)}h^2 = o(1), \end{cases} \quad (7)$$

where
$$\sigma^2(s)=\int_s^{1-s}\int_s^{1-s}(u\wedge v-uv)\,dH^-(u)dH^-(v).$$
Then as $n\to\infty$,

$$\begin{aligned}
&\tfrac{\sqrt{N}}{\sigma(k/N)}\{\hat m(x_0)-m(x_0)\}\\
= \quad & -\tfrac{\Delta_2\alpha_0}{(\alpha_0-1)^2}\int_0^1\sqrt{\tfrac{N}{k}}(\tfrac{B_N(\frac{k}{N}s)}{s}-B_N(\tfrac{k}{N}))\,ds-\tfrac{\Delta_2}{\alpha_0-1}\sqrt{\tfrac{N}{k}}B_n(\tfrac{k}{N})\\
&-\tfrac{\Delta_1\alpha_0}{(\alpha_0-1)^2}\int_0^1\sqrt{\tfrac{N}{k}}(\tfrac{B_N(1-\frac{k}{N}s)}{s}-B_N(1-\tfrac{k}{N}))\,ds-\tfrac{\Delta_1}{\alpha_0-1}\sqrt{\tfrac{N}{k}}B_N(1-\tfrac{k}{N})\\
&-\tfrac{\int_{k/N}^{1-k/N}B_N(s)\,dH^-(s)}{\sigma(k/N)}+o_p(1)\\
\stackrel{d}{\to} \quad & N(0,1+\{\tfrac{(2-\alpha_0)(2\alpha_0^2-2\alpha_0+1)}{2(\alpha_0-1)^4}+\tfrac{2-\alpha_0}{\alpha_0-1}\}I(\alpha_0<2)),
\end{aligned}$$

where $m(x_0)=E(Y_i|X_i=x_0)$,
$$\Delta_1=\{\frac{2-\alpha_0}{2(c_1^{2/\alpha_0}(x_0)+c_2^{2/\alpha_0}(x_0))}\}^{1/2}c_1^{1/\alpha_0}(x_0)I(\alpha_0<2)$$
and
$$\Delta_2=\{\frac{2-\alpha_0}{2(c_1^{2/\alpha_0}(x_0)+c_2^{2/\alpha_0}(x_0))}\}^{1/2}c_2^{1/\alpha_0}(x_0)I(\alpha_0<2).$$

Remark 1. *If $\alpha(x_0)>2$, then as $N_0\to\infty$*
$$\sigma^2(k/N)\stackrel{p}{\to}E(Y_i^2|X_i=x_0)-(E(Y_i|X_i=x_0))^2<\infty.$$
In this case, we require $\sqrt{nh}h^2\to 0$, which gives the same rate of convergence as the local smoothing estimator of a conditional mean without asymptotic bias. It also follows from the proof of the above theorem that the above $H(y)$ can be replaced by $F(y|x_0)$.

Remark 2. *It follows from the above theorem that a naive bootstrap method can be employed to construct a confidence interval for the conditional mean regardless of tail heaviness. We refer to Hall (1992) for an overview on bootstrap method. A review paper on applying bootstrap methods to extreme value statistics is Qi (2008).*

Next we consider estimating the difference between conditional mean and conditional median, i.e., $\theta(x)=E(Y_i|X_i=x)-F^-(1/2|x)$. Based on the above estimator for $m(x)$, the proposed estimator for θ is $\hat\theta(x)=\hat m(x)-\bar Y_{N,[N/2]}$, and its asymptotic limit is given in the theorem below.

Theorem 2. *Under conditions of Theorem 1 and that the conditional density function* $g(y|x) = \frac{dF(y|x)}{dy}$ *is positive and continuous at* $y = F^{-}(\frac{1}{2}|x_0)$ *and* $x = x_0$*, we have, as* $n \to \infty$*,*

$$\begin{aligned} & \frac{\sqrt{N}}{\sigma(k/N)}\{\hat{\theta}(x_0) - \theta(x_0)\} \\ = \quad & -\frac{\Delta_2\alpha_0}{(\alpha_0-1)^2}\int_0^1 \frac{N}{k}\left(\frac{B_N(\frac{k}{N}s)}{s} - B_N(\tfrac{k}{N})\right)ds - \frac{\Delta_2}{\alpha_0-1}\sqrt{\frac{N}{k}}B_n(\tfrac{k}{N}) \\ & -\frac{\Delta_1\alpha_0}{(\alpha_0-1)^2}\int_0^1 \sqrt{\frac{N}{k}}\left(\frac{B_N(1-\frac{k}{N}s)}{s} - B_N(1-\tfrac{k}{N})\right)ds - \frac{\Delta_1}{\alpha_0-1}\sqrt{\frac{N}{k}}B_N(1-\tfrac{k}{N}) \\ & -\frac{\int_{k/N}^{1-k/N} B_N(s)\,dH^{-}(s)}{\sigma(k/N)} - \frac{B_N(1/2)}{\sigma(k/N)g(F^{-}(\frac{1}{2}|x_0)|x_0)} + o_p(1) \\ \xrightarrow{d} \quad & N(0, \sigma_\theta^2), \end{aligned}$$

where σ_θ^2 *equals to the variance in Theorem 1 when* $\alpha_0 \le 2$*, and is*

$$\begin{aligned} & 1 + \frac{1}{4g^2(F^{-}(1/2|x_0)|x_0)\int_0^1\int_0^1 (u\wedge v - uv)\,dF^{-}(u|x_0)dF^{-}(v|x_0)} \\ & + \frac{\int_0^{1/2} u\,dF^{-}(u|x_0) + \int_{1/2}^1 (1-u)\,dF^{-}(u|x_0)}{g(F^{-}(1/2|x_0)|x_0)\int_0^1\int_0^1 (u\wedge v - uv)\,dF^{-}(u|x_0)dF^{-}(v|x_0)} \end{aligned}$$

when $\alpha_0 > 2$.

3 Simulation

We conduct a small scale simulation to illustrate the proposed method. To this end, we let $X_i's$ in (2) be independent $U(-1, 1)$ random variables, and consider

$$m(x) = x + 4\exp(-4x^2).$$

Furthermore in (2) we let ϵ_i be independent scaled t-distribution with d degrees of freedom for $d = 1.5$ and 3. Then $\alpha(x) = d$ in (1). We re-scale ϵ_i such that its standard deviation is 0.5. We set sample size $n = 1000$ or 3000, and choose $k = 5, 10, 20, 30, 40$ and 50. We use bandwidth $h = 0.2$ when $n = 1000$, and $h = 0.1$ when $n = 3000$. This effectively sets the sample sizes 200 and 300, respectively, in the local estimation for $m(x)$ for each given x.

We estimate $m(\cdot)$ on a regular grid of the 19 points between -0.9 and 0.9, and calculate the root mean square error:

$$\text{rMSE} = \left\{\frac{1}{19}\sum_{j=-9}^{9}\{\hat{m}(0.1j) - m(0.1j)\}^2\right\}^{1/2}. \tag{8}$$

For each setting, we replicate the exercise 500 times. To compare the performance with conventional nonparametric regression, we also calculate three nearest neighbor estimates, namely estimate $m(x)$ by the mean of Y_i's corresponding to those X_i's within, respectively, h-, $h/2$- and $h/4$-distance from x. Table 1 reports the mean and the standard deviation of rMSE for different settings over 500 replications. As we expected, the estimation error decreases when sample size n increases from 1000 to 3000, and the error also decreases when the tail index, reflected by the degrees of freedom (df), increases. With $t_{1.5}$-distributed errors, $k = 30$ gives a smallest standard deviation, and both $k = 20$ and $k = 30$ perform well. But with t_3-distributed errors, $k = 5$ leads to the most accurate estimates, which is in line with the theorem that tail parts do not play a role asymptotically in case of finite variance and so a smaller k is preferred. For the model with $t_{1.5}$-distributed errors, the nearest neighbor estimator is no longer asymptotically normal. Indeed our newly proposed estimator with either $k = 20$ or $k = 30$ performs better than the nearest neighbor estimator. However for the model with t_3-distributed errors, the nearest neighbor estimator is asymptotically normal and is indeed performs better than the new method.

(n, h, df)		New Estimator						NN Estimator		
		$k=5$	$k=10$	$k=20$	$k=30$	$k=40$	$k=50$	h	$h/2$	$h/4$
$(1000, 0.2, 1.5)$	Mean	3.674	0.381	0.188	0.200	0.236	0.279	0.218	0.250	0.340
	STD	47.66	1.836	0.291	0.053	0.071	0.090	0.447	0.621	0.701
$(3000, 0.1, 1.5)$	Mean	1.546	4.333	0.162	0.138	0.155	0.173	0.201	0.280	0.354
	STD	12.92	74.57	0.283	0.022	0.025	0.033	0.345	0.482	0.769
$(1000, 0.2, 3)$	Mean	0.134	0.154	0.208	0.274	0.348	0.428	0.122	0.080	0.103
	STD	0.021	0.023	0.033	0.045	0.059	0.072	0.021	.018	0.024
$(3000, 0.1, 3)$	Mean	0.059	0.067	0.105	0.151	0.202	0.255	0.050	0.058	0.079
	STD	0.040	0.011	0.016	0.022	0.028	0.036	0.010	0.011	0.015

Table 1: Mean and standard deviation (STD) of rMSE defined in (8) for the proposed new estimator and the nearest neighbor (NN) estimator in simulation with 500 replications.

4 Proofs

Proof of Theorem 1. Write

$$\begin{array}{rl} & \hat{m}_1(x_0)-\int_0^{k/N}H^-(v)\,dv \\ = & \frac{k}{N}H^-(U_{N,k})(\frac{\hat{\alpha}_1}{\hat{\alpha}_1-1}-\frac{\alpha_0}{\alpha_0-1}) \\ & +\frac{\alpha_0}{\alpha_0-1}\big(\frac{k}{N}H^-(U_{N,k})-\frac{k}{N}H^-(k/N)\big) \\ & +\big(\frac{k}{N}H^-(k/N)\frac{\alpha_0}{\alpha_0-1}-\int_0^{k/N}H^-(v)\,dv\big) \\ = & \frac{k}{N}H^-(U_{N,k})\frac{\hat{\alpha}_1\alpha_0}{(\hat{\alpha}_1-1)(\alpha_0-1)}\frac{1}{k}\sum_{i=1}^k\Big\{\log\frac{H^-(U_{N,i})}{H^-(U_{N,k})}-\log(U_{N,i}/U_{N,k})^{-1/\alpha_0}\Big\} \\ & +\frac{k}{N}H^-(U_{N,k})\frac{\hat{\alpha}_1\alpha_0}{(\hat{\alpha}_1-1)(\alpha_0-1)}\Big\{\frac{1}{k}\sum_{i=1}^k\log(U_{N,i}/U_{N,k})^{-1/\alpha_0}-1/\alpha_0\Big\} \\ & +\frac{k}{N}H^-(k/N)\frac{\alpha_0}{\alpha_0-1}\Big\{\frac{H^-(U_{N,k})}{H^-(k/N)}-(\frac{N}{k}U_{N,k})^{-1/\alpha_0}\Big\} \\ & +\frac{k}{N}H^-(k/N)\frac{\alpha_0}{\alpha_0-1}\big\{(\frac{N}{k}U_{N,k})^{-1/\alpha_0}-1\big\} \\ & +\Big\{\frac{k}{N}H^-(k/N)\frac{\alpha_0}{\alpha_0-1}-\int_0^{k/N}H^-(v)\,dv\Big\} \\ := & I_1+I_2+I_3+I_4+I_5, \end{array}$$

$$\begin{array}{rl} & \hat{m}_3(x_0)-\int_{1-k/N}^1H^-(v)\,dv \\ = & \frac{k}{N}H^-(U_{N,N-k+1})\frac{\hat{\alpha}_2\alpha_0}{(\hat{\alpha}_2-1)(\alpha_0-1)}\frac{1}{k}\sum_{i=1}^k\Big\{\log\frac{H^-(U_{N,N-i+1})}{H^-(U_{N,N-k+1})}-\log(\frac{1-U_{N,N-i+1}}{1-U_{N,N-k+1}})^{-1/\alpha_0}\Big\} \\ & +\frac{k}{N}H^-(U_{N,N-k+1})\frac{\hat{\alpha}_2\alpha_0}{(\hat{\alpha}_2-1)(\alpha_0-1)}\Big\{\frac{1}{k}\sum_{i=1}^k\log(\frac{1-U_{N,N-i+1}}{1-U_{N,N-k+1}})^{-1/\alpha_0}-1/\alpha_0\Big\} \\ & +\frac{k}{N}H^-(1-k/N)\frac{\alpha_0}{\alpha_0-1}\Big\{\frac{H^-(U_{N,N-k+1})}{H^-(1-k/N)}-\Big(\frac{N}{k}(1-U_{N,N-k+1})\Big)^{-1/\alpha_0}\Big\} \\ & +\frac{k}{N}H^-(1-k/N)\frac{\alpha_0}{\alpha_0-1}\Big\{\Big(\frac{N}{k}(1-U_{N,N-k+1})\Big)^{-1/\alpha_0}-1\Big\} \\ & +\big(\frac{k}{N}H^-(1-k/N)\frac{\alpha_0}{\alpha_0-1}-\int_{1-k/N}^1H^-(v)\,dv\big) \\ := & III_1+III_2+III_3+III_4+III_5 \end{array}$$

and

$$\begin{array}{rl} & \hat{m}_2(x_0)-\int_{k/N}^{1-k/N}H^-(v)\,dv \\ = & \int_{U_{N,k}}^{k/N}H^-(v)\,dG_N(v)+\int_{1-k/N}^{U_{N,N-k}}H^-(v)\,dG_N(v) \\ & +H^-(1-k/N)\{G_N(1-k/N)-1+k/N\}-H^-(k/N)\{G_N(k/N)-k/N\} \\ & -\int_{k/N}^{1-k/N}\{G_N(v)-v\}\,dH^-(v) \\ := & II_1+II_2+II_3+II_4+II_5. \end{array}$$

Using Conditions A1)–A2), (5) and the fact that $|y^{\delta_3h}-1|\le Mh\log y$ uniformly in $y\in[n^{\delta_1},n^{\delta_2}]$ for any given $0<\delta_1<\delta_2<1$ and $\delta_3>0$, where

$M > 0$ only depends on $\delta_1, \delta_2, \delta_3$, since $h \log n \to 0$, we have

$$\begin{array}{rl}
& |1 - H(y) - c_1(x_0) y^{-\alpha_0}| \\
= & |\frac{\int_{x_0-h}^{x_0+h}\{1-F(y|z)\}f_1(z)\,dz}{P(|X_1-x_0|\le h)} - c_1(x_0)y^{-\alpha(x_0)}| \\
\le & |\frac{\int_{x_0-h}^{x_0+h}\{1-F(y|z)-c_1(z)y^{-\alpha(z)}\}f_1(z)\,dz}{P(|X_1-x_0|\le h)}| \\
& +|\frac{\int_{x_0-h}^{x_0+h}\{c_1(z)y^{-\alpha(z)}-c_1(x_0)y^{-\alpha_0}\}f_1(z)\,dz}{P(|X_1-x_0|\le h)}| \\
\le & M_1 y^{-\alpha_0}\{y^{-\beta_0} + h^2(\log y)^2\}
\end{array} \tag{9}$$

uniformly in $y \in [n^{\delta_1}, n^{\delta_2}]$ for any given $0 < \delta_1 < \delta_2 < 1$, where $M_1 > 0$ is independent of y. Similarly

$$|H(-y) - c_2(x_0)y^{-\alpha_0}| \le M_2 y^{-\alpha_0}\{h^2(\log y)^2 + y^{-\beta_0}\} \tag{10}$$

uniformly in $y \in [n^{\delta_1}, n^{\delta_2}]$ for any given $0 < \delta_1 < \delta_2 < 1$, where $M_2 > 0$ is independent of y. Therefore

$$|H^-(1-t) - c_1^{1/\alpha_0}(x_0)t^{-1/\alpha_0}| \le M_3 t^{-1/\alpha_0}\{h^2(\log t)^2 + t^{\beta_0/\alpha_0}\} \tag{11}$$

and

$$|H^-(t) + c_2^{1/\alpha_0}(x_0)t^{-1/\alpha_0}| \le M_4 t^{-1/\alpha_0}\{h^2(\log t)^2 + t^{\beta_0/\alpha_0}\} \tag{12}$$

uniformly in $t \in [n^{-\delta_1}, n^{-\delta_2}]$ for any given $0 < \delta_2 < \delta_1 < 1$, where $M_3 > 0$ and $M_4 > 0$ are independent of t.

Note that

$$\frac{N}{nh} \xrightarrow{p} f_1(x_0), \quad P(\bar{Y}_{N,1} \ge -n^{-\delta}, \bar{Y}_{N,N} \le n^{\delta}) \to 1 \tag{13}$$

for $\delta \in (0,1)$ large enough.

Write

$$\begin{array}{rl}
\sigma^2(s) & = \int_{H^-(s)}^0 \int_{H^-(s)}^0 \{H(u)\wedge H(v) - H(u)H(v)\}\,dudv \\
& \quad + \int_0^{H^-(1-s)} \int_0^{H^-(1-s)} \{H(u)\wedge H(v) - H(u)H(v)\}\,dudv \\
& = 2\int_{H^-(s)}^0 \int_v^0 H(v)\{1-H(u)\}\,dudv \\
& \quad +2\int_0^{H^-(1-s)} \int_0^v H(u)\{1-H(v)\}\,dudv \\
& = -2\int_{H^-(s)}^0 vH(v)\,dv - \{\int_{H^-(s)}^0 H(u)\,du\}^2 \\
& \quad +2\int_0^{H^-(1-s)} v\{1-H(v)\}\,dv - \{\int_0^{H^-(1-s)}(1-H(u))\,du\}^2 \\
& = IV_1(s) + IV_2(s) + IV_3(s) + IV_4(s).
\end{array}$$

Then it follows from (11)–(13) that

$$\begin{cases} \frac{IV_1(k/N)}{(k/N)^{1-2/\alpha_0}} \xrightarrow{p} \frac{2c_2^{2/\alpha_0}(x_0)}{2-\alpha_0}, & \frac{IV_2(k/N)}{(k/N)^{1-2/\alpha_0}} \xrightarrow{p} 0, \\ \frac{IV_3(k/N)}{(k/N)^{1-2/\alpha_0}} \xrightarrow{p} \frac{2c_1^{2/\alpha_0}(x_0)}{2-\alpha_0}, & \frac{IV_4(k/N)}{(k/N)^{1-2/\alpha_0}} \xrightarrow{p} 0, \end{cases}$$

when $\alpha_0 < 2$, and

$$\sigma^2(k/N) \xrightarrow{p} \begin{cases} \sigma_0^2 < \infty & \text{if} \quad \alpha_0 > 2 \\ \infty & \text{if} \quad \alpha_0 = 2, \end{cases}$$

where $\sigma_0^2 = \int_0^1 \int_0^1 (u \wedge v - uv)\, dF^-(u|x_0) dF^-(v|x_0)$. Therefore,

$$\frac{(k/N)^{1-2/\alpha_0}}{\sigma^2(k/N)} \xrightarrow{p} \frac{2-\alpha_0}{2(c_1^{2/\alpha_0}(x_0) + c_2^{2/\alpha_0}(x_0))} I(\alpha_0 < 2). \tag{14}$$

Now using (6), (9)–(14) and (7), we can show that

$$\frac{\sqrt{N}}{\sigma(k/N)}\{|I_1| + |I_3| + |I_5| + |III_1| + |III_3| + |III_5|\} = o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} I_2 = -\Delta_2 \frac{\alpha_0}{(\alpha_0-1)^2} \int_0^1 \sqrt{\frac{N}{k}} \{\frac{B_N(\frac{k}{N}s)}{s} - B_N(\frac{k}{N})\}\, ds + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} I_4 = -\Delta_2 \frac{1}{\alpha_0-1} \sqrt{\frac{N}{k}} B_N(\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} III_2 = -\Delta_1 \frac{\alpha_0}{(\alpha_0-1)^2} \int_0^1 \sqrt{\frac{N}{k}} \{\frac{B_N(1-\frac{k}{N}s)}{s} - B_N(1-\frac{k}{N})\}\, ds + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} III_4 = -\Delta_1 \frac{1}{\alpha_0-1} \sqrt{\frac{N}{k}} B_N(1-\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} II_1 = -\Delta_2 \sqrt{\frac{N}{k}} B_N(\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} II_2 = -\Delta_1 \sqrt{\frac{N}{k}} B_N(1-\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} II_3 = \Delta_1 \sqrt{\frac{N}{k}} B_N(1-\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} II_4 = \Delta_2 \sqrt{\frac{N}{k}} B_N(\frac{k}{N}) + o_p(1),$$

$$\frac{\sqrt{N}}{\sigma(k/N)} II_5 = -\frac{\int_{k/N}^{1-k/N} B_N(s)\, dH^-(v)}{\sigma(k/N)} + o_p(1),$$

which implies that

$$\begin{array}{rl} & \frac{\sqrt{N}}{\sigma(k/N)}\{\hat{m}(x_0) - \int_0^1 H^-(v)\,dv\} \\ = & -\frac{\Delta_2\alpha_0}{(\alpha_0-1)^2}\int_0^1 \sqrt{\frac{N}{k}}(\frac{B_N(\frac{k}{N}s)}{s} - B_N(\frac{k}{N}))\,ds - \frac{\Delta_2}{\alpha_0-1}\sqrt{\frac{N}{k}}B_n(\frac{k}{N}) \\ & -\frac{\Delta_1\alpha_0}{(\alpha_0-1)^2}\int_0^1 \sqrt{\frac{N}{k}}(\frac{B_N(1-\frac{k}{N}s)}{s} - B_N(1-\frac{k}{N}))\,ds - \frac{\Delta_1}{\alpha_0-1}\sqrt{\frac{N}{k}}B_N(1-\frac{k}{N}) \\ & -\frac{\int_{k/N}^{1-k/N} B_N(s)\,dH^-(s)}{\sigma(k/N)} + o_p(1) \\ \xrightarrow{d} & N(0, 1 + \{\frac{(2-\alpha_0)(2\alpha_0^2-2\alpha_0+1)}{2(\alpha_0-1)^4} + \frac{2-\alpha_0}{\alpha_0-1}\}I(\alpha_0<2)) \end{array} \tag{15}$$

by noting that

$$\begin{array}{rl} & E\{\sqrt{\frac{N}{k}}B_N(\frac{k}{N})\frac{\int_{k/N}^{1-k/N} B_N(s)\,dH^-(s)}{\sigma(k/N)}|N\} \\ = & \frac{\sqrt{\frac{N}{k}}\int_{k/N}^{1-k/N}\frac{k}{N}(1-s)\,dH^-(s)}{\sigma(k/N)} \\ = & \frac{\sqrt{k/N}}{\sigma(k/N)}\{\int_{H^-(k/N)}^0 (1-H(u))\,du + \int_0^{H^-(1-k/N)}(1-H(u))\,du\} \\ \xrightarrow{p} & \{\frac{2-\alpha_0}{2(c_1^{2/\alpha_0}(x_0)+c_2^{2/\alpha_0}(x_0))}\}^{1/2}c_2^{1/\alpha_0}(x_0)I(\alpha_0<2) \end{array}$$

and

$$\begin{array}{rl} & E\{\sqrt{\frac{N}{k}}B_N(1-\frac{k}{N})\frac{\int_{k/N}^{1-k/N} B_N(s)\,dH^-(s)}{\sigma(k/N)}|N\} \\ = & \frac{\sqrt{\frac{N}{k}}\int_{k/N}^{1-k/N}\frac{k}{N}s\,dH^-(s)}{\sigma(k/N)} \\ = & \frac{\sqrt{k/N}}{\sigma(k/N)}\{\int_{H^-(k/N)}^0 H(u)\,du + \int_0^{H^-(1-k/N)} H(u)\,du\} \\ \xrightarrow{p} & \{\frac{2-\alpha_0}{2(c_1^{2/\alpha_0}(x_0)+c_2^{2/\alpha_0}(x_0))}\}^{1/2}c_1^{1/\alpha_0}(x_0)I(\alpha_0<2). \end{array}$$

It follows from A3) that

$$\begin{array}{rl} & \int_0^1 H^-(v)\,dv - \int_0^1 F^-(v|x_0)\,dv \\ = & \int_{-\infty}^0 H(v)\,dv + \int_0^\infty (1-H(v))\,dv - \int_{-\infty}^0 F(v|x_0)\,dv - \int_0^\infty (1-F(v|x_0))\,dv \\ = & \frac{\int_{x_0-h}^{x_0+h} f_1(z)\{\int_{-\infty}^0 F(y|z)\,dy + \int_0^\infty (1-F(y|z))\,dy - \int_{-\infty}^0 F(y|x_0)\,dy - \int_0^\infty (1-F(y|x_0))\,dy\}dz}{P(|X_1-x_0|\le h)} \\ = & O(h^2). \end{array} \tag{16}$$

Hence the theorem follows from (15), (16) and (7). □

Proof of Theorem 2. The theorem easily follows from the expansions in the proof of Theorem 1 and the fact that

$$\sqrt{N}\{\bar{Y}_{N,[N/2]} - H^-(\frac{1}{2})\} = \frac{\sqrt{N}(G_N(1/2)-1/2)}{g(F^-(\frac{1}{2}|x_0)|x_0)} + o_p(1).$$

□

References

[1] M. Csörgő, S. Csörgő, L. Horváth and D.M. Mason (1986). Weighted empirical and quantile processes. *Annals of Probability* 14, 31–85.

[2] G. Dierckx, Y. Goegebeur and A. Guillou (2014). Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. *Test 23, 330–355.*

[3] L. de Haan and A. Ferreira (2006). Extreme Value Theory: An Introduction. *Springer.*

[4] P. Hall (1992). The Bootstrap and Edgeworth Expansion. *Springer.*

[5] P. Hall and B. Jing (1998). Comparison of bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean. *Statistica Sinica 8, 887–906.*

[6] P. Hall, L. Peng and Q. Yao (2002). Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series Analysis 23, 313–331.*

[7] A.M. Hill (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3, 1163-1174.*

[8] A. Necir and D. Meraghni (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. *Insurance: Mathematics and Economics 45, 49–58.*

[9] Liang Peng (2001). Estimating the mean of a heavy tailed distribution. *Statistics & Probability Letters 52, 31–40.*

[10] Liang Peng (2004). Empirical likelihood confidence interval for a mean with a heavy tailed distribution. *Annals of Statistics 32, 1192–1214.*

[11] Y. Qi (2008). Bootstrap and empirical likelihood methods in extremes. *Extremes 11, 81–97.*