HIGH ORDER CONDITIONAL QUANTILE ESTIMATION BASED ON NONPARAMETRIC MODELS OF REGRESSION

By Carlos Martins-Filho, Feng Yao and Maximo Torero*

University of Colorado & IFPRI, West Virginia University and IFPRI

We consider the estimation of a high order conditional quantile associated with the distribution of the regressand in a nonparametric regression model. Our estimator is inspired by Pickands (1975) which has shown that arbitrary distributions which lie in the domain of attraction of an extreme value type have tails that, in the limit, behave as generalized Pareto distributions (GPD). Smith (1987) has studied the asymptotic properties of maximum likelihood (ML) estimators for the parameters of the GPD in this context, but in our paper the relevant random variables used in estimation are residuals from a first stage kernel based nonparametric estimator. We obtain convergence in probability and distribution of the residual based ML estimator for the parameters of the GPD as well as the asymptotic distribution for a suitably defined quantile estimator. A Monte Carlo study provides evidence that our estimator behaves well in finite samples and is easily implementable. Our results have direct application in finance, particularly in the estimation of conditional Value-at-Risk, but other researchers in applied fields such as insurance and hydrology will find the results useful.

1. Introduction. Consider the following nonparametric regression model

$$Y = m(X) + U \tag{1}$$

where m is a real valued function which belongs to a suitably restricted class (see section 3), $E(U|X = x) = 0$ and $V(U|X = x) = 1$.[1] We assume that U has a strictly increasing absolutely continuous distribution $F(u)$ which belongs to the domain of attraction of an extremal distribution (see Resnick (1987)). In this case, for $a \in (0, 1)$, the conditional a-quantile associated with the conditional distribution of Y given X, denoted by $q_{Y|X=x}(a)$, is given by

$$q_{Y|X=x}(a) = m(x) + q(a), \text{ where } q(a) \text{ is the } a\text{-quantile associated with } F.$$

If U were observed, $q(a)$ could be estimated from a random sample $\{U_i\}_{i=1}^n$ and combined with an estimator for m to obtain an estimator for $q_{Y|X=x}(a)$. In general, U is not observed, but given a random sample $\{(Y_i, X_i)\}_{i=1}^n$ and an estimator $\hat{m}(x)$ for $m(x)$ it is possible to obtain

$$\hat{U}_i = Y_i - \hat{m}(X_i) \text{ for } i = 1, \cdots, n. \tag{2}$$

The sequence of residuals $\{\hat{U}_i\}_{i=1}^n$ can then be used to produce an estimator $\hat{q}(a)$ for $q(a)$. We can define $\hat{q}_{Y|X=x}(a) = \hat{m}(x) + \hat{q}(a)$ as an estimator for $q_{Y|X=x}(a)$. In this paper, we are particularly interested in the case where a is very large, i.e., in the vicinity of 1. These high order conditional quantiles have become particularly important in empirical finance where they

*We thank Xiaodong Liu and participants of Econometrics Workshop at the University of Colorado, Boulder for helpful comments.

AMS 2000 subject classifications: Primary 62G32; secondary 62F10, 62G07, 62G08, 62G20.

Keywords and phrases: Conditional quantile, extreme value theory, generalized Pareto distribution, nonparametric regression.

[1]We assume that the conditional variance of U is 1 for simplicity. If $0 < V(U|X) = \theta < \infty$ our results would continue to hold as θ can be estimated consistently and at a parametric rate. See, *inter alia*, Martins-Filho and Yao (2006a).

are called conditional Value-at-Risk (VaR) (see McNeil and Frey (2000); Martins-Filho and Yao (2006b); Cai and Wang (2008)). It is interesting that the information that a is in the vicinity of 1 is helpful in the estimation of $q(a)$. Pickands (1975) showed that if F is in the domain of attraction of an extremal type distribution, denoted by $F(x) \in D(E)$, for some fixed k and function $\sigma(\xi)$

$$F(x) \in D(E) \iff lim_{\xi \to u_\infty} sup_{0<u<u_\infty - \xi} |F_\xi(u) - G(u; \sigma(\xi), k)| = 0, \tag{3}$$

where $F_\xi(u) = \frac{F(u+\xi)-F(\xi)}{1-F(\xi)}$, $u_\infty = l.u.b\{x : F(x) < 1\} \leq \infty$ is the upper endpoint of F, $u_\infty > \xi \in \Re$, G is a generalized Pareto distribution (GPD), i.e.,

$$G(y; \sigma, k) = \begin{cases} 1 - (1 - ky/\sigma)^{1/k} & \text{if } k \neq 0, \sigma > 0 \\ 1 - exp(-y/\sigma) & \text{if } k = 0, \sigma > 0 \end{cases}$$

with $0 < y < \infty$ if $k < 0$ and $0 < y < \sigma/k$ if $k > 0$.[2] It is evident that $F_\xi(u)$ represents the conditional distribution of the exceedances over ξ of an random variable U given that $U > \xi$.

The equivalence in (3) shows that G is a suitable parametric approximation for the upper tail of F provided that F belongs to the domain of attraction of an extremal type distribution. Intuitively, an estimator for $q(a)$ can be obtained from the estimation of the parameters k and $\sigma(\xi)$. Smith (1987) provides a comprehensive study of a maximum likelihood (ML) type estimator for k and $\sigma(\xi)$ when the sequence $\{U_i\}_{i=1}^n$ is observed. In this paper we extend Smith's results and study the asymptotic properties of ML type estimators for k and $\sigma(\xi)$ based on a sequence $\{\hat{U}_i\}_{i=1}^n$ obtained from a first stage nonparametric estimator $\hat{m}(x)$ for $m(x)$. The extension is desirable as many stochastic models of interest, in particular those used in insurance and finance, exhibit the conditional location-scale structure of equation (1) (see Embrechts, Kluppelberg and Mikosh (1997)) rather than the simpler formulation treated by Smith.

We have shown that, for the case where $F(x)$ belongs to the domain of attraction of a Fréchet distribution, the ML estimator for the parameters of the GPD based on the sequence $\{\hat{U}_i\}_{i=1}^n$ converge at a parametric rate to a normal distribution when suitably centered. The asymptotic distribution is similar to that obtained by Smith (1987), but although the use of nonparametric residuals does not impact the estimator's rate of convergence, it does increases its variance. We also study the asymptotic behavior of the estimator $\hat{q}(a)$ constructed from the ML estimators for the parameters of the GPD. In particular, we show that $\frac{\hat{q}(a)}{q(a)} - 1$ also converges in distribution to a normal at the parametric rate. These results, combined with known properties for suitably defined $\hat{m}(x)$ provide consistency of $\hat{q}_{Y|X}(a)$ as an estimator for $q_{Y|X}(a)$.

Besides the introduction, this paper has four more sections and two appendices. Section 2 provides definitions and discussions of the specific estimators we will consider. Section 3 provides the asymptotic characterization of our proposed estimators and the assumptions we used in our results. Section 4 contains a Monte Carlo study that sheds some light on the finite sample properties of the estimator under study and a comparison with a commonly used estimator proposed by Hill (1975) for the parameter k of the GPD distribution. Section 5 provides a conclusion and gives directions for further study. The appendices contain all proofs, supporting lemmas, tables and figures that summarize the Monte Carlo simulations.

2. Estimation.

The estimation procedure has two main stages. First, the definition of $\hat{U}_i$ in (2) requires a specific estimator for $m(x)$. For algebraic simplicity we will consider the Nadaraya-Watson (NW) estimator

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K_1\left(\frac{X_i - x}{h_{1n}}\right) Y_i}{\sum_{i=1}^{n} K_1\left(\frac{X_i - x}{h_{1n}}\right)}$$

[2] $l.u.b$ stands for least upper bound.

based on a random sample $\{(Y_i, X_i)\}_{i=1}^n$ of observations on $(Y, X) \in \Re^2$. Here, $K_1(\cdot)$ is a kernel function satisfying some standard properties (see section 3) and $0 < h_{1n}$ is a bandwidth.[3] It should be clear from what follows that other nonparametric estimators for $m(x)$ could be used to define $\hat{U}_i$. What is important is that they are uniformly asymptotically close to $m(x)$ in probability at a suitable rate. In particular, for the the NW estimator we have that under our assumptions

$$\sup_{x \in G} |\hat{m}(x) - m(x)| = O_p\left(\left(\frac{nh_{1n}}{\log n}\right)^{-1/2} + h_{1n}^2\right) \tag{4}$$

where G is any compact subset of $\Re$.

Given a sequence $\{U_i\}_{i=1}^n$ we define the (ascending) order statistics $\{U_{(i)}\}_{i=1}^n$. For any $N < n$ we define the excesses over $U_{(n-N)}$ by $\{Z_j\}_{j=1}^N = \{U_{(n-N+j)} - U_{(n-N)}\}_{j=1}^N$. In our context, it should be clear that since U_i is not observed, neither is Z_j. Order statistics can be viewed as estimators for a-quantiles associated with empirical distributions. As such, we can write

$$q_n(a) = \begin{cases} U_{(na)} & \text{if } na \in \mathbb{N} \\ U_{([na]+1)} & \text{if } na \notin \mathbb{N} \end{cases}$$

where $\mathbb{N}$ represents the positive integers, $q_n(a)$ is the a-quantile associated with the empirical distribution $F_n(u) = n^{-1}\sum_{i=1}^n \chi_{\{U_i \le u\}}$ with χ_A denoting the indicator function for the set A.

Consequently, for $a_n = 1 - \frac{N}{n}$ we can write $\{Z_j\}_{j=1}^N = \left\{U_{(n-N+j)} - q_n(a_n)\right\}_{j=1}^N$. It is well known from the unconditional distribution and quantile estimation literature (Azzalini (1981), Falk (1985), Yang (1985), Bowman, Hall and Prvan (1998), Martins-Filho and Yao (2007)) that smoothing beyond that attained by the empirical distribution can produce significant gains in finite samples with no impact on asymptotic rates of convergence. Consequently, to construct an estimated sequence of excesses $\{\tilde{Z}_j\}_{j=1}^N$, we define $\tilde{q}(z)$ as the solution for

$$\tilde{F}(\tilde{q}(z)) = z$$

where $\tilde{F}(u) = \int_{-\infty}^u \frac{1}{nh_{2n}} \sum_{i=1}^n K_2\left(\frac{\hat{U}_i - y}{h_{2n}}\right) dy$, $K_2(\cdot)$ is a kernel function and $0 < h_{2n}$ is a bandwidth satisfying certain regularity conditions. Therefore, we can define the *observed* sequence $\{\tilde{Z}_j\}_{j=1}^N = \left\{\hat{U}_{(n-N+j)} - \tilde{q}(a_n)\right\}_{j=1}^N$ to be used in the estimation of the parameters of the GPD in the second stage.

In the second stage of the estimation we first consider maximum likelihood estimators for σ and k based on the density $g(z;\sigma,k) = \frac{1}{\sigma}\left(1 - \frac{kz}{\sigma}\right)^{1/k-1}$ associated with the GPD distribution. In particular, we consider a solution $(\tilde{\sigma}_N, \tilde{k})$ for the following likelihood equations:

$$\frac{\partial}{\partial \sigma}\frac{1}{N}\sum_{i=1}^N \log g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0 \text{ and } \frac{\partial}{\partial k}\frac{1}{N}\sum_{i=1}^N \log g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0. \tag{5}$$

If $\{U_i\}_{i=1}^n$ were observed, for a threshold $\xi = U_{(n-N)}$ we could, based on (3), write

$$F_{U_{(n-N)}}(y) = \frac{F(y + U_{(n-N)}) - F(U_{(n-N)})}{1 - F(U_{(n-N)})} \approx 1 - \left(1 - \frac{ky}{\sigma_N}\right)^{1/k}$$

[3]The case where $(Y, X) \in \Re^{1+D}$ with $X \in \Re^D$ and $D > 1$ can be analyzed with arguments that are similar to those we have used. The only differences reside on how the kernel function is defined and the speed of convergence of the relevant bandwidths to zero.

where σ has a subscript N to make explicit the fact that it depends on the threshold $U_{(n-N)}$. Without loss of generality we can write for $a \in (0,1)$ that $q(a) = U_{(n-N)} + y_{N,a}$ where by construction $F(U_{(n-N)} + y_{N,a}) = a$. Hence, if $1 - F(U_{(n-N)})$ is estimated by N/n, we have

$$\frac{1-a}{N/n} \approx \left(1 - \frac{ky}{\sigma_N}\right)^{1/k}, \tag{6}$$

which suggests $y_{N,a} \approx \frac{\sigma_N}{k}\left(1 - \left(\frac{(1-a)n}{N}\right)^k\right)$. The approximation in (6) is the basis for our proposed estimator $\hat{q}(a)$ for $q(a)$, which is given by

$$\hat{q}(a) = \tilde{q}(a_n) + \hat{y}_{N,a} = \tilde{q}(a_n) + \frac{\tilde{\sigma}_N}{\tilde{k}}\left(1 - \left(\frac{(1-a)n}{N}\right)^{\tilde{k}}\right). \tag{7}$$

Lastly, an estimator for $q_{Y|X=x}(a)$ is given by $\hat{q}_{Y|X=x}(a) = \hat{m}(x) + \hat{q}(a)$. In the next section we provide asymptotic properties for $(\tilde{\sigma}_N, \tilde{k})$, $\hat{q}(a)$ and $\hat{q}_{Y|X=x}(a)$.

3. Asymptotic properties of the proposed estimators.

3.1. *Preliminaries.* We start by discussing some results in Smith (1987) as they are helpful in understanding our contribution and provide the basis for understanding our strategy for proving the main theorems. As mentioned above, contrary to our setting where the variables Y and X are related through a regression model, in Smith (1987) the estimation of $q(a)$ is conducted under the assumption that the sequence $\{Z_j\}_{i=1}^N$ is observed. As such, he proposes estimators $(\hat{\sigma}_N, \hat{k})$ that satisfy the first order conditions

$$\frac{\partial}{\partial \sigma}\frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \hat{\sigma}_N, \hat{k}) = 0 \text{ and } \frac{\partial}{\partial k}\frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \hat{\sigma}_N, \hat{k}) = 0 \tag{8}$$

associated with the likelihood function $L_N(\sigma, k) = \frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \sigma, k)$. Following Smith (1985) it will be convenient to reparametrize the likelihood function and represent arbitrary values σ and k as $\sigma = \sigma_N(1 + t\delta_N)$, $k = k_0 + \tau\delta_N$ for $t, \tau \in \Re$, $\delta_N \to 0$ as $N \to \infty$ and some σ_N and k_0. Hence, we can rewrite the likelihood function $L_N(\sigma, k)$ as $L_{TN}(t, \tau) = \frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \sigma_N(1 + t\delta_N), k_0 + \tau\delta_N)$. It is evident that: a) $L_{TN}(0,0) = L_N(\sigma_N, k_0)$; b) choosing $(\hat{\sigma}_N, \hat{k})$ such that equation (8) is satisfied is equivalent to choosing t^* and τ^* that satisfy

$$\frac{1}{\sigma_N\delta_N}\frac{\partial L_{TN}}{\partial t}(t^*, \tau^*) = 0 \text{ and } \frac{1}{\delta_N}\frac{\partial L_{TN}}{\partial \tau}(t^*, \tau^*) = 0.$$

Using Taylor's Theorem, for $\lambda_1, \lambda_2 \in [0,1]$, these first order conditions can be expanded around $(0,0)$ and can be written as

$$\begin{aligned}
\frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t,\tau) &= \frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial \sigma} log\, g(Z_i; \sigma_N, k_0)\frac{\sigma_N}{\delta_N} \\
&+ \frac{1}{N}\sum_{i=1}^N \frac{\partial^2}{\partial \sigma^2} log\, g(Z_i; \sigma_N(1 + \delta_N t\lambda_1), k_0 + \delta_N\tau\lambda_2)\sigma_N^2 t \\
&+ \frac{1}{N}\sum_{i=1}^N \frac{\partial^2}{\partial \sigma \partial k} log\, g(Z_i; \sigma_N(1 + \delta_N t\lambda_1), k_0 + \delta_N\tau\lambda_2)\sigma_N\tau = I_{1N} + I_{2N} + I_{3N}
\end{aligned}$$

and

$$\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau}L_{TN}(t,\tau) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial k}log\, g(Z_i;\sigma_N,k_0)\frac{1}{\delta_N}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial k\partial\sigma}log\, g(Z_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\sigma_N t$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial k^2}log\, g(Z_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\tau = I_{4N}+I_{5N}+I_{6N},$$

where the terms I_{lN} for $l=1,\cdots,6$ denote the corresponding average in the preceding equality.

Smith (1987) showed that if the class to which F belongs is restricted to satisfy,
FR1: $F\in D(\Phi_\alpha)$, that is, F belongs to the domain of attraction of a Fréchet distribution with index α,

FR2: $L(x)=x^\alpha(1-F(x))$ satisfies $\frac{L(tx)}{L(x)}=1+k(t)\phi(x)+o(\phi(x))$ as $x\to\infty$ for each $t>0$, where $0<\phi(x)\to 0$ as $x\to\infty$ is regularly varying with index $\rho\le 0$ and $k(t)=C\int_1^t u^{\rho-1}du$, for a constant C,

and its associated density $f(\cdot)$ is strictly positive, then for $\sigma_N=U_{(n-N)}/\alpha$, $0<\alpha=-1/k_0$ and $k_0<1/2$ we have:

$$E\left(\sigma_N\frac{\partial}{\partial\sigma}log\, g(Z;\sigma_N,k_0)\right)=\frac{C\phi(U_{(n-N)})}{(1+\alpha-\rho)}+o(\phi(U_{(n-N)}))$$
$$E\left(\frac{\partial}{\partial k}log\, g(Z;\sigma_N,k_0)\right)=-\frac{\alpha C\phi(U_{(n-N)})}{(\alpha-\rho)(1+\alpha-\rho)}+o(\phi(U_{(n-N)}))$$
$$E\left(\sigma_N^2\frac{\partial^2}{\partial\sigma^2}log\, g(Z;\sigma_N,k_0)\right)=-\frac{\alpha}{2+\alpha}+O(\phi(U_{(n-N)}))$$
$$E\left(\frac{\partial^2}{\partial k^2}log\, g(Z;\sigma_N,k_0)\right)=-\frac{2\alpha^2}{(1+\alpha)(2+\alpha)}+O(\phi(U_{(n-N)}))\text{ and}$$
$$E\left(\sigma_N\frac{\partial^2}{\partial\sigma\partial k}log\, g(Z;\sigma_N,k_0)\right)=\frac{\alpha^2}{(1+\alpha)(2+\alpha)}+O(\phi(U_{(n-N)})),$$

where all expectations are taken with respect to the unknown distribution $F_{U_{(n-N)}}$.

We note that condition FR1 is equivalent to $1-F(x)$ being regularly varying at ∞ with index $-\alpha$. In addition, by Karamata's Theorem $1-F(x)=c(x)exp\left(-\int_1^x t^{-1}\alpha(t)dt\right)$ for $x\ge 1$ and for measurable $c(x),\alpha(x):(1,\infty)\to\Re$ such that $lim_{x\to\infty}c(x)=c>0$, $lim_{x\to\infty}\alpha(x)=\alpha>0$ for some $c,\alpha>0$ (Resnick (1987)). In fact, given the density f and if in Karamata's Theorem $c(x)$ is a constant when x is sufficiently large, then $lim_{x\to\infty}\frac{xf(x)}{1-F(x)}=\alpha$, a result we (and Smith) use repeatedly. Restricting F to $D(\Phi_\alpha)$ is not entirely arbitrary. If $F\in D(\Psi_\alpha)$, the domain of attraction of a (reverse) Weibull distribution, then it must be that u_∞ is finite, a restriction which is not commonly placed on the regression error U. The only other possibility is F in the domain of attraction of a Gumbel distribution, $F\in D(\Lambda)$. In this case, whenever u_∞ is not finite we have that $1-F$ is rapidly varying, a case we will avoid.

In addition to the expectations listed above, Smith showed that $I_{1N}=O_p(N^{-1/2}\delta_N^{-1})$, $I_{4N}=O_p(N^{-1/2}\delta_N^{-1})$ and provided $N^{1/2}\delta_N\to\infty$ and $N^{1/2}\phi(U_{(n-N)})=O(1)$ we have I_{1N}, $I_{4N}=o_p(1)$. Furthermore, $I_{2N}=-\frac{\alpha}{1+\alpha}+o_p(1)$, $I_{3N}=\frac{\alpha^2}{(1+\alpha)(2+\alpha)}+o_p(1)$, $I_{5N}=\frac{\alpha^2}{(1+\alpha)(2+\alpha)}+o_p(1)$, $I_{6N}=-\frac{2\alpha^2}{(1+\alpha)(2+\alpha)}+o_p(1)$ uniformly on $S_T=\{(t,\tau):t^2+\tau^2<1\}$. Consequently, $\frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t,\tau)\xrightarrow{p}$

$t\left(-\frac{\alpha}{1+\alpha}\right)+\tau\left(\frac{\alpha^2}{(1+\alpha)(2+\alpha)}\right)$, $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau}L_{TN}(t,\tau)\xrightarrow{p} t\left(\frac{\alpha^2}{(1+\alpha)(2+\alpha)}\right)+\tau\left(-\frac{2\alpha^2}{(1+\alpha)(2+\alpha)}\right)$, which combined with the fact that $H=-\begin{pmatrix}-\frac{\alpha}{1+\alpha} & \frac{\alpha^2}{(1+\alpha)(2+\alpha)}\\ \frac{\alpha^2}{(1+\alpha)(2+\alpha)} & -\frac{2\alpha^2}{(1+\alpha)(2+\alpha)}\end{pmatrix}$ is assumed to be positive definite gives

$$\begin{pmatrix} t & \tau\end{pmatrix}\begin{pmatrix}\frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t,\tau)\\ \frac{1}{\delta_N^2}\frac{\partial}{\partial\tau}L_{TN}(t,\tau)\end{pmatrix}\xrightarrow{p}\begin{pmatrix} t & \tau\end{pmatrix}(-H)\begin{pmatrix} t\\ \tau\end{pmatrix}\leq 0 \text{ on } S_T. \tag{9}$$

Using Lemma 5 in Smith (1985) we can then conclude that $\frac{1}{\delta_N^2}L_{TN}(t,\tau)$ has, with probability approaching 1, a local maximum (t^*,τ^*) on $S_T=\{(t,\tau): t^2+\tau^2<1\}$ at which $\frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t^*,\tau^*)=0$ and $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau}L_{TN}(t^*,\tau^*)=0$. Put differently, there exists, with probability approaching 1, a local maximum $(\hat\sigma_N=\sigma_N(1+t^*\delta_N),\ \hat k=k_0+\tau^*\delta_N)$ on $S_R=\{(\sigma,k): \|(\frac{\sigma}{\sigma_N}-1,\ k-k_0)\|<\delta_N\}$ that satisfy the first order conditions in equation (8). Our first lemma establishes a similar result for the estimator $(\tilde\sigma_N,\tilde k)$ that satisfies the first order condition given in equation (5).[4] However, given that we must deal with estimated sequences $\tilde Z_j$, additional assumptions are needed.

3.2. *Assumptions.* As in Smith (1987) we retain FR1, FR2 and the assumption that $\{U_i\}_{i=1}^n$ forms an independent and identically distributed sequence of random variables with absolutely continuous and strictly increasing distribution $F\in D(\Phi_\alpha)$, with $\alpha=-1/k_0$ and $k_0<1/2$. Given equation (1) and the nonparametric estimation of m and q additional assumptions are needed.

Assumption A1: The kernel functions $K_i(x)$ for $i=1,2$ are symmetric, twice continuously differentiable functions $K_i(x): S_i\to\Re$, where S_i are bounded sets. They satisfy $\int_{S_i}K_i(s)ds=1$ and $\int_{S_i}sK_i(s)ds=0$. $\int_{S_1}s^2K_1(s)ds=\sigma_{K_1}^2<\infty$ and $\int s^jK_2(s)ds=0$ for $j=1,\cdots,m$ where $m>2$. We denote the j^{th} order derivative of K_i by $K_i^{(j)}$ and assume that $|K_i(u)-K_i(v)|\leq C|u-v|$ and $|K_i^{(1)}(u)-K_i^{(1)}(v)|\leq C|u-v|$ for some constant $C>0$.

The higher order m for K_2 is necessary in the proof of Lemma 2. All other assumption are common in the nonparametric estimation literature and are easily satisfied by a variety of commonly used kernels.

Assumption A2: The bandwidths $0<h_{in}\to 0$ as $n\to\infty$ for $i=1,2$. In addition, we assume that $h_{1n}\propto n^{-1/5}$, $h_{2n}\propto n^{-1/5+\delta}$ for $\delta>0$ and $\frac{n}{\sqrt{N}}h_{2n}^{m+1}\to 0$ as $n\to\infty$.

The last condition puts a restriction on the relative speed of N and h_{2n} as $n\to\infty$. Given the orders of h_{1n} and h_{2n} it suffices to choose $N\propto n^{4/5-\delta}$. In this case, all orders in A2 are satisfied and, as needed in Smith (1987), $N^{1/2}\delta_N\to\infty$ and $N^{1/2}\phi(U_{(n-N)})=O(1)$.

Assumption A3: $F(u)$ is absolutely continuous with density $0<f(u)$ for all $u<u_\infty=l.u.b\{u: F(u)<1\}$. f is m-times continuously differentiable with derivative function satisfying $|f^{(j)}(u)|<C$ for some constant C and $j=1,\cdots,m$.

The differentiability restrictions on f are necessary in the proof of Lemma 2.

Assumption A4: $\{(X_i,U_i)\}_{i=1,\cdots,n}$ is a sequence of independent and identically distributed random vectors with density equal to that of the vector (X,U) and given by $f_{XU}(x,u)$. We denote the marginal density of X by $f_X(x)$ and the conditional density of U given X by $f_{U|X}(u)$. We assume that $E(U|X)=0$ and $V(U|X)=1$ and that $\frac{f_{U|X}(u)}{f(u)}\to 1$ as $u\to\infty$.

The requirement that $\frac{f_{U|X}(u)}{f(u)}\to 1$ implies that U and X are asymptotically independent.

Assumption A5: $m(x)$ is twice continuously differentiable at all $x\in G$ and $f_X(x)$ is continuously differentiable at all $x\in G$, G compact.

Assumption A5 is sufficient for equation (4) to hold but can be relaxed at some cost. It is, however, standard in the nonparametric literature (Li and Racine (2007), Fan and Yao (2003)).

[4] $\|x\|$ denotes the Euclidean norm of the vector x.

3.3. *Existence of* $\tilde{\sigma}_N$ *and* $\tilde{k}$. We now establish the existence of $\tilde{\sigma}_N$ and $\tilde{k}$. The strategy of the proof is to show that the first order conditions associated with the likelihood function

$$\tilde{L}_{TN}(t,\tau) = \frac{1}{N}\sum_{j=1}^{N} log\, g(\tilde{Z}_j; \sigma_N(1+t\delta_N), k_0 + \tau\delta_N)$$

are asymptotically uniformly equivalent in probability to those associated with L_{TN} on the set S_T. Formally, we have

Lemma 1 *Let* $t, \tau \in \Re$, $0 < \delta_N \to 0$, $\delta_N N^{1/2} \to \infty$ *as* $N \to \infty$ *and denote arbitrary* σ *and* k *by* $\sigma = \sigma_N(1+t\delta_N)$ *and* $k = k_0 + \tau\delta_N$*. We define the log-likelihood function*

$$\tilde{L}_{TN}(t,\tau) = \frac{1}{N}\sum_{j=1}^{N} log\, g(\tilde{Z}_j; \sigma_N(1+t\delta_N), k_0 + \tau\delta_N),$$

where $\tilde{Z}_j = \hat{U}_{(n-N+j)} - \tilde{q}(a_n)$, $a_n = 1 - \frac{N}{n}$, $\tilde{q}(\cdot)$ *and* $\hat{U}_{(n-N+j)}$ *are as defined in section 2. Given conditions FR1, FR2 and assumptions A1-A5. Then, as* $n \to \infty$ $\frac{1}{\delta_N^2}\tilde{L}_{TN}(t,\tau)$ *has, with probability approaching* 1*, a local maximum* (t^*, τ^*) *on* $S_T = \{(t,\tau) : t^2 + \tau^2 < 1\}$ *at which* $\frac{1}{\delta_N^2}\frac{\partial}{\partial t}\tilde{L}_{TN}(t^*,\tau^*) = 0$ *and* $\frac{1}{\delta_N^2}\frac{\partial}{\partial \tau}\tilde{L}_{TN}(t^*,\tau^*) = 0$.

The vector (t^*, τ^*) implies a value $\tilde{\sigma}_N$ and $\tilde{k}$ which are solutions for the likelihood equations

$$\frac{\partial}{\partial \sigma}\frac{1}{N}\sum_{j=1}^{N} log\, g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0 \text{ and } \frac{\partial}{\partial k}\frac{1}{N}\sum_{j=1}^{N} log\, g(\tilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0.$$

Hence, there exists, with probability approaching 1, a local maximum ($\tilde{\sigma}_N = \sigma_N(1+t^*\delta_N)$, $\tilde{k} = k_0 + \tau^*\delta_N$) on $S_R = \{(\sigma, k) : \|(\frac{\sigma}{\sigma_N} - 1, k - k_0)\| < \delta_N\}$ that satisfy the first order conditions in equation (5).

The proof depends critically on two auxiliary results. First, there is a need for $\hat{m}$ to be uniformly asymptotically close to m at a certain order. Specifically, we need for a compact set G that $q_n(a_n)^{-1} sup_{x\in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$. This assures that the residuals $\hat{U}_i$ are in some sense *close* to the unobserved U_i. Second, in Lemma 2 (see appendix) $\tilde{q}(a_n)$ is shown to be asymptotically close to $q_n(a_n)$ by satisfying $\frac{\tilde{q}(a_n) - q_n(a_n)}{q_n(a_n)} = O_p(N^{-1/2})$.

It is important to emphasize that Lemma 1 (as Theorem 3.2 in Smith (1987)) does not provide a "consistency" result for the ML estimator. In fact, since the distribution $F_{U_{(n-N)}}$ is only approximately a GPD, there are no true values for the parameters of the GPD to which $\tilde{\sigma}$ and $\tilde{k}$ are approaching in probability. What the Lemma does state is that the solutions for the first order conditions listed in (5) correspond to a local maximum of the likelihood associated with the GPD in a shrinking neighborhood of the arbitrary point (σ_N, k_0).

3.4. *Asymptotic normality of* $\tilde{\gamma}' = (\tilde{\sigma}_N, \tilde{k})$. Smith (1987) showed that given conditions FR1, FR2 and provided $\{Z_j\}_{j=1}^N$ is an independent and identically distributed sequence from $F_{U_{(n-N)}}$, $N \to \infty$ and $\frac{C}{\alpha-\rho}N^{1/2}\phi(U_{(n-N)}) \to \mu \in \Re$, the local maximum $(\hat{\sigma}_N, \hat{k})$ of the GPD likelihood function, is such that for $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = \frac{q_n(a_n)}{\alpha}$

$$\sqrt{N}\begin{pmatrix} \frac{\hat{\sigma}_N}{\sigma_N} - 1 \\ \hat{k} - k_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \frac{\mu(1-k_0)(1+2k\rho)}{1-k_0+k_0\rho} \\ \frac{\mu(1-k_0)k_0(1+\rho)}{1-k_0+k_0\rho} \end{pmatrix}, H^{-1}\right)$$

where $H = \frac{1}{(1-2k_0)(1-k_0)}\begin{pmatrix} 1-k_0 & -1 \\ -1 & 2 \end{pmatrix}$.[5] Our first theorem provides a similar asymptotic result for the estimators $(\tilde{\sigma}_N, \tilde{k})$.

Theorem 1 *Suppose FR1, FR2, A1-A5 hold and that* $\frac{C}{\alpha-\rho}N^{1/2}\phi(U_{(n-N)}) \to \mu \in \Re$. *The local maximum* $(\tilde{\sigma}_N, \tilde{k})$ *of the GPD likelihood function, is such that for* $k_0 = -\frac{1}{\alpha}$ *and* $\sigma_N = \frac{q_n(a_n)}{\alpha}$

$$\sqrt{N}\begin{pmatrix} \frac{\tilde{\sigma}_N}{\sigma_N} - 1 \\ \tilde{k} - k_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \frac{\mu(1-k_0)(1+2k_0\rho)}{1-k_0+k_0\rho} \\ \frac{\mu(1-k_0)k(1+\rho)}{1-k_0+k_0\rho} \end{pmatrix}, H^{-1}V_2H^{-1}\right)$$

where $V_2 = \begin{pmatrix} \frac{k_0^2-4k_0+2}{(2k_0-1)^2} & \frac{-1}{k_0(k_0-1)} \\ \frac{-1}{k_0(k_0-1)} & \frac{2k_0^3-2k_0^2+2k_0-1}{2k_0^3-2k_0^2+2k_0-1} \end{pmatrix}$.

We note that the use of $\tilde{Z}_j$ instead of Z_j in the estimation impacts the variance of the asymptotic distribution. It is easy to verify that $H^{-1}V_2H^{-1} - H^{-1}$ is positive definite, implying an (expected) loss of efficiency that results from estimating U_i nonparametrically. However, any additional bias introduced by the nonparametric estimation is of second order effect as the asymptotic bias derived in Smith (1987) is precisely the same as the one we obtain in Theorem 1. An important note on the proof is that the fact that $\tilde{Z}_j$ is *not* iid as Z_j does not require the use of a CLT for dependent processes as justified in Lemma 3 in the appendix.

3.5. *Asymptotic normality of* $\hat{q}(a)$. The asymptotic distribution of the ML estimators given in Theorem 1 is the basis for obtaining a normality result for $\hat{q}(a)$ given in equation (7). The basic idea is to define, without loss of generality, $q(a) = q(a_n) + y_{N,a}$ for $a_n = 1 - N/n < a$ and estimate $q(a_n)$ by $\tilde{q}(a_n)$ and $y_{N,a}$ based on the estimated parameters of the GPD. It is important to note that, in Theorem 2, as $n \to \infty$ both a_n and a approach 1.

Theorem 2 *Suppose FR1, FR2 and assumptions A1-A5 hold. In addition, assume that,*
(i) $N^{1/2}C\phi(q(a_n))/(\alpha-\rho) \to \mu$ *with* $k_0 = -\frac{1}{\alpha}$ *and* $\sigma_N = q(a_n)/\alpha$ *and*
(ii) $n(1-a) \propto N$. *Then, for some* $z_a > 0$

$$\sqrt{n(1-a)}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \xrightarrow{d} N\left((-k_0)\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C} - c_b'H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right),\right.$$
$$\left. k_0^2\left(c_b'H^{-1}V_2H^{-1}c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\right).$$

where $c_b' = \begin{pmatrix} k_0^{-1}(z_a^{-1}-1) & k_0^{-2}log(z_a) + k_0^{-2}(z_a^{-1}-1) \end{pmatrix}$, $b_\sigma = E\left(\frac{\partial}{\partial\sigma}log\, g(Z_j;\sigma_N,k_0)\sigma_N\right)$ *and* $b_k = E\left(\frac{\partial}{\partial k}log\, g(Z_j;\sigma_N,k_0)\right)$.

Under the assumptions of Theorems 1 and 2 it is a direct consequence of the linear properties of limits that for all $a \in (0,1)$, $\hat{q}_{Y|X=x}(a) = \hat{m}(x) + \hat{q}(a) \xrightarrow{p} m(x) + q(a) = q_{Y|X=x}(a)$.

4. Simulations. We conduct a simulation study to implement our parameter estimators $\tilde{\gamma}' = (\tilde{\sigma}_N, \tilde{k})$ and quantile estimator $\hat{q}$, and compare them with some alternatives available in the literature. We generate data independently from

$$Y_i = m(X_i) + U_i,\ i = 1, \cdots, n$$

[5]Substituting $k_0 = -\alpha^{-1}$ shows that H is identical to the homonymous matrix in equation (9).

where X_i is distributed as a standard normal. We consider two nonlinear functions for $m(\cdot)$, $m_1(x) = 3sin(3x)$ and $m_2(x) = x^2$. U_i is generated independently from a distribution with density f that is in the domain of attraction of the Fréchet distribution Φ_α with index $\alpha = -1/k_0$.

The first distribution we considered is the log-gamma distribution, whose density is given by

$$f(u) = (log(u))^{\alpha-1}\frac{u^{-\frac{1}{\beta}-1}}{\beta^\alpha\Gamma(\alpha)}, \text{ for } u > 1, \alpha, \beta > 0.$$

It is easy to see that U_i is log-gamma distributed for $U_i > 1$ if and only if $log(U_i) > 0$ is gamma distributed with parameters $\alpha, \beta > 0$. Furthermore, one can show that $E(U_i) = (\frac{1}{1-\beta})^\alpha$, $V(U_i) = (\frac{1}{1-2\beta})^\alpha - (\frac{1}{1-\beta})^{2\alpha}$, and $k_0 = -\frac{1}{\beta}$. The Log-gamma distribution includes the Pareto distribution as a special case when $\alpha = 1$. We specifically let $(\alpha, \beta) = (1, 0.25)$, and $(1, 0.5)$, which correspond to $k_0 = -4$ and -2 respectively. Both the mean and variance of U_i exist for $(\alpha, \beta) = (1, 0.25)$, but the variance does not exist for $(\alpha, \beta) = (1, 0.5)$. U_i is demeaned since we use it as an error term in the regression model.

The second distribution we considered is the student-t distribution with v degrees of freedom. It can be shown that $k_0 = -\frac{1}{v}$, which is $k_0 = -1/3$ for $v = 3$, and $-1/2$ for $v = 2$ respectively. Here, when $v = 2$, the variance of U_i does not exist. Thus, the distributions we consider allow k_0 to take values in a wide range. We expect that the estimation will be relatively more difficult when the variance does not exist.

Implementation of our estimator requires the choice of bandwidths h_{1n} and h_{2n}. We select them using *rule-of-thumb* bandwidths $\hat{h}_{1n} = 1.25S(X)n^{-\frac{1}{5}}$ and $\hat{h}_{2n} = 0.79R(X)n^{-\frac{1}{5}}$, with a robust estimation for the variability of data as in (2.52) of Pagan and Ullah (1999), where $S(X)$ and $R(X)$ are the standard deviation and the sample interquartile range of X, respectively. We choose the second order Epanechnikov kernel for both the estimation of $m(x)$ and the smoothed sample quantile. The choice of bandwidths satisfies the restrictions imposed to obtain the asymptotic properties in Theorems 1 and 2. Our assumptions also call for the use of a higher order kernel in estimating the smoothed sample quantile. Here we investigate the robustness of our estimator with the popular second order Epanechnikov kernel for its simplicity.

In estimating the parameters, we include our estimator $\tilde{\gamma}$, Smith's estimator $\hat{\gamma} = (\hat{\sigma}_N, \hat{k})$, which utilizes the true U_i available in the simulation, and $\hat{k}^h$ for k_0, the estimator proposed by Hill (1975). Hill's estimator is designed for data from a heavy-tailed distribution with $k_0 < 0$ and has been studied extensively in the literature (Embrechts, Kluppelberg and Mikosh (1997)). It is generally the most efficient estimator of k_0 for sensible choices of N, though it is generally not the most efficient nor the most stable quantile estimator (McNeil and Frey (2000)). Since U_i is unknown in practice, we use $\hat{U}_i = Y_i - \hat{m}(x_i)$ to construct $\hat{k}^h = -\frac{1}{N}\sum_{j=1}^{N}(ln(\hat{U}_{(n-N+j)}) - ln(\hat{U}_{(n-N)}))$. The theoretical properties of $\hat{k}^h$ are unknown and here we investigate its finite sample performance relative to the estimator we propose. In estimating the a-quantile, we include our estimator $\hat{q}$, Smith's estimator q^s, Hill type estimator q^h and empirical quantile estimator q^e. Following (6.30) in Embrechts, Kluppelberg and Mikosh (1997), we construct $q^h = \hat{U}_{(n-N)}(\frac{1-a}{N/n})^{\hat{k}^h}$. q^e is simply the empirical quantile estimator based on $\{\hat{U}_i\}_{i=1}^n$. To give the reader a vivid picture of them in practice, we provide in Figure 1 a plot of different quantile estimates against different values of a, where q^s is omitted for ease of illustration. a ranges from 0.95 to 0.995 because we are interested in higher order quantiles. The data are generated with $m(x) = 3sin(3x)$, U_i is from the student-t distribution with $v = 2$ degree of freedom, and we select $n = 1000$ and $N = 100$. Both $\hat{q}$ and q^h are smooth functions of a, while q^e is not. All three estimators seem to capture the low order quantile well, though differences start to be more noticeable for a approaching one.

We fix the pairs $n = 500$ with $N = 50$, and $n = 1000$ with $N = 100$ in our simulation. We follow this simple choice, because the effective sample size N in the second stage estimation is doubled. We did not explicitly consider the choice of $N = O(n^{-4/5})$ as in our asymptotic analysis, since our proposed estimator seems to be relatively robust to the choice of N. On the other hand, the choice of N is critical for q^h, as its performance deteriorates quickly with N, as seen in Figures 2 and 3 and in the discussion below. Each experiment is repeated 5000 times. We summarize the performance of parameter estimators in terms of their mean (M), bias (B) (in the parameter k_0 only), and standard deviation (S) in Table 1 for both $m(x) = 3sin(3x)$ and $m(x) = x^2$ with log-gamma distributed U with $\alpha = 1$ and $\beta = 0.25$, in Table 2 with $\alpha = 1$ and $\beta = 0.5$, in Tables 3 and 4 with student-t distributed U with $v = 3$ and $v = 2$ respectively. We provide the performance of $0.95, 0.99$ and 0.995 quantile estimators in terms of the bias (B), standard deviation (S) and root mean squared error (R) in Tables 5-12. Specifically, results for log-gamma distributed U with $(\alpha, \beta) = (1, 0.25)$ and $m(x) = 3sin(3x)$ are detailed in Table 5, for $m(x) = x^2$ in Table 6, for log-gamma distributed U using $(\alpha, \beta) = (1, 0.5)$ in Tables 7 and 8, for student-t distributed U using $v = 3$ in Tables 9 and 10, for student-t distributed U using $v = 2$ in Tables 11 and 12.

In the case of estimating the parameters, we notice that $\hat{\gamma}$ and $\tilde{\gamma}$ tend to overestimate k_0. $\hat{k}^h$ carries a positive bias for the log-gamma distributed U, and a negative bias for the student-t distributed U. As N increases, all estimators' performance improve, in the sense that their standard deviation decreases and the bias of estimators of k_0 is also reduced. This seems to confirm the asymptotic results in the previous section. As we move from Table 1 to 2 and from Table 3 to 4, we find that the standard deviation of all estimators increase, and the bias of k_0 parameter estimators decreases. We think this is related to the bias and variance trade-off for the parameter estimation. As we have mentioned above, the variance of U does not exist for log-gamma distributions with $\beta = 0.5$ in Table 2 and for student-t distribution with $v = 2$ in Table 4. The distribution of U start to exhibit heavier tail behavior, thus more representative extreme observations have a higher probability to show up in a sample, which explains lower bias. Among three estimators for k_0, $\hat{k}^h$ in general has the best performance in terms of low bias and standard deviation, with exceptions in bias for the student-t distributed U. The two estimators $\hat{\gamma}$ and $\tilde{\gamma}$ estimate both σ_N and k_0 with very similar performances. Relative to $\hat{\gamma}$, $\tilde{\gamma}$ exhibits lower bias in estimating k_0, but slightly larger standard deviation in estimating both parameters. It seems to suggest our proposed estimator $\tilde{\gamma}$ is well supported by the NW estimator for the function $m(x)$.

In the case of estimating the quantile, we notice q^h carries a negative bias for estimating the 95% quantile, but positive bias for the 99% and 99.5% quantiles. q^e always underestimates the larger order quantiles. As N increases, all estimators' performances improve in terms of smaller bias, standard deviation and root mean squared error. The exception is q^h, whose bias increases with N. The distribution of U exhibits a heavier tail with $\beta = 0.5$ (Tables 7 and 8) relative to $\beta = 0.25$ (Tables 5 and 6) in the log-gamma distribution, with $v = 2$ (Tables 11 and 12) relative to $v = 3$ (Tables 9 and 10) in the student-t distribution. As we have mentioned above, the random variable U does not have a variance in these cases. We find it more difficult for all to estimate the quantiles across all experiment designs, with some exceptions in the bias. As expected, when we estimate higher order a-quantile, all estimators' performances deteriorate, with some exceptions in bias. When we estimate the 95% quantile, which is relatively close to the center of the distribution, there is no absolute dominance by any one of the methods. By counting the number of times of being the best estimator across all 16 experiments, q^s has smallest bias 9 times, q^h has the smallest root mean squared error 8 times. The advantage of the Hill type estimator does not seem to carry through in estimating the higher order-99%, and 99.5% quantiles. $\hat{q}$ and q^s are consistently the best with the smallest bias, standard deviation and root mean squared error, where $\hat{q}$ seems to have slightly larger bias, but smaller root mean squared error relative to q^s. q^h always carries the largest bias, being better than q^e in terms of root mean squared error only in the case U is student-t distributed.

The choice of N could be important since the number of residuals exceeding the threshold is based on $\hat{U}_{(n-N)}$. We need to choose large $\hat{U}_{(n-N)}$ to reduce the bias from approximating the tail distribution with GPD, but we need to keep N large (small $\hat{U}_{(n-N)}$) to control the variance of parameter estimates. We illustrate the impact from different N's on the performance of different estimators for the 99% quantile of U with a simulation, where we set $n = 1000$, $m(x) = 3sin(3x)$, and use a student-t distributed U. The bias and root mean squared error (RMSE) of the estimators q^s, $\hat{q}$, q^h, and q^e are plotted against $N = 20, 25, \cdots, 200$ in Figures 2 and 3 respectively. We notice q^e is negatively biased while q^s and $\hat{q}$ carry relatively small positive biases. These three estimators' biases are fairly stable across N. q^h's bias is influenced heavily by N, being smallest when N ranges from 20 to 60, largest with N greater than 70. The strong dependence of q^h's performance on N also exhibits in RMSE in Figure 2. Between $N = 20$ and 70, q^h performs best, but its performance deteriorates quickly when N is larger than 70. As expected, q^s and $\hat{q}$'s RMSE decrease with N from 20 to 70, but when N is larger than 70, their RMSE's are stable, close to each other, and are smaller than those of q^h and q^e. q^s and $\hat{q}$ almost always dominate q^e, which did not utilize the extreme value theory. The result indicates q^h's performance is sensitive to the choice of N, requiring a small N to control its bias, while q^s and $\hat{q}$ work well in a broader range of N's.

5. Summary and conclusions. The estimation of higher order quantiles associated with the distribution of a random variable Y is of great interest in many applied fields. It is also common for researchers in these fields to specify regression or location-scale models that relate Y to a set of covariates X. As such, they are often interested in the estimation of high order conditional quantiles associated with the conditional distribution of Y given X, i.e, $q_{Y|X=x}(a) = m(x) + q(a)$. The main difficulty in obtaining an estimator for $q_{Y|X=x}$ rests on the fact that the regression errors which could be used to estimate $q(a)$ are not observed. In this paper we have expanded the seminal work of Smith (1987), which considered the estimation of $q(a)$ when the associated random variable is observed, to the case where only regression residuals are available for the estimation of $q(a)$. Our results are based on a nonparametric estimation of the regression and a ML estimation of the distribution tail based on a GPD. We provide a full asymptotic characterization of the ML estimators for the parameters of the GPD and for the estimator $\hat{q}(a)$ for $q(a)$. It is encouraging to see that the asymptotic normality results of Smith are preserved albeit with a loss of estimation precision.

It should be emphasized that richer location-scale or regression models than the one we considered is an important extension of our work. For example, in empirical finance, the evolution of returns of a financial asset is normally modeled by dynamic location-scale models that require the estimation of both a regression and a conditional skedastic function. Furthermore, in this context the independent and identically distributed assumption we used throughout is normally inadequate. However, we are encouraged that our work has provided a framework in which these richer stochastic specifications can be studied.

Appendix 1 - Proofs. Throughout the proofs, C will represent an inconsequential and arbitrary constant that may take different values in different locations. χ_A denotes the indicator function for the set A, $P(A)$ denotes the probability of event A from the probability space $(\Omega, \mathcal{F}, P)$.

Lemma 1 .

PROOF. Given the results described in section 3.1 and Taylor's Theorem, for $\lambda_1, \lambda_2 \in [0, 1]$, we have

$$\frac{1}{\delta_N^2}\frac{\partial}{\partial t}\tilde{L}_{TN}(t,\tau) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\sigma}log\, g(\tilde{Z}_i;\sigma_N,k_0)\frac{\sigma_N}{\delta_N} + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial\sigma^2}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\sigma_N^2 t + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial\sigma\partial k}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\sigma_N\tau = \tilde{I}_{1N}+\tilde{I}_{2N}+\tilde{I}_{3N}$$

and

$$\frac{1}{\delta_N^2}\frac{\partial}{\partial \tau}\tilde{L}_{TN}(t,\tau) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial k}log\, g(\tilde{Z}_i;\sigma_N,k_0)\frac{1}{\delta_N} + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial k\partial\sigma}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\sigma_N t + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial k^2}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N t\lambda_1),k_0+\delta_N\tau\lambda_2)\tau = \tilde{I}_{4N}+\tilde{I}_{5N}+\tilde{I}_{6N}.$$

Let $a_n = 1-\frac{N}{n}$, $\tilde{E}_i = \{\hat{U}_i > \tilde{q}\,(a_n)\}$, $E_i = \{U_i > q_n\,(a_n)\}$, $Z_i = U_i - q_n(a_n)$ and $\tilde{Z}_i = \hat{U}_i - \tilde{q}(a_n)$ for $i = 1,\cdots,n$. Then, from section 3.1, we have

$$\tilde{I}_{1N} - I_{1N} = \frac{1}{\delta_N}(k_0^{-1}-1)\left(\frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_i}{\sigma_N}\chi_{\tilde{E}_i} - \left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\chi_{\tilde{E}_i}\right) + \frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\left(\chi_{\tilde{E}_i}-\chi_{E_i}\right)\right) = \frac{1}{\delta_N}(k_0^{-1}-1)(I_{11n}+I_{12n})$$

We first study I_{11n}. By the mean value theorem, for some $\lambda \in [0,1]$ and $Z_i^* = Z_i + \lambda(\tilde{Z}_i - Z_i)$ we have

$$\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_i}{\sigma_N}\chi_{\tilde{E}_i} - \left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\chi_{\tilde{E}_i} = \frac{k_0/\sigma_N}{\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^2}\chi_{\tilde{E}_i}(\tilde{Z}_i - Z_i),$$

and consequently we can write

$$I_{11n} = -\frac{1}{N}\sum_{i=1}^{n} q_n(a_n)\frac{k_0/\sigma_N}{\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^2}\chi_{\tilde{E}_i}\left(\frac{\hat{m}(X_i)-m(X_i)}{q_n(a_n)} + \frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)}\right).$$

From Lemma 2 $\frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)} = O_p(N^{-1/2})$. In addition, given that $q_n(a_n) \to \infty$ as $n \to \infty$, equation (4), and provided $N \propto n^{4/5-\delta}$ and $h_{1n} \propto n^{-1/5}$ we have $\frac{1}{q_n(a_n)}\sup_{x\in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$. Since $\sigma_N = -q_n(a_n)k_0$ we have that $I_{11n} \leq O_p(N^{-1/2})\left(\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^{-2}\chi_{\tilde{E}_i}\right)$. Note that $Z_i^* = Z_i + \frac{\lambda}{\sqrt{N}}O_p(1)$, hence

$$\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^{-2}\chi_{\tilde{E}_i} = \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0(Z_j+o_p(1))}{\sigma_N}\right)^{-2} = O_p(1), \tag{10}$$

and consequently $I_{11n} = O_p(N^{-1/2})$. We now consider I_{12n}, which can be written as

$$I_{12n} = \frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\left(\chi_{\tilde{E}_i}-\chi_{E_i}\right)\chi_{\tilde{E}_i\cup E_i}.$$

For $\delta_1, \delta_2 > 0$ we define the events $A = \left\{\omega : \frac{|\hat{U}_i - U_i|}{q_n(a_n)} < \delta_1\right\}$ and $B = \left\{\omega : \frac{|\tilde{q}(a_n)-q_n(a_n)|}{q_n(a_n)} < \delta_2\right\}$ and note that $C^c \subseteq A^c \cup B^c$, where $C = \{\omega : \chi_{\tilde{E}_i} - \chi_{E_i} = 0\}$. Hence, $\chi_{C^c} \leq \chi_{A^c} + \chi_{B^c}$ and

$$\begin{aligned} I_{12n} &\leq \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\chi_{A^c}\chi_{\tilde{E}_i\cup E_i} + \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\chi_{B^c}\chi_{\tilde{E}_i\cup E_i} \\ &= I_{121n} + I_{122n}. \end{aligned}$$

Since $\delta_1 > 0$ we have $\frac{|\hat{U}_i - U_i|}{\delta_1 q_n(a_n)} > 1$ on A^c and $\frac{|\tilde{q}(a_n)-q_n(a_n)|}{\delta_2 q_n(a_n)} > 1$ on B^c. Therefore,

$$I_{121n} < \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\frac{|\hat{U}_i - U_i|}{\delta_1 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i} \text{ and}$$

$$I_{122n} < \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\frac{|\tilde{q}(a_n)-q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i}.$$

Since $k_0 < 0$ and $\sigma_N > 0$ we have that $\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right| < C$. In addition, since $\frac{1}{q_n(a_n)}\sup_{x\in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ and the fact that $\chi_{\tilde{E}_i\cup E_i}$ limits the number of nonzero terms in the sum to at most $2N$, we have $I_{121n} < C\, o_p(N^{-1/2})$. Similarly, by Lemma 2 $I_{122n} < C\, O_p(N^{-1/2})$ and we have $I_{12n} = O_p(N^{-1/2})$. Combining the orders of I_{11n} and I_{12n} we conclude that $\tilde{I}_{1N} - I_{1N} = \frac{1}{\delta_N}(k_0^{-1}-1)O_p(N^{-1/2})$. Since, $\delta_N N^{1/2} \to \infty$ as $n \to \infty$ we have $\tilde{I}_{1N} - I_{1N} = o_p(1)$.

We now turn to establishing that $\tilde{I}_{4N} - I_{4N} = o_p(1)$. We write

$$\begin{aligned} \tilde{I}_{4N} - I_4 &= \frac{1}{\delta_N}\left(\frac{1}{N}\sum_{i=1}^{n}\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_i}{\sigma_N}\right.\right. \\ &- \left.\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right)\right)\chi_{\tilde{E}_i} \\ &+ \frac{1}{N}\sum_{i=1}^{n}\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right)\right)(\chi_{\tilde{E}_i}-\chi_{E_i}) \\ &= \frac{1}{\delta_N}(I_{41n}+I_{42n}). \end{aligned}$$

First, note that

$$I_{41n} = \frac{1}{k_0}\left(1-\frac{1}{k_0}\right)I_{11n} - \frac{1}{k_0^2}\frac{1}{N}\sum_{i=1}^{n}\left(log\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)-log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)\right)\chi_{\tilde{E}_i}.$$

Since we have already established that $I_{11n} = O_p(N^{-1/2})$, it suffices to investigate the order of the second term. By the mean value theorem, Lemma 2 and the fact that $\frac{1}{q_n(a_n)}\sup_{x\in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ we have

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{n}\left(log\left(1-\frac{k_0\tilde{Z}_i}{\sigma_N}\right)-log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)\right)\chi_{\tilde{E}_i} &< \frac{1}{k_0^2}o_p(N^{-1/2})\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^{-1}\chi_{\tilde{E}_i} \\ &+ \frac{1}{k_0^2}O_p(N^{-1/2})\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^{-1}\chi_{\tilde{E}_i}. \end{aligned}$$

Using the same arguments as in (10), we have $\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{k_0 Z_i^*}{\sigma_N}\right)^{-1}\chi_{\tilde{E}_i} = O_p(1)$, and consequently $I_{41n} = O_p(N^{-1/2})$. By the same arguments used when studying I_{12n}, we have

$$\begin{aligned} I_{42n} &< \frac{1}{N}\sum_{i=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\frac{|\hat{U}_i - U_i|}{\delta_1 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i} \\ &\quad + \frac{1}{N}\sum_{i=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right|\frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i} \\ &\quad = I_{421n} + I_{422n}, \end{aligned}$$

and provided $\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right| < C$ we have $I_{421n} = o_p(N^{-1/2})$ and $I_{422n} = O_p(N^{-1/2})$. To verify the bound it is sufficient to show that $\left|log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)\right| < C$ since $\left|\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right| < C$ from the study of I_{12n}. By the mean value theorem and for some $\lambda_i \in (0,1)$

$$\left|log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)\right| = \frac{1}{\lambda_i}\left|\left(1-\frac{\lambda_i k_0 Z_i}{\sigma_N}\right)^{-1}\frac{-k_0\lambda_i Z_i}{\sigma_N}\right| < C \tag{11}$$

provided λ_i is bounded away from zero for all i. Given that I_{41n}, $I_{42n} = O_p(N^{-1/2})$ we have $\tilde{I}_{4N} - I_{4N} = o_p(1)$, since $\delta_N N^{1/2} \to \infty$ as $n \to \infty$.

We now investigate the order of $\tilde{I}_{2N} - I_{2N}$. Consider arbitrary $\dot{\sigma}_N = \sigma_N(1+\delta_N t\lambda_1)$ and $\dot{k} = k_0 + \delta_N\tau\lambda_2$ and write

$$\begin{aligned} \tilde{I}_{2N} - I_{2N} = \frac{1}{(1+t\delta_N\lambda_1)^2}&\left((-2)\left(\frac{1}{\dot{k}}-1\right)\frac{1}{N}\sum_{j=1}^{N}\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}-\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right. \\ &\left.-\left(\frac{1}{\dot{k}}-1\right)\frac{1}{N}\sum_{j=1}^{N}\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^2-\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^2\right)\right). \end{aligned}$$

Hence, it suffices to examine

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N}&\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{l}-\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{l}\right) = \frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{\dot{k}\tilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_i}{\dot{\sigma}_N}\right)^{l}\right. \\ &\left.-\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\chi_{\tilde{E}_i} + \frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}(\chi_{\tilde{E}_i}-\chi_{E_i}) = I_{nl1} + I_{nl2} \end{aligned}$$

for $l = 1, 2$. By the mean value theorem, there exists $Z_i^* = \tilde{Z}_i + \lambda(\tilde{Z}_i - Z_i)$ for $\lambda \in (0,1)$ such that

$$I_{nl1} = l\frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}q_n(a_n)\left(-\frac{\hat{m}(X_i)-m(X_i)}{q_n(a_n)}-\frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)}\right)\chi_{\tilde{E}_i}.$$

Given Lemma 2 and the fact that $q_n(a_n)^{-1}\sup_{x\in G}|\hat{m}(x)-m(x)| = o_p(N^{-1/2})$ we can write

$$I_{nl1} \le O_p(N^{-1/2})\frac{l}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}q_n(a_n)\right|\chi_{\tilde{E}_i}.$$

Since $q_n(a_n) = -\sigma_N/k_0$ we have

$$\underset{S_T}{sup}\frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}q_n(a_n)\right|\chi_{\tilde{E}_i} = \underset{S_T}{sup}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\tilde{\sigma}}\right|\frac{1}{N}\sum_{i=1}^{n}\chi_{\tilde{E}_i}\underset{S_T}{sup}\left|\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}\right.$$
$$\left.\times\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\right|.$$

Now, given that $\delta_N \to 0$ we have for N sufficiently large $\underset{S_T}{sup}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\tilde{\sigma}}\right| < C$. In addition, this fact combined with $\frac{|\tilde{Z}_i - Z_i|}{q_n(a_n)} = o_p(1)$ uniformly (from equation (4) and Lemma 2) gives

$$\underset{S_T}{sup}\left|\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\right| < C$$

and consequently $I_{nl1} = O_p(N^{-1/2})$ uniformly in S_T. Now, as we have argued previously, we can write

$$|I_{nl2}| \le \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right|\chi_{A^c}\chi_{\tilde{E}_i\cup E_i} + \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right|\chi_{B^c}\chi_{\tilde{E}_i\cup E_i}$$
$$= I_{nl21} + I_{nl22}$$

where $I_{nl21} < \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right|\frac{|\hat{U}_i - U_i|}{\delta_1 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i}$ and $I_{nl22} < \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right| \times \frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i}$. Given that $\delta_N \to 0$ we have for N sufficiently large that $\dot{k} < 0$, $\dot{\sigma}_N > 0$ and $\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right| < C$. In addition, as argued above, since $\frac{1}{q_n(a_n)}\underset{x\in G}{sup}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ and the fact that $\chi_{\tilde{E}_i\cup E_i}$ limits the number of nonzero terms in the sum to at most $2N$, we have $I_{nl21} < C\,o_p(N^{-1/2})$. Similarly, by Lemma 2 $I_{nl22} < C\,O_p(N^{-1/2})$ and we have $I_{nl2} = O_p(N^{-1/2})$. Combining the orders of I_{nl1} and I_{nl2} we conclude that $\tilde{I}_{2N} - I_{2N} = o_p(1)$ uniformly on S_T. Now, note that $\tilde{I}_{3N} - I_{3N} = \tilde{I}_{5N} - I_{5N}$ and

$$\tilde{I}_{3N} - I_{3N} = \frac{1}{1+\delta_N t\lambda_1}\frac{1}{N}\sum_{j=1}^{N}\left(-\frac{1}{\dot{k}^2}\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N} - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right.$$
$$+\frac{1}{\dot{k}}\left(\frac{1}{\dot{k}}-1\right)\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{2} - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{2}\right)$$
$$\left.+\frac{1}{\dot{k}}\left(\frac{1}{\dot{k}}-1\right)\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right) - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right)\right)$$

and using the same arguments as in the case of $\tilde{I}_{2N} - I_{2N}$ we have $\tilde{I}_{3N} - I_{3N} = o_p(1)$ and $\tilde{I}_{5N} - I_{5N} = o_p(1)$ uniformly on S_T. Lastly, we investigate the order of $\tilde{I}_{6N} - I_{6N}$ which can be written as

$$\tilde{I}_{6N} - I_{6N} = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{2}{\dot{k}^3}\left(log\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right)\right.$$
$$+\frac{1}{\dot{k}}\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right) - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right)$$
$$+\frac{1}{\dot{k}^3}\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right) - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right)$$
$$\left.-\frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}}-1\right)\left(\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right)^2 - \left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^2\right)\right)$$
$$(12) \qquad = \frac{2}{\dot{k}^3}\frac{1}{N}\sum_{j=1}^{N}\left(log\left(1-\frac{\dot{k}\tilde{Z}_j}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right) + o_p(1) \text{ uniformly in } S_T.$$

The last equality follows form the arguments used above when investigating the order of $\tilde{I}_{2N} - I_{2N}$. The first term in equation (12) can be written as (excluding the constant $2/\dot{k}^3$)

$$(13) \qquad \frac{1}{N}\sum_{i=1}^{n}\left(log\left(1-\frac{\dot{k}\tilde{Z}_i}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right)\chi_{\tilde{E}_i} + \frac{1}{N}\sum_{i=1}^{n}log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)(\chi_{\tilde{E}_i} - \chi_{E_i}).$$

Using the mean value theorem, Lemma 2 and $\frac{1}{q_n(a_n)}\underset{x\in G}{sup}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ we have

$$\frac{1}{N}\sum_{i=1}^{n}\left(log\left(1-\frac{\dot{k}\tilde{Z}_i}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right)\chi_{\tilde{E}_i} \le \left(o_p(N^{-1/2}) + O_p(N^{-1/2})\right)$$
$$\times \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}}{\dot{\sigma}_N}\right|\chi_{\tilde{E}_i}$$
$$= \left(o_p(N^{-1/2}) + O_p(N^{-1/2})\right)O_p(1) = o_p(1)$$

uniformly in S_T using the same arguments given above. Lastly, the second term in (13) can be written as

$$\frac{1}{N}\sum_{i=1}^{n}log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)(\chi_{\tilde{E}_i} - \chi_{E_i})\chi_{\tilde{E}_i\cup E_i} \le \frac{1}{N}\sum_{i=1}^{n}\left|log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\chi_{A^c}\chi_{\tilde{E}_i\cup E_i}$$
$$+\frac{1}{N}\sum_{i=1}^{n}\left|log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\chi_{B^c}\chi_{\tilde{E}_i\cup E_i}$$
$$< \frac{1}{N}\sum_{i=1}^{n}\left|log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\frac{|\hat{U}_i - U_i|}{\delta_1 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i}$$
$$+\frac{1}{N}\sum_{i=1}^{n}\left|log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{\tilde{E}_i\cup E_i}$$

As argued above, given that $\delta_N \to 0$ as $N \to \infty$ there exists N sufficiently large such that $\dot{k} < 0$, $\dot{\sigma}_N > 0$ and we have that $\left|log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right| < C$ as in equation (11). In addition, by Lemma 2, $\frac{1}{q_n(a_n)}\underset{x\in G}{sup}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ and the fact that $\chi_{\tilde{E}_i\cup E_i}$ limits the number of nonzero terms in the sum to at most $2N$, we have that $\frac{1}{N}\sum_{i=1}^{n}log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)(\chi_{\tilde{E}_i} - \chi_{E_i})\chi_{\tilde{E}_i\cup E_i} = o_p(1)$. □

Lemma 2 *Under assumptions A1-A5 and conditions FR1 and FR2 we have*

$$N^{1/2}\left(\frac{\tilde{q}\left(a_n\right)-q_n\left(a_n\right)}{q\left(a_n\right)}\right)=O_p(1), \ \ where\ a_n=1-\tfrac{N}{n}.$$

Proof. We write

$$\sqrt{N}\left(\frac{\tilde{q}\left(a_n\right)-q_n\left(a_n\right)}{q\left(a_n\right)}\right)=\sqrt{N}\left(\frac{\tilde{q}\left(a_n\right)-q\left(a_n\right)}{q\left(a_n\right)}\right)-\sqrt{N}\left(\frac{q_n\left(a_n\right)-q\left(a_n\right)}{q\left(a_n\right)}\right)=T_{1n}-T_{2n}.$$

We first show that T_{2n} converges in distribution, which implies $T_{2n}=O_p(1)$. Note that

$$P(T_{2n}\leq z)=P\left(\frac{nk_0}{\sqrt{N}}(F(y_n)-a_n)\leq\frac{nk_0}{\sqrt{N}}(F_n(y_n)-F(y_n))\right)$$

with $y_n=q(a_n)+z\sigma_n$ and $\sigma_n=\frac{q(a_n)}{\sqrt{N}}$. By the mean value theorem, $F(y_n)=a_n+f(q^*(a_n))\sigma_n z$ where $q^*(a_n)=q(a_n)+\lambda\sigma_n z=q(a_n)(1+\lambda zN^{-1/2})$ for some $\lambda\in[0,1]$. Thus,

$$\frac{nk_0}{\sqrt{N}}(F(y_n)-a_n)=\frac{nk_0}{N}f(q^*(a_n))q(a_n)z=k_0\frac{(1-F(q^*(a_n)))n}{N}\frac{q(a_n)f(q^*(a_n))}{1-F(q^*(a_n))}z.$$

Since $q^*(a_n)=q(a_n)(1+o(1))$ we have that $\lim_{n\to\infty}\frac{(1-F(q^*(a_n)))n}{N}=1$. In addition, by Proposition 1.15 in Resnick (1987) $\lim_{n\to\infty}\frac{q(a_n)f(q^*(a_n))}{1-F(q^*(a_n))}=-\frac{1}{k_0}$, hence $\lim_{n\to\infty}\frac{nk_0}{\sqrt{N}}(F(y_n)-a_n)=-z$. We now show that $\frac{n}{\sqrt{N}}(F_n(y_n)-F(y_n))\xrightarrow{d}N(0,1)$. First, we observe that $\frac{n}{\sqrt{N}}-\frac{\sqrt{n(1-F(y_n))}}{1-F(y_n)}=o(1)$, hence we show that

$$\frac{\sqrt{n(1-F(y_n))}}{1-F(y_n)}(F_n(y_n)-F(y_n))=\sum_{i=1}^{n}Z_{in}\xrightarrow{d}N(0,1) \tag{14}$$

where $Z_{in}=\frac{(1-F(y_n))^{-1/2}}{\sqrt{n}}(\chi\{U_i\leq y_n\}-E(\chi\{U_i\leq y_n\}))$. It is readily verified that $E(Z_{in})=0$ and $V(Z_{in})=n^{-1}F(y_n)$. Hence, given that $\sum_{i=1}^{n}E(|Z_{in}|^3)\leq 2(n(1-F(y_n)))^{-1/2}=o(1)$ we have by Liapounov's CLT that $\frac{n}{\sqrt{N}}(F_n(y_n)-F(y_n))\xrightarrow{d}N(0,1)$. Hence, $T_{2n}\xrightarrow{d}N(0,k_0^2)$.

We now show that $T_{1n}=O_p(1)$ by establishing that T_{1n} converges in distribution. As above,

$$P(T_{1n}\leq z)=P\left(\frac{nk_0}{\sqrt{N}}(F(y_n)-a_n)\leq\frac{nk_0}{\sqrt{N}}(\tilde{F}(y_n)-F(y_n))\right), \tag{15}$$

and we establish that $\frac{n}{\sqrt{N}}(\tilde{F}(y_n)-F(y_n))\xrightarrow{d}N(0,1)$. We start by noting that for some $\lambda\in[0,1]$

$$\begin{aligned}\tilde{F}(y_n) &= \int_{-\infty}^{y_n}\frac{1}{nh_{2n}}\sum_{i=1}^{n}K_2\left(\frac{y-U_i}{h_{2n}}\right)dy-\int_{-\infty}^{y_n}\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}K_2^{(1)}\left(\frac{y-U_i}{h_{2n}}\right)dy(\hat{U}_i-U_i)\\ &+ \frac{1}{2}\int_{-\infty}^{y_n}\frac{1}{nh_{2n}^3}\sum_{i=1}^{n}K_2^{(2)}\left(\frac{y-U_i^*}{h_{2n}}\right)dy(\hat{U}_i-U_i)^2=Q_{1n}+Q_{2n}+Q_{3n}.\end{aligned}$$

where $U_i^*=\lambda U_i+(1-\lambda)\hat{U}_i$. Therefore,

$$\frac{n}{\sqrt{N}}(\tilde{F}(y_n)-F(y_n))=\frac{n}{\sqrt{N}}\left((Q_{1n}-F(y))+Q_{2n}+Q_{3n}\right).$$

We first examine Q_{3n}. Given (4) and the fact that $K_2^{(2)}$ is symmetric (A1) we have that

$$Q_{3n} \leq O_p\left(\frac{\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2\right)^2}{h_{2n}}\right)\frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n-U_i^*}{h_{2n}}\right)\right|.$$

Using Taylor's Theorem we can write for some $\lambda \in [0,1]$ and $U_i^{**} = \lambda U_i + (1-\lambda)U_i^*$ that

$$\frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n-U_i^*}{h_{2n}}\right)\right| \leq \frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n-U_i}{h_{2n}}\right)\right| + \frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n-U_i^{**}}{h_{2n}}\right)(U_i^*-U_i)\right|.$$

Provided that $|K_2^{(1)}(x)| < C$ on the bounded support S_2 (A1) and given that $f(y_n) \to 0$ as $n \to \infty$ we have $\frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n-U_i}{h_{2n}}\right)\right| = o_p(1)$. Given that $U_i^* - U_i = \lambda(m(X_i) - \hat{m}(X_i))$ and (4) we have $\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n-U_i^{**}}{h_{2n}}\right)(U_i^*-U_i)\right| \leq O_p\left(\frac{\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2}{h_{2n}^2}\right)\frac{1}{n}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n-U_i^{**}}{h_{2n}}\right)\right| \leq O_p\left(\frac{\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2}{h_{2n}^2}\right)$. Hence, this term is bounded in probability if $\frac{\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2}{h_{2n}^2} = O(1)$, which follows if $h_{1n} = O(h_{2n})$ and $nh_{1n}h_{2n}^4 = O(log\, n)$. These orders are satisfied by taking $h_{1n} \propto n^{-1/5}$ and $h_{2n} \propto n^{-1/5+\delta}$ for $\delta > 0$. Hence, under these conditions

$$\text{(16)} \qquad \frac{n}{\sqrt{N}}Q_{3n} = \frac{n}{\sqrt{N}}O_p\left(\frac{\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2\right)^2}{h_{2n}}\right) = o_p(1), \text{ provided } N \propto n^{4/5-\delta}.$$

$Q_{2n} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{2n}}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)(\hat{m}(X_i)-m(X_i))$, and using the fact that

$$\hat{m}(x) - m(x) - \frac{1}{nh_{1n}f_X(x)}\sum_{t=1}^{n}K_1\left(\frac{X_t-x}{h_{1n}}\right)Y_t^* = O_p\left(\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2\right)^2\right)$$

uniformly over a compact set G, with $Y_t^* = m^{(1)}(x)(X_t-x) + \frac{1}{2}m^{(2)}(x^*)(X_t-x)^2$, $x^* = \lambda X_t - (1-\lambda)x$ and $\lambda \in [0,1]$, we can write

$$\begin{aligned}
Q_{2n} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{h_{2n}f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)m^{(1)}(X_i)(X_t-X_i)\\
&+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{h_{2n}f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)\frac{1}{2}m^{(2)}(X_t^*)(X_t-X_i)^2\\
&+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{h_{2n}f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)U_t + O_p\left(\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2\right)^2\right)\\
&= Q_{21n} + Q_{22n} + Q_{23n} + O_p\left(\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2}+h_{1n}^2\right)^2\right).
\end{aligned}$$

We will obtain the order of each Q_{2jn} for $j = 1, 2, 3$ separately. Let

$$\psi_n(Z_i, Z_t) = \frac{1}{f_X(X_i)h_{2n}}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)U_t,$$

for $Z_i = (X_i, U_i)$ and write $Q_{23n} = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}(\psi_n(Z_t, Z_i) + \psi_n(Z_i, Z_t)) = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\phi_n(Z_i, Z_t)$ where $\phi_n(Z_t, Z_i)$ is a symmetric function. The partial sum for the case where $i = t$ is denoted by $Q'_{23n} = \frac{K_1(0)}{n^2 h_{2n} h_{1n}}\sum_{i=1}^{n}\frac{1}{f_X(X_i)}K_2\left(\frac{y_n - U_i}{h_{2n}}\right)U_i$, and given that $\frac{f_{U|X}(u)}{f(u)} \to 1$ as $u \to \infty$, Proposition 1.15 in Resnick (1987) and $\frac{1-F(y_n - h_{2n}u)}{1-F(y_n)} \to 1$ as $n \to \infty$, we have by Lebesgue's dominated convergence theorem that $\frac{n}{\sqrt{N}}Q'_{23n} = o_p(1)$. For the case where $i \neq t$ we write the remaining partial sums as

$$Q''_{23n} = \frac{1}{n}\sum_{t=1}^{n}E(\phi_n(Z_t, Z_i)|Z_t) - \frac{1}{2}E(\phi_n(Z_t, Z_i)) + O_p\left(n^{-1}(E(\phi_n^2(Z_t, Z_i)))^{1/2}\right)$$

Given that $E(U_i|X_i) = 0$, we have $E(\phi_n(Z_t, Z_i)) = 0$ and

$$E(\phi_n(Z_t, Z_i)|Z_t) = \frac{1}{n}\sum_{i=1}^{n}E\left(\frac{1}{f_X(X_i)}\frac{1}{h_{1n}}K_1\left(\frac{X_t - X_i}{h_{1n}}\right)\frac{1}{h_{2n}}K_2\left(\frac{y_n - U_i}{h_{2n}}\right)|X_t\right)U_t = \frac{1}{n}\sum_{i=1}^{n}Z_{tn}$$

with $E(Z_{tn}) = 0$. As above, using A4, Proposition 1.15 in Resnick (1987) and Lebesgue's dominated convergence theorem we have that $E\left(\frac{n y_n}{N}Z_{tn}^2\right) \to -k_0^{-1}$. Using similar arguments we have $n^{-1}E\left(\phi_n^2(Z_t, Z_i)\right)^{1/2} = O\left(n^{-1}\left(\frac{N}{n y_n h_{1n} h_{2n}}\right)^{1/2}\right)$. Consequently,

$$\frac{n}{\sqrt{N}}O_p\left(n^{-1}E\left(\phi_n^2(Z_t, Z_i)\right)^{1/2}\right) = O_p\left(\left(\frac{1}{n h_{1n} h_{2n} y_n}\right)^{1/2}\right) = o_p\left((n h_{1n} h_{2n})^{-1/2}\right) = o_p(1)$$

since $y_n \to \infty$ and $n h_{1n} h_{2n} \to \infty$. Hence, we can write that $\frac{n}{\sqrt{N}}\sqrt{y_n}Q''_{23n} = \frac{1}{n}\sum_{i=1}^{n}Z_{tn}\sqrt{y_n} + o_p(1)$. Since $E(Z_{tn}\sqrt{y_n}) = 0$ and $E\left(\frac{n y_n}{N}Z_{tn}^2\right) \to -k_0^{-1}$, by Liapounov's CLT we have $\frac{n}{\sqrt{N}}\sqrt{y_n}Q''_{23n} \overset{d}{\to} N(0, -k_0^{-1})$, and since $\sqrt{y_n} \to \infty$ as $n \to \infty$ we have that $\frac{n}{\sqrt{N}}Q''_{23n} = o_p(1)$.

Using similar arguments and manipulations we obtain $\frac{n}{\sqrt{N}}Q_{21n} = o_p\left(h_{1n}^2\sqrt{N}\right) + o_p(1)$ and $\frac{n}{\sqrt{N}}Q_{22n} = o_p\left(h_{1n}^2\sqrt{N}\right) + o_p(1)$. Hence, combining the orders for Q_{21n}, Q_{22n} and Q_{23n} we have

$$\frac{n}{\sqrt{N}}Q_{2n} = o_p\left(h_{1n}^2\sqrt{N}\right) + o_p(1) + \frac{n}{\sqrt{N}}O_p\left(\left(\left(\frac{n h_{1n}}{log\, n}\right)^{-1/2} + h_{1n}^2\right)^2\right). \tag{17}$$

Given that $h_{1n} \propto n^{-1/5}$ and $h_{2n} \propto n^{-1/5+\delta}$ for $\delta > 0$ and $N \propto n^{4/5-\delta}$, $\frac{n}{\sqrt{N}}Q_{2n} = o_p(1)$.

We now show that $\frac{n}{\sqrt{N}}(Q_{1n} - F(y_n)) \overset{d}{\to} N(0,1)$. First, we put $q_{1n} = \frac{1}{h_{2n}}\int_{-\infty}^{y_n}K_2\left(\frac{y - U_i}{h_{2n}}\right)dy$ and write

$$\begin{aligned}\frac{n}{\sqrt{N}}(Q_{1n} - F(y_n)) &= \sum_{i=1}^{n}\frac{1}{\sqrt{n(1 - F(y_n))}}(q_{1n} - E(q_{1n})) + \sum_{i=1}^{n}\frac{1}{\sqrt{n(1 - F(y_n))}}(E(q_{1n}) - F(y_n)) \\ &= I_{1n} + I_{2n}.\end{aligned}$$

Clearly, $E\left(\frac{1}{\sqrt{n(1-F(y_n))}}(q_{1n} - E(q_{1n}))\right) = 0$ and $V\left(\frac{1}{\sqrt{n(1-F(y_n))}}(q_{1n} - E(q_{1n}))\right) = \frac{s_n^2}{n(1-F(y_n))}$ where

$$s_n^2 = \int\frac{1}{h_{2n}}b\left(\frac{y_n - u}{h_{2n}}\right)F(u)du - \left(\int\frac{1}{h_{2n}}K_2\left(\frac{y_n - u}{h_{2n}}\right)F(u)du\right)^2$$

and $b(x) = 2K_2(x)\int_{-\infty}^{x}K_2(y)dy$. Define $s^2 = F(y_n)(1 - F(y_n))$ and write $\frac{s_n^2}{(1-F(y_n))} = \frac{s_n^2 - s^2}{1-F(y_n)} + F(y_n)$. Since, $\frac{s_n^2 - s^2}{1-F(y_n)} = o(h_{2n})$ and $F(y_n) \to 1$ as $n \to \infty$ we have $\frac{s_n^2}{1-F(y_n)} \to 1$. By Liapounov's

CLT, $I_{1n} \xrightarrow{d} N(0,1)$ provided that $E(|Z_{in}|^3) \to 0$ as $n \to \infty$, where

$$Z_{in} = \frac{1}{\sqrt{n(1-F(y_n))}} \left(\frac{1}{h_{2n}} \int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy - E\left(\frac{1}{h_{2n}} \int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy \right)\right).$$

The condition is verified by noting that

$$\left| \left(\frac{1}{h_{2n}} \int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy - E\left(\frac{1}{h_{2n}} \int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy \right)\right)\right| \leq 2$$

since $\frac{1}{h_{2n}} \int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy \leq 1$. Consequently, $|Z_{in}| \leq \frac{2}{\sqrt{n(1-F(y_n))}}$ and

$$E(|Z_{in}|^3) \leq \frac{2}{\sqrt{n(1-F(y_n))}} \frac{s_n^2}{n(1-F(y_n))} \to 0 \text{ as } n \to \infty.$$

Integrating by parts we have

$$|E(q_{1n}) - F(y_n)| = \left| \int (-h_{2n})\psi K_2(\psi) f(y_n) + \sum_{j=1}^{m-1} \frac{(-h_{2n}\psi)^{j+1}}{(j+1)!} f^{(j)}(y_n) + \frac{(-h_{2n}\psi)^{m+1}}{(m+1)!} f^{(m)}(y_n^*) d\psi \right|,$$

where $y_n^* = \lambda(y_n - h_{2n}\psi) + (1-\lambda)y_n$ for some $\lambda \in [0,1]$. Since K_2 is an m^{th}-order kernel and $|f^{(m)}(u)| < C$, we have that $|E(q_{1n}) - F(y_n)| \leq \frac{h_{2n}^{m+1}}{(m+1)!} \int |\psi^{m+1} K_2(\psi)| d\psi = O(h_{2n}^{m+1})$. Hence, $I_{2n} = O\left(\frac{n}{\sqrt{N}} h_{2n}^{m+1}\right) = o(1)$ and

$$\frac{n}{\sqrt{N}}(Q_{1n} - F(y_n)) \xrightarrow{d} N(0,1). \tag{18}$$

Equations (16), (17), and (18) show that $\frac{n}{\sqrt{N}}(\tilde{F}(y_n) - F(y_n)) \xrightarrow{d} N(0,1)$, and by consequence $T_{1n} = O_p(1)$ which completes the proof. □

Theorem 1

Proof. Let $\tilde{r}_N = \frac{\tilde{\sigma}_N}{\sigma_N} = 1 + \delta_N t^*$, $\tilde{k} = k_0 + \delta_N \tau^*$ and note that

$$\begin{pmatrix} \frac{1}{\delta_N^2}\frac{\partial}{\partial t} L_{TN}(t^*,\tau^*) \\ \frac{1}{\delta_N^2}\frac{\partial}{\partial \tau} L_{TN}(t^*,\tau^*) \end{pmatrix} = \frac{1}{\delta_N N} \begin{pmatrix} \sum_{j=1}^N \frac{\partial}{\partial r_N} log\, g(\tilde{Z}_j; \tilde{r}_N\sigma_N, \tilde{k}) \\ \sum_{j=1}^N \frac{\partial}{\partial k} log\, g(\tilde{Z}_j; \tilde{r}_N\sigma_N, \tilde{k}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{19}$$

For some λ_1, $\lambda_2 \in [0,1]$ let $k^* = \lambda_2 k_0 + (1-\lambda_2)\tilde{k}$, $r_N^* = \lambda_1 + (1-\lambda_1)\tilde{r}_N$,

$$H_N(r_N^*, k^*) = \frac{1}{N}\sum_{j=1}^N \begin{pmatrix} \frac{\partial^2}{\partial r_N^2} log\, g(\tilde{Z}_j; r_N^*\sigma_N, k^*) & \frac{\partial^2}{\partial k \partial r_N} log\, g(\tilde{Z}_j; r_N^*\sigma_N, k^*) \\ \frac{\partial^2}{\partial k \partial r_N} log\, g(\tilde{Z}_j; r_N^*\sigma_N, k^*) & \frac{\partial^2}{\partial k \partial k} log\, g(\tilde{Z}_j; r_N^*\sigma_N, k^*) \end{pmatrix} \text{ and}$$

$$v_N(1,k_0) = \sqrt{N} \begin{pmatrix} \frac{1}{N}\sum_{j=1}^N \frac{\partial}{\partial r_N} log\, g(\tilde{Z}_j; \sigma_N, k_0) \\ \frac{1}{N}\sum_{j=1}^N \frac{\partial}{\partial k} log\, g(\tilde{Z}_j; \sigma_N, k_0) \end{pmatrix} = \sqrt{N} \begin{pmatrix} \delta_N(\tilde{I}_{1N} - I_{1N}) + \delta_N I_{1N} \\ \delta_N(\tilde{I}_{4N} - I_{4N}) + \delta_N I_{4N} \end{pmatrix},$$

where $\tilde{I}_{1N}$, I_{1N}, $\tilde{I}_{4N}$, I_{4N} are as defined in Lemma 1. By a Taylor's expansion of the first order condition in (19) around $(1,k_0)$ we have

$$H_N(r_N^*, k^*)\sqrt{N} \begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} = v_N(1,k_0). \tag{20}$$

We start by investing the asymptotic properties of $v_N(1,k_0)$. Let

$$b_1 = -\frac{\alpha(1+\alpha)}{2+\alpha},\ b_2 = \left(-\frac{\alpha^2(1+\alpha)}{2+\alpha} + \frac{\alpha^3}{1+\alpha}\right)$$

and observe that from Lemma 2 and the fact that $\frac{q_n(a_n)}{q(a_n)} - 1 = o_p(1)$ we have that

$$\begin{aligned} v_N(1,k_0) &= \begin{pmatrix} b_1\sqrt{N}\frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)} + \delta_N\sqrt{N}I_{1N} + o_p(1) \\ b_2\sqrt{N}\frac{\tilde{q}(a_n)-q_n(a_n)}{q_n(a_n)} + \delta_N\sqrt{N}I_{4N} + o_p(1) \end{pmatrix} \\ &= \begin{pmatrix} b_1\sqrt{N}\left(\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)} - \frac{q_n(a_n)-q(a_n)}{q(a_n)}\right) + \delta_N\sqrt{N}I_{1N} + o_p(1) \\ b_2\sqrt{N}\left(\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)} - \frac{q_n(a_n)-q(a_n)}{q(a_n)}\right) + \delta_N\sqrt{N}I_{4N} + o_p(1) \end{pmatrix} \end{aligned}$$

By Lemma 3 and the fact that $N_1 - N = O_p(N^{1/2})$

$$\begin{pmatrix} \sqrt{N}\delta_N I_{1N} \\ \sqrt{N}\delta_N I_{4N} \end{pmatrix} = \begin{pmatrix} b_1\sqrt{N}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + \frac{1}{\sqrt{N}}\sum_{j=1}^N \frac{\partial}{\partial\sigma} log\, g(Z_j';\sigma_N,k_0)\sigma_N + o_p(1) \\ b_2\sqrt{N}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + \frac{1}{\sqrt{N}}\sum_{j=1}^N \frac{\partial}{\partial k} log\, g(Z_j';\sigma_N,k_0) + o_p(1) \end{pmatrix}$$

where $Z_j' = U_j - q(a_n)$ for $U_j > q(a_n)$. Hence, letting $b_\sigma = E\left(\frac{\partial}{\partial\sigma} log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)$ and $b_k = E\left(\frac{\partial}{\partial k} log\, g(Z_j';\sigma_N,k_0)\right)$ we have

$$v_N(1,k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} = \begin{pmatrix} b_1\sqrt{N}\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)} + \frac{1}{\sqrt{N}}\left(\sum_{j=1}^N \frac{\partial}{\partial\sigma} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_\sigma\right) + o_p(1) \\ b_2\sqrt{N}\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)} + \frac{1}{\sqrt{N}}\left(\sum_{j=1}^N \frac{\partial}{\partial k} log\, g(Z_j';\sigma_N,k_0) - b_k\right) + o_p(1) \end{pmatrix}.$$

Note that we can write

$$\begin{aligned} \frac{1}{\sqrt{N}}\left(\sum_{j=1}^N \frac{\partial}{\partial\sigma} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_\sigma\right) &= \sum_{i=1}^n N^{-1/2}\left(\frac{\partial}{\partial\sigma} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_\sigma\right)\chi_{\{U_i>q(a_n)\}} \\ &= \sum_{i=1}^n Z_{i1} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{N}}\left(\sum_{j=1}^N \frac{\partial}{\partial k} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_k\right) &= \sum_{i=1}^n N^{-1/2}\left(\frac{\partial}{\partial k} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_k\right)\chi_{\{U_i>q(a_n)\}} \\ &= \sum_{i=1}^n Z_{i2}. \end{aligned}$$

Also, from Lemma 2 we have that $\sqrt{N}\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)}$ is distributed asymptotically as $\sum_{i=1}^n(-k_0)(n(1-F(y_n)))^{-1/2}(q_{1n} - E(q_{1n})) + o_p(1) = \sum_{i=1}^n Z_{i3} + o_p(1)$ where $q_{1n} = \frac{1}{h_{2n}}\int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right)dy$ and $y_n = q(a_n)(1+N^{-1/2}z)$ for arbitrary z. It can be easily verified that $E(Z_{i1}) = E(Z_{i2}) = E(Z_{i3}) = 0$. In addition,

$$\begin{aligned} V(Z_{i1}) &= N^{-1}E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_\sigma\right)^2 P(\{U_i > q(a_n)\}) \\ &= n^{-1}E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_j';\sigma_N,k_0)\sigma_N - b_\sigma\right)^2 = n^{-1}\left(\frac{1}{1-2k_0} + o(1)\right) \end{aligned}$$

where the last equality follows from Smith (1987). Using similar arguments we obtain

$$V(Z_{i2}) = n^{-1}\left(\frac{2\alpha^2}{(1+\alpha)(2+\alpha)} + o(1)\right)$$

and from Lemma 2 we have that $V(Z_{i3}) = n^{-1}k_0^3 F(y_n) + o(h_{2n})$. We now define the vector $\psi_n = \sum_{i=1}^{n}(Z_{i1}, Z_{i2}, Z_{i3})'$ and for arbitrary $0 \neq \lambda \in \Re^3$ we consider $\lambda'\psi_n = \sum_{i=1}^{n}(\lambda_1 Z_{i1} + \lambda_2 Z_{i2} + \lambda_3 Z_{i3}) = \sum_{i=1}^{n} Z_{in}$. From above, we have that $E(Z_{in}) = 0$ and $V(Z_{in}) = \sum_{l=1}^{3}\lambda_d^2 E(Z_{id}^2) + 2\sum_{1\leq d<d'\leq 3}\lambda_d\lambda_{d'}E(Z_{id}Z_{id'})$. First, we consider $E(Z_{i1}Z_{i2})$ which can be written as

$$E(Z_{i1}Z_{i2}) = \frac{1}{n}T_{1n} - \frac{N}{n^2}b_\sigma b_k$$

where $T_{1n} = E\left(\frac{\partial}{\partial\sigma_N}log\, g(Z_j';\sigma_N,k_0)\sigma_N\frac{\partial}{\partial k}log\, g(Z_j';\sigma_N,k_0)\right)$. Since $b_\sigma = \frac{C\phi(U_{(n-N)})}{1+\alpha-\rho} + o(\phi(U_{(n-N)}))$ and $b_k = -\frac{C\alpha\phi(U_{(n-N)})}{(\alpha-\rho)(1+\alpha-\rho)} + o(\phi(U_{(n-N)}))$ we have that

$$E(Z_{i1}Z_{i2}) = \frac{1}{n}T_{1n} - O\left(\frac{(N^{1/2}\phi(U_{(n-N)}))^2}{n^2}\right) = \frac{1}{n}T_{1n} - n^{-2}O(1)$$

since $N^{1/2}\phi(U_{(n-N)}) = O(1)$. Now, note that

$$E(T_{1n}) = -b_k - \frac{1}{k_0}\left(\frac{1}{k_0}-1\right)^2 E\left(\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-2}\left(\frac{k_0Z_i'}{\sigma_N}\right)^2\right)$$
$$-\frac{1}{k_0^2}\left(\frac{1}{k_0}-1\right)E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_i'}{\sigma_N}\right)\right).$$

From Smith (1987) we have that $E\left(\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-2}\left(\frac{k_0Z_i'}{\sigma_N}\right)^2\right) = \frac{2}{(1+\alpha)(2+\alpha)} + O(\phi(U_{(n-N)}))$ and $b_k = O(\phi(U_{(n-N)}))$. From Lemma 4 we have that

$$E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_i'}{\sigma_N}\right)\right) = -\frac{1}{\alpha} + \frac{\alpha}{(1+\alpha)^2} + O(\phi(U_{(n-N)}))$$

which combined with the orders obtained for the other components of the expectation and the fact that $k_0 = -\alpha^{-1}$ give

$$E(Z_{i1}Z_{i2}) = -\frac{1}{n(k_0-1)(2k_0-1)} + \frac{1}{n}\phi(U_{(n-N)})O(1) - O(n^{-2}).$$

We now turn to $E(Z_{i1}Z_{i3})$ which can be written as

$$E(Z_{i1}Z_{i3}) = T_{2n} - E\left(N^{-1/2}\left(\frac{\partial}{\partial\sigma_N}log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)\chi_{U_i>q(a_n)}\right)E(q_{1n})(n(1-F(y_n)))^{-1/2},$$

where $T_{2n} = E\left(N^{-1/2}\left(\frac{\partial}{\partial\sigma_N}log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)\chi_{U_i>q(a_n)}(n(1-F(y_n)))^{-1/2}q_{1n}\right)$. We note that

$$E\left(N^{-1/2}\left(\frac{\partial}{\partial\sigma_N}log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)\chi_{\{U_i>q(a_n)\}}\right) = \frac{\sqrt{N}}{n}b_\sigma = \frac{\sqrt{N}}{n}O(\phi(U_{(n-N)})),$$

from Lemma 2 $E(q_{1n}) = F(y_n) + O(h_{2n}^{m+1}) = O(1)$ and since $(n(1-F(y_n)))^{-1/2}$ is asymptotically equivalent to $N^{-1/2}$, the second term in the covariance expression is of order

$$\frac{\sqrt{N}}{n}O(\phi(U_{(n-N)}))O(1)N^{-1/2} = n^{-1}O(\phi(U_{(n-N)})).$$

We now turn to T_{2n}, the first term in the covariance expression. Since $(n(1-F(y_n)))^{-1/2}$ is asymptotically equivalent to $N^{-1/2}$, we have by the Cauchy-Schwartz inequality

$$\begin{aligned} T_{2n} &= \frac{1}{n}E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_j';\sigma_N,k_0)\sigma_N q_{1n}\right) \\ &\leq \frac{1}{n}\left|E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_j';\sigma_N,k_0)\sigma_N q_{1n}\right)\right| \\ &\leq \frac{1}{n}\left(E\left(\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)^2\right)E(q_{1n}^2)\right)^{1/2} = n^{-1}o(1). \end{aligned}$$

Hence, $E(Z_{i1}Z_{i3}) = o(n^{-1})$. In a similar manner we obtain $E(Z_{i2}Z_{i3}) = o(n^{-1})$. Hence, $nV(Z_{in}) = \lambda' V_1\lambda + o(1)$, where

$$V_1 = \begin{pmatrix} \frac{1}{1-2k_0} & -\frac{1}{(k_0-1)(2k_0-1)} & 0 \\ -\frac{1}{(k_0-1)(2k_0-1)} & \frac{2}{(k_0-1)(2k_0-1)} & \\ 0 & 0 & k_0^2 \end{pmatrix}.$$

By Liapounov's CLT $\sum_{i=1}^n Z_{ni} \xrightarrow{d} N(0,\lambda' V_1\lambda)$ provided that $\sum_{i=1}^n E(|Z_{in}|^3) \to 0$. To verify this condition, it suffices to show that

$$(i)\sum_{i=1}^n E(|Z_{i1}|^3)\to 0;\ (ii)\sum_{i=1}^n E(|Z_{i2}|^3)\to 0;\ (iii)\sum_{i=1}^n E(|Z_{i3}|^3)\to 0.$$

(iii) was verified in Lemma 2, so we focus on (i) and (ii).

For (i), note that $\sum_{i=1}^n E(|Z_{1i}|^3) \leq \frac{1}{\sqrt{N}}E\left(\left|(1/k_0-1)(1-k_0Z_i'/\sigma_N)^{-1}k_0Z_i'/\sigma_N - 1\right|^3\right) \to 0$ provided $E(-(1-k_0Z_i'/\sigma_N)^{-3}(k_0Z_i'/\sigma_N)^3) < C$, which is easily verified by noting that

$$-(1-k_0Z_i'/\sigma_N)^{-3}(k_0Z_i'/\sigma_N)^3 < -(1-k_0Z_i'/\sigma_N)^{-3}(1-k_0Z_i'/\sigma_N)^3 = 1.$$

Lastly,

$$\sum_{i=1}^n E(|Z_{2i}|^3) \leq \frac{1}{\sqrt{N}}E\left(\left|-\frac{1}{k_0^2}log\left(1-k_0\frac{Z_i'}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-k_0\frac{Z_i'}{\sigma_N}\right)^{-1}k_0\frac{Z_i'}{\sigma_N}\right|^3\right) \to 0$$

provided $E\left(log(1-k_0Z_i'/\sigma_N)^3\right) < C$ give the bound we obtained in case (i). By FR2 and integrating by parts we have

$$\begin{aligned} E\left(log(1-k_0Z_i'/\sigma_N)^3\right) &= -\int_0^\infty log\left(1-k_0\frac{z}{\sigma_N}\right)^3 dF_{U_{(n-N)}(z)} \\ &= -\frac{1-F(U_{(n-N)}(1+\frac{z}{U_{(n-N)}}))}{1-F(U_{(n-N)})}(log(1+\frac{z}{U_{(n-N)}}))^3|_0^\infty \\ &+ \int_0^\infty \frac{L(U_{(n-N)}(1+\frac{z}{U_{(n-N)}}))}{L(U_{(n-N)})}(1+\frac{z}{U_{(n-N)}})^{-\alpha}3(log(1+\frac{z}{U_{(n-N)}}))^2 \\ &\times (1+\frac{z}{U_{(n-N)}})^{-1}(1/U_{(n-N)})dz = T_{1n}+T_{2n}. \end{aligned}$$

Three repeated applications of L'Hôpital's rule and Proposition 1.15 in Resnick (1987) gives $T_{1n} = 0$. For T_{2n} we have that given FR2 and again integrating by parts and letting $t = 1+z/U_{(n-N)}$

$$T_{2n} = \int_1^\infty 3(log(t))^2t^{-\alpha-1}dt + \phi(U_{(n-N)})\int_1^\infty 3(log(t))^2t^{-\alpha-1}\frac{C}{\rho}(t^\rho-1)dt + o(\phi(U_{(n-N)})).$$

It is easy to verify that $\int_1^\infty 3(log(t))^2 t^{-\alpha-1}dt = \frac{6}{\alpha^3}$ and consequently $T_{2n} = \frac{6}{\alpha^3} + O(\phi(U_{(n-N)}))$ which verifies (ii). By the Cramer-Wold theorem we have that $\psi_n \xrightarrow{d} N(0, V_1)$. Consequently, for any vector $\gamma \in \Re^2$ we have $\gamma'\left(v_N(\sigma_N, k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) \xrightarrow{d} N(0, \gamma' V_2 \gamma)$ where

$$V_2 = \begin{pmatrix} \frac{k_0^2-4k_0+2}{(2k_0-1)^2} & -\frac{1}{k_0(k_0-1)} \\ -\frac{1}{k_0(k_0-1)} & \frac{2k_0^3-2k_0^2+2k_0-1}{k_0^2(k_0-1)^2(2k_0-1)} \end{pmatrix}.$$

Again, by the Cramer-Wold theorem $\left(v_N(\sigma_N, k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) \xrightarrow{d} N(0, V_2)$. Hence, given equation (20), provided that $H_N(\sigma_N^*, k^*) \xrightarrow{p} H$ we have

$$\sqrt{N}\begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} - H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} = H^{-1}\left(v_N(\sigma_N, k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) \xrightarrow{d} N\left(0, H^{-1}V_2H^{-1}\right).$$

To see that $H_N(\sigma_N^*, k^*) \xrightarrow{p} H$, first observe that whenever $(t, \tau) \in S_T$ we have $(\tilde{r}_N, \tilde{k}) \in S_R$ and consequently $(r_N^*, k^*) \in S_R$. In addition, from Lemma 1 and the results from Smith (1987) we have $H_N(r_N, k) \xrightarrow{p} -H$ uniformly on S_R. By Theorem 21.6 in Davidson (1994) we conclude that $H_N(\sigma_N^*, k^*) \xrightarrow{p} H$. □

Lemma 3 *Let $a_n = 1 - \frac{N}{n}$ and for $j = 1, \cdots, N$ define $Z_j = U_j - q_n(a_n)$ whenever $U_j > q_n(a_n)$ and for $j = 1, \cdots, N_1$ define $Z_j' = U_j - q(a_n)$ whenever $U_j > q(a_n)$. If*

$$\Delta_\sigma = \frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial\sigma}log\, g(Z_j; \sigma_N, k_0)\sigma_N - \frac{1}{N}\sum_{j=1}^{N_1}\frac{\partial}{\partial\sigma}log\, g(Z_j'; \sigma_N, k_0)\sigma_N$$

and

$$\Delta_k = \frac{1}{N}\sum_{j=1}^{N}\frac{\partial}{\partial k}log\, g(Z_j; \sigma_N, k_0) - \frac{1}{N}\sum_{j=1}^{N_1}\frac{\partial}{\partial k}log\, g(Z_j'; \sigma_N, k_0),$$

then $N^{1/2}\Delta_\sigma = b_1\sqrt{N}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$ and $N^{1/2}\Delta_k = b_2\sqrt{N}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$, where $b_1 = -\frac{\alpha(1+\alpha)}{2+\alpha}$, $b_2 = \left(-\frac{\alpha^2(1+\alpha)}{2+\alpha} + \frac{\alpha^3}{1+\alpha}\right)$.

Proof. We first consider the case where $N = N_1$. Then,

$$\begin{aligned}\Delta_\sigma &= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial\sigma}log\, g(Z_j; \sigma_N, k_0)\sigma_N - \frac{\partial}{\partial\sigma}log\, g(Z_j'; \sigma_N, k_0)\sigma_N\right) \\ &= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{k_0} - 1\right)\left(\left(1 - \frac{k_0 Z_j}{\sigma_N}\right)^{-1}\frac{k_0 Z_j}{\sigma_N} - \left(1 - \frac{k_0 Z_j'}{\sigma_N}\right)^{-1}\frac{k_0 Z_j'}{\sigma_N}\right)\end{aligned}$$

By the mean value theorem, there exists $\lambda \in [0, 1]$ and $Z_j^* = Z_j' + \lambda(q(a_n) - q_n(a_n))$ such that

$$\Delta_\sigma = \frac{q_n(a_n) - q(a_n)}{q(a_n)}\left(\frac{1}{k_0} - 1\right)\frac{1}{N}\sum_{j=1}^{N}\left(1 - \frac{k_0}{\sigma_N}Z_j^*\right)^{-2}. \quad (21)$$

Again, using the mean value theorem, we have that for some $\theta \in [0,1]$ there is $Z_j^{**} = \theta Z_j' + (1-\theta)Z_j^* = Z_j' + \lambda(1-\theta)(q(a_n) - q_n(a_n))$ such that

$$\frac{1}{N}\sum_{j=1}^{N}\left(\left(1-\frac{k_0}{\sigma_N}Z_j^*\right)^{-2} - \left(1-\frac{k_0}{\sigma_N}Z_j'\right)^{-2}\right) = \frac{1}{N}\sum_{j=1}^{N}\frac{2k_0/\sigma_N}{\left(1-\frac{k_0}{\sigma_N}Z_j^{**}\right)^3}(Z_j^* - Z_j')$$
$$= -\theta\frac{q(a_n)-q_n(a_n)}{q(a_n)}\frac{2q(a_n)}{q_n(a_n)}\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j^{**}\right)^{-3}$$
$$= O_p(N^{-1/2})(1+o_p(1))\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j^{**}\right)^{-3}$$

where the last equality follows from the fact that $\frac{q(a_n)}{q_n(a_n)} = 1 + o_p(1)$ and Lemma 2. In addition,

$$\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j^{**}\right)^{-3} = \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j' + (\theta-\theta^2)(q(a_n)-q_n(a_n))\right)^{-3}$$
$$= \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j' + o_p(1)\right)^{-3} = O_p(1)$$

using the same arguments as in the proof of Lemma 1. Hence,

$$\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j^*\right)^{-2} = O_p(N^{-1/2}) + \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j'\right)^{-2}$$
$$= \frac{\alpha}{2+\alpha} - \frac{2C\phi(U_{(n-N)})}{(2+\alpha)(2+\alpha-\rho)} + o(\phi(U_{(n-N)})) + O_p(N^{-1/2})$$

where the last equality follows from Smith (1987). Consequently, since $\phi(U_{(n-N)}) = O(N^{-1/2})$ and substituting back in equation (21) we have that $N^{1/2}\Delta_\sigma = b_1 N^{1/2}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$.

We now turn to the case where $N_1 > N$. In this case we can write

$$N^{1/2}\Delta_\sigma = N^{1/2}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial\sigma}log\, g(Z_j;\sigma_N,k_0)\sigma_N - \frac{\partial}{\partial\sigma}log\, g(Z_j';\sigma_N,k_0)\sigma_N\right)$$
$$+ N^{1/2}\frac{1}{N}\sum_{j=1}^{N_1-N}\frac{\partial}{\partial\sigma}log\, g(Z_j';\sigma_N,k_0)\sigma_N.$$

The first term is $b_1 N^{1/2}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$ as in the case where $N = N_1$. As in Smith (1987) we have that the expectation of the second term is $\frac{N_1-N}{\sqrt{N}}\left(\frac{C\phi(U_{(n-N)})}{1+\alpha-\rho} + o(\phi(U_{(n-N)}))\right)$ which is $o_p(1)$ since $\phi(U_{(n-N)}) = O(N^{-1/2})$ and $\frac{N_1-N}{\sqrt{N}} = O_p(1)$. In addition its variance is $\frac{N_1-N}{N}O(1) = o_p(1)$. Hence, the last term is $o_p(1)$, and we can write for the case where $N_1 > N$ that $N^{1/2}\Delta_\sigma = b_1 N^{1/2}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$. Similar arguments give us the same order for $N^{1/2}\Delta_\sigma$ when $N > N_1$. The case for $N^{1/2}\Delta_k$ follows, *mutatis mutandis* , using exactly the same arguments. □

Lemma 4 $E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0 Z_i'}{\sigma_N}\right)\right) = -\frac{1}{\alpha} + \frac{\alpha}{(1+\alpha)^2} + O(\phi(U_{(n-N)}))$

Proof. We first observe that from the results in Smith (1987)

$$E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0 Z_i'}{\sigma_N}\right)\right) = -\alpha^{-1} + O(\phi(U_{(n-N)}))$$
$$+ E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\right).$$

Using the notation for $L(\cdot)$ in FR2 and given that

$$F_{U_{(n-N)}}(z) = 1 - \frac{L\left(\left(1+\frac{z}{U_{(n-N)}}\right)U_{(n-N)}\right)}{L(U_{(n-N)})}\left(1+\frac{z}{U_{(n-N)}}\right)^{-\alpha}$$

we can write $E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\right) = \int_0^\infty log\left(1-\frac{k_0 z}{\sigma_N}\right)\left(1-\frac{k_0 z}{\sigma_N}\right)^{-1} dF_{U_{(n-N)}}(z)$. Integrating by parts we have

$$\begin{aligned} E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\right) &= \int_0^\infty \frac{L\left(\left(1+z/U_{(n-N)}\right)U_{(n-N)}\right)}{L(U_{(n-N)})}\left(1+z/U_{(n-N)}\right)^{-\alpha} \\ &\times\left(\frac{1}{U_{(n-N)}}\left(1+z/U_{(n-N)}\right)^{-2}\right. \\ &\left.- \frac{1}{U_{(n-N)}} log\left(1+z/U_{(n-N)}\right)\left(1+z/U_{(n-N)}\right)^{-2}\right) dz. \end{aligned}$$

Setting $t = 1 + z/U_{(n-N)}$ we have that

$$E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\right) = \int_1^\infty \frac{L(tU_{(n-N)})}{L(U_{(n-N)})}(t^{-\alpha-2} - log(t)t^{-\alpha-2})dt$$

and by FR2

$$\begin{aligned} E\left(log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\right) &= \int_1^\infty (t^{-\alpha-2} - log(t)t^{-\alpha-2})dt \\ &+ C\phi(U_{(n-N)})\int_1^\infty (t^{-\alpha-2} - log(t)t^{-\alpha-2})\int_1^t u^{\rho-1} du dt \\ &+ o(\phi(U_{(n-N)})) = \frac{1}{\alpha+1} - \frac{1}{(1+\alpha)^2} + O(\phi(U_{(n-N)})) \end{aligned}$$

which combines with the order of the first equation in the proof to give the desired result. □

Theorem 2 .

Proof. Let $a \in (0,1)$ and $a_n = 1 - \frac{N}{n} < a$. We are interested in estimating $q(a)$ which we write as $q(a) = q(a_n) + y_{N,a}$. Estimating $q(a_n)$ by $\tilde{q}(a_n)$ and based on the GPD approximation we define an estimator $\hat{y}_{N,a}$ for $y_{N,a}$ as $\hat{y}_{N,a} = \frac{\tilde{\sigma}_N}{\tilde{k}}\left(1-\left(\frac{n(1-a)}{N}\right)^{\tilde{k}}\right)$. Note that, as defined, $\hat{y}_{N,a}$ satisfies

$$(22) \qquad 1 - \tilde{F}(\tilde{q}(a_n) + \hat{y}_{N,a}) = \frac{N}{n}\left(1 - \frac{\tilde{k}\hat{y}_{N,a}}{\tilde{\sigma}_N}\right)^{1/\tilde{k}}.$$

Let us pause and note that for a chosen N, equation (22) is satisfied with a distribution function $\dot{F}$ that is not necessarily $\tilde{F}$. However, given the continuity of $\tilde{F}$, there exists N satisfying the order relation $a > 1 - N/n$ for which (22) is satisfied by $\tilde{F}$. Hence, to avoid additional notation we proceed with $\tilde{F}$. We define the estimator for $q(a)$ as $\hat{q}(a) = \tilde{q}(a_n) + \hat{y}_{N,a}$. For $\sigma_n = q(a)(n(1-a))^{-1/2}$, arbitrary $0 < z$ and $V_n = -k_0\sqrt{n}/(1-a)^{1/2}$ we note that

$$\begin{aligned} P(\sigma_n(\hat{q}(a) - q(a)) \le z) &= P(1 - a \ge 1 - \tilde{F}(q(a_n) + y_{N,a} + \sigma_n z) \\ &= P(V_n((1-a) - (1 - F(q(a) + \sigma_n z)) \ge V_n((1 - \tilde{F}(q(a_n) + y_{N,a} + \sigma_n z)) \\ &- (1 - F(q(a) + \sigma_n z))). \end{aligned}$$

In addition, from the proof of Lemma 2 we have that $lim_{n\to\infty}V_n((1-a)-(1-F(q(a)+\sigma_n z)=z$. Now, let $W_n = V_n((1-\tilde{F}(q(a_n)+y_{N,a}+\sigma_n z))-(1-F(q(a)+\sigma_n z))$ and note that $\frac{n(1-F(q(a)))}{V_n(1-F(q(a)+\sigma_n z)}W_n = \sqrt{n(1-F(q(a)))}\left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)}-1\right) = -\frac{1}{k_0}W_n(1+o(1))$. We first establish that

$$\sqrt{n(1-F(q(a)))}\left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)}-1\right)$$

is asymptotically normally distributed. Without loss of generality consider $y_N = q(a_n)(Z_N-1)$ for $0 < Z_N \to z_a < \infty$. Note that if $Z_N = z_a$, then $y_{N,a} = y_N = q(a_n)(z_a-1)$. Then, $q(a)+\sigma_n z = q(a_n)z_a(1+z((1-a)n)^{-1/2}) = q(a_n)Z_N$. By FR2

$$\begin{aligned}\frac{(q(a_n)Z_N)^\alpha}{q(a_n)^\alpha}\frac{1-F(q(a_n)Z_N)}{1-F(q(a_n)} &= Z_N^{-1/k_0}\frac{1-F(q(a_n)Z_N)}{1-F(q(a_n)} \text{ since } \alpha = -1/k_0 \\ &= 1+k(Z_N)\phi(q(a_n))+o(\phi(q(a_n)))\end{aligned}$$

where $0 < \phi(q(a_n)) \to 0$ as $q(a_n) \to \infty$, $k(Z_N) = \frac{C(Z_N^\rho-1)}{\rho}$. Since we assume that $\frac{N^{1/2}C\phi(q(a_n))}{\alpha-\rho} \to \mu$, we have that as $Z_N \to z_a$, $k(Z_N)\phi(q(a_n)) - k(z_a)N^{-1/2}\frac{\mu(\alpha-\rho)}{C} \to 0$ and consequently

$$Z_N^{-1/k_0}\frac{1-F(q(a_n)Z_N)}{1-F(q(a_n)} = 1+k(z)N^{-1/2}\frac{\mu(\alpha-\rho)}{C}+o(N^{-1/2}). \tag{23}$$

We observe that for the function $h(\sigma,k,y) = -\frac{1}{k}log\left(1-\frac{ky}{\sigma}\right)$ we can write

$$\frac{1-\tilde{F}(\tilde{q}(a_n)+y_N)}{1-\tilde{F}(\tilde{q}(a_n))} = exp(-h(\tilde{\sigma}_N,\tilde{k},y_N))$$

and using the notation in Theorem 1 and the mean value theorem gives

$$h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N) = \begin{pmatrix} \sigma_N\frac{\partial}{\partial\sigma}h(\sigma_N^*,k^*,y_N) & \frac{\partial}{\partial k}h(\sigma_N^*,k^*,y_N)\end{pmatrix}\begin{pmatrix}\tilde{r}_N-1 \\ \tilde{k}-k_0\end{pmatrix}$$

for $\sigma_N^* = \lambda_1\tilde{\sigma}_N+(1-\lambda_1)\sigma_N$ and $k_N^* = \lambda_2\tilde{k}_N+(1-\lambda_2)k_0$ and $\lambda_1,\lambda_2 \in [0,1]$. It follows from $\sigma_N = -k_0q(a_n) = -\frac{k_0y_N}{Z_N-1}$ that $y_N = \frac{(1-Z_N)\sigma_N}{k_0}$ and from Theorem 1 we have

$$\sigma_N\frac{\partial}{\partial\sigma}h(\sigma_N^*,k^*,y_N) \xrightarrow{p} -k_0^{-1}(z_a^{-1}-1) \text{ and } \frac{\partial}{\partial k}h(\sigma_N^*,k^*,y_N) \xrightarrow{p} k_0^{-2}log(z_a)+k_0^{-2}(z_a^{-1}-1).$$

Hence, if $c_b' = \begin{pmatrix} -k_0^{-1}(z_a^{-1}-1) & k_0^{-2}log(z_a)+k_0^{-2}(z_a^{-1}-1)\end{pmatrix}$ and

$$\mu_p' = \begin{pmatrix} \frac{\mu(1-k_0)(1+2k\rho)}{1-k_0+k_0\rho} & \frac{\mu(1-k_0)k_0(1+\rho)}{1-k_0+k_0\rho}\end{pmatrix}$$

we can write

(24)

$$c_b'\sqrt{N}\begin{pmatrix}\tilde{r}_N-1 \\ \tilde{k}-k_0\end{pmatrix} \xrightarrow{d} N(c_b'\mu_p, c_b'H^{-1}V_2H^{-1}) \text{ and } \sqrt{N}(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N)) = O_p(1).$$

Now, we can conveniently write,

$$\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)-y_N)} = \frac{1-\tilde{F}(q(a_n)+y_N)}{1-\tilde{F}(\tilde{q}(a_n))}\frac{1-F(q(a_n))}{1-F(q(a_n)+y_N)}Z_N^{1/k_0}Z_N^{-1/k_0}.$$

Note that $\frac{1-\tilde{F}(q(a_n)+y_N)}{1-\tilde{F}(\tilde{q}(a_n))} = \left(1-\frac{\tilde{k}y_N}{\tilde{\sigma}_N}\right)^{1/\tilde{k}}\left(\frac{1-\tilde{F}(q(a_n))}{1-\tilde{F}(\tilde{q}(a_n))}\right)$ and

$$Z_N^{-1/k_0} = \left(1-\frac{k_0 y_N}{\sigma_N}\right)^{-1/k_0} = exp(h(\sigma_N, k_0, y_N)).$$

Furthermore from equation (23), $Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))} - 1 = N^{-1/2}\left(-k(z)\frac{\mu(\alpha-\rho)}{C}\right) + o(N^{-1/2})$. Hence,

$$\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} = Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}\frac{1-\tilde{F}(q(a_n))}{(1-\tilde{F}(\tilde{q}(a_n)))}exp(-h(\tilde{\sigma}_N,\tilde{k},y_N)+h(\sigma_N,k_0,y_N)).$$

Now, we given that

$$\frac{1-\tilde{F}(q(a_n))}{1-\tilde{F}(\tilde{q}(a_n))} - 1 = -\frac{\tilde{F}(q(a_n))-F(q(a_n))}{1-F(q(a_n))}$$

and from equation (14) in Lemma 2 we have

$$\frac{\sqrt{n(1-F(q(a_n))}}{1-F(q(a_n))}(1-\tilde{F}(q(a_n))-(1-F(q(a_n))) \xrightarrow{d} N(0,1)$$

as $q(a_n)\to\infty$. In particular, using the notation adopted in Lemma 2 we have that

$$\begin{aligned}\frac{\sqrt{n(1-F(q(a_n))}}{1-F(q(a_n))}(1-\tilde{F}(q(a_n))-(1-F(q(a_n))) &= -\sum_{i=1}^{n}\sqrt{n(1-F(q(a_n))}(q_{1n}-E(q_{1n}))+o_p(1)\\ &= \sum_{i=1}^{n} Z_{i4}+o_p(1).\end{aligned}$$

Hence,

$$\begin{aligned}\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1 = Z_N^{1/k_0}&\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}\frac{1-\tilde{F}(q(a_n))}{(1-\tilde{F}(\tilde{q}(a_n)))}exp(-h(\tilde{\sigma}_N,\tilde{k},y_N)\\ &+h(\sigma_N,k_0,y_N))-1\end{aligned}$$

and by equation (24) and the mean value theorem we have

$$exp(-h(\tilde{\sigma}_N,\tilde{k},y_N)+h(\sigma_N,k_0,y_N)) = 1-(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N))+o_p(N^{-1/2}).$$

Therefore, we write

$$\begin{aligned}\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)}-1\right) &= \sqrt{N}\left(Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}-1\right)\\ &+\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n))}{(1-\tilde{F}(\tilde{q}(a_n)))}-1\right)-\sqrt{N}(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N))\\ &+o_p(1).\end{aligned}$$

Since $\sqrt{N}\left(Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}-1\right)\to-\frac{k(z)\mu(\alpha-\rho)}{C}$ we focus on the joint distribution of the last two terms. By equation (24) we have that

$$\sqrt{N}(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N)) = c_b'\sqrt{N}\begin{pmatrix}\tilde{r}_N-1\\ \tilde{k}-k_0\end{pmatrix}+o_p(1) \tag{25}$$

and by Theorem 1 (adopting its notation) we have

$$\sqrt{N}\begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} = (H^{-1} + o_p(1))\left(v_N(1, k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right),$$

where the last vector in this equality depends on $\sqrt{N}\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)}$ which is asymptotically distributed as $\sum_{i=1}^n Z_{i3} + o_p(1)$, $\sum_{i=1}^n Z_{i2}$ and $\sum_{i=1}^n Z_{i1}$. Hence, we define $\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n))}{(1-\tilde{F}(\tilde{q}(a_n)))} - 1\right) = \sum_{i=1}^n Z_{i4}$, let $0 \neq d \in \Re^4$,

$$\varepsilon_n' = \begin{pmatrix} \sum_{i=1}^n Z_{i1} & \sum_{i=1}^n Z_{i2} & \sum_{i=1}^n Z_{i3} & \sum_{i=1}^n Z_{i4} \end{pmatrix}$$

and consider $d'\varepsilon_n = \sum_{i=1}^n \sum_{\delta=1}^4 Z_{i\delta} d_\delta = \sum_{i=1}^n Z_{ni}$. Note that Z_{ni} forms an iid sequence with $E(Z_{ni}) = 0$ and the asymptotic behavior of $\sum_{i=1}^n Z_{i1}$, $\sum_{i=1}^n Z_{i2}$ and $\sum_{i=1}^n Z_{i3}$ was studied in Theorem 1. In addition the asymptotic behavior of $\sum_{i=1}^n Z_{i4}$ was studied in Lemma 2. Recall that $E(Z_{i4}^2) = n^{-1}(F(y_n) + o(h_{2n}))$ and from Theorem 1 $E(Z_{i1}Z_{i4}) = o(n^{-1})$ and $E(Z_{i2}Z_{i4}) = o(n^{-1})$. Here we examine

$$\begin{aligned} E(Z_{i3}Z_{i4}) &= -\frac{k_0}{n((1-F(y_n))(1-F(q(a_n))))^{1/2}} E\left(q_{1n}\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy\right) \\ &- E(q_{1n}) E\left(\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy\right). \end{aligned}$$

By Lemma 2 $E(q_{1n}) - F(y_n) = O(h_{2n}^{m+1})$ and similarly we have $E\left(\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy\right) - F(q(a_n)) = O(h_{2n}^{m+1})$. Since in Lemma 2 we have $y_n = q(a_n) + \sigma_n z$, then for $\kappa(x) = h_{2n}^{-1}\int_{-\infty}^x K_2 \left(\frac{y-U_i}{h_{2n}}\right) dy$ we can write

$$E\left(q_{1n}\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy\right) = E(\kappa(q(a_n) + \sigma_n z)\kappa(q(a_n)))(\chi_{\{q(a_n)=y_n\}} + \chi_{\{q(a_n)\neq y_n\}}).$$

For $z > 0$ we have that $q(a_n) \neq y_n$ implies $y_n > q(a_n)$ so that

$$E(\kappa(q(a_n) + \sigma_n z)\kappa(q(a_n))\chi_{\{q(a_n)<y_n\}}) \leq C\chi_{\{q(a_n)<y_n\}} = C\left(F(q(a_n) + \sigma_n z) - F(q(a_n))\right).$$

By FR2 $lim_{n\to\infty}\frac{F(q(a_n)+\sigma_n z)-F(q(a_n))}{1-F(q(a_n))} = 0$, hence

$$(1 - F(q(a_n)))^{-1} E(\kappa(q(a_n) + \sigma_n z)\kappa(q(a_n))\chi_{\{q(a_n)=y_n\}}) = o(1)$$

and

$$E\left(q_{1n}\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)} K_2\left(\frac{y-U_i}{h_{2n}}\right) dy\right) = E\left(\kappa^2(q(a_n))\right) + o(1 - F(q(a_n))).$$

Consequently,

$$\begin{aligned} E(Z_{i3}Z_{i4}) &= -\frac{k_0}{n((1-F(y_n))(1-F(q(a_n))))^{1/2}}\left(E\left(\kappa^2(q(a_n))\right) + o(F(q(a_n)))\right) \\ &- F^2(q(a_n)) + O(h_{2n}^{m+1}) = -\frac{k_0}{n}(F(q(a_n)) + o(1)) \end{aligned}$$

and $V(Z_{in}) = \frac{1}{n}d'V_3 d + o(n^{-1})$ where $V_3 = \begin{pmatrix} \frac{1}{1-2k_0} & -\frac{1}{(k_0-1)(2k_0-1)} & 0 & 0 \\ -\frac{1}{(k_0-1)(2k_0-1)} & \frac{2}{(k_0-1)(2k_0-1)} & 0 & 0 \\ 0 & 0 & k_0^2 & -k_0 \\ 0 & 0 & -k_0 & 1 \end{pmatrix}$.

From the verification of Liapounov's condition in Theorem 1 we have that $d'\varepsilon_n \xrightarrow{d} N(0, d'V_3 d)$

and from the Cramer-Wold theorem $\varepsilon_n \xrightarrow{d} N(0, V_3)$. Now, from equation (25)

$$\sqrt{N}(h(\tilde{\sigma}_N, \tilde{k}, y_N) - h(\sigma_N, k_0, y_N)) = c_b' H^{-1}\left(v_N(1, k_0) - \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) + c_b' H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}$$

hence by letting $A_{.j}$ represent the j^{th} column of a matrix A, we write

$$\begin{aligned}
\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1\right) &= -\frac{k(z_a)\mu(\alpha-\rho)}{C} - \left(c_b' H_{.1}^{-1}\sum_{i=1}^n Z_{i1} + c_b' H_{.2}^{-1}\sum_{i=1}^n Z_{i2}\right. \\
&+ \left(c_b' H_{.1}^{-1} b_1 + c_b' H_{.2}^{-1} b_2\right)\sum_{i=1}^n Z_{i3} \\
&+ \left. c_b' H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) + \sum_{i=1}^n Z_{i4} + o_p(1) \\
&= -\frac{k(z_a)\mu(\alpha-\rho)}{C} - c_b' H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \\
&+ \begin{pmatrix} -c_b' H_{.1}^{-1} & -c_b' H_{.2}^{-1} & -c_b' H_{.1}^{-1} b_1 - c_b' H_{.2}^{-1} b_2 & 1 \end{pmatrix}\varepsilon_n + o_p(1).
\end{aligned}$$

Let $\eta' = \begin{pmatrix} -c_b' H_{.1}^{-1} & -c_b' H_{.2}^{-1} & -c_b' H_{.1}^{-1} b_1 - c_b' H_{.2}^{-1} b_2 & 1 \end{pmatrix}$, then from the results above we have $\eta'\varepsilon_n \xrightarrow{d} N(0, \eta' V_3 \eta)$ where simple algebraic manipulations give $\eta' V_3 \eta = c_b' H^{-1} V_2 H^{-1} c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1$. Consequently, if $\zeta \sim N\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C}, c_b' H^{-1} V_2 H^{-1} c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)$, then

$$\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1 - \left(-c_b' H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right)\right) \xrightarrow{d} \zeta,$$

and for $y_N = q(a_n)(Z_N - 1)$ with $Z_N \to z_a$ we immediately have

$$\sqrt{N}\left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)} - 1 - \left(-c_b' H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right)\right) \xrightarrow{d} \zeta.$$

Lastly, since $-W_n/k_0 + o(1) = \sqrt{n(1-F(q(a)))}\left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)} - 1\right)$ and if

$$\sqrt{n(1-F(q(a)))} = \sqrt{n(1-a)} \propto N^{1/2},$$

that is, $n(1-a) \to \infty$ at the same rate as N, then

$$\begin{aligned}
W_n \xrightarrow{d} N\Bigg((-k_0)\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C} - c_b' H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right), \\
k_0^2\left(c_b' H^{-1} V_2 H^{-1} c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\Bigg)
\end{aligned}$$

which immediately gives, $\sqrt{n(1-a)}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \xrightarrow{d} \zeta_1$ where

$$\begin{aligned}
\zeta_1 \sim N\Bigg((-k_0)\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C} - c_b' H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right), \\
k_0^2\left(c_b' H^{-1} V_2 H^{-1} c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\Bigg).
\end{aligned}$$

□

Appendix 2 - Tables and figures.

TABLE 1 MEAN(M), BIAS(B) AND STANDARD DEVIATION(S)FOR PARAMETER ESTIMATORS WITH LOG-GAMMA DISTRIBUTED U WITH $\alpha = 1,\ \beta = 0.25\ (k_0 = -4)$

$N = 50$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	.470	.113	-.200	3.800	.199	.469	.111	-.200	3.800	.199
$\tilde{\gamma}$	.430	.107	-.221	3.779	.203	.440	.108	-.220	3.780	.204
$\hat{k}^h$			-.562	3.438	.077			-.634	3.366	.082

$N = 100$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	.457	.074	-.223	3.777	.132	.456	.074	-.223	3.777	.133
$\tilde{\gamma}$	.431	.073	-.238	3.762	.135	.441	.073	-.233	3.767	.134
$\hat{k}^h$			-.602	3.398	.057			-.647	3.353	.058

TABLE 2 MEAN(M), BIAS(B) AND STANDARD DEVIATION(S)FOR PARAMETER ESTIMATORS WITH LOG-GAMMA DISTRIBUTED U WITH $\alpha = 1,\ \beta = 0.5\ (k_0 = -2)$

$N = 50$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	1.673	.445	-.454	1.546	.227	1.670	.443	-.452	1.548	.227
$\tilde{\gamma}$	1.572	.431	-.476	1.524	.228	1.589	.433	-.468	1.532	.232
$\hat{k}^h$			-.888	1.112	.133			-.896	1.104	.133

$N = 100$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	1.626	.301	-.478	1.522	.154	1.615	.300	-.483	1.517	.155
$\tilde{\gamma}$	1.569	.300	-.490	1.510	.158	1.563	.293	-.494	1.506	.157
$\hat{k}^h$			-.899	1.101	.096			-.900	1.100	.095

Table 3 Mean(M), Bias(B) and Standard deviation(S)for parameter estimators with student-t distributed U with $v = 3$ ($k_0 = -1/3$)

$N = 50$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	.992	.228	-.204	.129	.204	.990	.225	-.206	.128	.202
$\tilde{\gamma}$	.959	.228	-.210	.124	.206	.960	.225	-.211	.122	.205
$\hat{k}^h$			-.455	-.122	.063			-.461	-.128	.063

$N = 100$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	.961	.155	-.235	.099	.136	.968	.154	-.229	.104	.136
$\tilde{\gamma}$	.940	.152	-.238	.095	.137	.949	.156	-.233	.101	.137
$\hat{k}^h$			-.460	-.127	.045			-.463	-.130	.045

Table 4 Mean(M), Bias(B) and Standard deviation(S)for parameter estimators with student-t distributed U with $v = 2$ ($k_0 = -1/2$)

$N = 50$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	1.337	.332	-.405	.095	.222	1.335	.336	-.406	.094	.226
$\tilde{\gamma}$	1.294	1.110	-.416	.084	.227	1.293	.417	-.414	.086	.231
$\hat{k}^h$			-.565	-.065	.088			-.571	-.071	.089

$N = 100$	$m(x) = 3sin(3x)$					$m(x) = x^2$				
	σ_N		k_0			σ_N		k_0		
estimators	M	S	M	B	S	M	S	M	B	S
$\hat{\gamma}$	1.302	.226	-.430	.070	.152	1.301	.228	-.429	.071	.155
$\tilde{\gamma}$	1.272	.235	-.435	.065	.153	1.276	.271	-.434	.066	.158
$\hat{k}^h$			-.575	-.075	.064			-.577	-.077	.065

Table 5 Bias (×0.1)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = 3sin(3x)$, and log-gamma distributed U with $\alpha = 1$, $\beta = 0.25$ ($k_0 = -4$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.076	.099	.099	-.117	.286	.287	-.248	.494	.495
$\hat{q}$	1.047	.080	.132	.280	.260	.262	-.040	.465	.465
q^h	.402	.061	.073	2.190	.320	.388	6.223	.634	.888
q^e	.841	.082	.118	-.182	.309	.310	-.944	.498	.507

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.028	.067	.067	-.098	.198	.198	-.207	.337	.338
$\hat{q}$	.589	.056	.081	.091	.182	.183	-.136	.319	.320
q^h	-.048	.043	.043	2.285	.235	.328	7.111	.476	.856
q^e	.488	.059	.077	-.155	.223	.223	-.695	.380	.386

Table 6 Bias(×0.1)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = x^2$, and log-gamma distributed U with $\alpha = 1$, $\beta = 0.25$ ($k_0 = -4$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.059	.097	.097	-.162	.286	.287	-.304	.496	.497
$\hat{q}$	.265	.081	.085	-.305	.264	.266	-.512	.476	.479
q^h	-.445	.060	.075	2.358	.349	.421	8.088	.727	1.088
q^e	.103	.084	.084	-.691	.316	.323	-1.386	.510	.528

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.029	.069	.069	-.122	.202	.202	-.246	.342	.342
$\hat{q}$	.195	.058	.061	-.162	.189	.190	-.339	.329	.330
q^h	-.504	.044	.067	2.525	.249	.355	8.481	.515	.992
q^e	.112	.062	.063	-.383	.230	.233	-.916	.380	.391

Table 7 Bias($\times 0.1$)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = 3sin(3x)$, and log-gamma distributed U with $\alpha = 1$, $\beta = 0.5$ ($k_0 = -2$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.315	.420	.421	.183	1.919	1.919	1.188	4.059	4.060
$\hat{q}$	.362	.352	.354	-1.345	1.773	1.778	-.615	3.869	3.869
q^h	-2.431	.249	.348	15.508	2.564	2.996	60.118	6.869	9.128
q^e	-.206	.360	.360	-3.625	2.096	2.127	-7.617	4.149	4.218

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.160	.292	.293	.126	1.309	1.309	.554	2.672	2.673
$\hat{q}$	.220	.250	.251	-.799	1.220	1.222	-.554	2.586	2.586
q^h	-2.493	.180	.307	15.757	1.753	2.357	59.571	4.545	7.493
q^e	-.054	.266	.266	-2.174	1.475	1.490	-4.803	2.972	3.010

Table 8 Bias($\times 0.1$)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = x^2$, and log-gamma distributed U with $\alpha = 1$, $\beta = 0.5$ ($k_0 = -2$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.315	.416	.417	-.008	1.911	1.910	.799	4.028	4.028
$\hat{q}$	.313	.351	.352	-1.447	1.771	1.777	-.983	3.898	3.899
q^h	-2.564	.250	.358	16.226	2.542	3.015	62.467	6.735	9.185
q^e	-.248	.363	.364	-3.928	2.097	2.133	-7.384	4.169	4.233

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.115	.291	.291	.183	1.323	1.323	.959	2.714	2.716
$\hat{q}$	.160	.247	.248	-.681	1.230	1.232	-.111	2.621	2.621
q^h	-2.517	.179	.309	15.776	1.743	2.351	59.676	4.514	7.482
q^e	-.108	.260	.261	-1.836	1.495	1.506	-4.241	3.075	3.104

TABLE 9 BIAS($\times 0.1$)(B), STANDARD DEVIATION(S) AND ROOT MEAN SQUARED ERROR(R) FOR QUANTILE ESTIMATORS WITH $m(x) = 3sin(3x)$, AND STUDENT-T DISTRIBUTED U WITH $v = 3$ ($k_0 = -1/3$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.099	.202	.203	.155	.641	.641	-.227	1.136	1.136
$\hat{q}$	.504	.182	.188	-.002	.600	.600	-.628	1.086	1.088
q^h	-.658	.147	.161	2.528	.646	.694	7.728	1.175	1.406
q^e	.038	.188	.188	-1.616	.676	.695	-3.075	1.151	1.191

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	-.001	.141	.141	.299	.443	.444	.075	.771	.771
$\hat{q}$	.247	.123	.125	.185	.416	.417	-.196	.743	.743
q^h	-.753	.099	.124	2.529	.451	.517	7.738	.816	1.124
q^e	.013	.129	.129	-.918	.491	.500	-1.938	.848	.870

TABLE 10 BIAS($\times 0.1$)(B), STANDARD DEVIATION(S) AND ROOT MEAN SQUARED ERROR(R) FOR QUANTILE ESTIMATORS WITH $m(x) = x^2$, AND STUDENT-T DISTRIBUTED U WITH $v = 3$ ($k_0 = -1/3$)

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.116	.199	.199	.203	.638	.638	-.143	1.133	1.133
$\hat{q}$	.260	.175	.176	-.158	.602	.602	-.708	1.097	1.100
q^h	-.906	.140	.167	2.461	.647	.693	7.907	1.183	1.423
q^e	-.167	.180	.181	-1.622	.690	.708	-3.081	1.175	1.215

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.063	.144	.144	.325	.441	.442	-.000	.765	.765
$\hat{q}$	.222	.127	.129	.184	.420	.420	-.260	.741	.741
q^h	-.818	.103	.131	2.637	.459	.530	8.033	.829	1.154
q^e	-.017	.135	.135	-.867	.496	.503	-1.948	.852	.874

Table 11 Bias(×0.1)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = 3sin(3x)$, and student-t distributed U with $v = 2$ $(k_0 = -1/2)$

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.192	.318	.319	.579	1.362	1.363	.711	2.794	2.795
$\hat{q}$	.986	.958	.963	.326	1.807	1.807	.050	2.990	2.990
q^h	-.188	.716	.716	3.335	2.717	2.737	10.080	4.915	5.017
q^e	.386	.943	.944	-2.337	1.922	1.936	-4.831	3.139	3.176

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.071	.219	.219	.636	.957	.959	.545	1.919	1.920
$\hat{q}$	.549	.248	.254	.493	.929	.930	.125	1.872	1.872
q^h	-.458	.216	.221	3.289	.980	1.034	10.047	1.929	2.174
q^e	.230	.262	.263	-1.255	1.092	1.099	-3.146	2.198	2.220

Table 12 Bias(×0.1)(B), Standard deviation(S) and Root Mean Squared Error(R) for quantile estimators with $m(x) = x^2$, and student-t distributed U with $v = 2$ $(k_0 = -1/2)$

$N = 50$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.163	.315	.315	.535	1.380	1.381	.768	2.874	2.875
$\hat{q}$	.793	.673	.678	.205	1.456	1.456	.016	2.869	2.869
q^h	-.399	.614	.615	3.327	1.548	1.583	10.439	2.907	3.089
q^e	.201	.669	.669	-2.653	1.653	1.674	-5.002	3.203	3.241

$N = 100$	$\alpha = 0.95$			$\alpha = 0.99$			$\alpha = 0.995$		
estimators	B	S	R	B	S	R	B	S	R
q^s	.081	.220	.220	.565	.957	.958	.430	1.923	1.924
$\hat{q}$	.492	.255	.260	.450	.946	.947	.109	1.920	1.920
q^h	-.533	.208	.215	3.342	1.021	1.074	10.297	2.021	2.268
q^e	.173	.261	.262	-1.108	1.165	1.170	-3.326	2.208	2.233

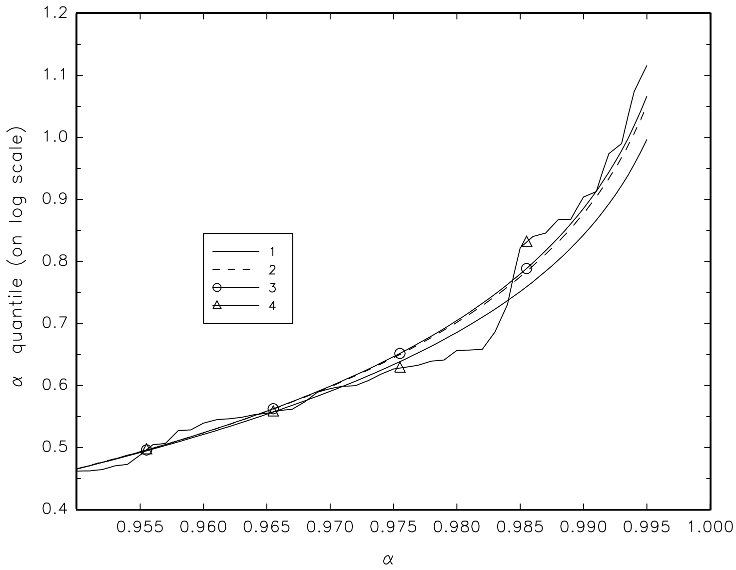

Fig 1. *Plot of quantile estimates across different* α, *with* $n = 1000$, $N = 100$, $m(x) = 3sin(3x)$ *and student-t distributed* U *with* $v = 2$. $1:$ *true quantile*, $2 : \hat{q}$, $3 : q^h$, *and* $4 : q^e$.

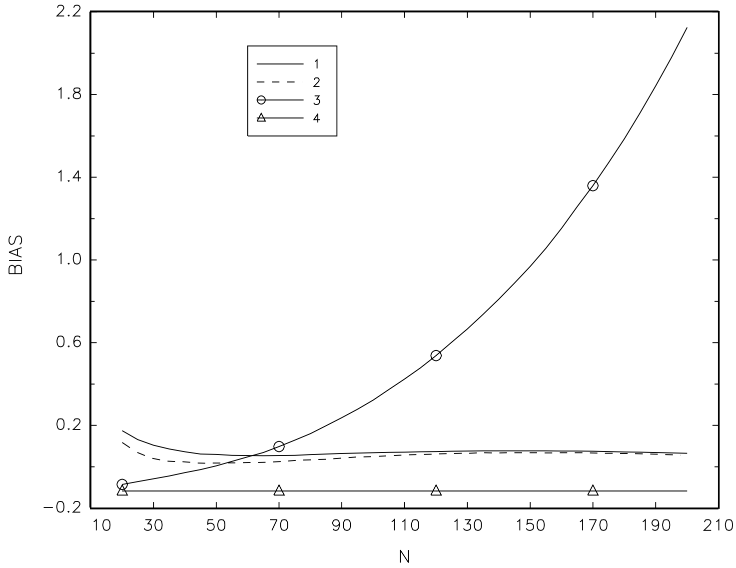

Fig 2. *Bias of* 99% *quantile estimators with different* N, *with* $n = 1000$, $m(x) = 3sin(3x)$ *and student-t distributed* U *with* $v = 2$. $1 : q^s$, $2 : \hat{q}$, $3 : q^h$, *and* $4 : q^e$.

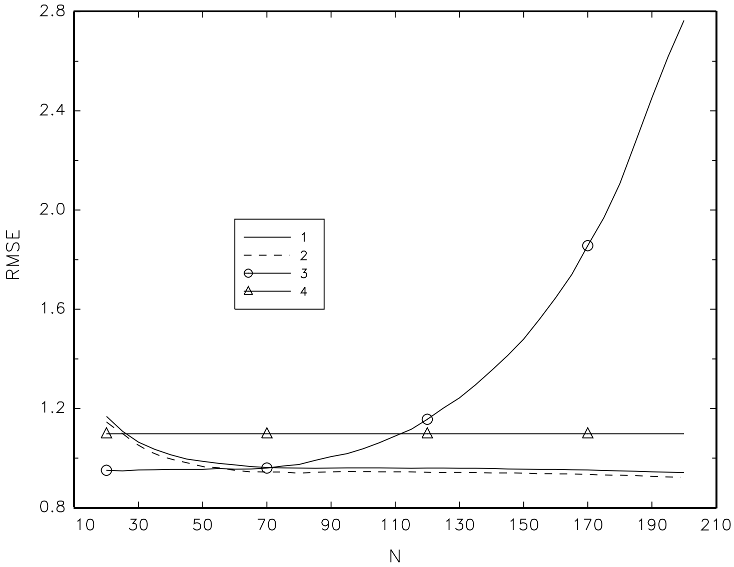

Fig 3. *Root mean squared error of* 99% *quantile estimators with different* N, *with* $n = 1000$, $m(x) = 3sin(3x)$ *and student-t distributed* U *with* $v = 2$. $1 : q^s$, $2 : \hat{q}$, $3 : q^h$, *and* $4 : q^e$.

References.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68** 326–328.

Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* 799–808.

Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics* **147** 120–130.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press, New York.

Embrechts, P., Kluppelberg, C. and Mikosh, T. (1997). *Modelling extremal events for insurance and finance*. Springer Verlag, Berlin.

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *Annals of Statistics* **13** 428–433.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer Verlag.

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

Li, Q. and Racine, J. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton, NJ.

Martins-Filho, C. and Yao, F. (2006a). A Note on the use of V and U statistics in nonparametric models of regression. *Annals of the Institute of Statistical Mathematics* **58** 389–406.

Martins-Filho, C. and Yao, F. (2006b). Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. *Studies in Nonlinear Dynamics & Econometrics* **10** Article 4.

Martins-Filho, C. and Yao, F. (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics* **141** 283–319.

McNeil, A. and Frey, B. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7** 271–300.

Pagan, A. and Ullah, A. (1999). *Nonparametric econometrics*. Cambridge University Press, Cambridge, UK.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3** 119–131.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer Verlag, New York.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics* **15** 1174–1207.

Yang, S.-S. (1985). A Smooth nonparametric estimator of a quantile function. *Journal of the American Satistical Association* **80** 1004–1011.

Department of Economics
University of Colorado
Boulder, CO 80305
E-mail: carlos.martins@colorado.edu
URL: http://spot.colorado.edu/∼martinsc

Department of Economics
West Virginia University
Morgantown, WV 80305
E-mail: feng.yao@mail.wvu.edu
URL: http://community.wvu.edu/∼fy006

IFPRI
2033 K Street, NW
Washington, DC
E-mail: m.torero@cgiar.org
URL: http://www.ifpri.org/staffprofile/maximo-torero