This article was downloaded by: [West Virginia University]
On: 08 February 2015, At: 08:08
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lecr20

High-Order Conditional Quantile Estimation Based on Nonparametric Models of Regression

Carlos Martins-Filho [a c] , Feng Yao [b] & Maximo Torero [c]
[a] Department of Economics , University of Colorado , Boulder , Colorado , USA
[b] Department of Economics , West Virginia University , Morgantown , West Virginia , USA
[c] IFPRI , Washington , District of Columbia , USA

Accepted author version posted online: 05 Sep 2014.Published online: 17 Dec 2014.

Click for updates

To cite this article: Carlos Martins-Filho , Feng Yao & Maximo Torero (2015) High-Order Conditional Quantile Estimation Based on Nonparametric Models of Regression, Econometric Reviews, 34:6-10, 906-957, DOI: 10.1080/07474938.2014.956612

To link to this article: http://dx.doi.org/10.1080/07474938.2014.956612

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Econometric Reviews, 34(6–10):906–957, 2015
Copyright © Taylor & Francis Group, LLC
ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2014.956612

High-Order Conditional Quantile Estimation Based on Nonparametric Models of Regression

Carlos Martins-Filho[1,3], Feng Yao[2], and Maximo Torero[3]
[1] *Department of Economics, University of Colorado, Boulder, Colorado, USA*
[2] *Department of Economics, West Virginia University, Morgantown, West Virginia, USA*
[3] *IFPRI, Washington, District of Columbia, USA*

We consider the estimation of a high order quantile associated with the conditional distribution of a regressand in a nonparametric regression model. Our estimator is inspired by Pickands (1975) where it is 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 standardized residuals from a first stage kernel based nonparametric estimation. 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 will also find the results useful.

Keywords Conditional quantile; Extreme value theory; Generalized Pareto distribution; Nonparametric regression.

JEL Classification C10; C14; C21.

1. INTRODUCTION

Consider the nonparametric regression model

$$Y = m(X) + \theta^{1/2}U, \tag{1}$$

Address correspondence to Prof. Carlos Martins-Filho, Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA; E-mail: carlos.martins@colorado.edu

where m is a real valued function that belongs to a suitably restricted class (see Section 3), $0 < \theta < \infty$ is an unknown parameter, $E(U \mid X = x) = 0$, and $V(U \mid X = x) = 1$. We assume that U has a strictly increasing absolutely continuous distribution $F(u)$ which belongs to the domain of attraction of an extremal distribution (Leadbetter et al., 1983; 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 $q_{Y \mid X=x}(a)$, is given by $q_{Y \mid X=x}(a) = m(x) + \theta^{1/2} q(a)$, where $q(a)$ is the a-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 estimators for m and θ to obtain an estimator for $q_{Y \mid X=x}(a)$. In general, U is not observed, but given a random sample $\{(Y_i, X_i)\}_{i=1}^n$ and estimators $\hat{m}(x)$ and $\hat{\theta}$ for $m(x)$ and θ, it is possible to obtain a sequence of standardized nonparametric residuals

$$\widehat{U}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\theta}^{1/2}} \qquad \text{for } i = 1, \ldots, n, \tag{2}$$

that can be used to produce an estimator $\hat{q}(a)$ for $q(a)$. Then, we can define $\hat{q}_{Y \mid X=x}(a) = \hat{m}(x) + \hat{\theta}^{1/2} \hat{q}(a)$ as an estimator for $q_{Y \mid X=x}(a)$.

In this article, we are particularly interested in cases 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 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 proves helpful in the estimation of $q(a)$. The seminal result comes from Pickands (1975), who 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 $\sigma(\xi)$,

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

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

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

with $\mu \leq y < \infty$ if $k \leq 0$, $\mu \leq y \leq \mu + \sigma/k$ if $k > 0$. It is evident that $F_\xi(u)$ is the conditional distribution of exceedances over a threshold ξ of a 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. Therefore, it is intuitively appealing to obtain an estimator for $q(a)$ 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 article, we extend Smith's results and study the asymptotic properties of ML type estimators for k and $\sigma(\xi)$ based on a sequence $\{\widehat{U}_i\}_{i=1}^n$ obtained from first stage estimators $\hat{m}(x)$ and $\hat{\theta}$. 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 Eq. (1) (see (Embrechts et al., 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 generalized pareto distribution (GPD) based on the sequence $\{\widehat{U}_i\}_{i=1}^n$ converges 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 standardized nonparametric residuals does not impact the estimator's rate of convergence, it does increase 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 $\hat{m}(x)$ and $\hat{\theta}$, provide weak consistency of $\hat{q}_{Y\,|\,X=x}(a)$ as an estimator for $q_{Y\,|\,X=x}(a)$.

Besides the introduction, this article 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

Our proposed estimation procedure has two main stages. First, the definition of $\widehat{U}_i$ in (2) requires specific estimators for $m(x)$ and θ. We consider the local linear (LL) estimator $\hat{m}(x) \equiv \hat{\alpha}$ where $(\hat{\alpha}, \hat{\beta}) \equiv \operatorname{argmin}_{\alpha,\beta} \sum_{i=1}^n (Y_i - \alpha - \beta(X_i - x))^2 K_1\left(\frac{X_i - x}{h_{1n}}\right)$ based on a random sample $\{(Y_i, X_i)\}_{i=1}^n$ of observations on $(Y, X) \in \Re^2$ and $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{m}(X_i))^2$. Here, $K_1(\cdot)$ is a kernel function satisfying some standard properties (see Section 3), and $0 < h_{1n}$ is a bandwidth.[1] It should be clear from what follows that other nonparametric estimators for $m(x)$ could be used to define $\widehat{U}_i$. What is important is that they are uniformly asymptotically close to $m(x)$ in probability at a suitable rate. In particular, under our assumptions, the LL estimator satisfies

---

[1] 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.

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

where $G \subset \Re$ is the compact support of m (see Assumption A5 in Subsection 3.2). It is worth mentioning that, if the regression model in (1) is parametric, i.e., $m(x) = m(x; \beta)$ for $\beta \in B \subset \Re^D$ where D is finite, then any estimator for β that converges at the parametric rate ($\sqrt{n}$) will satisfy (4), therefore preserving the asymptotic characterizations of Theorems 1 and 2 in Section 3.

Also critical is that $\hat{\theta}$ converges to θ sufficiently fast (but not necessarily at a parametric rate). Under our assumptions and given (4), it can be easily obtained that

$$\hat{\theta} - \theta = O_p\left(\left(\frac{nh_{1n}}{\log n}\right)^{-1/2} + h_{1n}^2\right), \tag{5}$$

which will prove sufficient for our asymptotic results. The second stage of estimation, which is based on the equivalence in (3), is more intricate and requires some additional notation and motivation. We first discuss estimation for the case where U_i is observed. Since the GPD is a suitable approximation for the upper tail of F, it is intuitively reasonable to use only sufficiently large values of U_i to estimate its parameters. Therefore, a key aspect of the estimation is the determination of a threshold value such that only its exceedances are used to estimate the parameters of the GPD. For any given sample size n, a specific threshold selection implicitly defines a number N of exceedances to be used in the second stage estimation. Alternatively, by choosing N an implicit threshold (not unique) is defined.

For an observed sequence $\{U_i\}_{i=1}^n$, we define the order statistics $\{U_{(i)}\}_{i=1}^n$ where $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$. For a fixed $N < n$, we define the excesses over $U_{(n-N)}$ by $\{Z_j\}_{j=1}^N = \{U_{(n-N+j)} - U_{(n-N)}\}_{j=1}^N$. Ascending 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 set of positive integers and $q_n(a)$ is the a-quantile associated with the empirical distribution $F_n(u) = n^{-1} \sum_{i=1}^n \chi_{\{U_i \leq u\}}$ with χ_A denoting the indicator function for the set A. Consequently, for $a_n = 1 - \frac{N}{n}$, we can write $\{Z_j\}_{j=1}^N = \{U_{(n-N+j)} - q_n(a_n)\}_{j=1}^N$. Thus, for a given sample size n and N (and consequently a_n), we can estimate the threshold $q_n(a_n)$ which will be exceeded by exactly the N largest elements of $\{U_i\}_{i=1}^n$. The sequence $\{Z_j\}_{j=1}^N$ could then be used to estimate the parameters of the GPD. Here, the threshold ξ in equivalence (3) is estimated by $q_n(a_n)$ which, given a sample of size n, determines the order statistics to be used in estimating the parameters of the GDP.

Since in our case we only observe $\{\widehat{U}_i\}_{i=1}^n$, we must produce an estimated sequence of exceedances with typical element given by $\widetilde{Z}_j$. Perhaps the most natural procedure would be to define $\widetilde{Z}_j = \widehat{U}_{(n-N+j)} - \hat{q}_n(a_n)$, where $\hat{q}_n(a_n)$ is the a_n-quantile associated with the empirical distribution of the nonparametric standardized residuals $\{\widehat{U}_i\}_{i=1}^n$. However, it is well known from the unconditional distribution and quantile estimation literature (Azzalini, 1981; Falk, 1985; Yang, 1985; Bowman et al., 1998; Martins-Filho and Yao, 2008) 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, we define $\tilde{q}(z)$ as the solution for $\widetilde{F}(\tilde{q}(z)) = z$, where $\widetilde{F}(u) = \int_{-\infty}^{u} \frac{1}{nh_{2n}} \sum_{i=1}^{n} K_2\left(\frac{y-\widehat{U}_i}{h_{2n}}\right) dy$, $K_2(\cdot)$ is a symmetric kernel function and $0 < h_{2n}$ is a bandwidth satisfying certain regularity conditions. Therefore, we define the *observed* sequence of exceedances to be used in the estimation of the parameters of the GPD in the second stage as $\{\widetilde{Z}_j\}_{j=1}^{N_s} = \left\{\widehat{U}_{(n-N_s+j)} - \tilde{q}(a_n)\right\}_{j=1}^{N_s}$. Note that here the number of residuals N_s that exceeds $\tilde{q}(a_n)$ may be different from N for any finite n. As will be seen in Section 3, this finite sample difference will be of no asymptotic consequence for the estimation.

Given the sequence $\{\widetilde{Z}_j\}_{j=1}^{N_s}$, we 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 when $\mu = 0$. In particular, we obtain a solution $(\tilde{\sigma}_N, \tilde{k})$ for the following likelihood equations:

$$\frac{\partial}{\partial \sigma}\frac{1}{N_s}\sum_{j=1}^{N_s} log\, g(\widetilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0 \quad \text{and} \quad \frac{\partial}{\partial k}\frac{1}{N_s}\sum_{j=1}^{N_s} log\, g(\widetilde{Z}_j; \tilde{\sigma}_N, \tilde{k}) = 0. \tag{6}$$

Now, if $\{U_i\}_{i=1}^n$ were observed, for a threshold $\xi = q_n(a_n)$ we could write, based on (3),

$$F_{q_n(a_n)}(y) = \frac{F(y + q_n(a_n)) - F(q_n(a_n))}{1 - F(q_n(a_n))} \approx 1 - \left(1 - \frac{ky}{\sigma_N}\right)^{1/k},$$

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

$$\frac{1-a}{1 - F(q_n(a_n))} \approx \left(1 - \frac{ky_{N,a}}{\sigma_N}\right)^{1/k}. \tag{7}$$

If F is approximated by the empirical distribution F_n, then $1 - F(q_n(a_n)) \approx \frac{N}{n}$, which suggests $y_{N,a} \approx \frac{\sigma_N}{k}\left(1 - \left(\frac{(1-a)n}{N}\right)^k\right)$ and $q(a) \approx q_n(a_n) + \frac{\sigma_N}{k}\left(1 - \left(\frac{(1-a)n}{N}\right)^k\right)$. The last approximation is the basis for our proposed estimator $\hat{q}(a)$ for $q(a)$, which is given by

$$\hat{q}(a) = \tilde{q}(a_n) + \frac{\tilde{\sigma}_N}{\tilde{k}}\left(1 - \left(\frac{(1-a)n}{N}\right)^{\tilde{k}}\right). \tag{8}$$

Lastly, the estimator for $q_{Y|X=x}(a)$ is given by $\hat{q}_{Y|X=x}(a) = \hat{m}(x) + \hat{\theta}^{1/2}\hat{q}(a)$. In the next section, we discuss the existence and 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 seminal results from Smith (1987) as they are helpful in 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 location-scale model, in Smith (1987) the estimation of $q(a)$ is conducted under the assumption that the sequence $\{Z_j\}_{j=1}^N$ is observed. As such, he proposes estimators $(\hat{\sigma}_N, \hat{k})$ that satisfy the first order conditions

$$\frac{\partial}{\partial\sigma}\frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \hat{\sigma}_N, \hat{k}) = 0 \quad \text{and} \quad \frac{\partial}{\partial k}\frac{1}{N}\sum_{j=1}^N log\, g(Z_j; \hat{\sigma}_N, \hat{k}) = 0 \tag{9}$$

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 $\sigma_N(1 + t\delta_N)$, $k_0 + \tau\delta_N$ 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) $L_{TN}(0,0) = L_N(\sigma_N, k_0)$ and b) choosing $(\hat{\sigma}_N, \hat{k})$ such that Eq. (9) is satisfied is equivalent to choosing t^* and τ^* that satisfy

$$\frac{1}{\sigma_N\delta_N}\frac{\partial L_{TN}}{\partial t}(t^*, \tau^*) = 0 \text{ and } \frac{1}{\delta_N}\frac{\partial L_{TN}}{\partial \tau}(t^*, \tau^*) = 0. \tag{10}$$

By a Taylor's Theorem expansion of the left-hand side of the derivatives in (10) about $(0,0)$, we have for $\lambda_1, \lambda_2 \in [0,1]$

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

and

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

where the terms I_{lN} for $l = 1,\dots,6$ denote the corresponding average in the preceding equality. The existence and characterization of a solution (t^*,τ^*) for Eq. (10) as maxima for L_{TN} depends on obtaining expressions for the expected value of the score vector and its derivative with respect to the unknown distribution F_ξ where $\xi\to\infty$. Smith (1987) has shown 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 $\alpha > 0$,

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,

then for a nonstochastic positive sequence $u_N\to\infty$ as $N\to\infty$ and for $\sigma_N = u_N/\alpha$, $0<\alpha=-1/k_0$ and $k_0<0$, we have $E(\sigma_N\frac{\partial}{\partial\sigma}log\, g(Z;\sigma_N,k_0)) = \frac{C\phi(u_N)}{(1+\alpha-\rho)} + o(\phi(u_N))$, $E(\frac{\partial}{\partial k}log\, g(Z;\sigma_N,k_0)) = -\frac{\alpha C\phi(u_N)}{(\alpha-\rho)(1+\alpha-\rho)} + o(\phi(u_N)$, $E(\sigma_N^2\frac{\partial^2}{\partial\sigma^2}log\, g(Z;\sigma_N,k_0)) = -\frac{\alpha}{2+\alpha} + O(\phi(u_N))$, $E(\frac{\partial^2}{\partial k^2}log\, g(Z;\sigma_N,k_0)) = -\frac{2\alpha^2}{(1+\alpha)(2+\alpha)} + O(\phi(u_N))$, and $E(\sigma_N\frac{\partial^2}{\partial\sigma\partial k}log\, g(Z;\sigma_N,k_0)) = \frac{\alpha^2}{(1+\alpha)(2+\alpha)} + O(\phi(u_N))$, where all expectations are taken with respect to the unknown distribution F_{u_N}. Evidently, these approximations are based on a sequence of thresholds u_N that approach the end point of the distribution F as the $N\to\infty$. In addition, it can easily be shown 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 $N^{1/2}\phi(u_N) = O(1)$, 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

$$\begin{pmatrix} t & \tau\end{pmatrix}\begin{pmatrix}\frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t,\tau)\\ \frac{1}{\delta_N^2}\frac{\partial}{\partial\tau}L_{TN}(t,\tau)\end{pmatrix}\xrightarrow{p}\begin{pmatrix} t & \tau\end{pmatrix}(-H)\begin{pmatrix} t\\ \tau\end{pmatrix}\le 0 \quad \text{on } S_T. \tag{11}$$

By 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 satisfies the first order conditions in Eq. (9).[2] Our first result (Lemma 1) shows that the solution for the first order conditions given in Eq. (6) corresponds to a local maximum of $\widetilde{L}_N(\sigma,k) = \frac{1}{N}\sum_{j=1}^{N_s} log\, g(\widetilde{Z}_j;\sigma,k)$. We do so by showing that the first order conditions given in (6) are asymptotically uniformly equivalent in probability to those in (9). However, given that we must deal with estimated sequences $\widetilde{Z}_j$, additional assumptions are needed.

3.2. Assumptions

As in Smith (1987), we retain 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. However, additional assumptions are needed. The first set of additional assumptions results from the fact that given Eq. (1) we must conduct nonparametric estimation of m, θ and q. Therefore, we need the following assumption.

Assumption A1. The kernel functions $K_i(x)$ for $i = 1,2$ are bounded, symmetric, twice continuously differentiable functions $K_i(x) : \Re \to \Re$ with compact support. They satisfy $\int K_i(s)ds = 1, \int sK_i(s)ds = 0$, $\int s^2K_1(s)ds = \sigma_{K_1}^2 < \infty$ and $\int s^jK_2(s)ds = 0$ for $j = 1,\ldots,m$, where $m > 2$. Furthermore, we assume that for any $u, v \in \Re$ with $u \neq v$, we have $|K_i(u) - K_i(v)| \leq C|u-v|$ for some constant $C > 0$.

The higher order m for K_2 is necessary in the proof of Lemma 2 in Appendix 1. All other assumptions 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 $n \to \infty$, where $m > 2$.

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 $N^{1/2}\phi(u_N) = O(1)$, where u_N is a positive nonstochastic sequence such that $u_N \to \infty$ as $N \to \infty$.

---

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

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

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

Assumption A4. $\{(X_i, U_i)\}_{i=1,\ldots,n}$ 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 $f_{U\,|\,X=x}(u\,|\,x)$. We assume that $E(U\,|\,X) = 0$, $E(U^2\,|\,X) = 1$, $E(U^4\,|\,X) = \mu_4 < \infty$, and that $\lim_{u\to\infty} \frac{f_{U\,|\,X=x}(u\,|\,x)}{f(u)} = 1$ almost surely.

The requirement that $\lim_{u\to\infty} \frac{f_{U\,|\,X=x}(u\,|\,x)}{f(u)} = 1$ almost surely implies that U and X are asymptotically independent in the tail of U.

We note that strong moment requirements on U are needed since we estimate conditional quantiles *via* the estimation of $m(x)$ and θ. An alternative approach, that bypasses these requirements, would be to extend the parametric quantile regression of Chernozhukov (2005) to a nonparametric setting.

Assumption A5. $m(x)$ and $f_X(x)$ are twice continuously differentiable with compact support given by G and $\inf_{x\in G} f_X(x) > 0$.

Assumptions A1, A2, A4, and A5 are sufficient for Eq. (4) to hold for both the LL and Nadaraya-Watson (NW) estimators, but can be relaxed at some cost. They are, however, standard in the nonparametric literature (Masry, 1996; Fan and Yao, 2003; Li and Racine, 2007).

The second additional assumption we need relates to the fact that, as mentioned in Section 2, our estimation procedure hinges on the estimation of the threshold u_N which appears in Section 3.1 by $\tilde{q}(\cdot)$. As a result, contrary to Smith (1987) (see his Theorem 3.2), we explicitly account for the stochastic nature of the estimated threshold. This added difficulty requires a further restriction on the class of distributions F we consider. Specifically, as in Davis and Resnick (1984), we assume

FR1': F has a strictly positive density denoted by f and for some $\alpha > 0$ we have $\lim_{x\to\infty} \frac{xf(x)}{1-F(x)} = \alpha$.

We note that by Corollary 1.12 and Proposition 1.15 c) in Resnick (1987), FR1' implies FR1, assuring that under FR1' $F \in D(\Phi_\alpha)$, with $\alpha = -1/k_0$ and $k_0 < 0$. Finally, let us note that 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.

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 $\widetilde{L}_{TN}(t,\tau) = \frac{1}{N}\sum_{j=1}^{N_s} log\, g(\widetilde{Z}_j; \sigma_N(1+t\delta_N), k_0+\tau\delta_N)$ are asymptotically uniformly equivalent in probability to those associated with L_{TN} on the set S_T. Formally, we have the following lemma.

Lemma 1. *Assume FR1' with* $\alpha > 1$, *FR2, and Assumptions A1–A5. 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*

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

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

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

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

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 satisfies the first order conditions in Eq. (6). The proof of Lemma 1 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 on the compact set G that $q_n(a_n)^{-1} sup_{x\in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$ and that $\hat{\theta} - \theta = O_p\left(\left(\frac{nh_{1n}}{log\, n}\right)^{-1/2} + h_{1n}^2\right)$. This assures that the residuals $\widehat{U}_i$ are in some sense *close* to the unobserved U_i. Second, in Lemma 2 $\tilde{q}(a_n)$ is shown to be asymptotically close to $q_n(a_n)$ by satisfying $\frac{\tilde{q}(a_n) - q_n(a_n)}{q(a_n)} = O_p(N^{-1/2})$. It is in Lemma 2 that the stochasticity of the threshold $\tilde{q}(\cdot)$ is explicitly handled and where assumption FR1' is used. 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 $F_{q(a_n)}$ 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 (6) 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, Theorem 3.2) showed that given conditions FR1, FR2 and provided $\{Z_j\}_{j=1}^{N}$ is an independent and identically distributed sequence from F_{u_N}, $N \to \infty$ and $\frac{C}{\alpha-\rho} N^{1/2} \phi(u_N) \to \mu \in \Re$, the local maximum $(\hat{\sigma}_N, \hat{k})$ of the GPD likelihood function, is such that for $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = \frac{u_N}{\alpha}$

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

where $H = \frac{1}{(1-2k_0)(1-k_0)} \begin{pmatrix} 1-k_0 & -1 \\ -1 & 2 \end{pmatrix}$.[3] Our first theorem provides a similar asymptotic result for the estimators $(\tilde{\sigma}_N, \tilde{k})$. A main difference between our result and Smith's derives from the fact that, as in Lemma 1, our Theorem 1 accounts for the stochasticity of the threshold $\tilde{q}(a_n)$ used to define the exceedances in our estimation. It is the stochasticity of the threshold that requires the stronger FR1' instead of FR1. In fact, even if the U_i were observed, given that exceedances Z_j depend on the stochastic $q_n(a_n)$, Theorem 3.2 as stated is not valid as it treats the threshold u_N as a nonstochastic sequence.

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

$$\sqrt{N} \begin{pmatrix} \frac{\tilde{\sigma}_N}{\sigma_N} - 1 \\ \tilde{k} - k_0 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \frac{\mu(1-k_0)(1+2k_0\rho)}{1-k_0+k_0\rho} \\ \frac{\mu(1-k_0)k(1+\rho)}{1-k_0+k_0\rho} \end{pmatrix}, H^{-1} V_2 H^{-1} \right),$$

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

It is important to emphasize that the equivalence in (3) requires the parameter σ on the GPD to be dependent on the threshold ξ that must be approaching the endpoint u_∞ of the distribution F. In our setting, the threshold is $q(a_n)$ (a function of N since $a_n = 1 - \frac{N}{n}$). Note that, as required, $q(a_n) \to u_\infty$ ($u_\infty = \infty$ when F belongs to the domain of attraction of a Fréchet distribution) if $(N/n) \to 0$. This observation justifies the notation $\sigma_N = \sigma(q(a_n))$. The fact that $\sigma_N = q(a_n)/\alpha$ in Lemma 1 and Theorem 1, where α is the coefficient of regular variation and the indexing parameter of a Fréchet distribution derives from the fact that we have restricted our study to a class of distributions F that are in the domain of attraction of a Fréchet distribution.

---

[3] Substituting $k_0 = -\alpha^{-1}$ shows that H is identical to the homonymous matrix in Eq. (11).

The use of $\widetilde{Z}_j$ 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 $\widetilde{Z}_j$ is *not* iid as Z_j does not require the use of a Central Limit Theorem (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 type estimators given in Theorem 1 is the basis for obtaining a normality result for $\hat{q}(a)$ given in Eq. (8). 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 to 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' with* $\alpha > 1$, *FR2, and assumptions A1–A5 hold. In addition, assume as follows:*

(i) $N^{1/2}\phi(q(a_n))\frac{C}{(\alpha-\rho)} \to \mu$ *with* $k_0 = -\frac{1}{\alpha}$ *and* $\sigma_N = q(a_n)/\alpha$;
(ii) $n(1-a) \propto N$. *Then, for some* $z_a > 0$,

$$\sqrt{n(1-a)}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \xrightarrow{d} N\left((-k_0)\left(-\frac{-(z_a^{\rho}-1)\mu(\alpha-\rho)}{\rho} - c_b' H^{-1} \lim_{n\to\infty} \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right),\right.$$
$$\left. k_0^2\left(c_b' H^{-1} V_2 H^{-1} c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\right),$$

where $c_b' = \left(-k_0^{-1}(z_a^{-1}-1) \;\; k_0^{-2}log(z_a)+k_0^{-2}(z_a^{-1}-1) \right)$, $b_\sigma = E\left(\frac{\partial}{\partial\sigma}log\, g(Z_j;\sigma_N,k_0)\sigma_N\right)$, *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 $a \to 1$ with $a \in (a_n, 1)$ we have $\hat{q}_{Y\,|\,X=x}(a) = \hat{m}(x) + \hat{q}(a) \xrightarrow{p} m(x) + q(a) = q_{Y\,|\,X=x}(a)$.

It is worth mentioning that the proofs of Theorems 1 and 2 depend only on nonparametric uniform convergence rates for $\hat{m}$ and $\hat{\theta}$. Specifically, even though $\hat{\theta}$ converges to θ at a parametric rate (Doksum and Samarov, 1995; Martins-Filho and Yao, 2006a), we have constructed our proofs using only lower speed in Eq. (5). This is encouraging as it points towards the possibility of obtaining a version of Theorem 2 based on the more flexible model $Y = m(X) + \theta^{1/2}(X)U$, where $\theta(X)$ is estimated nonparametrically.

The estimation of $q_{Y|X=x}(a)$ when a is in the vicinity of 0 (low order conditional quantile) has been considered by Chernozhukov (2005) when $q_{Y|X=x}(a) = x\beta(a)$, $\beta(a) \in \Re^d$. In this case, he provides a complete asymptotic characterization of the quantile regression estimator of $\beta(a)$ proposed by Koenker and Bassett (1978) for F in the domain of attraction of any extreme value distribution. Furthermore, his asymptotic results are obtained when the quantile order approaches 0 at a speed that is slower than or proportional to the sample size. Since our model for $q_{Y|X=x}(a)$ is nonparametric, it is in this sense more general than the one considered by Chernozhukov. However, our asymptotic results on the parameters of the GPD are limited to the case where F belongs to the domain of attraction of a Fréchet distribution and the quantile order (a) approaches 1 at a speed that is slower than the sample size ($n(1-a) \propto N \to \infty$ in Theorem 2). Furthermore, similar to Smith (1987) and Hall (1982), our proofs require the specification of the speed at which the tail $1 - F(x)$ behaves asymptotically as a power function. Condition (i) in Theorem 2 specifies this speed to be proportional to $\sqrt{N}$.

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

$$Y_i = m(X_i) + \theta^{1/2} U_i, \qquad i = 1, \ldots, n,$$

where X_i is uniformly distributed on $[-2, 2]$. We consider two nonlinear functions for $m(\cdot)$, $m_1(x) = 3\,sin(3x)$ and $m_2(x) = x^2$. $V_i = \theta^{1/2} U_i$ 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

$$f(u) = (log(u))^{\alpha_1 - 1} \frac{u^{-\frac{1}{\beta} - 1}}{\beta^{\alpha_1} \Gamma(\alpha_1)}, \qquad \text{for } u > 1, \quad \alpha_1, \beta > 0.$$

It is easy to see that V_i is log-gamma distributed for $V_i > 1$ if and only if $log(V_i) > 0$ is gamma distributed with parameters $\alpha_1, \beta > 0$. Furthermore, one can show that $E(V_i) = (\frac{1}{1-\beta})^{\alpha_1}$, $V(V_i) = \theta = (\frac{1}{1-2\beta})^{\alpha_1} - (\frac{1}{1-\beta})^{2\alpha_1}$, and $k_0 = -\beta$. The log-gamma distribution includes the Pareto distribution as a special case when $\alpha_1 = 1$. We specifically let $(\alpha_1, \beta) = (1, 0.25)$, and $(1, 0.4)$, which correspond to $k_0 = -0.25$ and -0.4 respectively. The variance of V_i when $(\alpha_1, \beta) = (1, 0.4)$ is ten times the variance of V_i when $(\alpha_1, \beta) = (1, 0.25)$. V_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 $-1/2.25$ for

$v = 2.25$, respectively. $V(V_i) = \theta = \frac{v}{v-2}$, so the variance of V_i for $v = 2.25$ is three times that the variance of V_i for $v = 3$. We expect that the estimation will be relatively more difficult when the variance is larger.

Implementation of our estimator requires the choice of bandwidths h_{1n} and h_{2n}. We select them using the *rule-of-thumb* bandwidths $\hat{h}_{1n} = 1.25\, S(X) n^{-\frac{1}{5}}$ and $\hat{h}_{2n} = 0.79\, R(\widehat{U}) n^{-\frac{1}{5}+0.01}$, with a robust estimation for the variability of data as in (2.52) of Pagan and Ullah (1999), where $S(X)$ and $R(\widehat{U})$ are the standard deviation of X and the sample interquartile range of $\widehat{U}$, 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 et al., 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 $\widehat{U}_i = \frac{Y_i - \hat{m}(x_i)}{\hat{\theta}^{1/2}}$ to construct $\hat{k}^h = -\frac{1}{N}\sum_{j=1}^{N}(ln(\widehat{U}_{(n-N+j)}) - ln(\widehat{U}_{(n-N)}))$. 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 (infeasible) estimator q^s, the Hill type estimator q^h, and the empirical quantile estimator q^e. Following (6.30) in Embrechts et al. (1997), we construct q^h as $q^h = \widehat{U}_{(n-N)}(\frac{1-a}{N/n})^{\hat{k}^h}$. q^e is simply the empirical quantile estimator based on $\{\widehat{U}_i\}_{i=1}^n$. To give the reader a vivid picture of these estimators in practice, we provide in Fig. 1 a plot of different quantile estimates against different values of a, where q^s is omitted for ease of illustration. We let a range from 0.95 to 0.995 as we are interested in higher order quantiles. The data are generated with $m(x) = 3\, sin(3x)$, $\theta^{1/2} U_i$ is from the log-gamma distribution with $(\alpha_1, \beta) = (1{,}0.4)$, and we select $n = 1{,}000$ and $N = round(\frac{1}{2} n^{0.8-0.01}) = 117$, where $round(\cdot)$ gives the nearest integer. 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 sample sizes n to be 500 and 1,000 in our simulation. We choose $N = round(c\, n^{0.8-0.01})$, where c is simply set to be $\frac{1}{2}$, and the choice of N satisfies the assumptions in our asymptotic analysis. Thus, N is 68 for $n = 500$ and 117 for $n = 1{,}000$, and the effective sample size N in the second stage estimation is increased, but it is less than doubled. 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 Figs. 2 and 3 and in the discussion below. Each experiment is repeated

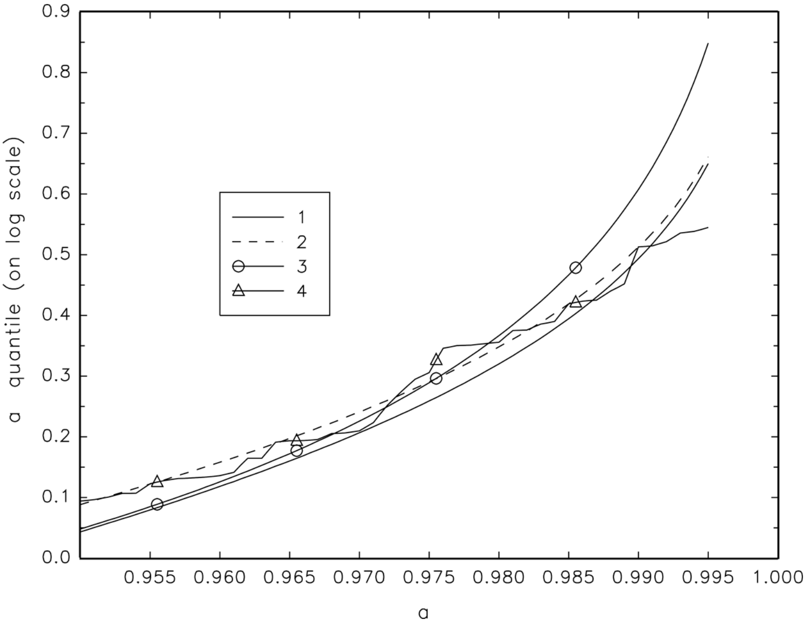

FIGURE 1 Plot of quantile estimates across different a, with $n = 1000$, $N = 117$, $m(x) = 3sin(3x)$, and log-gamma distributed $\theta^{\frac{1}{2}}U$ with $(\alpha_1, \beta) = (1{,}0.4)$. 1 : true quantile, $2 : \hat{q}$, $3 : q^h$, and $4 : q^e$

5,000 times. We summarize the performance of parameter estimators in terms of their bias (B), standard deviation (S), and root mean squared error (R) in Table 1 for both $m(x) = 3sin(3x)$ and $m(x) = x^2$ with log-gamma distributed U, and in Table 2 with student-t distributed U. We provide the performance for quantile estimators of $q(a)$ where $a = 0.95, 0.99$ and 0.995 by examining the bias (B), standard deviation (S), and root mean squared error (R) in Tables 3–6. Specifically, results for log-gamma distributed U with $(\alpha_1, \beta) = (1, 0.25)$ are detailed in Table 3, for log-gamma distributed U using $(\alpha_1, \beta) = (1, 0.4)$ in Table 4, and for student-t distributed U using $v = 3$ and $v = 2.25$ in Tables 5 and 6.

In the case of estimating the parameters, we notice that $\hat{\gamma}$ and $\tilde{\gamma}$ tend to overestimate (σ_N, k_0), while $\hat{\gamma}^h$ carries a negative bias. As N increases, the performance of all estimators improves, in the sense that they exhibit smaller bias and standard deviation, which seems to confirm the asymptotic result in previous section. As k_0 decreases (larger β in Table 1 and smaller v in Table 2), we find the biases of all estimators decrease with some exceptions for $\hat{k}^h$. The standard deviations of all estimators generally increase except for $\hat{\sigma}_N$, whose standard deviation decreases. The performance of estimators for σ_N in terms of root mean squared error actually improves as k_0 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 V_i is larger for the log-gamma distributions with $\beta = 0.4$ in Table 1

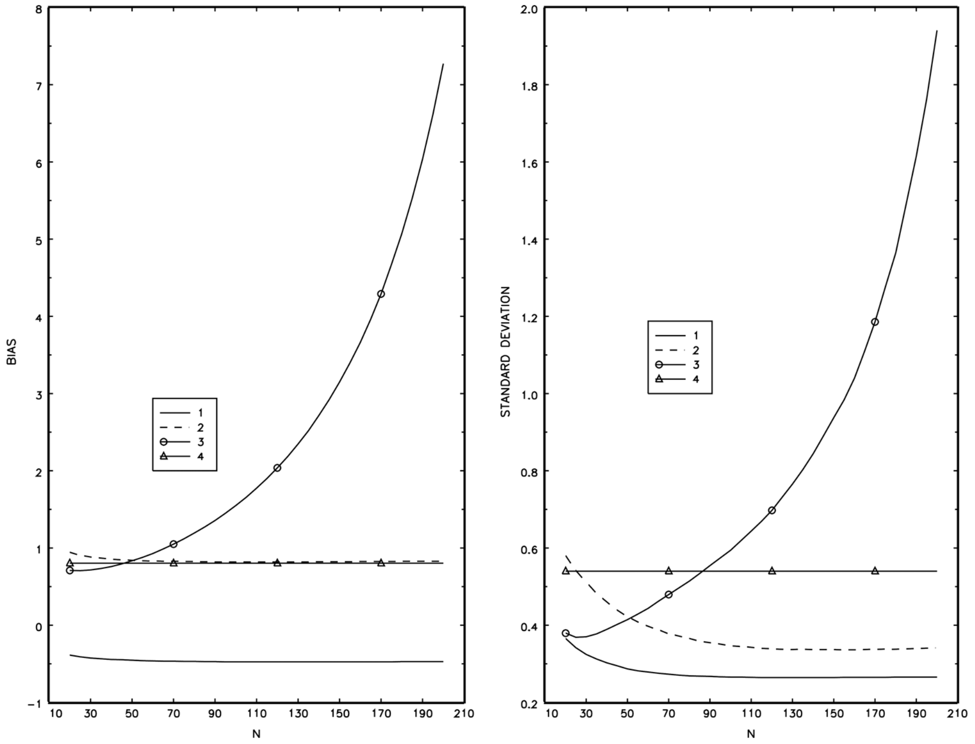

FIGURE 2 Bias and standard deviation of 99% quantile estimators with different N, with $n = 1000$, $m(x) = 3sin(3x)$ and log-gamma distributed $\theta^{\frac{1}{2}}U$ with $(\alpha_1, \beta) = (1, 0.4)$. $1 : q^s$, $2 : \hat{q}$, $3 : q^h$, and $4 : q^e$

and for student-t distribution with $v = 2.25$ in Table 2. The distribution of V_i starts to exhibit heavier tail behavior, thus more representative extreme observations have a higher probability to show up in a sample, which explains lower bias. $\tilde{\sigma}$'s performance is very similar to that of $\hat{\sigma}$, with slightly larger standard deviation and bias. Among the three estimators for k, $\hat{k}^h$ is worst in log-gamma distributed U in terms of the largest root mean squared error with largest bias and smallest standard deviation, but $\hat{k}^h$ outperforms the other two in student-t distributed U in terms of smallest root mean squared error, again with largest bias and smallest standard deviation. Relative to $\hat{k}$ which does not have to estimate $m(\cdot)$ and θ, $\tilde{k}$ exhibits larger standard deviation, similar or slightly smaller bias, and larger root mean squared error, though the difference is fairly small. It seems to suggest that our proposed estimator $\tilde{\gamma}$ is well supported by the local linear estimator for the function $m(\cdot)$ and θ.

In the case of estimating the quantiles, we notice that most estimators carry positive bias, except that for large k_0 ($k_0 = -0.25$ for log-gamma distributed V_i and $k_0 = -1/3$ for student-t distributed V_i), q^h is negatively biased in estimating 95% quantile, and q^e is negatively biased in estimating larger order (99% and 99.5%) quantiles. As N increases, all

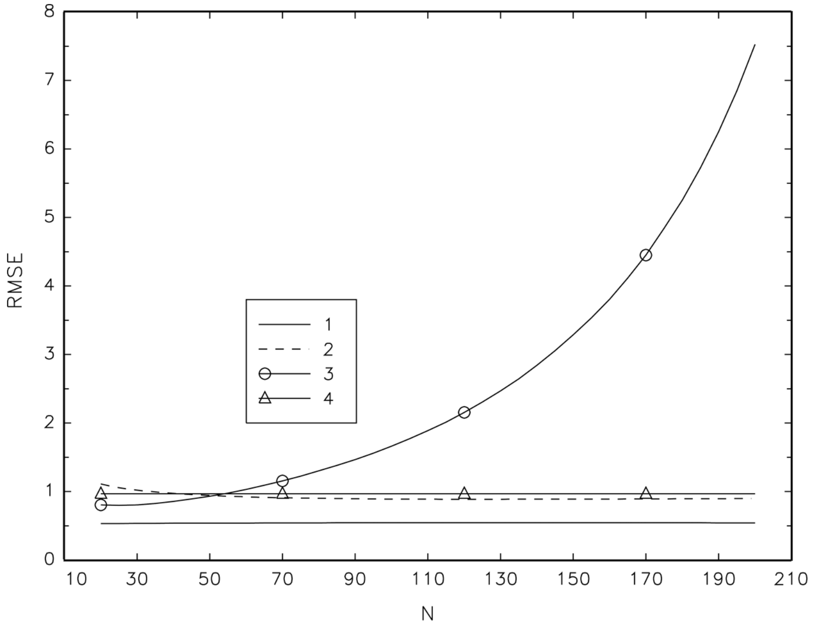

FIGURE 3 Root mean squared error of 99% quantile estimators with different N, with $n = 1000$, $m(x) = 3sin(3x)$ and log-gamma distributed $\theta^{\frac{1}{2}}U$ with $(\alpha_1, \beta) = (1, 0.4)$. $1 : q^s$, $2 : \hat{q}$, $3 : q^h$, and $4 : q^e$

TABLE 1
Bias (B), Standard Deviation (S), and Root Mean Squared Error (R) for Parameter Estimators with Log-Gamma Distributed $\theta^{\frac{1}{2}}U$ with $\alpha_1 = 1$, $k_0 = -\beta$, $n(\times 1000)$

				$m(x) = 3sin(3x)$						$m(x) = x^2$				
			σ_N			k_0			σ_N			k_0		
	β	σ_N	n	B	S	R	B	S	R	B	S	R	B	S	R
$\hat{\gamma}$	0.25	0.166	0.5	0.747	0.185	0.770	0.039	0.167	0.172	0.739	0.185	0.762	0.037	0.163	0.167
$\tilde{\gamma}$	0.25	0.166	0.5	0.750	0.226	0.784	0.030	0.170	0.172	0.766	0.224	0.798	0.029	0.165	0.168
$\hat{k}^h$	0.25	0.166	0.5				−0.539	0.093	0.547				−0.561	0.097	0.570
$\hat{\gamma}$	0.25	0.200	1	0.726	0.140	0.740	0.020	0.121	0.123	0.727	0.142	0.741	0.021	0.121	0.123
$\tilde{\gamma}$	0.25	0.200	1	0.733	0.175	0.754	0.015	0.121	0.122	0.744	0.179	0.765	0.016	0.122	0.123
$\hat{k}^h$	0.25	0.200	1				−0.478	0.064	0.482				−0.490	0.065	0.495
$\hat{\gamma}$	0.4	0.149	0.5	0.470	0.132	0.489	0.029	0.182	0.184	0.474	0.133	0.492	0.036	0.180	0.184
$\tilde{\gamma}$	0.4	0.149	0.5	0.633	0.266	0.686	0.019	0.186	0.187	0.643	0.268	0.697	0.026	0.184	0.186
$\hat{k}^h$	0.4	0.149	0.5				−0.555	0.124	0.568				−0.564	0.126	0.578
$\hat{\gamma}$	0.4	0.186	1	0.461	0.104	0.472	0.021	0.134	0.135	0.460	0.103	0.472	0.018	0.134	0.135
$\tilde{\gamma}$	0.4	0.186	1	0.606	0.219	0.644	0.015	0.135	0.136	0.604	0.223	0.644	0.012	0.135	0.136
$\hat{k}^h$	0.4	0.186	1				−0.474	0.084	0.481				−0.481	0.084	0.488

TABLE 2

Bias (B), Standard Deviation (S), and Root Mean Squared Error (R) for Parameter Estimators with Student-T Distributed $\theta^{\frac{1}{2}}U$, $k_0 = -1/v$, $n(\times 1000)$

				$m(x) = 3sin(3x)$						$m(x) = x^2$				
			σ_N			k_0			σ_N			k_0		
	v	σ_N	n	B	S	R	B	S	R	B	S	R	B	S	R
$\hat{\gamma}$	3	0.258	0.5	0.281	0.104	0.299	0.132	0.167	0.213	0.281	0.104	0.299	0.135	0.170	0.217
$\tilde{\gamma}$	3	0.258	0.5	0.296	0.127	0.323	0.128	0.170	0.213	0.296	0.127	0.322	0.130	0.171	0.215
$\hat{k}^h$	3	0.258	0.5				−0.181	0.061	0.191				−0.182	0.061	0.192
$\hat{\gamma}$	3	0.286	1	0.254	0.079	0.266	0.102	0.125	0.161	0.255	0.079	0.267	0.106	0.123	0.162
$\tilde{\gamma}$	3	0.286	1	0.263	0.100	0.282	0.098	0.126	0.160	0.266	0.099	0.284	0.102	0.124	0.161
$\hat{k}^h$	3	0.286	1				−0.156	0.044	0.163				−0.157	0.044	0.163
$\hat{\gamma}$	20.25	0.214	0.5	0.149	0.075	0.167	0.108	0.182	0.212	0.151	0.075	0.168	0.112	0.182	0.214
$\tilde{\gamma}$	20.25	0.214	0.5	0.269	0.147	0.307	0.107	0.186	0.214	0.270	0.143	0.305	0.107	0.186	0.214
$\hat{k}^h$	20.25	0.214	0.5				−0.139	0.076	0.159				−0.140	0.077	0.159
$\hat{\gamma}$	20.25	0.239	1	0.133	0.059	0.146	0.083	0.136	0.159	0.133	0.058	0.145	0.080	0.135	0.157
$\tilde{\gamma}$	20.25	0.239	1	0.240	0.122	0.269	0.080	0.138	0.159	0.236	0.122	0.266	0.075	0.137	0.156
$\hat{k}^h$	20.25	0.239	1				−0.118	0.056	0.130				−0.119	0.056	0.132

estimators' performances improve in terms of smaller bias, standard deviation, and root mean squared error, with some exceptions in bias. The distribution of V_i exhibits a heavier tail with $\beta = 0.4$ (Table 4) relative to $\beta = 0.25$ (Table 3) in the log-gamma distribution, with $v = 2.25$ (Table 6) relative to $v = 3$ (Table 5) in the student-t distribution. As we have mentioned above, the random variable V_i has a larger variance in these cases. We find it more difficult for almost all to estimate the quantiles across all experiment designs. The only exception is on q^s. When V_i has larger variance, q^s performs better in terms of smaller bias, standard deviation and root mean squared error when estimating the 95% and the 99% quantiles at least for the student distributed V_i, and q^s performs better in terms of smaller bias when estimating the 99.5% quantile. Without having to estimate $m(\cdot)$ and θ, q^s clearly benefits more from the presence of the more representative extreme observations. As expected, when we estimate higher order a-quantiles, all estimators' performances deteriorate, with some exceptions in bias. When we estimate the 95% quantile, which is relatively closer to the center of the distribution, q^s outperforms the others in terms of smallest bias, standard deviation, and root mean squared error. The second best is q^h, followed by q^e and $\hat{q}$. The advantage of the Hill type estimator does not seem to carry through in estimating the higher order-99%, and 99.5% quantiles, as q^h always carries the largest bias and standard deviation relative to the others. $\hat{q}$ and q^s are consistently the best with the smallest standard deviation and root mean squared error with some exceptions in small sample. In terms of root mean squared error and in large samples, for smaller k_0, q^s outperforms $\hat{q}$, which has larger bias and standard deviation. For larger k_0, $\hat{q}$ seems to have slightly larger bias, but smaller root mean squared error relative to q^s.

TABLE 3

Bias (×0.1) (B), Standard Deviation (S), and Root Mean Squared Error (R) for Quantile Estimators with $m_1(x) = 3sin(3x)$, $m_2(x) = x^2$, and Log-Gamma Distributed $\theta^{\frac{1}{2}}U$ with $\alpha_1 = 1$, $\beta = 0.25$ ($k_0 = -0.25$)

			$a = 0.95$			$a = 0.99$			$a = 0.995$		
	m	n	B	S	R	B	S	R	B	S	R
q^s	m_1	500	0.013	0.199	0.199	−0.028	0.609	0.610	−0.047	10.060	10.061
$\hat{q}$	m_1	500	0.066	0.248	0.257	0.003	0.289	0.289	−0.039	0.467	0.469
q^h	m_1	500	−0.126	0.193	0.231	10.594	0.870	10.816	40.390	20.009	40.828
q^e	m_1	500	0.033	0.255	0.257	−0.076	0.556	0.562	−0.226	0.731	0.765
q^s	m_1	1000	0.006	0.144	0.144	−0.016	0.425	0.425	−0.028	0.725	0.725
$\hat{q}$	m_1	1000	0.043	0.201	0.206	0.018	0.264	0.264	−0.005	0.343	0.343
q^h	m_1	1000	−0.136	0.160	0.210	10.034	0.556	10.174	30.017	10.160	30.232
q^e	m_1	1000	0.024	0.203	0.204	−0.021	0.449	0.449	−0.122	0.625	0.636
q^s	m_2	500	0.006	0.204	0.204	−0.047	0.604	0.606	−0.074	10.033	10.036
$\hat{q}$	m_2	500	0.067	0.244	0.252	0.045	0.277	0.281	0.024	0.454	0.455
q^h	m_2	500	−0.132	0.189	0.231	10.777	0.923	20.002	40.877	20.218	50.357
q^e	m_2	500	0.033	0.247	0.249	−0.043	0.540	0.542	−0.173	0.723	0.744
q^s	m_2	1000	0.004	0.146	0.146	−0.022	0.433	0.433	−0.038	0.733	0.734
$\hat{q}$	m_2	1000	0.040	0.202	0.205	0.037	0.256	0.258	0.026	0.331	0.332
q^h	m_2	1000	−0.144	0.159	0.215	10.105	0.569	10.242	30.208	10.208	30.428
q^e	m_2	1000	0.023	0.202	0.203	−0.007	0.446	0.446	−0.084	0.644	0.650

TABLE 4

Bias (×0.1) (B), Standard Deviation (S), and Root Mean Squared Error (R) for Quantile Estimators with $m_1(x) = 3sin(3x)$, $m_2(x) = x^2$, and Log-Gamma Distributed $\theta^{\frac{1}{2}}U$ with $\alpha_1 = 1$, $\beta = 0.4$ ($k_0 = -0.4$)

			$a = 0.95$			$a = 0.99$			$a = 0.995$		
	m	n	B	S	R	B	S	R	B	S	R
q^s	m_1	500	0.010	0.159	0.159	0.012	0.640	0.640	0.049	10.247	10.248
$\hat{q}$	m_1	500	0.323	0.359	0.483	0.729	0.596	0.942	0.999	0.816	10.290
q^h	m_1	500	0.146	0.294	0.329	20.689	10.407	30.035	60.911	30.396	70.700
q^e	m_1	500	0.298	0.362	0.469	0.654	0.860	10.081	0.727	10.091	10.311
q^s	m_1	1000	0.005	0.114	0.114	−0.010	0.447	0.447	−0.011	0.848	0.848
$\hat{q}$	m_1	1000	0.268	0.304	0.406	0.632	0.564	0.847	0.859	0.711	10.115
q^h	m_1	1000	0.107	0.253	0.275	10.806	0.948	20.040	40.580	10.942	40.975
q^e	m_1	1000	0.253	0.303	0.395	0.596	0.742	0.952	0.721	10.044	10.268
q^s	m_2	500	0.012	0.158	0.159	0.001	0.644	0.644	0.020	10.242	10.242
$\hat{q}$	m_2	500	0.329	0.358	0.486	0.741	0.588	0.946	10.005	0.800	10.284
q^h	m_2	500	0.148	0.292	0.328	20.782	10.412	30.120	70.167	30.419	70.941
q^e	m_2	500	0.305	0.362	0.473	0.658	0.850	10.075	0.751	10.109	10.339
q^s	m_2	1000	0.004	0.112	0.112	−0.004	0.439	0.439	0.004	0.838	0.838
$\hat{q}$	m_2	1000	0.258	0.312	0.405	0.624	0.598	0.865	0.856	0.761	10.145
q^h	m_2	1000	0.096	0.260	0.277	10.814	0.984	20.063	40.636	20.000	50.049
q^e	m_2	1000	0.244	0.311	0.395	0.592	0.774	0.974	0.726	10.076	10.298

TABLE 5
Bias (×0.1) (B), Standard Deviation (S), and Root Mean Squared Error (R) for Quantile Estimators with $m_1(x) = 3sin(3x)$, $m_2(x) = x^2$, and Student-T Distributed $\theta^{\frac{1}{2}}U$ with $v = 3$ ($k_0 = -1/3$)

			$a = 0.95$			$a = 0.99$			$a = 0.995$		
	m	n	B	S	R	B	S	R	B	S	R
q^s	m_1	500	0.006	0.114	0.114	0.012	0.358	0.358	−0.018	0.622	0.622
$\hat{q}$	m_1	500	0.072	0.162	0.178	0.102	0.297	0.314	0.081	0.481	0.488
q^h	m_1	500	-0.010	0.132	0.133	0.478	0.425	0.640	10.077	0.758	10.317
q^e	m_1	500	0.043	0.164	0.169	0.008	0.389	0.389	−0.055	0.590	0.593
q^s	m_1	1000	0.003	0.082	0.082	0.028	0.261	0.263	0.015	0.454	0.455
$\hat{q}$	m_1	1000	0.047	0.140	0.148	0.089	0.248	0.263	0.085	0.370	0.380
q^h	m_1	1000	-0.023	0.119	0.121	0.322	0.317	0.452	0.771	0.529	0.935
q^e	m_1	1000	0.031	0.140	0.143	0.028	0.317	0.318	−0.011	0.484	0.484
q^s	m_2	500	0.004	0.114	0.114	0.004	0.365	0.365	−0.031	0.639	0.639
$\hat{q}$	m_2	500	0.068	0.162	0.176	0.092	0.292	0.307	0.067	0.479	0.483
q^h	m_2	500	-0.014	0.133	0.134	0.471	0.418	0.630	10.068	0.746	10.303
q^e	m_2	500	0.040	0.164	0.168	−0.002	0.386	0.386	−0.066	0.585	0.589
q^s	m_2	1000	0.002	0.082	0.082	0.018	0.255	0.256	−0.003	0.439	0.439
$\hat{q}$	m_2	1000	0.047	0.136	0.144	0.087	0.246	0.261	0.079	0.370	0.378
q^h	m_2	1000	-0.024	0.115	0.118	0.323	0.315	0.451	0.776	0.530	0.940
q^e	m_2	1000	0.031	0.136	0.140	0.023	0.321	0.322	−0.018	0.491	0.491

TABLE 6
Bias (×0.1) (B), Standard deviation (S), and Root Mean Squared Error (R) for quantile estimators with $m_1(x) = 3sin(3x)$, $m_2(x) = x^2$, and student-T distributed $\theta^{\frac{1}{2}}U$ with $v = 2.25$ ($k_0 = -1/2.25$)

			$a = 0.95$			$a = 0.99$			$a = 0.995$		
	m	n	B	S	R	B	S	R	B	S	R
q^s	m_1	500	0.006	0.088	0.088	0.017	0.353	0.353	0.004	0.675	0.674
$\hat{q}$	m_1	500	0.325	0.246	0.408	0.655	0.483	0.814	0.836	0.736	10.114
q^h	m_1	500	0.258	0.214	0.336	0.988	0.643	10.179	10.758	10.113	20.081
q^e	m_1	500	0.296	0.244	0.384	0.566	0.590	0.818	0.681	0.874	10.108
q^s	m_1	1000	0.003	0.065	0.066	0.019	0.251	0.251	0.009	0.473	0.473
$\hat{q}$	m_1	1000	0.278	0.215	0.352	0.592	0.423	0.728	0.770	0.603	0.978
q^h	m_1	1000	0.224	0.192	0.295	0.792	0.510	0.942	10.385	0.827	10.614
q^e	m_1	1000	0.264	0.213	0.339	0.533	0.496	0.728	0.664	0.743	0.996
q^s	m_2	500	0.007	0.089	0.090	0.015	0.356	0.357	−0.002	0.682	0.682
$\hat{q}$	m_2	500	0.326	0.238	0.404	0.658	0.473	0.810	0.842	0.736	10.119
q^h	m_2	500	0.259	0.208	0.332	0.993	0.627	10.174	10.768	10.094	20.079
q^e	m_2	500	0.297	0.237	0.380	0.570	0.586	0.817	0.696	0.874	10.117
q^s	m_2	1000	0.002	0.064	0.064	0.023	0.249	0.250	0.019	0.474	0.474
$\hat{q}$	m_2	1000	0.269	0.218	0.347	0.584	0.441	0.732	0.768	0.634	0.996
q^h	m_2	1000	0.216	0.196	0.292	0.781	0.524	0.940	10.374	0.850	10.615
q^e	m_2	1000	0.254	0.216	0.333	0.530	0.518	0.741	0.665	0.767	10.015

The choice of N could be important since the number of residuals exceeding the threshold is based on $\tilde{q}(a_n)$. We need to choose large $\tilde{q}(a_n)$ to reduce the bias from approximating the tail distribution with GPD, but we need to keep N large (small $\tilde{q}(a_n)$) 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 $n = 1{,}000$, $m(x) = 3sin(3x)$, and use a log-gamma distributed U. The bias and standard deviation of the estimators q^s, $\hat{q}$, q^h, and q^e are plotted against $N = 20, 25, \ldots, 200$ in Figures 2, and the root mean squared error (RMSE) are provided in Fig. 3. We notice that q^s is negatively biased and always has the smallest bias, while $\hat{q}$ and q^e carry relatively small and positive biases of similar magnitudes. These three estimators' biases are fairly stable across N. q^h's bias is positive and is influenced heavily by N, being small when N ranges from 20 to 50, largest with N greater than 50. q^s's standard deviation is always smallest across different N, and it is decreasing with larger N. The standard deviation of $\hat{q}$ decreases similarly with larger N, being slightly larger than that of q^s but always smaller than that of q^e except for N smaller than 25. The standard deviation of q^h is heavily influenced by N and increases with N. Though it starts to be smaller than that of $\hat{q}$ when N is smaller than 50, and smaller than q^e when N is smaller than 90, its magnitude quickly outgrows the others when N is larger than 90. The strong dependence of q^h's performance on N also exhibits in RMSE in Fig. 3. Between $N = 20$ and 50, q^h performs better than $\hat{q}$ and q^e, but its performance deteriorates quickly and it is outperformed by the others when N is larger than 50. q^s always carries the smallest and stable RMSE across different N. As expected, $\hat{q}$'s RMSE decreases with N due to its decreasing standard deviation, but when N is larger than 110, its RMSE is stable. For N greater than 50, $\hat{q}$ always dominate q^e, which did not utilize the extreme value theory, and q^h. The result indicates q^h's performance is sensitive to the choice of N, requiring a small N to control its bias and standard deviation, while q^s and $\hat{q}$ work well and outperform q^e 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 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., $q_{Y|X=x}(a) = m(x) + \theta^{1/2}q(a)$. 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 article 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 standardized 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 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, and $P(A)$ denotes the probability of event A from the probability space $(\Omega, \mathcal{F}, P)$.

Proof of Lemma 1. Given the results described in Subsection 3.1 and Taylor's Theorem, for $\lambda_1, \lambda_2 \in (0,1)$, we have

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

and

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

Note that $\widetilde{I}_{jN}$ is defined as I_{jN} with Z_i replaced by $\widetilde{Z}_i$ for $j = 1, \ldots, 6$.

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

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

since $N - N_s = O_p(N^{1/2})$ (see Lemma 2) and $\delta_N N^{1/2} \to \infty$. We first study I_{11n}, which can be written as

$$I_{11n} = \frac{1}{N}\sum_{i=1}^{n}\left(\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N} - \left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right)$$
$$\times \chi_{E_i \cap \widetilde{E}_i} + \frac{1}{N}\sum_{i=1}^{n}\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\chi_{\widetilde{E}_i - E_i}$$
$$= I_{111n} + I_{112n} \qquad \text{where } i \in \widetilde{E}_i - E_i \text{ implies that } i \in \widetilde{E}_i \text{ and } i \notin E_i.$$

By the mean value theorem, for some $\lambda_i \in (0, 1)$ and $Z_i^* = Z_i + \lambda_i(\widetilde{Z}_i - Z_i)$, we have

$$\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\chi_{E_i \cap \widetilde{E}_i} - \left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\chi_{E_i \cap \widetilde{E}_i}\right| = \frac{|k_0|/\sigma_N}{\left(1 - \frac{k_0 Z_i^*}{\sigma_N}\right)^2}|\widetilde{Z}_i - Z_i|\chi_{E_i \cap \widetilde{E}_i}$$
$$= \frac{1}{\left(1 + \frac{Z_i^*}{q(a_n)}\right)^2}\frac{|\widetilde{Z}_i - Z_i|}{q(a_n)}\chi_{E_i \cap \widetilde{E}_i},$$

where the last equality follows from $\sigma_N = -q(a_n)k_0$ and $k_0 < 0$. Note that $\frac{|\widetilde{Z}_i - Z_i|}{q(a_n)} \le \theta^{1/2}|\hat{\theta}^{-1/2} - \theta^{-1/2}|\frac{|U_i|}{q(a_n)} + |\hat{\theta}^{-1/2}|\frac{|\hat{m}(X_i) - m(X_i)|}{q_n(a_n)}\frac{q_n(a_n)}{q(a_n)} + \frac{|\tilde{q}(a_n) - q_n(a_n)|}{q(a_n)}$. From Lemma 2, $\frac{\tilde{q}(a_n) - q_n(a_n)}{q(a_n)} = O_p(N^{-1/2})$. Furthermore, provided that $N \propto n^{4/5-\delta}$, $h_{1n} \propto n^{-1/5}$, and given that $q_n(a_n) \to \infty$ as $n \to \infty$ and Eq. (4) we have $\frac{1}{q_n(a_n)} sup_{x \in G}|\hat{m}(x) - m(x)| = o_p(N^{-1/2})$. Eq. (5) and the fact that $\theta > 0$ imply that $\hat{\theta}^{-1/2} - \theta^{-1/2} = O_p(L_{1n})$ and $\hat{\theta}^{-1/2} = O_p(1)$, where $L_{1n} = \left(\frac{nh_{1n}}{log\, n}\right)^{-1/2} + h_{1n}^2$. Consequently, since $q_n(a_n)/q(a_n) = 1 + o_p(1)$, we have $\frac{\widetilde{Z}_i - Z_i}{q(a_n)} \le \frac{|U_i|}{q(a_n)}O_p(L_{1n}) + O_p(N^{-1/2})$. Also, since $U_i = Z_i + q_n(a_n)$ direct substitution gives

$$\frac{Z_i^*}{q(a_n)} = \frac{Z_i}{q(a_n)}(1 + \lambda_i\theta^{1/2}(\hat{\theta}^{-1/2} - \theta^{-1/2})) + \lambda_i\left(\theta^{1/2}(\hat{\theta}^{-1/2} - \theta^{-1/2})\frac{q_n(a_n)}{q(a_n)}\right.$$
$$\left. - \hat{\theta}^{-1/2}\frac{(\hat{m}(X_i) - m(X_i))}{q_n(a_n)}\frac{q_n(a_n)}{q(a_n)} - \frac{\tilde{q}(a_n) - q_n(a_n)}{q(a_n)}\right),$$

and since $\lambda_i < 1$ for all i, we have $\frac{Z_i^*}{q(a_n)} = \frac{Z_i}{q(a_n)}(1 + o_p(1)) + o_p(1)$ uniformly and

$$I_{111n} \leq \frac{1}{N}\sum_{i=1}^{n} \frac{1}{\left(1 + \frac{Z_i}{q(a_n)}(1 + o_p(1)) + o_p(1)\right)^2}\left(\left(\frac{Z_i}{q(a_n)} + O_p(1)\right)O_p(L_{1n}) + O_p(N^{-1/2})\right)\chi_{E_i \cap \widetilde{E}_i}.$$

Given that $Z_i > 0$ whenever $i \in \chi_{E_i \cap \widetilde{E}_i}$ and $\frac{x}{(1+x)^2} < 1$ for $x > 0$, we have $I_{111n} \leq (O_p(L_{1n}) + O_p(N^{-1/2}))\frac{1}{N}\sum_{i=1}^{n} \chi_{E_i \cap \widetilde{E}_i} = O_p(L_{1n}) + O_p(N^{-1/2})$ since $\frac{1}{N}\sum_{i=1}^{n} \chi_{E_i \cap \widetilde{E}_i} = O_p(1)$. We now consider I_{112n}, which can be written as $I_{112n} = \frac{1}{N}\sum_{i=1}^{n}\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\left(\chi_{\widetilde{E}_i} - \chi_{E_i}\right)\chi_{\widetilde{E}_i - E_i}$. For $\delta_1, \delta_2 > 0$, we define the events $A = \left\{\omega : \frac{|\widehat{U}_i - U_i|}{q_n(a_n)} < \delta_1\right\}$ 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 $C = \{\omega : \chi_{\widetilde{E}_i} - \chi_{E_i} = 0\}$. Hence, $\chi_{C^c} \leq \chi_{A^c} + \chi_{B^c}$ and

$$\begin{aligned} I_{112n} &\leq \frac{1}{N}\sum_{i=1}^{n}\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right|\chi_{A^c}\chi_{\widetilde{E}_i - E_i} + \frac{1}{N}\sum_{i=1}^{n}\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right|\chi_{B^c}\chi_{\widetilde{E}_i - E_i} \\ &= I_{1121n} + I_{1122n}. \end{aligned}$$

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

$$\begin{aligned} I_{1121n} &< \frac{1}{N}\sum_{i=1}^{n}\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right|\frac{|\widehat{U}_i - U_i|}{\delta_1 q_n(a_n)}\chi_{\widetilde{E}_i - E_i} \qquad \text{and} \\ I_{1122n} &< \frac{1}{N}\sum_{i=1}^{n}\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right|\frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{\widetilde{E}_i - E_i}. \end{aligned}$$

Since $k_0 < 0$, $\sigma_N > 0$, and $\widetilde{Z}_i > 0$ whenever $i \in \widetilde{E}_i - E_i$, we have that $\left|\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right| < C$. From Lemma 2 we can immediately conclude that $I_{1122n} \leq \frac{1}{\delta_2}O_p(N^{-1/2})\frac{1}{N}\sum_{i=1}^{n}\chi_{\widetilde{E}_i - E_i}$, and since $\frac{1}{N}\sum_{i=1}^{n}\chi_{\widetilde{E}_i - E_i} = O_p(1)$, we have $I_{1122n} = O_p(N^{-1/2})$. Now, $\frac{|\widehat{U}_i - U_i|}{q_n(a_n)} \leq \frac{|U_i|}{q_n(a_n)}O_p(L_{1n})$ $+o_p(N^{-1/2})$, and therefore, $I_{1121n} \leq O_p(L_{1n})\frac{1}{N\delta_1}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i - E_i} + o_p(N^{-1/2})\frac{1}{N\delta_1}\sum_{i=1}^{n}\chi_{\widetilde{E}_i - E_i}$. The second term following the inequality is $o_p(N^{-1/2})$ since $\frac{1}{N\delta_1}\sum_{i=1}^{n}\chi_{\widetilde{E}_i - E_i} = O_p(1)$. For the first term, note that $\left|\frac{U_i}{q_n(a_n)}\right| = \left|\frac{Z_i}{q_n(a_n)} + 1\right|$ and for $i \in \widetilde{E}_i - E_i$, $U_i \leq q_n(a_n)$ and consequently if $U_i > 0$, we have $|\frac{Z_i}{q_n(a_n)}| < \frac{|U_i|}{q_n(a_n)} + 1 < 2$. If $U_i \leq 0$ for $i \in \widetilde{E}_i - E_i$, then $\widehat{U}_i = \frac{m(X_i) - \hat{m}(X_i)}{\hat{\theta}^{1/2}} + \frac{\theta^{1/2}}{\hat{\theta}^{1/2}}U_i \geq 0$. Since, $\sup_x \frac{|m(X_i) - \hat{m}(X_i)|}{\hat{\theta}^{1/2}} = o_p(1)$ and $\frac{\theta^{1/2}}{\hat{\theta}^{1/2}} = 1 + o_p(1)$, it must be that $U_i > 0$ with probability approaching 1. Consequently, for N sufficiently large and $i \in \widetilde{E}_i - E_i$, we have $|\frac{Z_i}{q_n(a_n)}| < 2$ and $I_{1121n} = O_p(L_{1n}) + o_p(N^{-1/2})$. Combining the orders of I_{1121n} and I_{1122n}, we have $I_{112n} = O_p(L_{1n}) + O_p(N^{-1/2})$ and $I_{11n} = O_p(L_{1n} + O_p(N^{-1/2})$.

We now consider I_{12n} and note that

$$I_{12n} \leq \frac{1}{N\delta_1}\sum_{i=1}^{n}\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right|\frac{|\widehat{U}_i-U_i|}{q_n(a_n)}\chi_{E_i}$$
$$+\frac{1}{N\delta_2}\sum_{i=1}^{n}\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right|\frac{|\tilde{q}(a_n)-q_n(a_n)|}{q_n(a_n)}\chi_{E_i}.$$

By Lemma 2 and the fact that $\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right| < C$, the second term following the inequality is $O_p(N^{-1/2})$ given that $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i}=O_p(1)$. Again, using $\frac{|\widehat{U}_i-U_i|}{q_n(a_n)} = \frac{|U_i|}{q_n(a_n)}O_p(L_{1n})+o_p(N^{-1/2})$ we have that the first term after the inequality is bounded by $O_p(L_{1n})\frac{1}{N\delta_1}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{E_i}+o_p(N^{-1/2})\frac{1}{N\delta_1}\sum_{i=1}^{n}\chi_{E_i}$. Since $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i}=O_p(1)$, we need only investigate the order of $\frac{1}{N}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{E_i}$. Note that $\frac{1}{N}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{E_i} \leq C\frac{1}{N}\sum_{i=1}^{n}\frac{Z_i}{q(a_n)}\chi_{E_i}+O_p(1)$ since $\frac{q(a_n)}{q_n(a_n)}=O_p(1)$, $Z_i>0$ whenever $i \in E_i$ and $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i}=O_p(1)$. Furthermore, $E\left(\frac{Z_i}{q(a_n)}\chi_{E_i}\right)=E\left(\frac{Z_i}{q(a_n)}\chi_{E_i}|\chi_{E_i}=1\right)P(\chi_{E_i}=1)=\left(\frac{1}{\alpha-1}+O(\phi(q(a_n)))\right)O_p(\frac{N}{n})$ for $\alpha>1$. Thus, $\frac{1}{N}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{E_i}=O_p(1)$, and we conclude that $I_{12n}=O_p(L_{1n})+O_p(N^{-1/2})$. Combining the orders of I_{11n} and I_{12n} we have $\widetilde{I}_{1N}-I_{1N}=\frac{1}{\delta_N}(k_0^{-1}-1)(O_p(L_{1n})+O_p(N^{-1/2}))$. Since $\delta_N N^{1/2}\to\infty$ and $\sqrt{N}L_{1n}\to 0$ as $n\to\infty$ whenever $N \propto n^{4/5-\delta}$ for $0<\delta$ and $h_{1n}\propto n^{-1/5}$, we have $\widetilde{I}_{1N}-I_{1N}=o_p(1)$.

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

$$\widetilde{I}_{4N}-I_{4N}=\frac{1}{\delta_N}\left(\frac{1}{N}\sum_{i=1}^{n}\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0\widetilde{Z}_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0\widetilde{Z}_i}{\sigma_N}\right)^{-1}\frac{k_0\widetilde{Z}_i}{\sigma_N}\right.\right.$$
$$\left.-\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right)\chi_{E_i}\right)\chi_{\widetilde{E}_i}$$
$$\left.+\frac{1}{N}\sum_{i=1}^{n}\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right)\right)\chi_{E_i}(\chi_{\widetilde{E}_i}-\chi_{E_i})$$
$$=\frac{1}{\delta_N}(I_{41n}+I_{42n}).$$

First, note that

$$I_{41n}=-\frac{1}{k_0^2}\left(\frac{1}{N}\sum_{i=1}^{n}\left(log\left(1-\frac{k_0\widetilde{Z}_i}{\sigma_N}\right)-log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right)\chi_{E_i\cap\widetilde{E}_i}+\frac{1}{N}\sum_{i=1}^{n}log\left(1-\frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\chi_{\widetilde{E}_i-E_i}\right)$$
$$+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)I_{11n}=-\frac{1}{k_0^2}(I_{411n}+I_{412n})+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)I_{11n}.$$

Since we have already established that $I_{11n} = O_p(L_{1n}) + O_p(N^{-1/2})$, it suffices to investigate the order of I_{411n} and I_{412n}. By the mean value theorem for some $\lambda_i \in (0, 1)$ and $Z_i^* = Z_i + \lambda_i(\widetilde{Z}_i - Z_i)$, we have

$$\left| log\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right) - log\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)\right| \chi_{E_i \cap \widetilde{E}_i} = \left(1 + \frac{Z_i^*}{q(a_n)}\right)^{-1} \frac{|\widetilde{Z}_i - Z_i|}{q(a_n)} \chi_{E_i \cap \widetilde{E}_i},$$

where $\frac{\widetilde{Z}_i - Z_i}{q(a_n)} \le \frac{|U_i|}{q(a_n)} O_p(L_{1n}) + O_p(N^{-1/2})$ and $\frac{Z_i^*}{q(a_n)} = \frac{Z_i}{q(a_n)}(1 + o_p(1)) + o_p(1)$ uniformly. Consequently,

$$\begin{aligned} I_{411n} < {} & \frac{1}{N}\sum_{i=1}^{n}\left(1 + \frac{Z_i}{q(a_n)}(1 + o_p(1)) + o_p(1)\right)^{-1} \\ & \times \left(\left(\frac{Z_i}{q(a_n)} + O_p(1)\right) O_p(L_{1n}) + O_p(N^{-1/2})\right) \chi_{E_i \cap \widetilde{E}_i}, \end{aligned}$$

Given that $Z_i > 0$ whenever $i \in \chi_{E_i \cap \widetilde{E}_i}$ and $\frac{x}{(1+x)} < 1$ for $x > 0$, we have $I_{411n} \le (O_p(L_{1n}) + O_p(N^{-1/2}))\frac{1}{N}\sum_{i=1}^{n} \chi_{E_i \cap \widetilde{E}_i} = O_p(L_{1n}) + O_p(N^{-1/2})$ since $\frac{1}{N}\sum_{i=1}^{n} \chi_{E_i \cap \widetilde{E}_i} = O_p(1)$. Given that $\chi_{\widetilde{E}_i - E_i} = \chi_{\widetilde{E}_i}(\chi_{\widetilde{E}_i} - \chi_{E_i})$ and $\chi_A = \chi_A^2$, we have for $\delta_1, \delta_2 > 0$

$$\begin{aligned} I_{412n} & \le \frac{1}{N}\sum_{i=1}^{n} log\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\left(\frac{|\widehat{U}_i - U_i|}{\delta_1 q_n(a_n)} + \frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\right)\chi_{\widetilde{E}_i - E_i} \\ & \le O_p(L_{1n})\frac{1}{\delta_1 N}\sum_{i=1}^{n} log\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i - E_i} + O_p(N^{-1/2})\frac{1}{N}\sum_{i=1}^{n} log\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\chi_{\widetilde{E}_i - E_i}. \end{aligned}$$

Note that by the mean value theorem, $\left|log\left(1 - \frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\right| = \left|\left(1 - \frac{k_0 Z_i^*}{\sigma_N}\right)^{-1} \frac{-k_0\widetilde{Z}_i}{\sigma_N}\right| < \frac{-k_0}{\sigma_N}\widetilde{Z}_i$ since $\widetilde{Z}_i > 0$ whenever $i \in \widetilde{E}_i - E_i$, $Z_i^* = \lambda_i\widetilde{Z}_i > 0$ for some $0 < \lambda_i < 1$. Hence,

$$I_{412n} \le O_p(L_{1n})\frac{1}{\delta_1 N}\sum_{i=1}^{n}\frac{-k_0}{\sigma_N}\widetilde{Z}_i\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i - E_i} + O_p(N^{-1/2})\frac{1}{N}\sum_{i=1}^{n}\frac{-k_0}{\sigma_N}\widetilde{Z}_i\chi_{\widetilde{E}_i - E_i}.$$

Since $\frac{q(a_n)}{q_n(a_n} = 1 + o_p(1)$, $\frac{|U_i|}{q_n(a_n)} \le \frac{|Z_i|}{q_n(a_n)} + 1$, and $\frac{\widetilde{Z}_i}{q(a_n)} \le \frac{|Z_i|}{q_n(a_n)}(1 + O_p(L_{1n})) + O_p(N^{-1/2})$, we have for the first term following the inequality

$$\frac{1}{N}\sum_{i=1}^{n}\frac{-k_0}{\sigma_N}\widetilde{Z}_i\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i - E_i} \le \frac{1}{N}\sum_{i=1}^{n}\left(\frac{|Z_i|}{q_n(a_n)}(1 + O_p(L_{1n})) + O_p(N^{-1/2})\right)\left(1 + \frac{|Z_i|}{q_n(a_n)}\right)\chi_{\widetilde{E}_i - E_i}.$$

Given that $\frac{|Z_i|}{q_n(a_n)} < 2$ whenever $i \in \widetilde{E}_i - E_i$ for N sufficiently large, we have that $\frac{1}{N}\sum_{i=1}^{n}\frac{|Z_i|^2}{q_n(a_n)^2}\chi_{\widetilde{E}_i - E_i} \le \frac{4}{N}\sum_{i=1}^{n}\chi_{\widetilde{E}_i - E_i} = O_p(1)$. The second term following the inequality can be bounded using the similar arguments, and we obtain $I_{412n} = O_p(L_{1n}) + O_p(N^{-1/2})$. We now investigate the order of I_{42n}. Note that

$$\begin{aligned}
I_{42n} &< \frac{1}{N}\sum_{i=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right|\frac{|\widehat{U}_i-U_i|}{\delta_1 q_n(a_n)}\chi_{E_i}\\
&\quad+\frac{1}{N}\sum_{i=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_i}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right|\frac{|\tilde{q}(a_n)-q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_{E_i}\\
&= I_{421n}+I_{422n}.
\end{aligned}$$

Since $\frac{|\widehat{U}_i-U_i|}{q_n(a_n)} \le \frac{|U_i|}{q_n(a_n)}O_p(L_{1n}) + o_p(N^{-1/2})$, we write

$$\begin{aligned}
I_{421n} \le{}& O_p(L_{1n})\frac{1}{k_0^2\delta_1 N}\sum_{i=1}^{n}\left|log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right|\frac{|U_i|}{q_n(a_n)}\chi_{E_i}\\
&+O_p(L_{1n})\left|\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\right|\frac{1}{\delta_1 N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_i}{\sigma_N}\right|\frac{|U_i|}{q_n(a_n)}\chi_{E_i}\\
&+o_p(N^{-1/2})\frac{1}{k_0^2\delta_1 N}\sum_{i=1}^{n}\left|log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right|\chi_{E_i}\\
&+o_p(N^{-1/2})\left|\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\right|\frac{1}{\delta_1 N}\sum_{i=1}^{n}\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_i}{\sigma_N}\right|\chi_{E_i}.
\end{aligned}$$

Since $k_0 < 0$ and $Z_i > 0$ for all $i \in E_i$, we have that $\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_i}{\sigma_N}\right| < C$, and the second and fourth terms following the inequality are $O_p(L_{1n})$ and $o_p(N^{-1/2})$ since $\frac{1}{N}\sum_{i=1}^{n}\frac{|U_i|}{q_n(a_n)}\chi_{E_i} = O_p(1)$ (see the order of I_{12n}) and $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i} = O_p(1)$.

Note that for all $i \in E_i$ and given that $\frac{q(a_n)}{q_n(a_n)} = 1 + o_p(1)$ we have that $\frac{1}{N}\sum_{i=1}^{n}\left|log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right|\frac{|U_i|}{q_n(a_n)}\chi_{E_i} \le \frac{1}{N}\sum_{i=1}^{n} log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\left(1-\frac{k_0Z_i}{\sigma_N}O_p(1)\right)\chi_{E_i}$. In addition, if $\alpha > 1$, $E\left(log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right) = \frac{\alpha}{(\alpha-1)^2} + O(\phi(q(a_n))) = O(1)$ and, consequently, $\frac{1}{N}\sum_{i=1}^{n} log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\left(1-\frac{k_0Z_i}{\sigma_N}\right)\chi_{E_i} = O_p(1)$, which establishes that the first term after the inequality is $O_p(L_{1n})$. Similarly, for $\alpha > 1$, $E\left(log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\right) = \frac{1}{\alpha} + O(\phi(q(a_n))) = O(1)$ which establishes that $\frac{1}{N}\sum_{i=1}^{n} log\left(1-\frac{k_0Z_i}{\sigma_N}\right)\chi_{E_i} = O_p(1)$ and the order of the third term to be $O_p(N^{-1/2})$.

We now examine the order of I_{422n}. Given $\frac{\tilde{q}_n(a_n)-q_n(a_n)}{q_n(a_n)} = O_p(N^{-1/2})$ and $\left|\left(1-\frac{k_0Z_i}{\sigma_N}\right)^{-1}\frac{k_0Z_i}{\sigma_N}\right| < C$, we write $I_{422n} \le \frac{1}{\delta_2}O_p(N^{-1/2})(\frac{1}{k_0^2\delta_2}\frac{1}{N}\sum_{i=1}^{n}\left|log(1-\frac{k_0Z_i}{\sigma_N})\right|\chi_{E_i} + C\left|\frac{1}{k_0}(1-\frac{1}{k_0})\right|\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i})$. Since we have already established that $\frac{1}{N}\sum_{i=1}^{n}$

$\left|log\left(1-\frac{k_0 Z_i}{\sigma_N}\right)\right|\chi_{E_i} = O_p(1)$ and $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i} = O_p(1)$, we conclude that $I_{422n} = O_p(N^{-1/2})$. Combining all orders obtained, we have that $I_{41n} + I_{42n} = O_p(L_{1n}) + O_p(N^{-1/2})$ and, consequently, $\widetilde{I}_{4N} - I_{4N} = o_p(1)$, since $\delta_N N^{1/2} \to \infty$ as $n \to \infty$.

We now investigate the order of $\widetilde{I}_{2N} - I_{2N}$. 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

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

Hence, it suffices to examine

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

for $l = 1, 2$. First, note that $I_{nl1} = I_{nl11} + I_{nl12}$, where

$$I_{nl11} = \frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{l} - \left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\chi_{E_i\cap\widetilde{E}_i} \quad \text{and}$$

$$I_{nl12} = \frac{1}{N}\sum_{i=1}^{n}\left(1-\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{l}\chi_{\widetilde{E}_i - E_i}.$$

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

$$I_{nl11} \le l\frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}q_n(a_n)\right|\left(O_p(L_{1n})\left(\frac{Z_i}{q_n(a_n)}+1\right) + O_p(N^{-1/2})\right)\chi_{E_i\cap\widetilde{E}_i}. \tag{12}$$

Since $q(a_n) = -\sigma_N/k_0$ and $\frac{q(a_n)}{q_n(a_n)} = O_p(1)$, we have

$$\sup_{S_T}\frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}\frac{Z_i}{q_n(a_n)}q_n(a_n)\right|\chi_{E_i\cap\widetilde{E}_i} \le O_p(1)\sup_{S_T}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\dot{\sigma}_N}\right|$$
$$\times\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i\cap\widetilde{E}_i}\sup_{S_T}\left|\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{Z_i}{q_n(a_n)}\right|.$$

Now, given that $\delta_N \to 0$ we have for N sufficiently large $sup_{S_T}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\dot{\sigma}_N}\right| < C$ and $sup_{S_T}\left|\left(\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{l-1}\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-l}\right| < C$. Hence, to establish the order of the left-hand side of the inequality, it suffices to obtain the order of $\upsilon_n = \frac{1}{N}\sum_{i=1}^n \chi_{E_i\cap\widetilde{E}_i} sup_{S_T}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\frac{Z_i}{q_n(a_n)}\right|$. Note that

$$\begin{aligned}
\upsilon_n &\le C\frac{1}{N}\sum_{i=1}^n \chi_{E_i\cap\widetilde{E}_i} \sup_{S_T}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right| \sup_{S_T}\left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)} \\
&\le C\frac{1}{N}\sum_{i=1}^n \chi_{E_i\cap\widetilde{E}_i} \sup_{S_T}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right| \text{ since } \sup_{S_T}\left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)} < C \\
&\le C\frac{1}{N}\sum_{i=1}^n \chi_{E_i\cap\widetilde{E}_i} \sup_{S_T}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}q(a_n)\left(\frac{Z_i}{q(a_n)}(1+o_p(1))+o_p(1)\right)\right)^{-1}\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right| = O_p(1)
\end{aligned}$$

since $\frac{Z_i^*}{q(a_n)} = \left(\frac{Z_i}{q(a_n)}(1+o_p(1))+o_p(1)\right)$, $sup_{S_T}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}q(a_n)\left(\frac{Z_i}{q(a_n)}(1+o_p(1))+o_p(1)\right)\right)^{-1}\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right| < C$, and $\frac{1}{N}\sum_{i=1}^n \chi_{E_i\cap\widetilde{E}_i} = O_p(1)$. Consequently, $I_{nl11} = O_p(L_{1n}) + O_p(N^{-1/2})$ as all remaining terms in (12) are $O_p(1)$. Now, we write

$$I_{nl12} \le \frac{1}{N}\sum_{i=1}^n \left|\left(1-\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{l}\right| \chi_{\widetilde{E}_i-E_i}\left(O_p(L_{1n})\frac{1}{\delta_1}\left(\frac{|U_i|}{q_n(a_n)}+\frac{1}{q_n(a_n)}\right)+\frac{1}{\delta_2}O_p(N^{-1/2})\right)$$

and obtain the order of $v_n = \frac{1}{N}\sum_{i=1}^n \left|\left(1-\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{l}\right|\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i-E_i}$. Note that

$$v_n \le \frac{1}{N}\sum_{i=1}^n \sup_{S_T}\left|\left(1-\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)^{l}\right|\frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i-E_i} \le C\frac{1}{N}\sum_{i=1}^n \frac{|U_i|}{q_n(a_n)}\chi_{\widetilde{E}_i-E_i} = O_p(1)$$

from the study of the order of I_{1121n}. Consequently, $I_{nl12} = O_p(L_{1n}) + O_p(N^{-1/2})$ which combined with the order of I_{nl11} gives $I_{nl1} = O_p(L_{1n}) + O_p(N^{-1/2})$. Now, as argued previously, we can write

$$I_{nl2} \le \frac{1}{N}\sum_{i=1}^n \left(\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\left(O_p(L_{1n})\frac{1}{\delta_1}\left(\frac{|U_i|}{q_n(a_n)}+\frac{1}{q_n(a_n)}\right)+\frac{1}{\delta_2}O_p(N^{-1/2})\right)\chi_{E_i}. \tag{13}$$

Letting $T_n = \frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\frac{|U_i|}{q_n(a_n)}\chi_{E_i}$, we note that

$$T_n \leq \frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)}+1\right)\chi_{E_i}$$
$$\leq \frac{1}{N}\sum_{i=1}^{n}\left(\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right)\left(-\frac{\dot{k}Z_i}{\dot{\sigma}_N}O_p(1)+1\right)\chi_{E_i},\ \text{since}\ \left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)} = O_p(1),$$

and given that $\delta_N \to 0$, for N sufficiently large, we have $\dot{k} < 0$, $\dot{\sigma}_N > 0$ and $\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right| < C$. Thus, it suffices to establish the order of $\frac{1}{N}\sum_{i=1}^{n} -\frac{\dot{k}Z_i}{\dot{\sigma}_N}\chi_{E_i} \leq C\frac{1}{N}\sum_{i=1}^{n}\frac{Z_i}{q(a_n)}\chi_{E_i}$, which is $O_p(1)$ from the study of I_{12n}. Consequently, $T_n = O_p(1)$ uniformly on S_T and $I_{nl2} = O_p(L_{1n}) + O_p(N^{-1/2})$, since all other terms in inequality (13) are $O_p(1)$ given $\left|\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{l}\right| < C$ and the fact that $\frac{1}{N}\sum_{i=1}^{n}\chi_{E_i} = O_p(1)$. Combining the orders of I_{nl1} and I_{nl2}, we conclude that $\widetilde{I}_{2N} - I_{2N} = o_p(1)$ uniformly on S_T. Now, note that $\widetilde{I}_{3N} - I_{3N} = \widetilde{I}_{5N} - I_{5N}$ and

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

and using the same arguments as in the case of $\widetilde{I}_{2N} - I_{2N}$, we have $\widetilde{I}_{3N} - I_{3N} = o_p(1)$ and $\widetilde{I}_{5N} - I_{5N} = o_p(1)$ uniformly on S_T.

Lastly, we investigate the order of $\widetilde{I}_{6N} - I_{6N}$, which can be written as

$$\widetilde{I}_{6N} - I_{6N} = \frac{1}{N}\sum_{j=1}^{N_s}\left(\frac{2}{\dot{k}^3}log\left(1-\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)+\frac{1}{\dot{k}}\left(1-\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)+\frac{1}{\dot{k}^3}\left(1-\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)\right.$$
$$\left.-\frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right)^{2}\right)$$
$$-\frac{1}{N}\sum_{j=1}^{N}\left(\frac{2}{\dot{k}^3}log\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)+\frac{1}{\dot{k}}\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)+\frac{1}{\dot{k}^3}\left(1-\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right.$$

$$- \frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}} - 1\right)\left(1 - \frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)^{2}\Bigg)$$

$$= \frac{2}{\dot{k}^3}\left(\frac{1}{N}\sum_{j=1}^{N_s} log\left(1 - \frac{\dot{k}\widetilde{Z}_j}{\dot{\sigma}_N}\right) - \frac{1}{N}\sum_{j=1}^{N} log\left(1 - \frac{\dot{k}Z_j}{\dot{\sigma}_N}\right)\right) + o_p(1) \text{ uniformly in } S_T. \quad (14)$$

The last equality follows form the arguments used above when investigating the order of $\widetilde{I}_{2N} - I_{2N}$. The first term in Eq. (14) can be written as (excluding the constant $2/\dot{k}^3$) $I_{61n} + I_{62n}$, where

$$I_{61n} = \frac{1}{N}\sum_{i=1}^{n}\left(log\left(1 - \frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right) - log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\chi_{E_i}\right)\chi_{\widetilde{E}_i} \quad \text{and}$$

$$I_{62n} = \frac{1}{N}\sum_{i=1}^{n} log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\chi_{E_i}(\chi_{\widetilde{E}_i} - \chi_{E_i}).$$

Now,

$$I_{62n} \leq \frac{1}{N}\sum_{i=1}^{n}\left|log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\chi_{E_i}\left(\frac{O_p(L_{1n})}{\delta_1}\left(\frac{|U_i|}{q_n(a_n)} + 1\right) + \frac{1}{\delta_2}O_p(N^{-1/2})\right),$$

and we consider $\frac{1}{N}\sum_{i=1}^{n}\left|log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\frac{|U_i|}{q_n(a_n)}\chi_{E_i}$. Note that $Z_i > 0$ whenever $i \in E_i$ and as $N \to \infty$ $\delta_N \to 0$, $\dot{k} \to k_0$ and $\frac{\dot{\sigma}_N}{\sigma_N} \to 1$. Consequently, given that $\frac{q(a_n)}{q_n(a_n)} = O_p(1)$, we have

$$\frac{1}{N}\sum_{i=1}^{n}\left|log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\frac{|U_i|}{q_n(a_n)}\chi_{E_i}$$

$$\leq \frac{1}{N}\sum_{i=1}^{n}\left|log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right|\left(1 - \frac{k_0 Z_i}{\sigma_N}O_p(1)\right)\chi_{E_i}$$

$$= \frac{1}{N}\sum_{i=1}^{n}\left|log\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)\right|\left(1 - \frac{k_0 Z_i}{\sigma_N}O_p(1)\right)\chi_{E_i} + o_p(1) = O_p(1),$$

where the last equality follows from the order of I_{421n}. Hence, $I_{62n} = O_p(L_{1n}) + O_p(N^{-1/2})$ uniformly on S_T. We write $I_{61n} = I_{611n} + I_{612n}$, where $I_{611n} = \frac{1}{N}\sum_{i=1}^{n}\left(log\left(1 - \frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right) - log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right)\chi_{E_i \cap \widetilde{E}_i}$, and $I_{612n} = \frac{1}{N}\sum_{i=1}^{n} log\left(1 - \frac{\dot{k}\widetilde{Z}_i}{\dot{\sigma}_N}\right)\chi_{E_i - \widetilde{E}_i}$. Then,

$$I_{611n} = \frac{1}{N}\sum_{i=1}^{n}\left(1 - \frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}}{\dot{\sigma}_N}(\widetilde{Z}_i - Z_i)\chi_{E_i \cap \widetilde{E}_i}$$

$$\leq \frac{1}{N}\sum_{i=1}^{n}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\right|\left|\frac{\dot{k}}{\dot{\sigma}_N}q_n(a_n)\right|\left(O_p(L_{1n})\left(\frac{Z_i}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_{E_i\cap\widetilde{E}_i}$$

$$\leq \sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}q_n(a_n)\right|\frac{1}{N}\sum_{i=1}^{n}\sup_{S_T}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\right|\left(O_p(L_{1n})\left(\frac{Z_i}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_{E_i\cap\widetilde{E}_i}$$

$$= O_p(1)\frac{1}{N}\sum_{i=1}^{n}\sup_{S_T}\left|\left(1-\frac{\dot{k}Z_i^*}{\dot{\sigma}_N}\right)^{-1}\right|\left(O_p(L_{1n})\left(\frac{Z_i}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_{E_i\cap\widetilde{E}_i}$$

$= O_p(L_{1n}) + O_p(N^{-1/2})$ uniformly on S_T given the order of v_n.

Since $\widetilde{Z}_i > 0$ whenever $i \in \widetilde{E}_i - E_i$ and since as $N \to \infty$ $\delta_N \to 0$, $\dot{k} \to k_0$, and $\frac{\dot{\sigma}_N}{\sigma_N} \to 1$, we have $I_{612n} = \frac{1}{N}\sum_{i=1}^{n} log\left(1-\frac{k_0\widetilde{Z}_i}{\sigma_N}\right)\chi_{\widetilde{E}_i-E_i} + o_p(1)$. From the order of I_{412n}, we have $I_{612n} = O_p(L_{1n}) + O_p(N^{-1/2})$ and, consequently, $I_{61n} = O_p(L_{1n}) + O_p(N^{-1/2})$, which combined with the order of I_{62n} gives $\widetilde{I}_{6N} - I_{6N} = o_p(1)$ uniformly on S_T.

Lemma 2. *Under Assumptions A1–A5 and conditions FR1' and FR2, if $\alpha \geq 1$, we have*

$$\sqrt{N}\left(\frac{\tilde{q}(a_n)-q_n(a_n)}{q(a_n)}\right) = O_p(1), \qquad \text{where } a_n = 1-\frac{N}{n}.$$

Proof. We write

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

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 $q^*(a_n) = q(a_n) + \lambda\sigma_n z = q(a_n)(1+\lambda z N^{-1/2})$ 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, given FR1' and by Proposition 1.15 in Resnick (1987) we have $lim_{n\to\infty}\frac{q(a_n)f(q^*(a_n))}{1-F(q^*(a_n))} = -\frac{1}{k_0}$, and 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

$$\frac{\sqrt{n(1-F(y_n))}}{1-F(y_n)}(F_n(y_n) - F(y_n)) = \sum_{i=1}^{n} Z_{in} \overset{d}{\rightarrow} N(0,1), \tag{15}$$

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

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

$$P(T_{1n} \le z) = P\left(\frac{nk_0}{\sqrt{N}}(F(y_n) - a_n) \le -\frac{nk_0}{\sqrt{N}}(\widetilde{F}(y_n) - F(y_n))\right), \tag{16}$$

and we establish that $\frac{n}{\sqrt{N}}(\widetilde{F}(y_n) - F(y_n)) \overset{d}{\rightarrow} N(0,1)$. We start by noting that for some $\lambda_i \in (0,1)$

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

where $U_i^* = \lambda_i U_i + (1-\lambda_i)\widehat{U}_i$. Therefore, $\frac{n}{\sqrt{N}}(\widetilde{F}(y_n) - F(y_n)) = \frac{n}{\sqrt{N}}((Q_{1n} - F(y_n)) + Q_{2n} + Q_{3n})$. Letting $L_{1n} = \left(\frac{log\, n}{nh_{1n}}\right)^{1/2} + h_{1n}^2$, we observe that under Assumptions A1–A5, $\hat{\theta} - \theta = O_p(L_{1n})$, and consequently, given that $\theta > 0$, $\hat{\theta}^{-1/2} - \theta^{1/2} = O_p(L_{1n})$. Since $\widehat{U}_i - U_i = \theta^{1/2}(\hat{\theta}^{-1/2} - \theta^{-1/2})U_i - \hat{\theta}^{-1/2}(\hat{m}(X_i) - m(X_i))$, we have

$$|\widehat{U}_i - U_i| \le (1 + \theta^{1/2}|U_i|)\ O_p(L_{1n}) \qquad \text{and} \qquad (\widehat{U}_i - U_i)^2 \le (1 + 2\theta|U_i| + \theta U_i^2)\ O_p\left(L_{1n}^2\right). \tag{17}$$

We now examine the orders of Q_{jn}, $j = 2, 3$. Given A1 and (17), we have $Q_{3n} \le O_p\left(\frac{L_{1n}^2}{h_{2n}}\right)\sum_{j=1}^{3} Q_{3jn}$, where $Q_{3jn} = \frac{1}{2nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n - U_i^*}{h_{2n}}\right)\right||U_i|^{j-1}$. Using Taylor's Theorem, we can write for some $\lambda_i \in (0,1)$ and $U_i^{**} = \lambda_i U_i + (1-\lambda_i)U_i^*$ that

$$Q_{31n} \le \frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n - U_i}{h_{2n}}\right)\right| + \frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right||\widehat{U}_i - U_i|.$$

Since $|K_2^{(1)}(x)| < C$ by (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 (17)

$$\begin{aligned}\frac{1}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| |\widehat{U}_i - U_i| &\le \frac{1}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| \left(1 + \theta^{1/2}|U_i|\right)\ O_p(L_{1n})\\ &\le \frac{1}{h_{2n}^2} O_p(L_{1n}) \left(1 + \theta^{1/2}\ \frac{1}{n}\sum_{i=1}^n |U_i|\right) = O_p\left(\frac{L_{1n}}{h_{2n}^2}\right),\end{aligned}$$

where the last inequality follows from the fact that $|K_2^{(2)}| < C$ by (A1) and the last equality follows from A4 by using Kolmogorov's law of large numbers to obtain $\frac{1}{n}\sum_{i=1}^n |U_i| = O_p(1)$. Consequently, $Q_{31n} = o_p(1) + O_p\left(\frac{L_{1n}}{h_{2n}^2}\right)$ which is bounded in probability provided $h_{1n} = O(h_{2n})$ 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 $\delta > 0$. Using Taylor's Theorem again, we can write for some $\lambda_i \in (0,1)$ and $U_i^{**} = \lambda_i U_i + (1-\lambda_i)U_i^*$ that

$$\begin{aligned}Q_{32n} &\le \frac{\theta^{1/2}}{nh_{2n}}\sum_{i=1}^n \left|K_2^{(1)}\left(\frac{y_n - U_i}{h_{2n}}\right)\right| |U_i| + \frac{\theta^{1/2}}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| |U_i||\widehat{U}_i - U_i|\\ &= Q_{321n} + Q_{322n}. \end{aligned} \tag{18}$$

We note that Q_{321n} is $o_p(1)$ since

$$\begin{aligned}E\left(Q_{321n}\right) &= \theta^{1/2}\int |y_n - \phi h_{2n}||K_2^{(1)}(\phi)|f(y_n - \phi h_{2n})d\phi\\ &= \theta^{1/2}(1-F(y_n))\int \frac{|y_n - \phi h_{2n}|f(y_n - \phi h_{2n})}{1 - F(y_n - \phi h_{2n})}\frac{1-F(y_n - \phi h_{2n})}{1-F(y_n)}K_2^{(1)}(\phi)|d\phi = o(1)\end{aligned}$$

by Lebesgue's dominated convergence theorem, Proposition 1.15 in Resnick (1987), the fact that $1 - F(y_n) \to 0$ as $n \to \infty$ and $\int |K_2^{(1)}(\phi)|d\phi < C$. Given (17)

$$\begin{aligned}Q_{322n} &\le \frac{\theta^{1/2}}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| |U_i|(1 + \theta^{1/2}|U_i|)O_p(L_{1n})\\ &= O_p(L_{1n})\left(\theta^{1/2}\frac{1}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| |U_i| + \theta\frac{1}{nh_{2n}^2}\sum_{i=1}^n \left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| U_i^2\right)\\ &\le O_p\left(\frac{L_{1n}}{h_{2n}^2}\right) C\left(\theta^{1/2}\frac{1}{n}\sum_{i=1}^n |U_i| + \theta\frac{1}{n}\sum_{i=1}^n U_i^2\right) \quad \text{since } |K_2^{(2)}(\phi)| < C\\ &= O_p\left(\frac{L_{1n}}{h_{2n}^2}\right),\end{aligned}$$

where the last equality follows from A4 by using Kolmogorov's law of large numbers to obtain $\frac{1}{n}\sum_{i=1}^{n}|U_i| = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}U_i^2 = O_p(1)$. Consequently, $Q_{32n} = o_p(1) + O_p\left(\frac{L_{1n}}{h_{2n}^2}\right) = O_p(1)$ given $h_{1n} \propto n^{-1/5}$ and $h_{2n} \propto n^{-1/5+\delta}$ for $\delta > 0$. A similar use of Taylor's Theorem gives

$$\begin{aligned} Q_{33n} &\le \frac{\theta}{2}\frac{1}{nh_{2n}}\sum_{i=1}^{n}\left|K_2^{(1)}\left(\frac{y_n - U_i}{h_{2n}}\right)\right| U_i^2 + \frac{\theta}{2}\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| U_i^2|\widehat{U}_i - U_i| \\ &= Q_{331n} + Q_{332n}. \end{aligned} \tag{19}$$

We note that Q_{331n} is $O_p(1)$ since

$$\begin{aligned} E\left(Q_{331n}\right) &= \frac{\theta}{2}\int (y_n - \phi h_{2n})^2|K_2^{(1)}(\phi)|f(y_n - \phi h_{2n})d\phi \\ &= \frac{\theta}{2}(1 - F(y_n))y_n \\ &\quad \times \int \frac{|y_n - \phi h_{2n}|}{y_n}\frac{|y_n - \phi h_{2n}|f(y_n - \phi h_{2n})}{1 - F(y_n - \phi h_{2n})}\frac{1 - F(y_n - \phi h_{2n})}{1 - F(y_n)}K_2^{(1)}(\phi)|d\phi = O(1) \end{aligned}$$

by Lebesgue's dominated convergence theorem, Proposition 1.15 in Resnick (1987), the fact that $y_n(1 - F(y_n)) \to C$ as $n \to \infty$ when $\alpha \ge 1$ and $\int |K_2^{(1)}(\phi)|d\phi < C$. Given (17)

$$\begin{aligned} Q_{332n} &\le \frac{\theta}{2}\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| U_i^2(1 + \theta^{1/2}|U_i|)O_p(L_{1n}) \\ &= O_p(L_{1n})\left(\frac{\theta}{2}\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| U_i^2 + \frac{\theta^{3/2}}{2}\frac{1}{nh_{2n}^2}\sum_{i=1}^{n}\left|K_2^{(2)}\left(\frac{y_n - U_i^{**}}{h_{2n}}\right)\right| |U_i|^3\right) \\ &\le O_p\left(\frac{L_{1n}}{h_{2n}^2}\right)C\left(\frac{\theta}{2}\frac{1}{n}\sum_{i=1}^{n}U_i^2 + \frac{\theta^{3/2}}{2}\frac{1}{n}\sum_{i=1}^{n}|U_i|^3\right) \text{ since } |K_2^{(2)}(\phi)| < C \\ &= O_p\left(\frac{L_{1n}}{h_{2n}^2}\right), \end{aligned}$$

where the last equality follows from A4 by using Kolmogorov's law of large numbers to obtain $\frac{1}{n}\sum_{i=1}^{n}|U_i| = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}|U_i|^3 = O_p(1)$. Consequently, $Q_{33n} = O_p(1) + O_p\left(\frac{L_{1n}}{h_{2n}^2}\right) = O_p(1)$ given $h_{1n} \propto n^{-1/5}$ and $h_{2n} \propto n^{-1/5+\delta}$ for $\delta > 0$. Hence,

$$\frac{n}{\sqrt{N}}Q_{3n} = \frac{n}{\sqrt{N}}O_p\left(\frac{L_{1n}^2}{h_{2n}}\right) = o_p(1), \text{ provided } N \propto n^{4/5-\delta}. \tag{20}$$

$$
\begin{aligned}
Q_{2n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{2n}}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)(\widehat{U}_i-U_i)\\
&= \theta^{1/2}(\hat{\theta}^{-1/2}-\theta^{-1/2})\frac{1}{nh_{2n}}\sum_{i=1}^{n}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)U_i\\
&\quad -\frac{1}{\hat{\theta}^{1/2}}\frac{1}{nh_{2n}}\sum_{i=1}^{n}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)(\hat{m}(X_i)-m(X_i))\\
&= Q_{21n}+Q_{22n}.
\end{aligned}
$$

Since $\hat{m}(x)-m(x)-\frac{1}{nh_{1n}f_X(x)}\sum_{t=1}^{n}K_1\left(\frac{X_t-x}{h_{1n}}\right)Y_t^*=O_p\left(L_{1n}^2\right)$[4] uniformly over the compact set G, with $Y_t^*=\frac{1}{2}m^{(2)}(x^*)(X_t-x)^2+\theta^{1/2}U_t$, $x^*=\lambda X_t-(1-\lambda)x$ for $\lambda\in(0,1)$, and $\hat{\theta}^{-1/2}=O_p(1)$, we can write

$$
\begin{aligned}
Q_{22n} &= O_p(1)\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{h_{2n}f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)\frac{1}{2}m^{(2)}(X_t^*)(X_t-X_i)^2\right.\\
&\quad \left.+\,\theta^{1/2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{h_{2n}f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)\frac{1}{h_{1n}}K_1\left(\frac{X_t-X_i}{h_{1n}}\right)U_t\right)+O_p\left(L_{1n}^2\right)\\
&= O_p(1)(Q_{221n}+\theta^{1/2}Q_{222n})+O_p\left(L_{1n}^2\right).
\end{aligned}
$$

We will obtain the order of each Q_{22jn} for $j=1,2$ 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 $Q_{222n}=\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}(\psi_n(Z_t,Z_i)+\psi_n(Z_i,Z_t))=\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\phi_n(Z_i,Z_t)$, where $\phi_n(Z_t,Z_i)$ is a symmetric function. The partial sum for the case where $i=t$ is denoted by $Q'_{222n}=\frac{K_1(0)}{n^2h_{2n}h_{1n}}\sum_{i=1}^{n}\frac{1}{f_X(X_i)}K_2\left(\frac{y_n-U_i}{h_{2n}}\right)U_i$, and given that $\frac{f_{U\,|\,X=x}(u|x)}{f(u)}\to 1$ as $u\to\infty$, FR1', 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 $\frac{n}{\sqrt{N}}Q'_{222n}=o_p(1)$. For the case where $i\neq t$, we write the remaining partial sums as

$$
Q''_{222n}=\frac{1}{n}\sum_{t=1}^{n}E(\phi_n(Z_t,Z_i)|Z_t)-\frac{1}{2}E(\phi_n(Z_t,Z_i))+O_p\left(n^{-1}(E(\phi_n^2(Z_t,Z_i)))^{1/2}\right).
$$

---

[4] If m were estimated by a Nadaraya–Watson estimator then $Y_t^*=m^{(1)}(x)(X_t-x)+\frac{1}{2}m^{(2)}(x^*)(X_t-x)^2+\theta^{1/2}U_t$. The additional terms that involve $m^{(1)}(x)(X_t-x)$ can be shown to be negligible in probability at the desired rate $\frac{n}{\sqrt{N}}$.

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

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

with $E(Z_{tn}) = 0$. As above, using A4, FR1', and Lebesgue's dominated convergence theorem, we have that $E\left(\frac{ny_n}{N}Z_{tn}^2\right) \to -k_0^{-1}\theta$. Using similar arguments, we have $n^{-1}E\left(\phi_n^2(Z_t, Z_i)\right)^{1/2} = O\left(n^{-1}\left(\frac{N}{ny_nh_{1n}h_{2n}}\right)^{1/2}\right)$. Consequently,

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

since $y_n \to \infty$ and $nh_{1n}h_{2n} \to \infty$. Hence, we can write that $\frac{n}{\sqrt{N}}\sqrt{y_n}Q''_{222n} = \frac{n}{\sqrt{N}}\frac{1}{n}\sum_{i=1}^{n} Z_{tn}\sqrt{y_n} + o_p(1)$. Since $E(Z_{tn}\sqrt{y_n}) = 0$ and $E\left(\frac{ny_n}{N}Z_{tn}^2\right) \to -k_0^{-1}\theta$, by Liapounov's CLT, we have $\frac{n}{\sqrt{N}}\sqrt{y_n}Q''_{23n} \overset{d}{\to} N(0, -k_0^{-1}\theta)$, and since $\sqrt{y_n} \to \infty$ as $n \to \infty$, we have that $\frac{n}{\sqrt{N}}Q''_{222n} = o_p(1)$.

Using similar arguments and manipulations, we obtain $\frac{n}{\sqrt{N}}Q_{221n} = o_p\left(h_{1n}^2\sqrt{N}\right) + o_p(1)$. Hence, combining the orders for Q_{221n} and Q_{222n}, we have

$$\frac{n}{\sqrt{N}}Q_{22n} = o_p\left(h_{1n}^2\sqrt{N}\right) + o_p(1) + \frac{n}{\sqrt{N}}O_p\left(L_{1n}^2\right) = o_p(1), \tag{21}$$

where the last equality holds provided that $h_{1n} \propto n^{-1/5}$ and $h_{2n} \propto n^{-1/5+\delta}$ for $\delta > 0$ and $N \propto n^{4/5-\delta}$.

Since $\hat{\theta}^{-1/2} - \theta^{1/2} = O_p(L_{1n})$, $Q_{21n} = O_p(L_{1n})\frac{1}{nh_{2n}}\sum_{i=1}^{n} K_2\left(\frac{y_n - U_i}{h_{2n}}\right)U_i$. Note that

$$E\left(\frac{1}{h_{2n}}K_2\left(\frac{y_n - U_i}{h_{2n}}\right)U_i\right)$$
$$= (1 - F(y_n))\int K_2(\phi)\frac{(y_n - h_{2n}\phi)f(y_n - h_{2n}\phi)}{1 - F(y_n - h_{2n}\phi)}\frac{1 - F(y_n - h_{2n}\phi)}{1 - F(y_n)}d\phi,$$

and by Lebesgue's dominated convergence theorem, Proposition 1.15 in Resnick (1987), and the fact that $\int K_2(\phi)d\phi = 1$, we have $Q_{21n} = O_p(L_{1n})(1 - F(y_n))$. Consequently,

$$\frac{n}{\sqrt{N}}Q_{21n} = \frac{n}{\sqrt{N}}(1 - F(y_n))O_p(L_{1n})$$
$$= \left(\frac{n}{\sqrt{N}}\left(1 - F(y_n) - \frac{N}{n}\right) + \sqrt{N}\right)O_p(L_{1n}).$$

Note that $\frac{n}{\sqrt{N}}\left(1-F(y_n)-\frac{N}{n}\right)=O(1)$ and $\sqrt{N}L_{1n}=\left(\frac{N\log n}{nh_{1n}}\right)^{1/2}+N^{1/2}h_{1n}^2=o(1)$ given that $h_{1n}\propto n^{-1/5}$ and $N\propto n^{4/5-\delta}$ for $0<\delta<4/5$. Hence, we have that $\frac{n}{\sqrt{N}}Q_{21n}=o_p(1)$ and combined with (21) gives $\frac{n}{\sqrt{N}}Q_{2n}=o_p(1)$.

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

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

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

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

Integrating by parts, we have

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

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

$$\frac{n}{\sqrt{N}}(Q_{1n} - F(y_n)) \xrightarrow{d} N(0,1). \tag{22}$$

Equations (20), (21), and (22) show that $\frac{n}{\sqrt{N}}(\widetilde{F}(y_n) - F(y_n)) \xrightarrow{d} N(0,1)$, and by consequence, $T_{1n} = O_p(1)$ which completes the proof.

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

$$\begin{pmatrix} \frac{1}{\delta_N^2}\frac{\partial}{\partial t}L_{TN}(t^*,\tau^*) \\ \frac{1}{\delta_N^2}\frac{\partial}{\partial \tau}L_{TN}(t^*,\tau^*) \end{pmatrix} = \frac{1}{\delta_N N}\begin{pmatrix} \sum_{j=1}^{N_s}\frac{\partial}{\partial r_N}log\, g(\widetilde{Z}_j;\tilde{r}_N\sigma_N,\tilde{k}) \\ \sum_{j=1}^{N_s}\frac{\partial}{\partial k}log\, g(\widetilde{Z}_j;\tilde{r}_N\sigma_N,\tilde{k}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{23}$$

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

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

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

where $\widetilde{I}_{1N}$, I_{1N}, $\widetilde{I}_{4N}$, I_{4N} are as defined in Lemma 1. By a Taylor's expansion of the first order condition in (23) around $(1,k_0)$, we have

$$H_N(r_N^*,k^*)\sqrt{N}\begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} = v_N(1,k_0). \tag{24}$$

We start by investigating 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 $N_s - N = O_p(N^{1/2})$

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

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 $\sum_{i=1}^n(-k_0)(n(1-F(y_n)))^{-1/2}(q_{1in} - E(q_{1in})) + o_p(1) = \sum_{i=1}^n Z_{i3} + o_p(1)$ where $q_{1in} = \frac{1}{h_{2n}}\int_{-\infty}^{y_n} K_2\left(\frac{y-U_i}{h_{2n}}\right)dy$ 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 $b_\sigma = \frac{C\phi(q(a_n))}{1+\alpha-\rho} + o(\phi(q(a_n)))$ and $b_k = -\frac{C\alpha\phi(q(a_n))}{(\alpha-\rho)(1+\alpha-\rho)} + o(\phi(q(a_n)))$, we have that

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

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

$$\begin{aligned} E\left(T_{1n}\right) = &-b_k - \frac{1}{k_0}\left(\frac{1}{k_0} - 1\right)^2 E\left(\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)^{-2}\left(\frac{k_0 Z_i'}{\sigma_N}\right)^2\right) \\ &- \frac{1}{k_0^2}\left(\frac{1}{k_0} - 1\right) E\left(log\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0 Z_i'}{\sigma_N}\right)\right). \end{aligned}$$

From Smith (1987), we have that $E\left(\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)^{-2}\left(\frac{k_0 Z_i'}{\sigma_N}\right)^2\right) = \frac{2}{(1+\alpha)(2+\alpha)} + O(\phi(q(a_n)))$ and $b_k = O(\phi(q(a_n)))$. From Lemma 4, we have that

$$E\left(log\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0 Z_i'}{\sigma_N}\right)\right) = -\frac{1}{\alpha} + \frac{\alpha}{(1+\alpha)^2} + O(\phi(q(a_n))),$$

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

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

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

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

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

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

from Lemma 2 $E(q_{1in}) = F(y_n) + O(h_{2n}^{m+1}) = O(1)$ 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 $\frac{\sqrt{N}}{n} O(\phi(q(a_n)))O(1)N^{-1/2} = n^{-1}O(\phi(q(a_n)))$. 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

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

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

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

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

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

$$-(1 - k_0 Z_i'/\sigma_N)^{-3}(k_0 Z_i'/\sigma_N)^3 < -(1 - k_0 Z_i'/\sigma_N)^{-3}(1 - k_0 Z_i'/\sigma_N)^3 = 1.$$

Lastly,

$$\begin{aligned}
\sum_{i=1}^n E(|Z_{2i}|^3) \le \frac{1}{\sqrt{N}} E\Big(\Big|&-(1/k_0^2)log(1 - k_0 Z_i'/\sigma_N) \\
&+ (1/k_0)(1 - 1/k_0)(1 - k_0 Z_i'/\sigma_N)^{-1} k_0 z_i'/\sigma_N\Big|^3\Big) \to 0
\end{aligned}$$

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

$$\begin{aligned}E\left(log(1-k_0Z_i'/\sigma_N)^3\right) &= -\int_0^\infty log(1-k_0z/\sigma_N)^3 dF_{q(a_n)}(z)\\ &= -\frac{1-F(q(a_n)(1+z/q(a_n)))}{1-F(q(a_n))}(log(1+z/q(a_n)))^3|_0^\infty\\ &\quad+\int_0^\infty \frac{L(q(a_n)(1+z/q(a_n)))}{L(q(a_n))}(1+z/q(a_n))^{-\alpha}3(log(1+z/q(a_n)))^2\\ &\quad\times(1+z/q(a_n))^{-1}(1/q(a_n))dz = T_{1n}+T_{2n}.\end{aligned}$$

Three repeated applications of L'Hôpital's rule and FR1' gives $T_{1n}=0$. For T_{2n}, we have that given FR2 and again integrating by parts and letting $t=1+z/q(a_n)$

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

It is easy to verify that $\int_1^\infty 3(log(t))^2t^{-\alpha-1}dt=\frac{6}{\alpha^3}$ and, consequently, $T_{2n}=\frac{6}{\alpha^3}+O(\phi(q(a_n)))$, 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 $\gamma'\left(v_N(\sigma_N,k_0)-\sqrt{N}\binom{b_\sigma}{b_k}\right)\xrightarrow{d}N(0,\gamma'V_2\gamma)$ 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 $\left(v_N(\sigma_N,k_0)-\sqrt{N}\binom{b_\sigma}{b_k}\right)\xrightarrow{d}N(0,V_2)$. Hence, given Eq. (24), 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* $j=1,\ldots,N$ *define* $Z_j=U_j-q_n(a_n)$ *whenever* $U_j>q_n(a_n)$, *and for* $j=1,\ldots,N_1$ *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

$$\begin{aligned}\Delta_\sigma &= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial}{\partial\sigma}log\, g(Z_j;\sigma_N,k_0)\sigma_N - \frac{\partial}{\partial\sigma}log\, g(Z'_j;\sigma_N,k_0)\sigma_N\right)\\ &= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{k_0}-1\right)\left(\left(1-\frac{k_0Z_j}{\sigma_N}\right)^{-1}\frac{k_0Z_j}{\sigma_N} - \left(1-\frac{k_0Z'_j}{\sigma_N}\right)^{-1}\frac{k_0Z'_j}{\sigma_N}\right).\end{aligned}$$

By the mean value theorem, there exists $\lambda_j \in (0,1)$ and $Z^*_j = Z'_j + \lambda_j(q(a_n) - q_n(a_n))$ such that

$$\Delta_\sigma = \frac{q_n(a_n)-q(a_n)}{q(a_n)}\left(\frac{1}{k_0}-1\right)\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z^*_j\right)^{-2}. \tag{25}$$

Again, using the mean value theorem, we have that for some $\theta_j \in (0,1)$ there is $Z^{**}_j = \theta_jZ'_j + (1-\theta_j)Z^*_j = Z'_j + \lambda_j(1-\theta_j)(q(a_n) - q_n(a_n))$ such that

$$\begin{aligned}&\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z^*_j\right)^{-2} - \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z'_j\right)^{-2}\\ &= \frac{1}{N}\sum_{j=1}^{N}\frac{2k_0/\sigma_N}{\left(1-\frac{k_0}{\sigma_N}Z^{**}_j\right)^3}(Z^*_j - Z'_j)\\ &= -\theta\frac{q(a_n)-q_n(a_n)}{q(a_n)}\frac{2q(a_n)}{q_n(a_n)}\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z^{**}_j\right)^{-3}\\ &= O_p(N^{-1/2})(1+o_p(1))\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z^{**}_j\right)^{-3},\end{aligned}$$

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,

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z^{**}_j\right)^{-3} &= \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z'_j + (\theta-\theta^2)(q(a_n)-q_n(a_n))\right)^{-3}\\ &= \frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z'_j + o_p(1)\right)^{-3} = O_p(1),\end{aligned}$$

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

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j^*\right)^{-2} &= O_p(N^{-1/2})+\frac{1}{N}\sum_{j=1}^{N}\left(1-\frac{k_0}{\sigma_N}Z_j'\right)^{-2}\\ &= \frac{\alpha}{2+\alpha}-\frac{2C\phi(q(a_n))}{(2+\alpha)(2+\alpha-\rho)}+o(\phi(q(a_n)))+O_p(N^{-1/2}),\end{aligned}$$

where the last equality follows from Smith (1987). Consequently, since $\phi(q(a_n)) = O(N^{-1/2})$ and substituting back in Eq. (25) 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

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

The first term is $b_1N^{1/2}\frac{q_n(a_n)-q(a_n)}{q(a_n)}+o_p(1)$ as in the case where $N = N_1$. As in Smith (1987), we have that the expectation of the second term is $\frac{N_1-N}{\sqrt{N}}\left(\frac{C\phi(q(a_n))}{1+\alpha-\rho}+o(\phi(q(a_n)))\right)$ which is $o_p(1)$ since $\phi(q(a_n)) = O(N^{-1/2})$ 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 $N^{1/2}\Delta_\sigma = b_1N^{1/2}\frac{q_n(a_n)-q(a_n)}{q(a_n)} + o_p(1)$. 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. $E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_i'}{\sigma_N}\right)\right) = -\frac{1}{\alpha}+\frac{\alpha}{(1+\alpha)^2}+O(\phi(q(a_n)))$.

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

$$\begin{aligned}E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_i'}{\sigma_N}\right)\right) = -\alpha^{-1}+O(\phi(q(a_n)))\\ +E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\right).\end{aligned}$$

Using the notation for $L(\cdot)$ in FR2 and given that $F_{q(a_n)}(z) = 1-\frac{L\left(\left(1+\frac{z}{q(a_n)}\right)q(a_n)\right)}{L(q(a_n))}\left(1+\frac{z}{q(a_n)}\right)^{-\alpha}$, we can write $E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\right) = \int_0^\infty log\left(1-\frac{k_0z}{\sigma_N}\right)\left(1-\frac{k_0z}{\sigma_N}\right)^{-1}dF_{q(a_n)}(z)$.

Integrating by parts, we have

$$E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\right) = \int_0^\infty \frac{L\left(\left(1+z/q(a_n)\right)q(a_n)\right)}{L(q(a_n))}\left(1+z/q(a_n)\right)^{-\alpha}$$
$$\times\left(\frac{1}{q(a_n)}\left(1+z/q(a_n)\right)^{-2}\right.$$
$$\left.-\frac{1}{q(a_n)}log\left(1+z/q(a_n)\right)\left(1+z/q(a_n)\right)^{-2}\right)dz.$$

Setting $t = 1 + z/q(a_n)$, we have that

$$E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\right) = \int_1^\infty \frac{L(tq(a_n))}{L(q(a_n))}(t^{-\alpha-2} - log(t)t^{-\alpha-2})dt,$$

and by FR2

$$\begin{aligned} &E\left(log\left(1-\frac{k_0Z_i'}{\sigma_N}\right)\left(1-\frac{k_0Z_i'}{\sigma_N}\right)^{-1}\right)\\ &= \int_1^\infty (t^{-\alpha-2} - log(t)t^{-\alpha-2})dt\\ &\quad + C\phi(q(a_n))\int_1^\infty (t^{-\alpha-2} - log(t)t^{-\alpha-2})\int_1^t u^{\rho-1}du dt + o(\phi(q(a_n)))\\ &= \frac{1}{\alpha+1} - \frac{1}{(1+\alpha)^2} + O(\phi(q(a_n))), \end{aligned}$$

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

Proof of Theorem 2. 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

$$1 - \widetilde{F}(\tilde{q}(a_n) + \hat{y}_{N,a}) = \frac{N}{n}\left(1 - \frac{\tilde{k}\hat{y}_{N,a}}{\tilde{\sigma}_N}\right)^{1/\tilde{k}}. \tag{26}$$

Let us pause and note that for a chosen N, Eq. (26) is satisfied with a distribution function $\dot{F}$ that is not necessarily $\widetilde{F}$. However, given the continuity of $\widetilde{F}$, there exists N satisfying the order relation $a > 1 - N/n$ for which (26) is satisfied by $\widetilde{F}$. Hence, to avoid additional notation, we proceed with $\widetilde{F}$. 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

$$\begin{aligned} P(\sigma_n^{-1}(\hat{q}(a) - q(a)) \leq z) &= P(1 - a \geq 1 - \widetilde{F}(q(a_n) + y_{N,a} + \sigma_n z) \\ &= P(V_n((1-a) - (1 - F(q(a) + \sigma_n z)) \\ &\geq V_n((1 - \widetilde{F}(q(a_n) + y_{N,a} + \sigma_n z)) \\ &\quad - (1 - F(q(a) + \sigma_n z))). \end{aligned}$$

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 $W_n = V_n((1 - \widetilde{F}(q(a_n) + y_{N,a} + \sigma_n z)) - (1 - F(q(a) + \sigma_n z))$ and note that $\frac{n(1-F(q(a)))}{V_n(1-F(q(a)+\sigma_n z)} W_n = \sqrt{n(1-F(q(a)))}\left(\frac{1-\widetilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)} - 1\right) = -\frac{1}{k_0} W_n(1 + o(1))$. We first establish that

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

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,

$$Z_N^{-1/k_0} \frac{1 - F(q(a_n)Z_N)}{1 - F(q(a_n))} = 1 + k(z_a)N^{-1/2}\frac{\mu(\alpha - \rho)}{C} + o(N^{-1/2}). \tag{27}$$

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

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

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

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

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 $\sigma_N = -k_0 q(a_n) = -\frac{k_0 y_N}{Z_N - 1}$ that $y_N = \frac{(1-Z_N)\sigma_N}{k_0}$ and from Theorem 1 we have

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

Hence, if $c_b' = \left(\begin{smallmatrix} -k_0^{-1}(z_a^{-1}-1) & k_0^{-2}log(z_a)+k_0^{-2}(z_a^{-1}-1) \end{smallmatrix} \right)$ and $\mu_p' = \left(\begin{smallmatrix} \frac{\mu(1-k_0)(1+2k\rho)}{1-k_0+k_0\rho} & \frac{\mu(1-k_0)k_0(1+\rho)}{1-k_0+k_0\rho} \end{smallmatrix} \right)$, we can write

$$c_b'\sqrt{N}\begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} \overset{d}{\rightarrow} N(c_b'\mu_p, c_b'H^{-1}V_2H^{-1}) \text{ and } \sqrt{N}(h(\tilde{\sigma}_N, \tilde{k}, y_N) - h(\sigma_N, k_0, y_N)) = O_p(1). \tag{28}$$

Now, we can conveniently write $\frac{1-\widetilde{F}(q(a_n)+y_N)}{1-F(q(a_n)-y_N)} = \frac{1-\widetilde{F}(q(a_n)+y_N)}{1-\widetilde{F}(\tilde{q}(a_n))}\frac{1-F(q(a_n))}{1-F(q(a_n)+y_N)} Z_N^{1/k_0} Z_N^{-1/k_0}$. Note that $\frac{1-\widetilde{F}(q(a_n)+y_N)}{1-\widetilde{F}(\tilde{q}(a_n))} = \left(1 - \frac{\tilde{k}y_N}{\tilde{\sigma}_N}\right)^{1/\tilde{k}} \left(\frac{1-\widetilde{F}(q(a_n))}{1-\widetilde{F}(\tilde{q}(a_n))}\right)$ 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 Eq. (27), $Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))} - 1 = N^{-1/2}\left(-k(z_a)\frac{\mu(\alpha-\rho)}{C}\right) + o(N^{-1/2})$. Hence,

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

Now, we note that $\frac{1-\widetilde{F}(q(a_n))}{1-\widetilde{F}(\tilde{q}(a_n))} - 1 = -\frac{\widetilde{F}(q(a_n))-F(q(a_n))}{1-F(q(a_n))}$, and from Eq. (15) in Lemma 2, we have $\frac{\sqrt{n(1-F(q(a_n))}}{1-F(q(a_n))}(1-\widetilde{F}(q(a_n)) - (1-F(q(a_n))) \overset{d}{\rightarrow} N(0,1)$ as $q(a_n) \rightarrow \infty$. In particular, using the notation adopted in Lemma 2, we have that

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

Hence,

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

and by Eq. (28) 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

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

Since $\sqrt{N}\left(Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}-1\right) \to -\frac{k(z_a)\mu(\alpha-\rho)}{C}$, we focus on the joint distribution of the last two terms. By Eq. (28), we have that

$$\sqrt{N}(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N)) = c_b'\sqrt{N}\begin{pmatrix}\tilde{r}_N-1\\ \tilde{k}-k_0\end{pmatrix}+o_p(1), \tag{29}$$

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 $\sqrt{N}\left(\frac{1-\widetilde{F}(q(a_n))}{(1-\widetilde{F}(\tilde{q}(a_n)))}-1\right)=\sum_{i=1}^n Z_{i4}$. Let $0\neq d\in\Re^4$, $\varepsilon_n'=(\,\sum_{i=1}^n Z_{i1}\;\sum_{i=1}^n Z_{i2}\;\sum_{i=1}^n Z_{i3}\;\sum_{i=1}^n Z_{i4}\,)$, 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 independent and identically distributed (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

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

By Lemma 2, $E(q_{1in})-F(y_n)=O(h_{2n}^{m+1})$, 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 $\kappa_i(x)=h_{2n}^{-1}\int_{-\infty}^{x}K_2\left(\frac{y-U_i}{h_{2n}}\right)dy$, we can write $E\left(q_{1in}\frac{1}{h_{2n}}\int_{-\infty}^{q(a_n)}K_2\left(\frac{y-U_i}{h_{2n}}\right)dy\right)=E(\kappa_i(q(a_n)+\sigma_n z)\kappa_i(q(a_n)))(\chi_{\{q(a_n)=y_n\}}+\chi_{\{q(a_n)\neq y_n\}})$. For $z>0$, we have that $q(a_n)\neq y_n$

implies $y_n > q(a_n)$ so that

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

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

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

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

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

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 $d'\varepsilon_n \overset{d}{\to} N(0, d'V_3 d)$ and from the Cramer–Wold theorem $\varepsilon_n \overset{d}{\to} N(0, V_3)$. Now, from Eq. (29)

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

Hence by letting $A_{.j}$ represent the jth column of a matrix A, we write

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

Let $\eta' = (\,-c_b'H_{.1}^{-1}\ -c_b'H_{.2}^{-1}\ -c_b'H_{.1}^{-1}b_1 - c_b'H_{.2}^{-1}b_2\ 1\,)$. Then from the results above, we have $\eta'\varepsilon_n \overset{d}{\to} N(0, \eta'V_3\eta)$, where simple algebraic manipulations give $\eta'V_3\eta = c_b'H^{-1}V_2H^{-1}c_b +$

$2c_b'\binom{2-k_0}{1-k_0}+1$. Consequently, if $\zeta \sim N\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C}, c_b'H^{-1}V_2H^{-1}c_b + 2c_b'\binom{2-k_0}{1-k_0}+1\right)$, then

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

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

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

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

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

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

$$\zeta_1 \sim N\Bigg((-k_0)\left(-\frac{k(z_a)\mu(\alpha-\rho)}{C} - c_b'H^{-1}\lim_{n\to\infty}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right),$$
$$k_0^2\left(c_b'H^{-1}V_2H^{-1}c_b + 2c_b'\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\Bigg).$$

ACKNOWLEDGMENTS

We thank Essie Maasoumi, an Associate Editor, and two anonymous referees for comments that improved the paper substantially. Any remaining errors are the authors' responsibility.

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., Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* 85:799–808.

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

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics* 33:806–839.

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

Davis, R., Resnick, S. I. (1984). Tail Estimates Motivated by Extreme Value Theory. *Annals of Statistics* 12:1467–1487.

Doksum, K., Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Annals of Statistics* 23:1443–1473.

Embrechts, P., Kluppelberg, C., Mikosh, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer Verlag.

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

Fan, J., Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Verlag.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of Royal Satistical Society Series B* 44:37–42.

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

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica* 46:33–50.

Leadbetter, M. R., Lindgren, G., Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.

Li, Q., Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton University Press.

Martins-Filho, C., 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., 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., Yao, F. (2008). A smoothed conditional quantile frontier estimator. *Journal of Econometrics* 143:317–333.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17:571–599.

McNeil, A., 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., Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge, UK: Cambridge University Press.

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*. New York: Springer Verlag.

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.