Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

Maximo Torero

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: m.torero@cgiar.org
Voice: + 1 202 862 5662

December, 2013

Abstract. We propose nonparametric estimators for conditional value-at-risk (VaR) and expected shortfall (ES) associated with conditional distributions of a series of returns on a financial asset. The return series and the conditioning covariates, which may include lagged returns and other exogenous variables, are assumed to be strong mixing and follow a fully nonparametric conditional location-scale model. First stage nonparametric estimators for location and scale are combined with a generalized Pareto approximation for distribution tails proposed by Pickands (1975) to give final estimators for conditional VaR and ES. We provide consistency and asymptotic normality of the proposed estimators under suitable normalization. We also present the results of a Monte Carlo study that sheds light on their finite sample performance. Empirical viability of the model and estimators is investigated through a backtesting exercise using returns on future contracts for five agricultural commodities.

Keywords and phrases. Value-at-risk, expected shortfall, extreme value theory, nonparametric location-scale models, strong mixing.

JEL classifications. C14, C15, C22, G10.

AMS-MS classifications. Primary: 62G32, 62G07, 62G08, 62G20.

1 Introduction

Conditional value-at-risk (CVaR) and conditional expected shortfall (CES) are two of the most used synthetic measures of market risk (Duffie and Singleton (2003), McNeill et al. (2005), Danielsson (2011)). From a statistical perspective these risk measures have straightforward definitions. Let $\{Y_t\}$ denote a stochastic process representing the returns[1] on a given portfolio, stock or market index, where $t \in \mathbb{Z}$ indexes a discrete measure of time and $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$ denotes the conditional distribution of Y_t given $\mathbf{X}_t = \mathbf{x}$. The stochastic vector $\mathbf{X}_t \in \mathbb{R}^d$ normally includes lags $\{Y_{t-\ell}\}_{1\leq \ell \leq m}$ for some $m \in \mathbb{N}$ as well as other relevant conditioning random variables that reflect economic and market conditions. Then, for $a \in (0,1)$, a-CVaR($\mathbf{x}$) is defined to be the a-quantile associated with $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$ and a-CES($\mathbf{x}$) is defined to be the conditional expectation of Y_t given that Y_t exceeds the a-CVaR($\mathbf{x}$) associated with $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$, i.e., a-CES($\mathbf{x}$)$=E(Y_t|Y_t >$ a-CVaR($\mathbf{x}$)).

Usually, practical interest focuses on estimating a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$) for extreme values of a which are in the vicinity of 1. For example, the Capital Adequacy Directive from the Bank of International Settlements requires the risk capital of a bank to be sufficient to cover losses on its portfolio (over a 10-day holding period) with a probability $a = 0.99$. In the conditional (regression) quantile literature this is commonly referred to as estimation of *extremal* or *high-order* quantiles and has been studied by Chernozhukov and Umantsev (2001), Chernozhukov (2005) and Martins-Filho et al. (2014), among others. In such cases, a-CVaR($\mathbf{x}$), a-CES($\mathbf{x}$) and their estimation depend critically on the nature of the upper (deep) tail of $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$.

A seminal contribution to the study of distribution upper tail behavior is Theorem 7 in Pickands (1975). There, it is shown that for any distribution F in the domain of attraction of an extremal distribution (Leadbetter et al. (1983), Resnick (1987)), denoted here by $F \in D(E)$, for some fixed k and function $\sigma(\xi)$

$$F \in D(E) \iff \lim_{\xi \to u_\infty} \sup_{0<u<u_\infty - \xi} |F_\xi(u) - G(u; 0, \sigma(\xi), k)| = 0, \tag{1}$$

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, $u_\infty > \xi \in \mathbb{R}$ and G

[1] Let P_t denote the price of a financial asset at time t. Throughout this paper, a "return" Y_t is given by $Y_t = -\log \frac{P_t}{P_{t-1}}$. We adopt this definition because in practice, regulators, portfolio and risk managers are mostly concerned with the distribution of losses, i.e., negative values of $\log \frac{P_t}{P_{t-1}}$.

is a generalized Pareto distribution (GPD), i.e.,

$$G(u;\mu,\sigma,k) = \begin{cases} 1-(1-k(u-\mu)/\sigma)^{1/k} & \text{if } k \neq 0, \sigma > 0 \\ 1-exp(-(u-\mu)/\sigma) & \text{if } k = 0, \sigma > 0 \end{cases} \tag{2}$$

with $\mu \leq u < \infty$ if $k \leq 0$ and $\mu \leq u \leq \sigma/k$ if $k > 0$. The equivalence in (1) suggests, as argued in Davis and Resnick (1984) and Smith (1987), that if $F \in D(E)$ it is reasonable to estimate its upper tail and associated functionals, such as extremal quantiles, based on the parametric approximation provided by G.

In this paper we take this approach and use (1) to motivate estimators for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$) based on a location-scale model of $\{Y_t\}$. As such, we assume that the process $\{Y_t\}$ can be written as

$$Y_t = m(\mathbf{X}_t) + h^{1/2}(\mathbf{X}_t)\varepsilon_t, \tag{3}$$

where m and h are nonparametric functions defined on the range of $\mathbf{X}_t$, ε_t is independent of $\mathbf{X}_t$ and $\{\varepsilon_t\}$ is an independent and identically distributed (IID) process with distribution $F \in D(E)$ such that $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = 1$. Specifically, we assume that F is in the domain of attraction of a Fréchet distribution, i.e., the class of distributions satisfying

$$1 - F(x) = x^{-\alpha}L(x) \text{ for } \alpha > 0, \tag{4}$$

where $L(x)$ a slowly varying function at infinity (Gnedenko (1943)) with $\alpha = -1/k$. Restricting F to the domain of attraction of a Fréchet distribution is not entirely arbitrary. If F belonged to the domain of attraction of a (reverse) Weibull distribution, then it must be that its endpoint u_∞ is finite, a restriction which is not commonly placed on the regression error ε. The only other possibility is F in the domain of attraction of a Gumbel distribution. In this case, whenever u_∞ is not finite we have that $1 - F$ is rapidly varying, a case we will avoid.

The model can be viewed as a nonparametric generalization of certain autoregressive conditionally heteroscedastic (ARCH) structures and has been studied by Masry and Tjstheim (1995), Hardle and Tsybakov (1997), Masry and Fan (1997) and Fan and Yao (1998), among others. Under (3), for $a \in (0, 1)$,

$$a\text{-CVaR}(\mathbf{x}) \equiv q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a) = m(\mathbf{x}) + h^{1/2}(\mathbf{x})q(a) \tag{5}$$

and

$$a\text{-CES}(\mathbf{x}) \equiv E\left(Y_t | Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right) = m(\mathbf{x}) + h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t > q(a)), \tag{6}$$

where $q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)$ denotes the conditional a-quantile associated with $F_{Y_t|\mathbf{X}_t=\mathbf{x}}$ and $q(a)$ is the a-quantile associated with F. If a random sample $\{\varepsilon_t\}_{t=1}^n$ were observed, $q(a)$ could be estimated by $\hat{q}(a)$ based on the parametric approximation provided by G using the maximum likelihood estimator proposed and studied by Smith (1987). In this case, $\hat{q}(a)$ could be combined with nonparametric estimators $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ to produce estimators for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$). In practice, $\{\varepsilon_t\}_{t=1}^n$ is not observed, but given a sample $\{(Y_t, \mathbf{X}_t^T)\}_{t=1}^n$ ($\mathbf{x}^T$ indicates the transposition of the vector $\mathbf{x}$) and estimators $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$, it is possible to construct a sequence of standardized nonparametric residuals

$$\hat{\varepsilon}_t = \begin{cases} \frac{Y_t - \hat{m}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)}, & \text{if } \hat{h}(\mathbf{X}_t) > 0 \\ 0, & \text{if } \hat{h}(\mathbf{X}_t) \leq 0 \end{cases} \tag{7}$$

for $t = 1, \cdots, n$ that can be used to produce feasible estimators for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$).

This two-stage estimation procedure - first, obtain standardized residuals from the estimation of m and h; second, use the residuals to obtain estimators of $q(a)$ and $E(\varepsilon_t|\varepsilon_t > q(a))$ that can then be used to produce estimators for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$) - was, to our knowledge, first proposed by McNeill and Frey (2000) in the case where m and h are parametrically indexed by a finite dimensional parameter. They provided no asymptotic characterization or finite sample properties for the resulting estimators of conditional value-at-risk or expected shortfall. However, their backtesting exercise on several time series of selected market indexes provided encouraging evidence of the estimators' performance. Martins-Filho and Yao (2006) generalized the estimation framework of McNeil and Frey to the case where m and h are nonparametric functions. They demonstrate via an extensive Monte Carlo simulation, and through backtesting, that accounting for nonlinearities in m and h can be important in improving the estimators' finite sample performance. Martins-Filho et al. (2014) provide an asymptotic characterization of the two stage estimation procedure for a high-order conditional quantile in a model with constant and unknown variance ($h(\mathbf{x}) = \theta$) and a process $\{(Y_t, \mathbf{X}_t^T)\}_{t\in\mathbb{Z}}$ that is IID. Their results, however, are of limited use in empirical finance where the IID assumption is untenable and $h(\cdot)$ is not adequately modeled as a constant function of t. Furthermore, by restricting attention to the case where the conditioning variables belong to $\mathbb{R}$, they failed to elucidate the restrictions that the dimension d may impose on nonparametric estimation of conditional value-at-risk and expected shortfall.

Here, we extend Martins-Filho et al. (2014) in three important directions: a) we relax the assumption that $\{(Y_t, \mathbf{X}_t^T)\}_{t\in\mathbb{Z}}$ is an IID process and instead consider the case where the process is strictly stationary and strong mixing of a suitable order. This allows for the presence of lagged values of Y_t in the conditioning vector $\mathbf{X}_t$, a possibility not covered in our earlier paper and of significant practical interest; b) we allow the conditional variance h to be a nonconstant function of $\mathbf{X}_t$; c) we consider the estimation of a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$). We establish consistency and asymptotic normality of the maximum likelihood estimators for $q(a)$ and $E(\varepsilon_t|\varepsilon_t > q(a))$ based on the GPD approximation in (1) and use these results to obtain consistency and asymptotic normality of our proposed estimators for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$).

Nonparametric estimators of a-CVaR and a-CES have recently been proposed and studied in several papers. Since a-CVaR($\mathbf{x}$) is a conditional quantile, estimation can naturally proceed using nonparametric regression quantiles as in Yu and Jones (1998), Cai (2002) or Cai and Wang (2008). These estimators for a-CVaR($\mathbf{x}$) can then be used to produce nonparametric estimators for a-CES($\mathbf{x}$) as in Scaillet (2005), Cai and Wang (2008) and Kato (2012).[2] Our approach differs from that of the extant literature in that we explore the approximation provided in (1) under the location-scale structure in an attempt to improve estimation and to treat cases where a is in the vicinity of 1. Reliance on the assumption that the stochastic process $\{Y_t\}$ can be described by (3) implies that the estimators defined in the aforementioned papers may be used for processes where our procedure may not. Nonetheless, the benefits of using the *additional* information provided by (1) regarding tail behavior when a is large are clearly revealed in our Monte Carlo study (see section 4), where our estimation procedure is shown, for example, to outperform that of Cai and Wang (2008).

Besides this introduction, this paper has five more sections and two appendices. Section 2 provides a

[2] Also, there exists a large literature on the nonparametric estimation of *unconditional* value-at-risk and expected shortfall (Scaillet (2004), Chen and Tang (2005), Chen (2008), Linton and Xiao (2013), Hill (2014)) that is indirectly related to our work.

detailed description and discussion of the estimation procedure. Section 3 contains a list of assumptions and the main Theorems that describe the asymptotic behavior of the relevant estimators. Section 4 contains a Monte Carlo study that sheds light on the finite sample behavior of the estimators and contrasts its performance with the estimators proposed by Cai and Wang (2008). Section 5 provides an empirical application in which a-CVaR and a-CES are estimated using time series of returns on future contracts for five widely traded agricultural commodities. A backtesting exercise is also conducted for each of the time series. Section 6 provides concluding remarks and gives some directions for future research. Tables and figures associated with the Monte Carlo study and the empirical exercise are provided in Appendix 1. All proofs and supporting lemmas are provided in Appendix 2.

2 Estimation of a-CVaR and a-CES

As suggested in the introduction, our estimation procedure has two main stages. In the first stage, specific estimators for $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ are required to define of $\hat{\varepsilon}_t$ in (7). Given a sample $\{(Y_t, \mathbf{X}_t^T)\}_{t=1}^n$ we consider the local linear (LL) estimator $\hat{m}(\mathbf{x}) \equiv \hat{\beta}_0$ where $(\hat{\beta}_0, \hat{\beta}) \equiv \underset{\beta_0,\beta}{\operatorname{argmin}} \sum_{t=1}^n \left(Y_t - \beta_0 - (\mathbf{X}_t^T - \mathbf{x}^T)\beta\right)^2 K_1\left(\frac{\mathbf{X}_t - \mathbf{x}}{h_{1n}}\right)$, $K_1(\cdot)$ is a multivariate kernel function and $h_{1n} > 0$ is a bandwidth.[3] For the estimation of h we follow the procedure proposed in Fan and Yao (1998). First, obtain a sequence $\{\hat{U}_t \equiv Y_t - \hat{m}(\mathbf{X}_t)\}_{t=1}^n$ and define $\hat{h}(\mathbf{x}) \equiv \hat{\eta}$ where $(\hat{\eta}, \hat{\eta}_1) \equiv \underset{\eta,\eta_1}{\operatorname{argmin}} \sum_{t=1}^n \left(\hat{U}_t^2 - \eta - (\mathbf{X}_t^T - \mathbf{x}^T)\eta_1\right)^2 K_2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h_{2n}}\right)$, $K_2(\cdot)$ is a multivariate kernel function and $h_{2n} > 0$ is a bandwidth, both potentially different from those used in the definition of $\hat{m}$. The estimators $\hat{m}(\mathbf{X}_t)$ and $\hat{h}(\mathbf{X}_t)$ are used to produce the sequence of standardized nonparametric residuals given in (7).

The second stage, which is based on the equivalence in (1), is more intricate and requires additional notation and motivation. In particular, it is useful to draw a parallel to the work of Smith (1987) by discussing estimation for the case where the ε_t are 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 ε_t to estimate its parameters. Therefore, a key aspect of the estimation is the determination of either a threshold value ξ,

[3] Since $\mathbf{X}_t$ may contain up to m lagged values of Y_t, the effective sample size used in estimation is $n - m$. However, for notational ease, we assume that $Y_0, Y_{-1}, \cdots$ are observed as needed to define the relevant sums of length n.

such that only its exceedances are used to estimate the parameters of the GPD, or more directly, a number $N < n$ of the largest values of ε_t to be used in the estimation.[4] For an observed sequence $\{\varepsilon_t\}_{t=1}^n$, define the ascending order statistics $\{\varepsilon_{(t)}\}_{t=1}^n$ and for some fixed (nonstochastic) $N < n$ define the excesses over $\varepsilon_{(n-N)}$ by $\{Z_i\}_{i=1}^N = \{\varepsilon_{(n-N+i)} - \varepsilon_{(n-N)}\}_{i=1}^N$. Ascending order statistics can be viewed as a-quantiles associated with empirical distributions. As such, we can write

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

where $q_n(a)$ is the a-quantile associated with $F_n(u) = \frac{1}{n}\sum_{t=1}^n \chi(\varepsilon_t)$ where $\chi(\varepsilon) = \begin{cases} 1 & \text{if } \varepsilon \leq u \\ 0 & \text{if } \varepsilon > u \end{cases}$. Consequently, by defining $a_n = 1 - \frac{N}{n}$ we can write

$$\{Z_i\}_{i=1}^N = \left\{\varepsilon_{(n-N+i)} - q_n\left(a_n\right)\right\}_{i=1}^N. \tag{8}$$

Thus, for a given sample size n and a choice of N (or equivalently a_n) we consider the threshold $q_n(a_n)$ which will be exceeded by exactly the N largest elements of $\{\varepsilon_t\}_{t=1}^n$. The sequence $\{Z_i\}_{i=1}^N$ can then be used to estimate the parameters of the GPD. Note that in this setting, for a given sample, the choice of N uniquely determines the threshold $q_n(a_n)$.

In our case we only observe $\{\hat{\varepsilon}_t\}_{t=1}^n$, therefore we must produce an estimated sequence of exceedances with typical element that will be denoted by $\tilde{Z}_i$. Perhaps the most natural procedure would be to define $\tilde{Z}_i = \hat{\varepsilon}_{(n-N+i)} - \hat{q}_n\left(a_n\right)$ where $\hat{q}_n(a_n)$ is the a_n-quantile associated with the empirical distribution of the nonparametric residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$. 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 use the sequence of standardized nonparametric residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$ to estimate F by integrating a Rosenblatt kernel estimator for the density f associated with F, i.e.,

$$\tilde{F}(u) = \frac{1}{nh_{3n}}\sum_{t=1}^n \int_{-\infty}^u K_3\left(\frac{\hat{\varepsilon}_t - y}{h_{3n}}\right) dy \tag{9}$$

[4] It should be clear that when a threshold is chosen, for any given sample, the number of exceedances is uniquely determined, but the choice of N does not uniquely determine a threshold.

where $K_3(\cdot)$ is a univariate kernel and $h_{3n} > 0$ is a bandwidth. We define a preliminary estimator $\tilde{q}(a)$ for $q(a)$ as the solution for $\tilde{F}(\tilde{q}(a)) = a$. Therefore, we construct the *observed* sequence of exceedances to be used in the estimation of the parameters of the GPD in the second stage as $\{\tilde{Z}_i\}_{i=1}^{N_s} = \left\{\hat{\varepsilon}_{(n-N_s+i)} - \tilde{q}(a_n)\right\}_{i=1}^{N_s}$. It should be noted that as in the case where ε_t is observed, N (or a_n) is fixed and the threshold $\tilde{q}(a_n)$ is stochastic. However, here the number of residuals N_s that exceed $\tilde{q}(a_n)$ may be different from N for any finite n and is stochastic (sample dependent). As will be seen in section 3, this stochasticity is fully accounted for in our results and the discrepancy between N and N_s is shown to be of no consequence for the asymptotic properties of our proposed estimators. Throughout the paper, the study of the estimators' asymptotic behavior will require that $0 < a_n < a < 1$ and that $a_n \to 1$ as $n \to \infty$. Put differently, N should be chosen such that $1 - N/n < a$ and $N/n \to 0$ as $n \to \infty$.

The sequence $\{\tilde{Z}_i\}_{i=1}^{N_s}$ is used to obtain maximum likelihood estimators for σ and k based on (1). Since F is restricted to satisfy (4), the relevant generalized Pareto density is $g(z;\sigma,k) = \frac{1}{\sigma}\left(1 - \frac{kz}{\sigma}\right)^{1/k-1}$ where $\mu = 0$. In particular, we consider a solution $(\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})$ for the likelihood equations

$$\frac{\partial}{\partial\sigma}\frac{1}{N}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i;\sigma,k) = 0 \quad \text{and} \quad \frac{\partial}{\partial k}\frac{1}{N}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i;\sigma,k) = 0. \tag{10}$$

Based on (1) we can write $F_{\tilde{q}(a_n)}(y) = \frac{F(y+\tilde{q}(a_n))-F(\tilde{q}(a_n))}{1-F(\tilde{q}(a_n))} \approx 1 - \left(1 - \frac{ky}{\sigma_{\tilde{q}(a_n)}}\right)^{1/k}$ where σ has a subscript $\tilde{q}(a_n)$ to make explicit the fact that it depends on the threshold $\tilde{q}(a_n)$. Without loss of generality, for $a \in (a_n, 1)$, we write $q(a) = \tilde{q}(a_n) + y_{\tilde{q}(a_n),a}$ where by construction $F(\tilde{q}(a_n) + y_{\tilde{q}(a_n),a}) = a$. Hence, we have

$$\frac{1-a}{1-F(\tilde{q}(a_n))} \approx \left(1 - \frac{k\; y_{\tilde{q}(a_n),a}}{\sigma_{\tilde{q}(a_n)}}\right)^{1/k}. \tag{11}$$

If F is estimated by $\tilde{F}$, and noting that $1 - \tilde{F}(\tilde{q}(a_n)) = 1 - a_n$, we write $y_{\tilde{q}(a_n),a} \approx \frac{\sigma_{\tilde{q}(a_n)}}{k}\left(1 - \left(\frac{1-a}{1-a_n}\right)^k\right)$. The approximation in (11) is the basis for our proposed estimator $\hat{q}(a)$ for $q(a)$, which is given by

$$\hat{q}(a) = \tilde{q}(a_n) + \frac{\tilde{\sigma}_{\tilde{q}(a_n)}}{\tilde{k}}\left(1 - \left(\frac{1-a}{1-a_n}\right)^{\tilde{k}}\right). \tag{12}$$

We note that if the exceedances $\varepsilon_t - q(a)$ over the quantile $q(a)$ were distributed *exactly* as $g(z;\sigma,k)$, then integration by parts would give $E(\varepsilon_t|\varepsilon_t > q(a)) = q(a) + \frac{\sigma}{1+k}$. In the general case where the exceedances are not distributed as $g(z;\sigma,k)$, but F satisfies conditions FR1 and FR2 with $\alpha > 1$ in section 3, it can be easily shown (Lemma 7) that $E(\varepsilon_t|\varepsilon_t > q(a)) = \frac{q(a)}{1+k}(1+o(1))$. This motivates our proposed estimator for $E(\varepsilon_t|\varepsilon_t > q(a))$ which is given by

$$\widehat{E}(\varepsilon_t|\varepsilon_t > q(a)) = \frac{\hat{q}(a)}{1+\tilde{k}}. \tag{13}$$

Combining the estimators $\hat{m}$, $\hat{h}$, (12), (13) into equations (5), (6) we define the estimators $\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\hat{q}(a)$ and $\widehat{E}\left(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right) = \hat{m}(\mathbf{x}) + \hat{h}^{1/2}(\mathbf{x})\widehat{E}(\varepsilon_t|\varepsilon_t > q(a))$ for a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$) associated with the series $\{Y_t\}$ and the conditioning set $\{\mathbf{X}_t = \mathbf{x}\}$. In the next section we study the asymptotic behavior of these estimators.

3 Asymptotic characterization of the proposed estimators

3.1 Preliminaries

We start by discussing the seminal results in Smith (1985, 1987) which are the basis for deriving the asymptotic properties of our estimators and understanding our method of proof. Consider nonstochastic N and threshold $q(a_n)$ used to define a sequence of exceedances $\{Z_i'\}_{i=1}^{N_1}$ where $Z_i' = \varepsilon_{(n-N_1+i)} - q(a_n)$.[5] In addition, define $(\check{\sigma}_{q(a_n)}, \check{k})$ as a solution for the likelihood equations

$$\frac{\partial}{\partial\sigma}\frac{1}{N}\sum_{i=1}^{N_1} log\, g(Z_i';\sigma,k) = 0 \text{ and } \frac{\partial}{\partial k}\frac{1}{N}\sum_{i=1}^{N_1} log\, g(Z_i';\sigma,k) = 0 \tag{14}$$

associated with $L_N'(\sigma,k) = \frac{1}{N}\sum_{i=1}^{N_1} log\, g(Z_i';\sigma,k)$. Smith (1987, Theorem 3.2) showed that if F satisfies FR1: $L(x) = x^{\alpha}(1-F(x))$ for $\alpha > 0$ is slowly varying at infinity with $\frac{L(tx)}{L(x)} = 1+k(t)\phi(x)+o(\phi(x))$ as $x \to \infty$ for each $t > 0$, where $0 < \phi(x) \to 0$ as $x \to \infty$ is regularly varying with index $\rho \leq 0$ and $k(t) = C\int_1^t u^{\rho-1}du$, for a constant C,

and $\{Z_i'\}_{i=1}^{N_1}$ is an independent and identically distributed sequence from $F_{q(a_n)}$, then provided that

$$\frac{C}{\alpha-\rho}N^{1/2}\phi(q(a_n)) \to \mu \in \mathbb{R},$$

the estimator $(\check{\sigma}_{q(a_n)}, \check{k})$ is such that for $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = \frac{q(a_n)}{\alpha}$

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

[5] Note that N_1 may be different from N.

as N, $q(a_n) \to \infty$, where $H = \frac{1}{(1-2k_0)(1-k_0)} \begin{pmatrix} 1-k_0 & -1 \\ -1 & 2 \end{pmatrix}$.[6] As observed by Smith, the use of this theorem normally involves taking either N or $q(a_n)$ as being stochastic and the other as being nonstochastic. Throughout this paper, as in example 2 in Smith (1987, pp. 1180-1181), we take N as nonstochastic and let the threshold be sample dependent (stochastic). When $\{\varepsilon_t\}_{t=1}^n$ is observed and the threshold $q(a_n)$ is estimated by the empirical quantile $q_n(a_n)$, the estimation of the parameters of the GPD is conducted by using the sequence $\{Z_i\}_{i=1}^N$ in (8). In this case the estimators $(\hat{\sigma}_{q_n(a_n)}, \hat{k})$ are defined as solutions for

$$\frac{\partial}{\partial \sigma} \frac{1}{N} \sum_{i=1}^{N} log\, g(Z_i; \sigma, k) = 0 \text{ and } \frac{\partial}{\partial k} \frac{1}{N} \sum_{i=1}^{N} log\, g(Z_i; \sigma, k) = 0 \tag{15}$$

associated with the likelihood function $L_N(\sigma, k) = \frac{1}{N} \sum_{i=1}^{N} log\, g(Z_i; \sigma, k)$. Accounting for the stochasticity of $q_n(a_n)$ requires a further restriction on the class of distributions F we consider. Specifically, as in Davis and Resnick (1984), we assume

FR2: 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 all F that satisfy FR2 also satisfy (4) (see Proposition 1.15 in Resnick (1987)) with $\alpha = -1/k_0$ and $k_0 < 0$. Since our conditional location-scale model in (3) requires that $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = 1$, a commonly used F (in empirical Finance) that satisfy both FR1 and FR2 is a zero-centered and suitably scaled Student-t distribution with degree of freedom $v > 2$. In this case, FR2 is satisfied with $\alpha = v$.[7]

It will be convenient to reparametrize the likelihood functions and represent arbitrary values σ and k as $\sigma = \sigma_N(1 + \tau_1 \delta_N)$, $k = k_0 + \tau_2 \delta_N$ for $\tau_1, \tau_2 \in \mathbb{R}$ with $\delta_N \to 0$ as $N \to \infty$ and some σ_N and k_0. Hence, we will write the likelihood function $L'_N(\sigma, k)$ as $L'_{TN}(\tau_1, \tau_2) = \frac{1}{N} \sum_{i=1}^{N_1} log\, g(Z'_i; \sigma_N(1+\tau_1\delta_N), k_0 + \tau_2\delta_N)$. It is evident that $L'_{TN}(0,0) = L'_N(\sigma_N, k_0)$ and for $(\check{\sigma}_{q(a_n)}, \check{k})$ that satisfies (14), there are $\check{\tau}_1$ and $\check{\tau}_2$ that satisfy

$$\frac{1}{\sigma_N \delta_N} \frac{\partial L'_{TN}}{\partial \tau_1}(\tau_1, \tau_2) = 0 \text{ and } \frac{1}{\delta_N} \frac{\partial L'_{TN}}{\partial \tau_2}(\tau_1, \tau_2) = 0. \tag{16}$$

Similarly, we write $L_{TN}(\tau_1, \tau_2) = \frac{1}{N} \sum_{i=1}^{N} log\, g(Z_i; \sigma_N(1+\tau_1\delta_N), k_0 + \tau_2\delta_N)$ and observe that for $(\hat{\sigma}_{q(a_n)}, \hat{k})$ that satisfies (15), there are $\hat{\tau}_1$ and $\hat{\tau}_2$ that satisfy

$$\frac{1}{\sigma_N \delta_N} \frac{\partial L_{TN}}{\partial \tau_1}(\tau_1, \tau_2) = 0 \text{ and } \frac{1}{\delta_N} \frac{\partial L_{TN}}{\partial \tau_2}(\tau_1, \tau_2) = 0. \tag{17}$$

[6] Substituting $k_0 = -\alpha^{-1}$ shows that H is identical to the homonymous matrix in equation (19).

[7] Other distributions that satisfy FR2 and equation (4) are the Pareto, log-gamma and Burr.

The existence and characterization of a solution $(\hat{\tau}_1, \hat{\tau}_2)$ for equation (17) as maxima for L_{TN} is accomplished through Theorem 1 where we show that the likelihood equations in (16) and (17) are uniformly asymptotically equivalent in probability.

Theorem 1. *Assume FR1 and FR2. Then, as $n \to \infty$*

$$\frac{1}{\sigma_N \delta_N} \frac{\partial L_{TN}}{\partial \tau_1}(\tau_1, \tau_2) - \frac{1}{\sigma_N \delta_N} \frac{\partial L'_{TN}}{\partial \tau_1}(\tau_1, \tau_2) = o_p(1) \ \textit{and} \ \frac{1}{\delta_N} \frac{\partial L'_{TN}}{\partial \tau_2}(\tau_1, \tau_2) - \frac{1}{\delta_N} \frac{\partial L_{TN}}{\partial \tau_2}(\tau_1, \tau_2) = o_p(1)$$

Using Theorem 1 and Lemma 5 in Smith (1985) we can conclude that $\frac{1}{\delta_N^2} L_{TN}(\tau_1, \tau_2)$ has, with probability approaching 1, a local maximum $(\hat{\tau}_1, \hat{\tau}_2)$ on $S_T = \{(\tau_1, \tau_2) : \tau_1^2 + \tau_2^2 < 1\}$ at which $\frac{1}{\delta_N^2} \frac{\partial}{\partial \tau_1} L_{TN}(\hat{\tau}_1, \hat{\tau}_2) = 0$ and $\frac{1}{\delta_N^2} \frac{\partial}{\partial \tau_2} L_{TN}(\hat{\tau}_1, \hat{\tau}_2) = 0$. Put differently, there exists, with probability approaching 1, a local maximum $(\hat{\sigma}_{q_n(a_n)} = \sigma_N(1 + \hat{\tau}_1 \delta_N), \hat{k} = k_0 + \hat{\tau}_2 \delta_N)$ on $S_R = \{(\sigma, k) : \|(\frac{\sigma}{\sigma_N} - 1, k - k_0)\|_E < \delta_N\}$ that satisfies the first order conditions in equation (15).[8] Hence, Theorem 3.2 in Smith (1987) holds when the threshold $q(a_n)$ is estimated by $q_n(a_n)$. We note that Smith (1987) makes the same claim in the aforementioned example 2, but provides no detailed proof.

The first step in the study of the asymptotic behavior of our estimators is to establish that a solution for equation (10) exists and corresponds to a local maximum of the likelihood function. Our strategy will be to show that the likelihood equations associated with the reparametrized $\tilde{L}_{TN} = \frac{1}{N} \sum_{i=1}^{N_s} log\, g(\tilde{Z}_i; \sigma_N(1 + \tau_1 \delta_N), k_0 + \tau_2 \delta_N)$ are uniformly asymptotically equivalent in probability to those associated with L_{TN}. The proof is similar to that of Theorem 1, but with the added complication that the nonparametric residual sequence $\{\hat{\varepsilon}_t\}_{t=1}^n$ is used to construct the exceedances in $\tilde{L}_{TN}$. Before we give a statement of the theorem, in the next subsection we list a series of assumptions that will hold throughout the rest of the paper. Any additional conditions that are needed in specific theorems or lemmas are listed in their enunciation.

3.2 Assumptions

Besides FR1 and FR2 and the assumption that $\{\varepsilon_t\}$ is an IID process with $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = 1$, additional assumptions are needed to assure that the nonparametric estimators $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ converge uniformly in probability to $m(\mathbf{x})$ and $h(\mathbf{x})$ at suitable rates. We adopt the following notation in our assumptions

[8] $\|\mathbf{x}\|_E$ denotes the Euclidean norm of the vector $\mathbf{x}$.

and proofs: a) $0 < C < \infty$ will represent an inconsequential and arbitrary constant taking different values; b) $\mathcal{G}$ denotes a compact subset of $\mathbb{R}^d$; c) $[x]$ denotes the integer part of $x \in \mathbb{R}$; d) $P(A)$ denotes the probability of event A associated with a probability space $(\Omega, \mathcal{F}, P)$ or a probability measure, depending on the context; e) for any function $m : \mathbb{R}^d \to \mathbb{R}$ where s order partial derivatives exist, we denote by $D_i m(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ the first order partial derivatives of m with respect to its i^{th} argument for $i = 1, \cdots, d$ and the s-order partial derivatives are denoted by $D_{i_1 \cdots i_s} m(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ for $i_1, \cdots, i_s = 1, \cdots, d$. The gradient of the the function m is denoted by $m^{(1)}(\mathbf{x})$ and its Hessian by $m^{(2)}(\mathbf{x})$; f) the joint density of the vector of conditioning variables $\mathbf{X}_t$ is denoted by $f_{\mathbf{X}}(\mathbf{x})$.

Assumption A1: $K(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is a product kernel $K(\mathbf{x}) = \prod_{j=1}^d \mathcal{K}(x_j)$ with $\mathcal{K}(x) : \mathbb{R} \to \mathbb{R}$ such that: 1) $|\mathcal{K}(x)| \leq C$ for all $x \in \mathbb{R}$; 2) For some C, $\mathcal{K}(x) = 0$ whenever $|x| > C$; 3) $\int \mathcal{K}(x)dx = 1$, $\int x^j \mathcal{K}(x)dx = 0$ for $j = 1, \cdots, s-1$, $\int x^s \mathcal{K}(x)dx = \mu_{\mathcal{K},s} < \infty$; 4) $\mathcal{K}$ satisfies a Lipschitz condition of order 1, that is, for all $x, y \in \mathbb{R}$ with $x \neq y$ $|\mathcal{K}(x) - \mathcal{K}(y)| \leq C|x - y|$ for some C; 5) The kernel K_3 satisfies 1), 2) is symmetric and twice continuously differentiable in $\mathbb{R}$, $\int K_3(x)dx = 1$, $\int x^j K_3(x)dx = 0$ for $j = 1, \cdots, m_1 - 1$, $\int x^{m_1} K_3(x)dx < \infty$ and for all $x, y \in \mathbb{R}$ with $x \neq y$ we have $\left| \frac{d}{dx} K_3(x) - \frac{d}{dx} K_3(y) \right| \leq C\,|x - y|$ for some $C > 0$.

The kernel $\mathcal{K}(x)$ is used to construct $K_i(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ where $K_i(\mathbf{x}) = \prod_{j=1}^d \mathcal{K}(x_j)$ for $i = 1, 2$. Furthermore, for $j = 1, \cdots, d$ we have $\int_{\mathbb{R}^d} K_i(\mathbf{x})d\mathbf{x} = 1$, $\int x_j^l K_i(\mathbf{x})d\mathbf{x} = 0$ for $l = 1, \cdots, s-1$, $\int x_j^s K_i(\mathbf{x})d\mathbf{x} = \mu_{\mathcal{K},s}$ and $\int x_{i_1} \cdots x_{i_r} K_i(\mathbf{x})d\mathbf{x} = 0$ whenever $r < s$ or $i_j \neq i_k$ for some $j, k \leq s$. The order s for K_1 and K_2 are needed to establish that the biases for $\hat{m}$ and $\hat{h}$ are, respectively, of order $O(h_{in}^s)$ for $i = 1, 2$ in Lemmas 2 and 3. The order m_1 for K_3 is necessary in the proof of Lemma 4. All other assumption are common in the nonparametric estimation literature and are easily satisfied by a variety of commonly used kernels.

Assumption A2: 1) $\{\mathbf{X}_t\}_{t=1,2,\cdots}$ is a strictly stationary α-mixing process with $\alpha(l) \leq C\,l^{-B}$ for some $B > 2$; 2) $f_{\mathbf{X}}(\mathbf{x})$ and all of its partial derivatives of order < s are differentiable and uniformly bounded on $\mathbb{R}^d$; 3) $0 < \inf_{\mathbf{x} \in \mathcal{G}} f_{\mathbf{X}}(\mathbf{x})$.

A2 1) implies that for some $\delta > 2$ and $a > 1 - \frac{2}{\delta}$, $\sum_{j=1}^{\infty} j^a \alpha(j)^{1-\frac{2}{\delta}} < \infty$, a fact that is needed in our proofs. We note that α-mixing is the weakest of the mixing concepts (Doukhan (1994)) and its use here is only possible due to Lemma A.2 in Gao (2007), which plays a critical role in the proof of Lemma 4.

Assumption A3: 1) $m(\mathbf{x})$ and all of its partial derivatives of order $< s$ are differentiable on $\mathbb{R}^d$. The partial derivatives are uniformly bounded on $\mathbb{R}^d$; 2) $0 < h(\mathbf{x})$ and all of its partial derivatives of order $< s$ are differentiable and uniformly bounded on $\mathbb{R}^d$; 3) $E(h(\mathbf{X})^{\zeta}) < \infty$ for some $\zeta > 2$.

The degree of smoothness s of m, h and $f_{\mathbf{X}}$ (in A2 3)), the dimension d and the mixing size B are, as expected, tightly connected with the speed at which $\hat{m}$ and $\hat{h}$ converge (uniformly) to m and h. However, these parameters also interact in specific ways to determine the asymptotic behavior of $\hat{q}(a)$ and $\hat{E}(\varepsilon_t|\varepsilon_t > q(a))$.

Assumption A4: 1) The density f is m_1-times continuously differentiable with $\left|\frac{d^j}{du^j} f(u)\right| < C$ for some constant C and $j = 1, \cdots, m_1$; 2) $E(\varepsilon_t^{4+\epsilon}) < \infty$ for some $\epsilon > 0$

FR1 and A4 2 imply that $\alpha > 4$ in (4). Thus, nonparametric estimation of m and h requires a restriction on F to a class of distributions that can have thick tails, but not so thick as to prevent the existence of moments slightly larger than four. Also, A3 3 and A4 2 imply, by the c_r-Inequality, that $E\left(|Y_t|^{4+\epsilon}\right) < \infty$. Estimation of unconditional a-CES when $E(|Y_t|) < \infty$ and $V(Y_t) = \infty$ has been considered by Linton and Xiao (2013) and Hill (2014). The differentiability restrictions on f are necessary in the proof of Lemma 4.

Assumption A5: 1) The joint density of $\mathbf{X}_i, \mathbf{X}_t, \varepsilon_i$, denoted by $f_{\mathbf{X}_i,\mathbf{X}_t,\varepsilon_i}(\mathbf{X}_i, \mathbf{X}_t, \varepsilon_i)$ is continuous; 2) The joint density of $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_t, \varepsilon_i\varepsilon_j, \varepsilon_t$, denoted by $f_{\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_t,\varepsilon_i,\varepsilon_j,\varepsilon_t}(\mathbf{X}_i, \mathbf{X}_t, \mathbf{X}_t, \varepsilon_i, \varepsilon_j, \varepsilon_t)$ is continuous.

Assumption A5 is necessary in Lemma 4 and is directly related to the verification of existence of bounds required to use Lemma A.2 in Gao (2007).

Assumption A6 : 1) $h_{1n} \propto n^{-\frac{1}{2s+d}}$, $h_{2n} \propto n^{-\frac{1}{2s+d}}$, $h_{3n} \propto n^{-\frac{s}{2(2s+d)}+\delta}$, $N \propto n^{\frac{2s}{2s+d}-\delta}$ for some $\delta > 0$ and $s \geq 2d$.

3.3 Existence of $\tilde{\sigma}_N$ and $\tilde{k}$

We now establish the existence of $\tilde{\sigma}_{\tilde{q}(a_n)}$ and $\tilde{k}$ and characterize them as a local maximum. As mentioned earlier, the strategy of the proof is to show that the first order conditions associated with the likelihood function $\tilde{L}_{TN}(\tau_1, \tau_2) = \frac{1}{N}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i; \sigma_N(1 + \tau_1\delta_N), k_0 + \tau_2\delta_N)$ are asymptotically uniformly equivalent in probability to those associated with L_{TN} on the set S_T. For concreteness, we take $a_n = 1 - \frac{N}{n}$, and we formally have

Theorem 2. *Assume that FR1, FR2 with $\alpha > 4$ and A1-A6. Let $\tau_1, \tau_2 \in \mathbb{R}$, $0 < \delta_N \to 0$, $\delta_N N^{1/2} \to \infty$ as $N \to \infty$ and denote arbitrary σ and k by $\sigma = \sigma_N(1 + \tau_1\delta_N)$ and $k = k_0 + \tau_2\delta_N$. We define the log-likelihood function $\tilde{L}_{TN}(\tau_1, \tau_2) = \frac{1}{N}\sum_{i=1}^{N_s} log\, g(\tilde{Z}_i; \sigma_N(1+\tau_1\delta_N), k_0+\tau_2\delta_N)$, where $\tilde{Z}_i = \hat{\varepsilon}_{(n-N_s+i)} - \tilde{q}(a_n)$, $a_n = 1-\frac{N}{n}$, $\tilde{q}(\cdot)$ and $\hat{\varepsilon}_{(n-N_s+i)}$ are as defined in section 2. Then, as $n \to \infty$, $\frac{1}{\delta_N^2}\tilde{L}_{TN}(\tau_1, \tau_2)$ has, with probability approaching 1, a local maximum (τ_1^*, τ_2^*) on $S_T = \{(\tau_1, \tau_2) : \tau_1^2 + \tau_2^2 < 1\}$ at which $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_1}\tilde{L}_{TN}(\tau_1^*, \tau_2^*) = 0$ and $\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_2}\tilde{L}_{TN}(\tau_1^*, \tau_2^*) = 0$.*

The vector (τ_1^*, τ_2^*) implies a value $\tilde{\sigma}_{\tilde{q}(a_n)}$ and $\tilde{k}$ which are solutions for the likelihood equations

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

Hence, there exists, with probability approaching 1, a local maximum $(\tilde{\sigma}_{\tilde{q}(a_n)} = \sigma_N(1+\tau_1^*\delta_N),\ \tilde{k} = k_0+\tau_2^*\delta_N)$ on $S_R = \{(\sigma, k) : \|(\frac{\sigma}{\sigma_N} - 1,\ k - k_0)\|_E < \delta_N\}$ that satisfy the first order conditions in equation (10).

The proof of Theorem 2 depends critically on two sets of results. First, since ε_t is unobserved and is estimated by $\hat{\varepsilon}_t$ we must obtain convergence of both $\hat{m}(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ to the true $m(\mathbf{x})$ and $h(\mathbf{x})$ uniformly in $\mathcal{G}$ at suitable rates. Lemmas 2 and 3 in Appendix 2 give conditions under which we obtain

$$\sup_{\mathbf{x}\in\mathcal{G}}|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p\left(L_{1n}\right) \text{ and } \sup_{\mathbf{x}\in\mathcal{G}}|\hat{h}(\mathbf{x}) - h(\mathbf{x})| = O_p\left(L_{2n}\right),$$

where $L_{1n} = \left(\frac{log\, n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$ and $L_{2n} = \left(\frac{log\, n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^s$. These orders are sufficient to obtain that the difference between estimated residuals and true errors is given by $|\hat{\varepsilon}_t - \varepsilon_t| = O_p(L_{1n}) + (O_p(L_{1n}) + O_p(L_{2n}))|\varepsilon_t|$ uniformly in $\mathcal{G}$. Second, Lemma 4 shows that $\tilde{q}(a_n)$ is asymptotically close to $q_n(a_n)$ by satisfying $\frac{\tilde{q}(a_n)-q_n(a_n)}{q(a_n)} = O_p(N^{-1/2})$. In addition to making full use of the probability orders of $\hat{m}$ and $\hat{h}$, it is in Lemma 4 that the stochasticity of the estimated threshold $\tilde{q}$ is explicitly handled and where the restrictions (FR1, FR2 and $\alpha > 4$) on the class of functions to which F belongs are needed. It is also in Lemma 4 that the stochasticity of N_s and the fact it may differ from N in finite samples is handled by showning that $\frac{N_s-N}{N^{1/2}} = O_p(1)$. Furthermore, the proof of Lemma 4 requires that the relative speed of decay of h_{1n}, h_{2n} and h_{3n} and the speed at which $N \to \infty$ be carefully controlled. Assumption A6 1) in Theorem 2 provides polynomial functions of n that assure the orders for these sequences produce the desired $\frac{\tilde{q}(a_n)-q_n(a_n)}{q(a_n)} = O_p(N^{-1/2})$. In addition, as needed in Smith (1987), $N^{1/2}\delta_N \to \infty$ and $N^{1/2}\phi(q(a_n)) = O(1)$, where $q(a_n)$ is a positive nonstochastic sequence such that $q(a_n) \to \infty$ as $N \to \infty$.

The influence of the dimension d of the conditioning space manifests itself on the asymptotic results in a strong manner via the requirement that the degree of smoothness of the functions m and h be such that $s \geq 2d$. We believe that alleviation of this strong requirement can only result from further constraints on the class of functions containing m and h.

3.4 Asymptotic normality of $\tilde{\gamma}^T = (\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})$

The following theorem shows that under suitable normalization the estimators $(\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})$ are asymptotically distributed as a normal random variable.

Theorem 3. *Suppose FR1, FR2 with $\alpha > 4$, A1-A6 hold and that $\frac{C}{\alpha-\rho}N^{1/2}\phi(q(a_n)) \to \mu \in \mathbb{R}$. Then, the local maximum $(\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})$ of the GPD likelihood function, is such that for $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = \frac{q(a_n)}{\alpha}$*

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

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

This theorem shows that the use of $\tilde{Z}_i$ instead of Z_i to define the exceedances used in the estimation of the parameters of the GPD 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 ε_t 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 3.

3.5 Asymptotic normality of $\hat{q}(a)$, $\hat{E}(\varepsilon_t|\varepsilon_t > q(a))$, a-CVaR(x) and a-CES(x)

The asymptotic distribution of the ML estimators given in Theorem 3 is the basis for obtaining the asymptotic distributions of $\hat{q}(a)$ and $\hat{E}(\varepsilon_t|\varepsilon_t > q(a))$. The basic idea in the case of $\hat{q}(a)$ is to define, without loss of generality, $q(a) = q(a_n) + y_{\tilde{q}(a_n),a}$ for $a_n = 1 - N/n < a$ and estimate $q(a_n)$ by $\tilde{q}(a_n)$ and $y_{\tilde{q}(a_n),a}$ based on the estimated parameters of the GPD. Since $\hat{E}(\varepsilon_t|\varepsilon_t > q(a)) = \frac{\hat{q}(a)}{1+\tilde{k}}$, its asymptotic distribution can be derived directly from the results for $\hat{q}(a)$ and $\tilde{k}$. It is important to note that in Theorems 4, 5 and 6 below, both a_n and a approach 1 as $n \to \infty$ since $a_n < a$. The fact that a is not fixed and $a \to 1$ as $n \to \infty$ is only part of how we envision the asymptotic experiment guiding our theorems. Clearly, for any fixed sample size n and choice of a, the estimators (12), (13), a-CVaR($\mathbf{x}$) and a-CES($\mathbf{x}$) are unambiguously defined.

Theorem 4. *Suppose FR1, FR2 with $\alpha > 4$, A1-A6 and $\frac{C}{\alpha-\rho}N^{1/2}\phi(q(a_n)) \to \mu$ with $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = q(a_n)/\alpha$. Then, if $n(1-a) \propto N$, for some $\mathcal{Z} > 0$*

$$\sqrt{n(1-a)}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \quad \stackrel{d}{\to} \quad \mathcal{N}(\mu_1, \Sigma_1),$$

where $\mu_1 = k_0\left(\frac{(\mathcal{Z}^\rho - 1)\mu(\alpha-\rho)}{\rho} + c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right)$, $\Sigma_1 = k_0^2\left(c_b^T H^{-1} V_2 H^{-1} c_b + 2c_b^T\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)$, $c_b^T = \begin{pmatrix} -k_0^{-1}(\mathcal{Z}^{-1}-1) & k_0^{-2}log(\mathcal{Z}) + k_0^{-2}(\mathcal{Z}^{-1}-1) \end{pmatrix}$, $b_\sigma = E\left(\frac{\partial}{\partial\sigma} log\, g(Z_i;\sigma_N,k_0)\sigma_N\right)$ *and* $b_k = E\left(\frac{\partial}{\partial k} log\, g(Z_i;\sigma_N,k_0)\right)$.

Theorem 5. *Suppose FR1, FR2 with $\alpha > 4$, A1-A6 and $\frac{C}{\alpha-\rho}N^{1/2}\phi(q(a_n)) \to \mu$ with $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = q(a_n)/\alpha$. Then, if $n(1-a) \propto N$, for some $\mathcal{Z} > 0$*

$$\sqrt{n(1-a)}\left(\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a))}{\frac{q(a)}{1+k_0}} - 1\right) \quad \stackrel{d}{\to} \quad \mathcal{N}(\mu_2, \Sigma_2),$$

where $\mu_2 = k_0\frac{(\mathcal{Z}^\rho - 1)\mu(\alpha-\rho)}{\rho} + k_0 c_b^T H^{-1} \lim_{n\to\infty}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} - \frac{1}{1+k_0}\lim_{n\to\infty}\sqrt{N}\begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}$, c_b, b_σ, b_k *are as defined in Theorem 4,*

$$\Sigma_2 = k_0^2\left(c_b^T H^{-1} V_2 H^{-1} c_b + 2c_b^T\begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right) + 2\frac{k_0}{1+k_0}\eta^T V_3 \theta + \frac{1}{(1+k_0)^2}\theta^T V_3 \theta,$$

with

$$\eta^T = \begin{pmatrix} -c_b^T H^{-1} & -c_b^T H^{-1}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} & 1 \end{pmatrix}, \theta^T = \begin{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} & \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} & 0 \end{pmatrix},$$

$$V_3 = \begin{pmatrix} \frac{1}{1-2k_0} & -\frac{1}{(k_0-1)(2k_0-1)} & 0 & 0 \\ -\frac{1}{(k_0-1)(2k_0-1)} & \frac{2}{(k_0-1)(2k_0-1)} & 0 & 0 \\ 0 & 0 & k_0^2 & -k_0 \\ 0 & 0 & -k_0 & 1 \end{pmatrix}, \ b_1 = \frac{1-k_0}{k_0(2k_0-1)} \ \textit{and} \ b_2 = \frac{1}{k_0^2}\left(\frac{k_0-1}{2k_0-1} - \frac{1}{k_0-1}\right).$$

From Theorems 3 and 4 we obtain the asymptotic normality and consistency of a-CVaR($\mathbf{x}$)$\equiv \hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)$ and a-CES($\mathbf{x}$)$\equiv \hat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))$ in the following theorem.

Theorem 6. *Suppose FR1, FR2 with $\alpha > 4$, A1-A6 and $\frac{C}{\alpha-\rho}N^{1/2}\phi(q(a_n)) \to \mu$ with $k_0 = -\frac{1}{\alpha}$ and $\sigma_N = q(a_n)/\alpha$. Then, if $n(1-a) \propto N$, for some $\mathcal{Z} > 0$ we have*

a) $\sqrt{n(1-a)}\left(\frac{\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)}{q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)} - 1\right) \xrightarrow{d} \mathcal{N}(\mu_1, \Sigma_1)$, *where μ_1 and Σ_1 are as defined in Theorem 4;*

b) $\sqrt{n(1-a)}\left(\frac{\hat{E}(Y_t|Y_t>q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))}{E(Y_t|Y_t>q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))} - 1\right) \xrightarrow{d} \mathcal{N}\left(\mu_2 - \frac{\mu(\rho-\alpha)\mathcal{Z}^\rho}{(\rho-\alpha+1)\alpha}, \Sigma_2\right)$, *where μ_2 and Σ_2 are as defined in Theorem 5.*

As a direct consequence of Theorem 6 we have

$$\frac{\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)}{q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)} = 1 + o_p(1) \quad \text{and} \quad \frac{\hat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))}{E(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))} = 1 + o_p(1)$$

as $n(1-a) \to \infty$, therefore establishing consistency of the estimators.

4 Monte Carlo study

We perform a Monte Carlo study to investigate the finite sample properties of the parameter estimator $\tilde{\gamma} = (\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})^T$, the a-CVaR($\mathbf{x}$) estimator $\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)$ and the a-CES($\mathbf{x}$) estimator $\widehat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a))$. To simplify the notation, throughout this section we put $\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a) \equiv \hat{q}$, $\widehat{E}(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)) \equiv \hat{E}$ with corresponding true values given by q and E. The underlying values of a and $\mathbf{x}$ will be clear in context.

We generate data from the following location-scale model

$$Y_t = m(Y_{t-1}) + h(t)^{1/2}\varepsilon_t, t = 1, \cdots, n. \tag{18}$$

We choose $m(Y_{t-1})$ to be $sin(0.5Y_{t-1})$ and consider $h(t) = h_i(Y_{t-1}) + \theta h(t-1)$ for $i = 1, 2$, where $h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$ and $h_2(Y_{t-1}) = 1 - 0.9exp(-2Y_{t-1}^2)$. The quadratic type heteroskedasticity function $h_1(\cdot)$ has been considered in Cai and Wang (2008), where we add the $sin(\cdot)$ function to make the nonlinearity more prominent, and $h_2(\cdot)$ is considered in Martins-Filho and Yao (2006). θ is set to be 0 or 0.5. Our estimators are based on a model where $\theta = 0$, but the model with $\theta = 0.5$ and $h_1(\cdot)$ without the $sin(\cdot)$ function correspond to the popular GARCH model, and it would be interesting to investigate the performance of our estimators under this structure. Note also that h_1 is unbounded, therefore violating assumption A3. Initial values of Y_t and $h(t)$ are set to be zero and Y_t is generated recursively according to equation (18). We discard the first 1000 observations so that the samples are not heavily influenced by the choice of initial values.

We generate ε_t independently from a Student-t distribution with v degree of freedom. It can be easily shown that $k_0 = -\frac{1}{v}$, so we have $k_0 = -0.4$ for $v = 2.5$, $k_0 = -1/3$ for $v = 3$, and $k_0 = -1/20$ for $v = 20$. We note that only the case where $v = 20$ conforms to the assumptions needed to establish asymptotic normality of our estimators, but we consider the other cases to investigate the behavior of the estimators when our asymptotic results may not hold. Here, the variance of ε_t is largest with $v = 2.5$ and we expect that in this case estimation will be relatively more difficult. On the other hand, when $v = 20$ the student-t distribution resembles the normal distribution. For identification purpose, we standardize ε_t so that it has unit variance.[9]

Implementation of our estimator requires the choice of bandwidths h_{1n}, h_{2n} and h_{3n}. Since h_{1n} and h_{2n} are utilized to estimate the conditional mean and variance, we select them using the *rule-of-thumb* data driven plug-in method of Ruppert et al. (1995) and denote them by $\hat{h}_{1n}$ and $\hat{h}_{2n}$. Specifically, $\hat{h}_{1n}$ and $\hat{h}_{2n}$ are obtained from the following regressand and regressor sequences $\{Y_t, Y_{t-1}\}_{t=1}^n$ and $\{(Y_t - \hat{m}(Y_{t-1}))^2, Y_{t-1}\}_{t=1}^n$ respectively. We select h_{3n} by using the *rule-of-thumb* bandwidth $\hat{h}_{3n} = 0.79R(Y_{t-1})n^{-1/5+\delta}$ as in (2.52) of Pagan and Ullah (1999), where $R(y_{t-1})$ is the sample interquartile range of Y_{t-1} and we set $\delta = 0.01$ so that it satisfies our assumption on the bandwidth. The second order Epanechnikov kernel is used for our estimators.

In estimating the parameters, we consider both our estimators $\tilde{\gamma} = (\tilde{\sigma}_{\tilde{q}(a_n)}, \tilde{k})^T$ and Smith type esti-

[9]We have also performed our study with the log-gamma distribution, a density that is also in the domain of attraction of the Fréchet distribution. Since its support is bounded from below, it is much less commonly used to model the financial return. Though the relative rankings regarding estimators changes somewhat in specific experiment designs, we do not report these result to save space and focus on the more popular Student-t distribution for more detailed exposition.

mators $\hat{\gamma} = (\hat{\sigma}_{q_n(a_n)}, \hat{k})^T$, which utilize the true conditional mean $m(\cdot)$, variance $h(\cdot)$ and ε_t available in the simulation. Without having to estimate $m(\cdot)$ and $h(\cdot)$, we expect that Smith's estimators will perform best and serve as a benchmark to evaluate our estimator. In estimating the conditional value-at-risk (q) and expected shortfall (E), we include our estimators ($\hat{q}$, $\hat{E}$), the Smith type estimator (q^s, E^s), and the estimators ($\dot{q}, \dot{E}$) proposed by Cai and Wang (2008). We follow their instruction for implementation and utilize the theoretical optimal bandwidths available in the simulation for ($\dot{q}, \dot{E}$) to minimize the noise.

To give the readers a vivid picture of them in practice, we provide in Figure 1 a plot of the conditional value-at-risk and expected shortfall estimates evaluated at the sample mean of Y_{t-1} across different values of a. a ranges from 0.95 to 0.999 because we are interested in higher order quantiles. The estimation utilizes 1000 sample data points generated from equation (18) with $h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$, $\theta = 0$ and student-t distributed ε_t with $v = 3$ degree of freedom. We use $N = round(1000^{0.8-0.01}) = 234$ in constructing our estimates, where $round(\cdot)$ gives the nearest integer. We note that all estimators are smooth functions of a, and they seem to capture the shape of the true value-at-risk and expected shortfall well. It seems more difficult to estimate expected shortfall than value-at-risk as the gap between the estimates and the true is noticeably larger for the expected shortfall.

The performance of our estimator is fairly robust to the choice of N and we follow a simple choice of $N = round(n^{0.8-0.01})$ in the simulation for $n = 1000, 2000$ and 4000, which gives $N = 234, 405$, and 701 respectively. Thus, with n being doubled, the effective sample size N in the second stage of our estimation is less than doubled, as required by the assumption on N. Each experiment is repeated 2000 times, except for $n = 4000$ where we set the number of repetitions to be 1000. We summarize the performance of all parameter estimators in terms of their bias (B), standard deviation (S) and root mean squared error (R) in Table 1 for $\theta = 0$ and in Table 2 for $\theta = 0.5$. We consider the performance of the $a-$conditional value-at-risk and expected shortfall estimators for $a = 0.95, 0.99, 0.995$ and 0.999 evaluated at Y_n, the most recent observation in the sample. Specifically, the performances in terms of the bias (B), standard deviation (S) and relative root mean squared error (R) for $\theta = 0$ with $h_1(Y_{t-1})$ and $h_2(Y_{t-1})$ are detailed in Tables 3 and 4, and those for $\theta = 0.5$ are summarized in Tables 5 and 6. To facilitate comparison, we report the relative root mean squared error as the ratio of the root mean squared error of each estimator over that of the estimator with the smallest root mean squared error in each experiment design. To reduce the impact of extreme experiment runs, we truncate the smallest and largest 2.5% estimates from the repetitions for all estimators. As the results for $n = 2000$ are qualitatively similar, we only report detailed results for $n = 1000$ and $n = 4000$.

In the case of estimating parameters, we notice both $\hat{\gamma}$ and $\tilde{\gamma}$ overestimate (σ_N, k_0). As the sample size increases, both estimators' performance improve, in the sense that their bias, standard deviation and root mean squared error decrease. This confirms the asymptotic results in the previous section. When k_0 is decreased (smaller v in Tables 1 and 2), we generally find the bias of both $\hat{\gamma}$ and $\tilde{\gamma}$ decrease, and the standard deviation of $\tilde{\gamma}$ increases, but there is no definite pattern on the standard deviation of $\hat{\gamma}$. We think this is related to the bias and variance trade-off for parameter estimation. As mentioned above, the variance of ε_t without standardization is larger with smaller k_0, and the distribution of ε_t exhibits heavier tail behavior, thus the more representative extreme observations have a larger probability to show up in a sample, which explains the lower bias. It is generally harder to estimate σ_N than k_0, as estimates of σ_N exhibit larger root mean squared error. When $v = 2.5$ and 3, $\hat{\gamma}$ generally outperforms $\tilde{\gamma}$ in terms of smaller bias, standard deviation and root mean squared error, though the difference diminishes with larger sample size. When $v = 20$, $\tilde{\gamma}$ exhibits smaller bias, standard deviation of similar or sometimes smaller magnitudes, and its performance is very similar to $\hat{\gamma}$. The results suggest that our proposed estimator $\tilde{\gamma}$ is well supported by the nonparametric kernel estimators for the functions $m(Y_{t-1})$ and $h(Y_{t-1})$.

In the case of estimating the conditional value-at-risk and expected shortfall, we observe that performances of all estimators generally improve with the sample sizes in terms of smaller bias, standard deviation and root mean squared error, with some exceptions on the bias. It confirms our asymptotic results that our estimator for the conditional value-at-risk and expected shortfall are consistent. In the case of estimating conditional value-at-risk, $\hat{q}$ and q^s carry positive bias for $a = 0.95$ and 0.99, but exhibit negative bias for larger values of a. $\dot{q}$ shows a similar pattern for bias, with more positive bias occurrences for larger a. In the case of estimating expected shortfall, all estimators are generally negatively biased. As k_0 increases, the performance of the estimators for (q^s, E^s) and $(\hat{q}, \hat{E})$ generally improves in terms of smaller standard deviation and root mean squared error for $a \neq 0.95$. This is expected since the distribution of ε_t exhibits less heavy tails with larger k_0. However, the performance of $(\dot{q}, \dot{E})$ does not seem to depend on k_0 in a clear fashion. With a few exceptions, we notice that it is more difficult to estimate the conditional expected shortfall relative to the value-at-risk, judged by the larger bias, standard deviation and root mean squared error for all estimators across different experiment designs. It is also harder to estimate higher order conditional value-at-risk and expected shortfall, as demonstrated by the larger bias, standard deviation and root mean squared error for all estimators, with some exceptions for the bias.

Across all experiment designs, the best estimators for (q, E) are (q^s, E^s), with a few exceptions in estimating E with $a = 0.95$. Thus, the root mean square errors are constructed for the other two estimators relative to (q^s, E^s). When $a > 0.95$, we observe that $(\hat{q}, \hat{E})$ consistently outperforms $(\dot{q}, \dot{E})$ in terms of smaller standard deviation and root mean squared error, with only a few exceptions when $\theta = 0.5$. When $a = 0.95$, the advantage of $\hat{q}$ over $\dot{q}$ generally persists. In terms of estimating E, $\dot{E}$ shows smaller bias, larger standard deviation, and its root mean squared error is generally smaller than that of $\hat{E}$. We notice that the finite sample improvement could be sizable when $a > 0.95$. To illustrate, we plot in Figure 2 the relative root mean squared error of $\frac{\dot{q}}{\hat{q}}$ and that of $\frac{\dot{E}}{\hat{E}}$ across sample sizes 1000 and 4000 for $\theta = 0$. We observe that the relative root mean squared errors are all greater than one. Furthermore, as the sample size increases, the relative root mean squared error generally becomes larger, illustrating the finite sample improvement of $(\hat{q}, \hat{E})$ over $(\dot{q}, \dot{E})$ gets magnified with sample sizes. As v is increased, the advantage of $(\hat{q}, \hat{E})$ over $(\dot{q}, \dot{E})$ is more prominent. For example, in the case of estimating q, the relative root mean squared error of $\frac{\dot{q}}{\hat{q}}$ is over 3 for $v = 20$, so the reduction in the root mean squared error of $\hat{q}$ over $\dot{q}$ is more than 66%. In the case of estimating E, the relative root mean squared error $\frac{\dot{E}}{\hat{E}}$ is over 1.8 for $v = 20$, so the reduction in the root mean squared error of $\hat{E}$ over $\dot{E}$ is more than 44%.

We conclude that our estimators $(\hat{q}, \hat{E})$ have good finite sample performance and can be especially useful when estimating higher order conditional value-at-risk and expected shortfall. The results of the estimators do not change qualitatively across different values of θ, which suggest that accounting for the nonlinearity in the conditional mean and variance functions is important for estimating the high order q and E. Overall, the study suggests that utilizing the extreme value theory and properly accounting for the nonlinearity in the estimation seems to pay off in the finite samples.

The choice of N could be an important issue because the number of residuals exceeding the threshold is based on $\tilde{q}(a_n)$. We need to choose a large $\tilde{q}(a_n)$ to reduce the bias from approximating the tail distribution with GPD, but we need to keep N large (or $\tilde{q}(a_n)$ small) to control the variance of the estimates.[10] We suggested before that our estimators are relatively robust to the choice of N, and here we specifically illustrate the impact from different N's on the performance of our estimators for the 99% conditional value-at-risk and expected shortfall with a simulation. We set $n = 1000$, $\sigma_1^2(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$, $\theta = 0$ and use a student-t distributed ε_t with $v = 3$. We graph the bias and root mean squared error of $\hat{q}$ and $\hat{E}$ against $N = 20, 25, \cdots, 300$ in Figure 3. The other experiment designs give graphs of similar general pattern. We observe that $\hat{q}$ carries a small positive bias and $\hat{E}$ is generally negatively biased. As we have mentioned above, it is harder to estimate the conditional expected shortfall than the value-at-risk, judged with the larger bias and root mean squared error of $\hat{E}$. The performance of $\hat{q}$ is fairly robust with the range of N considered, with slight improvement when N is greater than 20. The bias of $\hat{E}$ seems to be smallest when N is around 40, but its magnitude is enlarged with smaller N, and grows steadily with larger N. The root mean squared error of $\hat{E}$ decreases sharply from $N = 20$ to 60 and drops further gradually until $N = 120$. It remains fairly stable for a wide range of N and eventually increases slowly for N greater than 220.

5 Empirical illustration with backtesting

We illustrate the empirical applicability of our estimators using five historical daily series $\{Y_t\}$ on the following log returns of future prices (contracts expiring between 1 and 3 months): (1) Maize from August 10, 1998 to July 28, 2004. (2) Rice from August 1, 2002 to July 18, 2008. (3) Soybean from July 25, 2006 to July 6, 2012. (4) Soft wheat (wheatcbot) from August 15, 1996 to July 31, 2002. The data are obtained from the Chicago Board of Trade. We also obtain (5) Hard wheat (wheatkcbt) of August 1, 1996 to July 18, 2002 from Kansas City Board of Trade.

[10]Note that the number of exceedances N_s over $\tilde{q}(a_n)$ is asymptotically of the same order as N, since $\sqrt{N}\left(\frac{N_s - N}{N}\right) = O_p(1)$ (Lemma 4).

To backtest on a data set $\{Y_1, Y_2, \cdots, Y_m\}$, we utilize the previous n observations $\{Y_{t-n+1}, Y_{t-n+2}, \cdots, Y_t\}$ to estimate the a-CVaR by $\hat{q}_{Y|X=Y_t}(a)$ and the a-CES by $\hat{E}(Y|Y > q_{Y|X=Y_t}(a))$ for $a = 0.95, 0.99$, and 0.995, where $0 < n < m$, $t \in T = \{n, n+1, \cdots, m-1\}$. We fix $m = 1500$, $n = 1000$, let $N = round(n^{0.8-0.01}) = 234$ and implement our estimators as in the simulation study. We provide in Figure 4 the plot of log returns of Maize futures prices against time together with the 95% conditional value-at-risk and expected shortfall estimates. Clearly our estimates respond quickly to the changing volatility in the market.

To backtest the a-CVaR estimator, we define a violation as the event $\{Y_{t+1} > q_{Y|X=Y_t}(a)\}$. Under the null hypothesis that the return dynamics of Y_t are correctly specified, $I_t \equiv \chi_{\{Y_{t+1} > q_{Y|X=Y_t}(a)\}} \sim Bernoulli(1-a)$ where χ_A is the indicator function. Consequently, $W = \sum_{t \in T} I_t \sim Binomial(m-n, 1-a)$. We perform a two sided test with the alternative hypothesis that the quantile is not correctly estimated with too many or too few violations. Since $q_{Y|X=Y_t}(a)$ is not observed, we estimate it with $\hat{q}_{Y|X=Y_t}(a)$ and construct the empirical version of the test statistic as $\hat{W} = \sum_{t \in T} \chi_{\{Y_{t+1} > \hat{q}_{Y|X=Y_t}(a)\}}$. Under the null hypothesis, the standardized test statistic $\frac{\hat{W} - (m-n)(1-a)}{\sqrt{(m-n)(1-a)a}}$ is distributed asymptotically as a standard normal. We report the violation numbers together with the p-values based on the normal distribution for our estimator on the left half of Table 7. For all five daily series and across all values of a considered, the actual number of violations are fairly close to the expected number, with large p-values indicating no rejection of the null hypothesis. The only relatively large deviation of the violation numbers from expected is for $a = 0.95$ on Maize, but its p-value is still larger than 0.1.

To backtest the a-CES we consider the normalized difference between Y_{t+1} and $E(Y|Y > q_{Y|X=Y_t}(a))$ as $r_{t+1} = \frac{Y_{t+1} - E(Y|Y > q_{Y|X=Y_t}(a))}{h^{1/2}(Y_t)} = \varepsilon_{t+1} - E(\varepsilon|\varepsilon > q(a))$. If the return dynamics are correctly specified, given that $Y_{t+1} > q_{Y|X=Y_t}(a)$, r_{t+1} is independent and identically distributed with mean zero. Since $E(Y|Y > q_{Y|X=Y_t}(a))$ is not observed, we use the estimated residuals $\{\hat{r}_{t+1} : t \in T \text{ and } Y_{t+1} > \hat{q}_{Y|X=Y_t}(a)\}$, where $\hat{r}_{t+1} = \frac{Y_{t+1} - \hat{E}(Y|Y > q_{Y|X=Y_t}(a))}{\hat{h}^{1/2}(Y_t)}$. Without making specific distributional assumptions on the residuals, we perform a one-sided bootstrap test as described in Efron and Tibshirani (1993) pp.224-227 to test the null hypothesis that the mean of the residuals is zero against the alternative that the mean is greater than zero, since underestimating $a-$conditional expected shortfall is likely to be the direction of interest. The p-values of the test for the five series across all values of a are provided on the right half of Table 7. Given 5% significance level for the test, the null hypothesis for our a-conditional expected shortfall estimator is not rejected for $a = 0.99$ and 0.995 for all the series, but it is rejected for $a = 0.95$. The empirical result seems to confirm our Monte Carlo study that our estimators can be especially useful in estimating higher order conditional value-at-risk and expected shortfall.

6 Summary and conclusion

The estimation of conditional value-at-risk and conditional expected shortfall has been the subject of much interest in both empirical finance and theoretical econometrics. Perhaps the interest is driven by the usefulness of these measures for regulators, portfolio managers and other professionals interested in an effective and synthetic tool for measuring risk. Most stochastic models used and estimators proposed for conditional value-at-risk and expected shortfall are hampered in their use by tight parametric specifications that most certainly impact performance usability. In this paper we have proposed fully nonparametric estimators for value-at-risk and expected shortfall, showed their consistency and obtained their asymptotic distribution. Our Monte Carlo study has revealed that our estimators outperform those proposed by Cai and Wang (2008) indicating that the use of the approximations provided by Extreme Value Theory may indeed prove beneficial.

We see an important direction for future research related to the contribution in this paper. The fact that we require $s \geq 2d$ presents a strong requirement on the smoothness of the location and scale functions. This perverse manifestation of the curse of dimensionality requires a solution. Perhaps restricting m and h to belong to a class of additive functions, such that $m(\mathbf{x}) = \sum_{u=1}^{d} m_u(x_u)$ and $h(\mathbf{x}) = \sum_{u=1}^{d} h_u(x_u)$ may be sufficient to relax substantially the restriction that $s \geq 2d$.

Appendix 1 - Tables and figures

TABLE 1 BIAS (B), STANDARD DEVIATION (S) AND ROOT MEAN SQUARED ERROR (R) FOR PARAMETER ESTIMATORS WITH SAMPLE SIZE $n(\times 1000)$ AND $\theta = 0$, WHERE $k_0 = -1/v$.

			$h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$						$h_2(Y_{t-1}) = 1 - 0.9exp(-2Y_{t-1}^2)$					
				σ_N			k_0			σ_N			k_0	
	v	n	B	S	R	B	S	R	B	S	R	B	S	R
$\hat{\gamma}$	2.5	1	.261	.041	.265	.137	.089	.163	.263	.041	.266	.141	.092	.168
$\tilde{\gamma}$	2.5	1	.318	.087	.330	.136	.095	.166	.283	.080	.294	.136	.110	.175
$\hat{\gamma}$	2.5	4	.219	.025	.220	.106	.052	.118	.219	.026	.220	.103	.050	.114
$\tilde{\gamma}$	2.5	4	.258	.060	.265	.099	.058	.115	.215	.052	.221	.089	.070	.113
$\hat{\gamma}$	3	1	.350	.048	.353	.152	.083	.173	.348	.049	.351	.151	.085	.173
$\tilde{\gamma}$	3	1	.374	.071	.380	.154	.086	.176	.323	.072	.331	.152	.099	.181
$\hat{\gamma}$	3	4	.296	.029	.298	.116	.050	.126	.295	.028	.296	.115	.048	.125
$\tilde{\gamma}$	3	4	.309	.046	.313	.111	.056	.124	.256	.041	.260	.109	.062	.126
$\hat{\gamma}$	20	1	.673	.069	.677	.230	.083	.244	.674	.070	.678	.231	.083	.245
$\tilde{\gamma}$	20	1	.669	.069	.673	.235	.086	.250	.614	.065	.618	.215	.084	.231
$\hat{\gamma}$	20	4	.589	.036	.590	.182	.047	.188	.593	.039	.594	.186	.051	.193
$\tilde{\gamma}$	20	4	.585	.036	.586	.183	.047	.189	.549	.035	.550	.172	.048	.179

TABLE 2 BIAS (B), STANDARD DEVIATION (S) AND ROOT MEAN SQUARED ERROR (R) FOR PARAMETER ESTIMATORS WITH SAMPLE SIZE $n(\times 1000)$ AND $\theta = 0.5$, WHERE $k_0 = -1/v$.

			$h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$						$h_2(Y_{t-1}) = 1 - 0.9exp(-2Y_{t-1}^2)$					
				σ_N			k_0			σ_N			k_0	
	v	n	B	S	R	B	S	R	B	S	R	B	S	R
$\hat{\gamma}$	2.5	1	.263	.040	.266	.140	.087	.165	.261	.041	.265	.140	.089	.166
$\tilde{\gamma}$	2.5	1	.321	.082	.332	.141	.091	.168	.306	.082	.317	.133	.104	.169
$\hat{\gamma}$	2.5	4	.218	.024	.219	.103	.051	.115	.218	.025	.219	.104	.048	.115
$\tilde{\gamma}$	2.5	4	.260	.057	.266	.099	.055	.113	.246	.057	.252	.094	.060	.112
$\hat{\gamma}$	3	1	.347	.049	.350	.149	.084	.171	.349	.050	.352	.154	.085	.176
$\tilde{\gamma}$	3	1	.371	.074	.378	.151	.088	.175	.357	.071	.364	.152	.093	.178
$\hat{\gamma}$	3	4	.295	.030	.296	.113	.048	.123	.296	.029	.297	.115	.049	.125
$\tilde{\gamma}$	3	4	.308	.047	.311	.111	.051	.122	.294	.040	.297	.109	.055	.123
$\hat{\gamma}$	20	1	.670	.070	.674	.226	.087	.242	.673	.071	.677	.230	.087	.246
$\tilde{\gamma}$	20	1	.665	.069	.669	.225	.087	.241	.653	.069	.657	.214	.088	.231
$\hat{\gamma}$	20	4	.591	.037	.593	.184	.049	.190	.592	.037	.594	.186	.051	.193
$\tilde{\gamma}$	20	4	.587	.036	.588	.178	.050	.185	.581	.037	.582	.169	.049	.176

Table 3 Bias (B), Standard deviation (S) and Relative Root Mean Squared Error (R) for conditional value-at-risk (q) and expected shortfall (E) estimators with $h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$, sample size $n(\times 1000)$, and $\theta = 0$.

$v = 2.5$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.023	.061	1	.025	.228	1	-.048	.391	1	-.631	1.149	1
$\hat{q}$	1	.040	.125	1.998	.066	.344	1.528	.009	.542	1.377	-.512	1.441	1.167
$\dot{q}$	1	.032	.266	4.082	.103	.544	2.412	.149	.773	2.000	-.204	1.932	1.482
q^s	4	.009	.031	1	.021	.115	1	-.028	.202	1	-.483	.627	1
$\hat{q}$	4	.020	.089	2.833	.064	.217	1.931	.039	.331	1.637	-.319	.872	1.172
$\dot{q}$	4	-.004	.176	5.471	-.004	.378	3.234	.019	.551	2.712	.017	1.356	1.712
E^s	1	-.446	.212	1	-.710	.649	1	-1.003	.998	1	-2.533	2.466	1
$\hat{E}$	1	-.422	.292	1.039	-.644	.822	1.086	-.907	1.239	1.085	-2.319	2.977	1.067
$\dot{E}$	1	.125	.564	1.170	.089	1.298	1.352	.014	1.936	1.368	-1.006	5.055	1.458
E^s	4	-.411	.126	1.081	-.614	.363	1	-.852	.562	1	-2.153	1.423	1
$\hat{E}$	4	-.381	.204	1.087	-.523	.516	1.030	-.716	.769	1.029	-1.832	1.853	1.010
$\dot{E}$	4	.087	.388	1	.041	.943	1.322	-.011	1.369	1.341	-.419	3.399	1.326
$v = 3$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.026	.066	1	.032	.216	1	-.038	.348	1	-.559	.906	1
$\hat{q}$	1	.039	.126	1.871	.053	.315	1.463	-.012	.470	1.343	-.524	1.105	1.148
$\dot{q}$	1	.037	.301	4.288	.146	.597	2.810	.176	.795	2.324	-.046	1.673	1.572
q^s	4	.011	.033	1	.037	.111	1	-.000	.187	1	-.370	.535	1
$\hat{q}$	4	.024	.091	2.718	.077	.205	1.870	.059	.304	1.657	-.232	.745	1.199
$\dot{q}$	4	-.008	.194	5.611	.061	.481	4.148	.080	.597	3.221	.175	1.302	2.018
E^s	1	-.521	.192	1.089	-.743	.525	1	-.988	.772	1	-2.200	1.716	1
$\hat{E}$	1	-.510	.262	1.126	-.723	.650	1.069	-.964	.932	1.069	-2.166	1.992	1.055
$\dot{E}$	1	.092	.502	1	-.124	1.090	1.206	-.245	1.554	1.254	-.778	3.790	1.386
E^s	4	-.475	.126	1.246	-.621	.321	1	-.802	.478	1	-1.757	1.104	1
$\hat{E}$	4	-.446	.201	1.241	-.543	.456	1.014	-.689	.653	1.017	-1.516	1.431	1.004
$\dot{E}$	4	.094	.383	1	-.105	.732	1.057	-.216	1.111	1.213	-.618	2.663	1.317
$v = 20$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.015	.053	1	.006	.103	1	-.027	.144	1	-.178	.267	1
$\hat{q}$	1	.011	.094	1.735	-.013	.147	1.429	-.053	.186	1.318	-.221	.305	1.175
$\dot{q}$	1	.184	.487	9.520	.517	.706	8.490	.594	.774	6.642	.628	.916	3.461
q^s	4	.003	.027	1	.020	.048	1	.009	.070	1	-.075	.141	1
$\hat{q}$	4	.004	.054	1.996	.017	.077	1.511	.004	.097	1.383	-.087	.166	1.176
$\dot{q}$	4	.153	.446	17.36	.521	.709	16.76	.625	.763	14.03	.802	.881	7.456
E^s	1	-.685	.136	1.313	-.752	.220	1	-.812	.271	1	-1.021	.408	1
$\hat{E}$	1	-.697	.181	1.355	-.781	.266	1.053	-.848	.316	1.057	-1.073	.449	1.058
$\dot{E}$	1	-.072	.527	1	-.759	1.293	1.913	-1.018	1.728	2.343	-1.559	3.191	3.229
E^s	4	-.639	.112	1.544	-.661	.145	1	-.694	.172	1	-.832	.254	1
$\hat{E}$	4	-.639	.142	1.559	-.668	.180	1.022	-.703	.207	1.025	-.847	.288	1.028
$\dot{E}$	4	-.043	.418	1	-.624	1.116	1.889	-.898	1.404	2.331	-1.444	2.476	3.294

Table 4 Bias (B), Standard deviation (S) and Relative Root Mean Squared Error (R) for conditional value-at-risk (q) and expected shortfall (E) estimators with $h_2(Y_{t-1}) = 1 - 0.9exp(-2Y_{t-1}^2)$, sample size $n(\times 1000)$, and $\theta = 0$.

$v = 2.5$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.011	.029	1	.009	.113	1	-.028	.192	1	-.313	.568	1
$\hat{q}$	1	.017	.125	4.051	.036	.299	2.659	.014	.438	2.253	-.207	1.052	1.654
$\dot{q}$	1	.009	.183	5.893	.155	.451	4.211	.191	.612	3.297	.049	1.257	1.940
q^s	4	.005	.016	1	.015	.054	1	-.005	.092	1	-.202	.290	1
$\hat{q}$	4	.009	.121	7.339	.047	.268	4.852	.050	.376	4.103	-.054	.836	2.371
$\dot{q}$	4	-.018	.119	7.302	.107	.401	7.396	.193	.559	6.392	.344	1.106	3.278
E^s	1	-.213	.123	1	-.343	.334	1	-.485	.510	1	-1.223	1.259	1
$\hat{E}$	1	-.200	.268	1.356	-.288	.638	1.462	-.400	.925	1.433	-1.000	2.149	1.350
$\dot{E}$	1	-.015	.292	1.186	-.428	.616	1.568	-.685	.904	1.611	-1.288	2.765	1.738
E^s	4	-.187	.081	1	-.272	.185	1	-.376	.281	1	-.955	.718	1
$\hat{E}$	4	-.169	.251	1.486	-.198	.533	1.728	-.257	.745	1.677	-.646	1.632	1.468
$\dot{E}$	4	.013	.223	1.095	-.384	.476	1.859	-.727	.654	2.081	-1.548	1.834	2.008
$v = 3$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.012	.037	1	.013	.120	1	-.025	.195	1	-.303	.531	1
$\hat{q}$	1	.003	.157	4.063	.002	.330	2.743	-.037	.455	2.317	-.317	.946	1.633
$\dot{q}$	1	.007	.278	7.196	.202	.576	5.067	.250	.735	3.940	.134	1.338	2.201
q^s	4	.005	.018	1	.016	.060	1	-.004	.099	1	-.195	.285	1
$\hat{q}$	4	.005	.143	7.616	.035	.283	4.610	.027	.378	3.842	-.126	.746	2.192
$\dot{q}$	4	-.013	.198	10.56	.220	.536	9.383	.324	.697	7.794	.480	1.245	3.866
E^s	1	-.278	.147	1	-.397	.333	1	-.528	.483	1	-1.167	1.080	1
$\hat{E}$	1	-.295	.314	1.370	-.416	.621	1.442	-.546	.839	1.400	-1.183	1.655	1.279
$\dot{E}$	1	-.077	.364	1.182	-.710	.690	1.909	-1.006	.989	1.972	-1.893	2.652	2.049
E^s	4	-.245	.107	1	-.320	.200	1	-.414	.285	1	-.906	.649	1
$\hat{E}$	4	-.244	.281	1.390	-.293	.516	1.570	-.369	.678	1.536	-.803	1.289	1.363
$\dot{E}$	4	-.046	.324	1.224	-.687	.552	2.331	-1.043	.680	2.478	-2.039	1.701	2.383
$v = 20$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.010	.034	1	.003	.064	1	-.018	.090	1	-.112	.172	1
$\hat{q}$	1	-.007	.124	3.519	-.008	.181	2.828	-.024	.213	2.334	-.104	.305	1.576
$\dot{q}$	1	.131	.370	11.12	.526	.577	12.15	.651	.647	10.00	.826	.792	5.593
q^s	4	.003	.018	1	.013	.033	1	.006	.047	1	-.050	.096	1
$\hat{q}$	4	-.014	.092	5.088	-.001	.128	3.591	-.006	.147	3.117	-.050	.204	1.950
$\dot{q}$	4	.104	.323	18.55	.509	.574	21.45	.657	.642	19.49	.928	.789	11.31
E^s	1	-.424	.165	1	-.465	.209	1	-.502	.241	1	-.632	.336	1
$\hat{E}$	1	-.428	.252	1.090	-.461	.322	1.103	-.493	.360	1.096	-.606	.466	1.068
$\dot{E}$	1	-.285	.429	1.132	-.772	.932	2.373	-.997	1.139	2.717	-1.613	1.333	2.922
E^s	4	-.403	.151	1	-.418	.165	1	-.439	.181	1	-.527	.238	1
$\hat{E}$	4	-.410	.221	1.081	-.419	.263	1.102	-.436	.286	1.100	-.511	.355	1.077
$\dot{E}$	4	-.184	.416	1.056	-.654	.888	2.455	-.833	1.114	2.930	-1.323	1.753	3.799

Table 5 Bias (B), Standard deviation (S) and Relative Root Mean Squared Error (R) for conditional value-at-risk (q) and expected shortfall (E) estimators with $h_1(Y_{t-1}) = 1 + 0.01y_{t-1}^2 + 0.5sin(Y_{t-1})$, sample size $n(\times 1000)$, and $\theta = 0.5$.

$v = 2.5$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.034	.087	1	.032	.322	1	-.079	.542	1	-.952	1.553	1
$\hat{q}$	1	.065	.197	2.224	.104	.518	1.634	.018	.792	1.447	-.773	2.004	1.179
$\dot{q}$	1	.006	.266	2.852	.098	.646	2.021	.123	.954	1.757	-.408	2.580	1.434
q^s	4	.013	.045	1	.039	.169	1	-.022	.294	1	-.631	.913	1
$\hat{q}$	4	.036	.158	3.466	.113	.370	2.236	.087	.548	1.880	-.392	1.357	1.273
$\dot{q}$	4	-.021	.174	3.754	.017	.453	2.621	.051	.668	2.268	.123	1.848	1.668
E^s	1	-.645	.286	1	-1.043	.875	1	-1.480	1.339	1	-3.756	3.269	1
$\hat{E}$	1	-.604	.422	1.044	-.941	1.148	1.091	-1.342	1.704	1.086	-3.480	3.991	1.063
$\dot{E}$	1	.246	.893	1.313	.097	1.704	1.254	-.074	2.466	1.236	-1.827	6.562	1.368
E^s	4	-.575	.172	1	-.842	.520	1	-1.162	.809	1	-2.932	2.056	1
$\hat{E}$	4	-.530	.325	1.035	-.711	.805	1.804	-.972	1.179	1.079	-2.520	2.749	1.041
$\dot{E}$	4	.178	.713	1.225	.080	1.292	1.307	-.001	1.766	1.247	-.712	4.541	1.283
$v = 3$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.031	.093	1	.043	.306	1	-.050	.500	1	-.762	1.324	1
$\hat{q}$	1	.050	.211	2.215	.074	.487	1.594	-.016	.716	1.426	-.726	1.670	1.192
$\dot{q}$	1	.027	.324	3.313	.169	.700	2.331	.190	.966	1.960	-.279	2.150	1.419
q^s	4	.014	.048	1	.058	.160	1	.011	.267	1	-.487	.750	1
$\hat{q}$	4	.033	.177	3.597	.110	.361	2.219	.083	.503	1.910	-.345	1.113	1.303
$\dot{q}$	4	-.028	.188	3.795	.027	.474	2.793	.077	.699	2.636	.152	1.617	1.816
E^s	1	-.738	.267	1	-1.040	.759	1	-1.376	1.125	1	-3.040	2.527	1
$\hat{E}$	1	-.725	.407	1.060	-1.019	.988	1.102	-1.353	1.411	1.100	-3.013	3.019	1.079
$\dot{E}$	1	.185	.898	1.168	-.130	1.475	1.150	-.247	2.062	1.168	-1.136	5.038	1.306
E^s	4	-.670	.159	1	-.863	.435	1	-1.107	.654	1	-2.422	1.524	1
$\hat{E}$	4	-.641	.323	1.043	-.785	.692	1.083	-1.000	.966	1.081	-2.209	2.020	1.046
$\dot{E}$	4	.179	.714	1.069	-.096	1.061	1.102	-.251	1.405	1.109	-.618	3.638	1.289
$v = 20$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.023	.077	1	.013	.148	1	-.032	.208	1	-.239	.391	1
$\hat{q}$	1	.022	.184	2.311	.008	.277	1.866	-.038	.337	1.609	-.245	.517	1.250
$\dot{q}$	1	.124	.420	5.471	.496	.718	5.878	.584	.796	4.687	.554	.991	2.480
q^s	4	.006	.038	1	.032	.074	1	.016	.107	1	-.110	.213	1
$\hat{q}$	4	.008	.148	3.807	.043	.212	2.677	.033	.245	2.300	-.073	.347	1.481
$\dot{q}$	4	.013	.219	5.642	.439	.664	9.848	.565	.728	8.556	.726	.879	4.768
E^s	1	-.976	.164	1.539	-1.067	.301	1	-1.150	.380	1	-1.440	.586	1
$\hat{E}$	1	-.977	.267	1.576	-1.072	.411	1.035	-1.155	.490	1.036	-1.444	.696	1.032
$\dot{E}$	1	-.091	.637	1	-.885	1.828	1.831	-1.208	2.482	2.280	-1.753	4.830	3.306
E^s	4	-.923	.117	2.559	-.957	.177	1	-1.006	.220	1	-1.209	.343	1.009
$\hat{E}$	4	-.912	.218	2.578	-.932	.297	1.005	-.974	.341	1.002	-1.155	.465	1
$\dot{E}$	4	.014	.364	1	-.810	1.479	1.732	-1.150	1.949	2.197	-1.804	3.586	3.222

Table 6 Bias (B), Standard deviation (S) and Relative Root Mean Squared Error (R) for conditional value-at-risk (q) and expected shortfall (E) estimators with $h_2(Y_{t-1}) = 1 - 0.9exp(-2Y_{t-1}^2)$, sample size $n(\times 1000)$, and $\theta = 0.5$.

$v = 2.5$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.019	.056	1	.015	.208	1	-.054	.354	1	-.597	1.031	1
$\hat{q}$	1	.032	.249	4.235	.076	.576	2.781	.039	.825	2.310	-.365	1.901	1.626
$\dot{q}$	1	-.018	.298	5.031	.075	.615	2.964	.083	.842	2.366	-.325	1.966	1.673
q^s	4	.007	.027	1	.020	.099	1	-.020	.171	1	-.406	.530	1
$\hat{q}$	4	.023	.240	8.528	.104	.518	5.228	.110	.713	4.181	-.101	1.501	2.252
$\dot{q}$	4	-.026	.211	7.539	.050	.522	5.194	.110	.739	4.333	.115	1.536	2.307
E^s	1	-.403	.209	1	-.651	.595	1	-.922	.909	1	-2.323	2.239	1
$\hat{E}$	1	-.375	.490	1.359	-.534	1.148	1.436	-.740	1.646	1.394	-1.873	3.708	1.288
$\dot{E}$	1	.060	.592	1.310	-.386	1.147	1.372	-.664	1.724	1.427	-1.575	4.949	1.610
E^s	4	-.356	.129	1	-.527	.321	1	-.730	.490	1	-1.844	1.239	1
$\hat{E}$	4	-.317	.448	1.447	-.372	.948	1.650	-.496	1.309	1.592	-1.307	2.759	1.374
$\dot{E}$	4	.083	.485	1.297	-.304	.939	1.599	-.639	1.222	1.568	-1.436	3.460	1.686

$v = 3$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.020	.067	1	.018	.218	1	-.054	.353	1	-.575	.925	1
$\hat{q}$	1	.043	.280	4.026	.081	.574	2.648	.031	.777	2.179	-.413	1.565	1.486
$\dot{q}$	1	.001	.336	4.777	.142	.691	3.220	.160	.897	2.552	-.157	1.838	1.693
q^s	4	.009	.034	1	.034	.110	1	-.002	.184	1	-.357	.524	1
$\hat{q}$	4	.028	.257	7.437	.106	.510	4.516	.102	.672	3.688	-.146	1.277	2.027
$\dot{q}$	4	-.026	.244	7.056	.095	.591	5.192	.159	.776	4.299	.205	1.494	2.378
E^s	1	-.516	.216	1	-.742	.549	1	-.986	.803	1	-2.181	1.784	1
$\hat{E}$	1	-.489	.488	1.237	-.660	.994	1.293	-.870	1.345	1.259	-1.946	2.656	1.169
$\dot{E}$	1	-.044	.679	1.217	-.643	1.246	1.519	-.910	1.850	1.621	-1.509	4.966	1.842
E^s	4	-.463	.148	1	-.604	.327	1	-.778	.479	1	-1.701	1.102	1
$\hat{E}$	4	-.434	.448	1.282	-.498	.851	1.436	-.624	1.119	1.403	-1.389	2.111	1.246
$\dot{E}$	4	-.090	.605	1.257	-.669	.999	1.750	-.996	1.300	1.793	-1.663	3.155	1.759

$v = 20$		$a = 0.95$			$a = 0.99$			$a = 0.995$			$a = 0.999$		
	n	B	S	R	B	S	R	B	S	R	B	S	R
q^s	1	.019	.059	1	.008	.123	1	-.029	.171	1	-.199	.318	1
$\hat{q}$	1	.014	.228	3.666	.026	.337	2.749	.002	.392	2.259	-.130	.549	1.503
$\dot{q}$	1	.055	.332	5.398	.316	.583	5.399	.386	.659	4.398	.430	.806	2.432
q^s	4	.006	.031	1	.023	.057	1	.009	.082	1	-.094	.168	1
$\hat{q}$	4	-.003	.187	5.923	.040	.272	4.477	.041	.309	3.761	-.017	.409	2.128
$\dot{q}$	4	.003	.238	7.523	.267	.512	9.415	.382	.592	8.492	.560	.749	4.858
E^s	1	-.778	.170	1.426	-.853	.271	1	-.920	.333	1	-1.156	.498	1
$\hat{E}$	1	-.764	.320	1.483	-.811	.456	1.040	-.863	.526	1.033	-1.059	.707	1.012
$\dot{E}$	1	-.126	.544	1	-.661	1.673	2.010	-.932	2.169	2.413	-1.910	3.091	2.887
E^s	4	-.732	.133	2.109	-.760	.174	1	-.800	.206	1	-.963	.304	1.009
$\hat{E}$	4	-.719	.270	2.177	-.715	.361	1.027	-.737	.405	1.018	-.852	.524	1
$\dot{E}$	4	-.024	.352	1	-.534	1.348	1.858	-.785	1.866	2.450	-1.552	3.104	3.468

Table 7 Backtest results for $a-$conditional Value-at-Risk (q) and expected shortfall(E) on $m-n=500$ observations, expected violations $=(m-n)(1-a)$. q: Number of violations and p-value (in brackets). E: p-value for exceedance residuals to have zero mean.

	q			E		
	$a=0.95$	$a=0.99$	$a=0.995$	$a=0.95$	$a=0.99$	$a=0.995$
	Expected violations					
	25	5	2.5			
Maize	18 (.151)	5(1)	2(.751)	0	.161	.735
Rice	29(.412)	4(.653)	2(.751)	0	.081	.248
Soybean	21(.412)	3(.369)	2(.751)	0	.302	.244
Wheatcbot	30(.305)	6(.653)	2(.751)	.001	.339	.273
Wheatkcbt	25(1)	5(1)	2(.751)	0	.082	.239

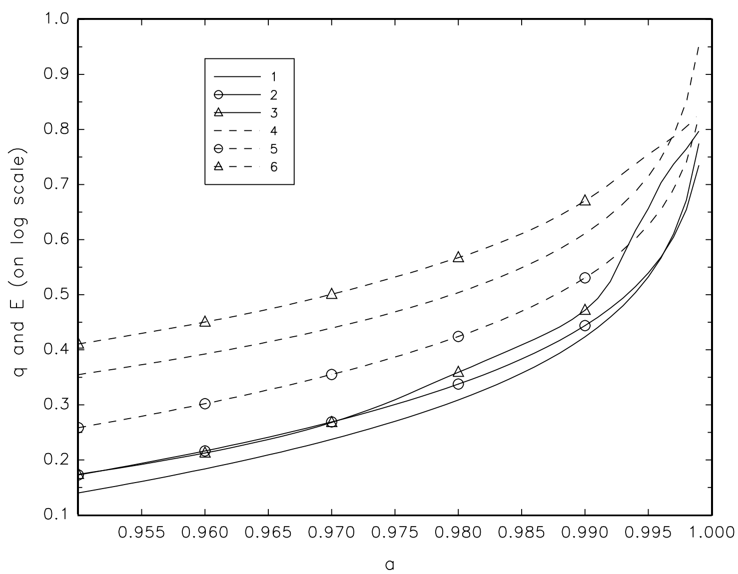

Figure 1: Plot of conditional value-at-risk (q) and expected shortfall (E) estimates evaluated at sample mean across different a, with $n=1000$, $h_1(Y_{t-1})=1+0.01Y_{t-1}^2+0.5sin(Y_{t-1})$, $\theta=0$ and student-t distributed ε_t with $v=3$. 1 : true q, 2 : $\hat{q}$, 3 : $\dot{q}$, 4 : true E, 5 : $\hat{E}$, and 6 : $\dot{E}$.

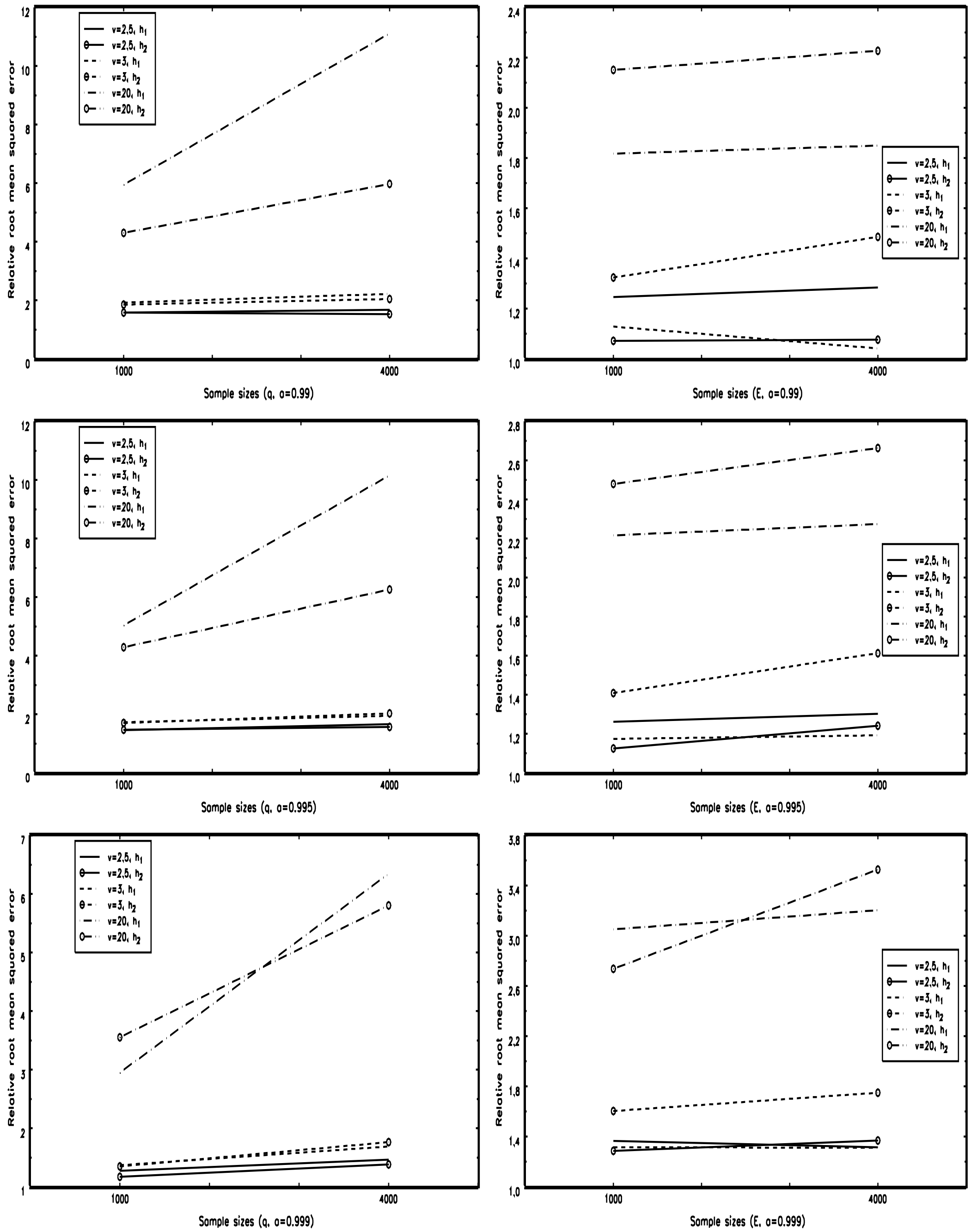

Figure 2: Relative root mean squared error of $\frac{\dot{q}}{\hat{q}}$ (left) and $\frac{\dot{E}}{\hat{E}}$ (right) across sample sizes 1000 and 4000 for $\theta = 0$, student-t distributed ϵ_t with v degree of freedom and $h_i(Y_{t-1})$ for $i = 1, 2$.

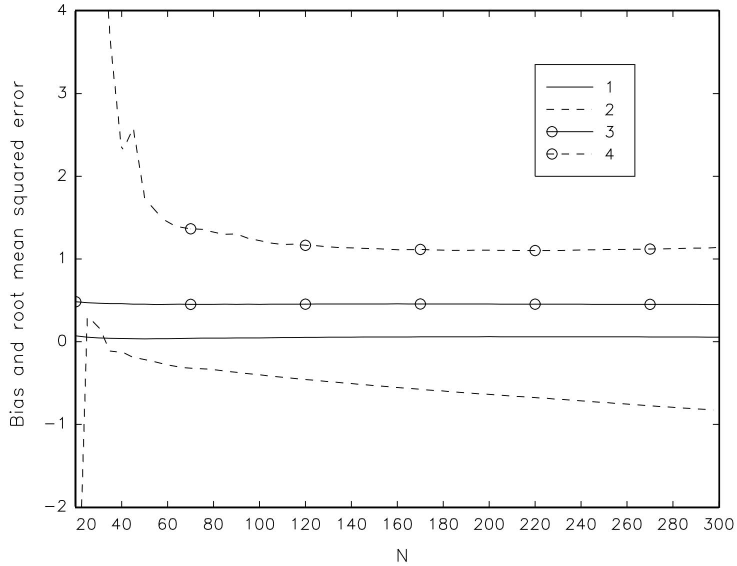

Figure 3: Bias and root mean squared error (R) of 99% conditional value-at-risk ($\hat{q}$) and expected shortfall ($\hat{E}$) estimators with different N with $n = 1000$, $h_1(y_{t-1})$, $\theta = 0$ and student-t distributed ϵ_t with $v = 3$. 1 : bias of $\hat{q}$, 2: bias of $\hat{E}$, 3: R of $\hat{q}$, and 4: R of $\hat{E}$.

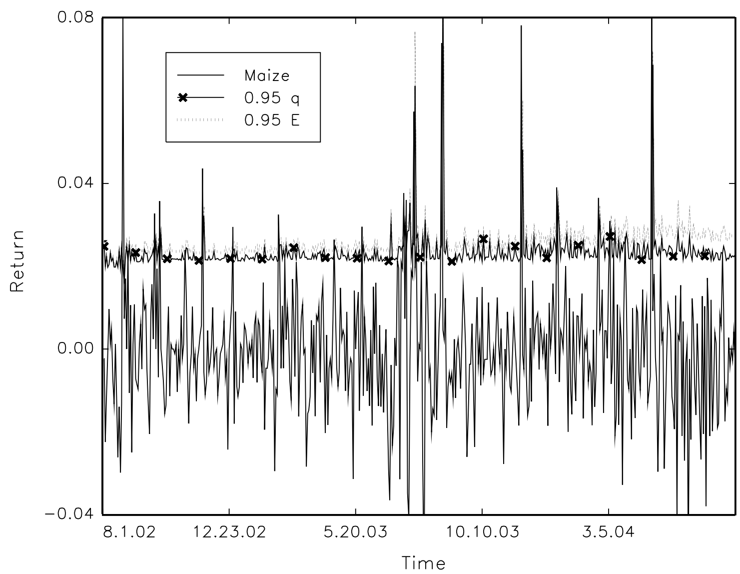

Figure 4: Plot of the log return for Maize future prices from Aug. 1, 2002 to July 28, 2004, together with the 95% conditional value-at-risk (dashed line) and expected shortfall (dotted line) estimates.

Appendix 2 - Lemmas and proofs

We rely, throughout the proofs, on some results from Smith (1985) and Smith (1987). For a nonstochastic positive sequence $q(a_n) \to \infty$ as $N \to \infty$ and for $\sigma_N = q(a_n)/\alpha$, $0 < \alpha = -1/k_0$ and $k_0 < 0$ we have $E\left(\sigma_N \frac{\partial}{\partial\sigma} log\, g(Z;\sigma_N,k_0)\right) = \frac{C\phi(q(a_n))}{(1+\alpha-\rho)} + o(\phi(q(a_n)))$, $E\left(\frac{\partial}{\partial k} log\, g(Z;\sigma_N,k_0)\right) = -\frac{\alpha C\phi(q(a_n))}{(\alpha-\rho)(1+\alpha-\rho)} + o(\phi(q(a_n))$, $E\left(\sigma_N^2 \frac{\partial^2}{\partial\sigma^2} log\, g(Z;\sigma_N,k_0)\right) = -\frac{\alpha}{2+\alpha} + O(\phi(q(a_n)))$, $E\left(\frac{\partial^2}{\partial k^2} log\, g(Z;\sigma_N,\ k_0)\right) = -\frac{2\alpha^2}{(1+\alpha)(2+\alpha)} + O(\phi(q(a_n)))$ and $E\left(\sigma_N \frac{\partial^2}{\partial\sigma\partial k} log\, g(Z;\sigma_N,k_0)\right) = \frac{\alpha^2}{(1+\alpha)(2+\alpha)} + O(\phi(q(a_n)))$, where all expectations are taken with respect to the unknown distribution $F_{q(a_n)}$. Evidently, these approximations are based on a sequence of thresholds $q(a_n)$ that approach the end point of the distribution F as the $N \to \infty$.

Theorem 1.

Proof. For a sample $\{\varepsilon_t\}_{t=1}^n$ and nonstochastic $N < n$ such that $a_n = 1 - \frac{N}{n}$ we denote $E = \{t : \varepsilon_t > q_n(a_n)\}$ and $E' = \{t : \varepsilon_t > q(a_n)\}$. The number of elements in E and E' are denoted by N and N_1. Using Taylor's Theorem we expand (16) around $(0,0)$ such that,

$$\begin{aligned}
\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_1}L'_{TN}(\tau_1,\tau_2) &= \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial}{\partial\sigma}log\, g(Z_i';\sigma_N,k_0)\frac{\sigma_N}{\delta_N} \\
&+ \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial^2}{\partial\sigma^2}log\, g(Z_i';\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N^2\tau_1 \\
&+ \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial^2}{\partial\sigma\partial k}log\, g(Z_i';\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N\tau_2 = I'_{1N} + I'_{2N} + I'_{3N}
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_2}L'_{TN}(\tau_1,\tau_2) &= \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial}{\partial k}log\, g(Z_i';\sigma_N,k_0)\frac{1}{\delta_N} \\
&+ \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial^2}{\partial k\partial\sigma}log\, g(Z_i';\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N\tau_1 \\
&+ \frac{1}{N}\sum_{i=1}^{N_1}\frac{\partial^2}{\partial k^2}log\, g(Z_i';\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\tau_2 = I'_{4N} + I'_{5N} + I'_{6N},
\end{aligned}$$

where $\lambda_1, \lambda_2 \in (0,1)$ and the terms I'_{lN} for $l = 1, \cdots, 6$ denote the corresponding averages in the preceding equality. Similar expressions are defined as I_{lN} for $l = 1, \cdots, 6$ by replacing Z_i' with Z_i and N_1 with N. 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 $S_T = \{(\tau_1, \tau_2) : \tau_1^2 + \tau_2^2 < 1\}$. Consequently, $\frac{1}{\delta_N^2}\frac{\partial}{\partial \tau_1} L'_{TN}(\tau_1, \tau_2) \xrightarrow{p} \tau_1\left(-\frac{\alpha}{1+\alpha}\right) + \tau_2\left(\frac{\alpha^2}{(1+\alpha)(2+\alpha)}\right)$, $\frac{1}{\delta_N^2}\frac{\partial}{\partial \tau_2} L'_{TN}(\tau_1, \tau_2) \xrightarrow{p} \tau_1\left(\frac{\alpha^2}{(1+\alpha)(2+\alpha)}\right) + \tau_2\left(-\frac{2\alpha^2}{(1+\alpha)(2+\alpha)}\right)$, 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

$$\begin{pmatrix} \tau_1 & \tau_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\delta_N^2}\frac{\partial}{\partial \tau_1} L'_{TN}(\tau_1, \tau_2) \\ \frac{1}{\delta_N^2}\frac{\partial}{\partial \tau_2} L'_{TN}(\tau_1, \tau_2) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \tau_1 & \tau_2 \end{pmatrix} (-H) \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \le 0 \text{ on } S_T. \tag{19}$$

We will establish that $I_{lN} - I'_{lN} = o_p(1)$ for each $l = 1, \cdots, 6$.

$$\begin{aligned} I_{1N} - I'_{1N} &= \frac{1}{\delta_N}(k_0^{-1} - 1)\left(\frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N} - \frac{1}{N}\sum_{i=1}^{N_1}\left(1 - \frac{k_0 Z'_i}{\sigma_N}\right)^{-1}\frac{k_0 Z'_i}{\sigma_N}\right) + \frac{1}{\delta_N}\left(\frac{N_1 - N}{N}\right) \\ &= \frac{1}{\delta_N}(k_0^{-1} - 1) I_{11n} + \frac{1}{\delta_N} I_{12n} \end{aligned}$$

Since $\frac{N_1 - N}{N^{1/2}} = O_p(1)$ (see Lemma 4) and $\delta_N N^{1/2} \to \infty$, $\frac{1}{\delta_N} I_{12n} = o_p(1)$.

Case 1: $q_n(a_n) < q(a_n)$. Then, $E' \subset E$, $N_1 \le N$ and

$$\frac{1}{\delta_N} I_{11n} = \frac{1}{\delta_N}\left(\frac{1}{N}\sum_{i=1}^{N_1}\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N} - \frac{1}{N}\sum_{i=1}^{N_1}\left(1 - \frac{k_0 Z'_i}{\sigma_N}\right)^{-1}\frac{k_0 Z'_i}{\sigma_N} + \frac{1}{N}\sum_{i=N_1+1}^{N}\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right).$$

Let $\Psi(z) = \left(1 - \frac{k_0}{\sigma_N} z\right)^{-1}\frac{k_0 z}{\sigma_N}$. By the Mean Value Theorem there exists $\lambda_i \in (0, 1)$ and $Z_i^* = Z_i + \lambda_i(Z'_i - Z_i)$ such that $\Psi(Z_i) - \Psi(Z'_i) = \left(1 - \frac{k_0}{\sigma_N} Z_i^*\right)^{-2}\frac{q_n(a_n) - q(a_n)}{q(a_n)}$ since $q(a_n) = -\sigma_N/k_0$. From Lemma 4, $\frac{q_n(a_n) - q(a_n)}{q(a_n)} = O_p(N^{-1/2})$ and $\left(1 - \frac{k_0}{\sigma_N} Z_i^*\right) > 1$ since $Z_i^* > 0$. Hence, $\Psi(Z_i) - \Psi(Z'_i) = O_p(N^{-1/2})$. Now, $\left|\frac{1}{N}\sum_{i=N_1+1}^{N}\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right| \le \frac{1}{N}\sum_{i=N_1+1}^{N}\left|\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N}\right| \le \frac{N - N_1}{N} = N^{-1/2} O_p(1)$ since for all $Z_i > 0$, $\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N} < 1$. Consequently, $\frac{1}{\delta_N} I_{11n} = \frac{N_1}{N\delta_N} O_p(N^{-1/2}) + \frac{1}{\delta_N} O_p(N^{-1/2}) = o_p(1)$ since $\delta_N N^{1/2} \to \infty$.

Case 2: $q(a_n) < q_n(a_n)$. Then, $E \subset E'$, $N \le N_1$ and

$$\frac{1}{\delta_N} I_{11n} = \frac{1}{\delta_N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N} - \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{k_0 Z'_i}{\sigma_N}\right)^{-1}\frac{k_0 Z'_i}{\sigma_N} - \frac{1}{N}\sum_{i=N+1}^{N_1}\left(1 - \frac{k_0 Z'_i}{\sigma_N}\right)^{-1}\frac{k_0 Z'_i}{\sigma_N}\right).$$

Using the same arguments as in case 1, we have $\frac{1}{\delta_N} I_{11n} = \frac{1}{\delta_N} O_p(N^{-1/2}) = o_p(1)$.

$$\tilde{I}_{4N} - I_{4N} = -\frac{1}{\delta_N}\frac{1}{k_0^2}\left(\frac{1}{N}\sum_{i=1}^{N} log\left(1-\frac{k_0 Z_i}{\sigma_N}\right) - \frac{1}{N}\sum_{i=1}^{N_1} log\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)\right)$$
$$+\frac{1}{k_0}\frac{1}{\delta_N}\left(1-k_0^{-1}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{k_0 Z_i}{\sigma_N}\right)^{-1}\frac{k_0 Z_i}{\sigma_N} - \frac{1}{N}\sum_{i=1}^{N_1}\left(1-\frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\frac{k_0 Z_i'}{\sigma_N}\right)$$
$$= -\frac{1}{\delta_N}\frac{1}{k_0^2} I_{41n} + \frac{1}{k_0}\frac{1}{\delta_N}\left(1-k_0^{-1}\right) I_{11n}.$$

From the study of $I_{1N} - I'_{1N}$, $\frac{1}{\delta_N} I_{11n} = o_p(1)$.

Case 1: $q_n(a_n) < q(a_n)$. Then, $E' \subseteq E$, $N_1 \leq N$. Let $Q(z) = log\left(1-\frac{k_0}{\sigma_N}z\right)$ and write $I_{41n} = \frac{1}{N}\sum_{i=1}^{N_1}(Q(Z_i) - Q(Z_i')) + \frac{1}{N}\sum_{i=N_1+1}^{N} Q(Z_i)$. By the Mean Value Theorem there exists $\lambda_i \in (0,1)$ and $Z_i^* = Z_i + \lambda_i(Z_i' - Z_i)$ such that $Q(Z_i) - Q(Z_i') = \left(1-\frac{k_0}{\sigma_N}Z_i^*\right)^{-1}\frac{q_n(a_n)-q(a_n)}{q(a_n)} = \left(1+\frac{Z_i^*}{q(a_n)}\right)^{-1} O_p(N^{-1/2}) = O_p(N^{-1/2})$, since $Z_i^* > 0$. Now, given that $q_n(a_n) < \varepsilon_i < q(a_n)$

$$\frac{1}{N}\sum_{i=N_1+1}^{N} Q(Z_i) = \frac{1}{N}\sum_{i=N_1+1}^{N} log\left(1+\frac{Z_i}{q(a_n)}\right) \leq \frac{1}{N}\sum_{i=N_1+1}^{N}\frac{Z_i}{q(a_n)} \leq \frac{1}{N}\sum_{i=N_1+1}^{N}\left(\frac{q(a_n)-q_n(a_n)}{q(a_n)}\right)$$
$$= \frac{N-N_1}{N} O_p(N^{-1/2}).$$

Case 2: $q(a_n) < q_n(a_n)$. Then, $E \subseteq E'$, $N \leq N_1$ and $I_{41n} = \frac{1}{N}\sum_{i=1}^{N}(Q(Z_i) - Q(Z_i')) - \frac{1}{N}\sum_{i=N+1}^{N_1} Q(Z_i') = O_p(N^{-1/2}) - \frac{1}{N}\sum_{i=N+1}^{N_1} Q(Z_i')$. $\frac{1}{N}\sum_{i=N+1}^{N_1} Q(Z_i') \leq \frac{1}{N}\sum_{i=N+1}^{N_1}\frac{Z_i'}{q(a_n)} = O_p(N^{-1/2})$. Hence, $\frac{I_{41n}}{\delta_N} = o_p(1)$.

$$I_{2N} - I'_{2N} = \frac{\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\frac{N-N_1}{N}$$
$$-\frac{\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\left(\frac{1}{\dot{k}}-1\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^2 - \frac{1}{N}\sum_{i=1}^{N_1}\left(1-\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^2\right)$$
$$-\frac{2\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\left(\frac{1}{\dot{k}}-1\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right) - \frac{1}{N}\sum_{i=1}^{N_1}\left(1-\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)\right)$$
$$= \frac{\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\frac{N-N_1}{N} - \frac{\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\left(\frac{1}{\dot{k}}-1\right) I_{1n} - \frac{2\tau_1}{(1+\delta_N\tau_1\lambda_1)^2}\left(\frac{1}{\dot{k}}-1\right) I_{2n}.$$

$\frac{\tau_1}{N^{1/2}(1+\delta_N\tau_1\lambda_1)^2}\frac{N-N_1}{N^{1/2}} = o_p(1)$ since $\delta_N \to 0$ and $\frac{N-N_1}{N^{1/2}} = O_p(1)$. Let $\zeta_l(z) = \left(1-\frac{\dot{k}z}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}z}{\dot{\sigma}_N}\right)^l$ for $l = 1, 2$. Then, it suffices to establish that $I_{ln} = \frac{1}{N}\sum_{i=1}^{N}\zeta_l(Z_i) - \frac{1}{N}\sum_{i=1}^{N_1}\zeta_l(Z_i) = o_p(1)$ for $l = 1, 2$ uniformly in S_T.

Case 1: $q_n(a_n) < q(a_n)$. Then, $E' \subseteq E$, $N_1 \leq N$ and $I_{ln} = \frac{1}{N}\sum_{i=1}^{N_1}(\zeta_l(Z_i) - \zeta_l(Z_i')) + \frac{1}{N}\sum_{i=N_1+1}^{N}\zeta_l(Z_i)$.

By the Mean Value Theorem there exists $\lambda_i \in (0,1)$ and $Z_i^* = Z_i + \lambda_i(Z_i' - Z_i)$ such that

$$\begin{aligned}
\zeta_l(Z_i) - \zeta_l(Z_i') &= l\frac{\dot{k}}{\dot{\sigma}_N}\left(1 - \frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-l-1}\left(\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{l-1}\frac{q(a_n) - q_n(a_n)}{q(a_n)}q(a_n) \\
&= l\left(1 - \frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-l-1}\left(\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{l-1}\frac{-\dot{k}}{\dot{\sigma}_N}\frac{\sigma_N}{k_0}O_p(N^{-1/2}) \\
|\zeta_l(Z_i) - \zeta_l(Z_i')| &\leq l\sup_{S_T}\left|\left(1 - \frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-l-1}\left(\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{l-1}\right|\sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}\frac{\sigma_N}{k_0}\right|O_p(N^{-1/2}) \\
&= O_p(N^{-1/2}) \text{ for } l = 1, 2
\end{aligned}$$

since $\sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}\frac{\sigma_N}{k_0}\right| < C$ and $\sup_{S_T}\left|\left(1 - \frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-l-1}\left(\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{l-1}\right| < C$ for n sufficiently large. Now,

$$N^{-1}\sum_{i=N_1+1}^{N}\zeta_l(Z_i) = N^{-1}\sum_{i=N_1+1}^{N}\left(1 - \frac{\dot{k}}{\dot{\sigma}_N}Z_i\right)^{-l}\left(\frac{\dot{k}}{\dot{\sigma}_N}Z_i\right)^{l} \leq \frac{N - N_1}{N}C = O_p(N^{-1/2}).$$

Hence, $I_{ln} = \frac{N_1}{N}O_p(N^{-1/2}) + \frac{1}{N^{1/2}}O_p(1) = o_p(1)$.

Case 2: $q(a_n) < q_n(a_n)$. Then, $E \subseteq E'$, $N \leq N_1$ and $I_{ln} = \frac{1}{N}\sum_{i=1}^{N}\left(\zeta_l(Z_i) - \zeta_l(Z_i')\right) - \frac{1}{N}\sum_{i=N+1}^{N_1}\zeta_l(Z_i')$ and we have $I_{ln} = o_p(1)$ uniformly on S_T following the same arguments as in Case 1.

Now, we note that $I_{3N} - I_{3N}' = I_{5N} - I_{5N}'$ and write

$$\begin{aligned}
I_{3N} - I_{3N}' = \;& \frac{\tau_2}{1 + \delta_N\tau_1\lambda_1}\frac{1}{N}\sum_{i=1}^{N}\left(-\frac{1}{\dot{k}}\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_i}{\dot{\sigma}_N} + \frac{1}{\dot{k}}\left(\frac{1}{\dot{k}} - 1\right)\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^2\right) \\
& + \frac{1}{1 + \delta_N\tau_1\lambda_1}\frac{1}{N}\sum_{i=1}^{N_1}\left(\frac{1}{\dot{k}}\left(1 - \frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_i'}{\dot{\sigma}_N} - \frac{1}{\dot{k}}\left(\frac{1}{\dot{k}} - 1\right)\left(1 - \frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^2\right) \\
= \;& o_p(1)
\end{aligned}$$

uniformly on S_T from the arguments used to study $I_{2N} - I_{2N}'$. Lastly, we write

$$\begin{aligned}
I_{6N} - I_{6N}' = \tau_2 & \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{2}{\dot{k}^3}log\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right) + \frac{2}{\dot{k}^3}\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right)\right. \\
& - \frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}} - 1\right)\left(1 - \frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)^2\Bigg) - \frac{1}{N}\sum_{i=1}^{N_1}\left(\frac{2}{\dot{k}^3}log\left(1 - \frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right) + \frac{2}{\dot{k}^3}\left(1 - \frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)\right.
\end{aligned}$$

$$- \frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_i'}{\dot{\sigma}_N}\right)^2\Bigg)\Bigg) = \frac{2\tau_2}{\dot{k}^3}\left(\frac{1}{N}\sum_{i=1}^{N} log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right) - \frac{1}{N}\sum_{i=1}^{N_1} log\left(1-\frac{\dot{k}Z_i}{\dot{\sigma}_N}\right)\right) + o_p(1)$$

uniformly in S_T by the study of $I_{2N} - I'_{2N}$.

$$= \frac{2\tau_2}{\dot{k}^3} I_{6n} + o_p(1)$$

Case 1: $q_n(a_n) < q(a_n)$. Then, $E' \subseteq E$, $N_1 \leq N$ and letting $\dot{Q}(z) = log(1 - \frac{\dot{k}}{\dot{\sigma}}z)$ we write $I_{6n} = \frac{1}{N}\sum_{i=1}^{N_1}(\dot{Q}(Z_i) - \dot{Q}(Z_i')) + \frac{1}{N}\sum_{i=N_1+1}^{N}\dot{Q}(Z_i)$. By the Mean Value Theorem there exists $\lambda_i \in (0,1)$ and $Z_i^* = Z_i + \lambda_i(Z_i' - Z_i) > 0$ such that

$$|\dot{Q}(Z_i) - \dot{Q}(Z_i')| \leq \sup_{S_T}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-1}\right| \sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}\frac{\sigma_N}{k_0}\right| \frac{q(a_n) - q_n(a_n)}{q(a_n)} = O_p(N^{-1/2})$$

since $\sup_{S_T}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}Z_i^*\right)^{-1}\right| < C$ and $\sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}\frac{\sigma_N}{k_0}\right| < C$ for n sufficiently large. Hence, $\frac{1}{N}\sum_{i=1}^{N_1}(\dot{Q}(Z_i) - \dot{Q}(Z_i')) \leq \frac{N_1}{N}O_p(N^{-1/2}) = o_p(1)$. Now,

$$\frac{1}{N}\sum_{i=N_1+1}^{N}\dot{Q}(Z_i) \leq \frac{1}{N}\sum_{i=N_1+1}^{N}\left|\frac{\dot{k}}{\dot{\sigma}_N}\right| Z_i \leq C\frac{1}{N}\sum_{i=N_1+1}^{N}\frac{Z_i}{q(a_n)} = C\frac{N-N_1}{N}O_p(N^{-1/2}) = O_p(N^{-1})$$

from the study of $I_{4N} - I'_{4N}$. Similar arguments establish the the desired order in case 2 when $q(a_n) < q_n(a_n)$. □

Theorem 2.

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

$$\begin{aligned}\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_1}\tilde{L}_{TN}(\tau_1,\tau_2) &= \frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial}{\partial\sigma} log\, g(\tilde{Z}_i;\sigma_N,k_0)\frac{\sigma_N}{\delta_N} \\ &+ \frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial^2}{\partial\sigma^2} log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N^2\tau_1 \\ &+ \frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial^2}{\partial\sigma\partial k} log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N\tau_2 = \tilde{I}_{1N} + \tilde{I}_{2N} + \tilde{I}_{3N}\end{aligned}$$

and

$$\frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_2}\tilde{L}_{TN}(\tau_1,\tau_2) = \frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial}{\partial k}log\, g(\tilde{Z}_i;\sigma_N,k_0)\frac{1}{\delta_N}$$
$$+\frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial^2}{\partial k\partial\sigma}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\sigma_N\tau_1$$
$$+\frac{1}{N}\sum_{i=1}^{N_s}\frac{\partial^2}{\partial k^2}log\, g(\tilde{Z}_i;\sigma_N(1+\delta_N\tau_1\lambda_1),k_0+\delta_N\tau_2\lambda_2)\tau_2 = \tilde{I}_{4N}+\tilde{I}_{5N}+\tilde{I}_{6N}.$$

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

$$\chi(\varepsilon)=\begin{cases}1 & \text{if } \varepsilon-q_n(a_n)>0\\ 0 & \text{if } \varepsilon-q_n(a_n)\le 0\end{cases},\tilde{\chi}(\varepsilon)=\begin{cases}1 & \text{if } \varepsilon-\tilde{q}(a_n)>0\\ 0 & \text{if } \varepsilon-\tilde{q}(a_n)\le 0\end{cases},$$

$$\chi_I(\varepsilon)=\begin{cases}1 & \text{if } \varepsilon-\tilde{q}(a_n)>0,\ \varepsilon-q_n(a_n)>0\\ 0 & \text{otherwise}\end{cases} \text{ and } \chi_D(\varepsilon)=\begin{cases}1 & \text{if } \varepsilon-\tilde{q}(a_n)>0,\ \varepsilon-q_n(a_n)\le 0\\ 0 & \text{otherwise}\end{cases}.$$

Also, let $\tilde{E}=\{t:\hat{\varepsilon}_t>\tilde{q}\,(a_n)\}$, $E=\{t:\varepsilon_t>q_n\,(a_n)\}$, $Z_t=\varepsilon_t-q_n(a_n)$ and $\tilde{Z}_t=\hat{\varepsilon}_t-\tilde{q}(a_n)$ for $t=1,\cdots,n$.

Then, we have

$$\tilde{I}_{1N}-I_{1N}=\frac{1}{\delta_N}\left(1-\frac{N_s}{N}\right)+\frac{1}{\delta_N}(k_0^{-1}-1)\left(\frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}-\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\chi(\varepsilon_t)\right)\tilde{\chi}(\varepsilon_t)\right.$$
$$\left.+\frac{1}{N}\sum_{t=1}^{n}\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\chi(\varepsilon_t)\left(\tilde{\chi}(\varepsilon_t)-\chi(\varepsilon_t)\right)\right)=o_p(1)+\frac{1}{\delta_N}(k_0^{-1}-1)(I_{11n}+I_{12n})$$

since $N-N_s=O_p(N^{1/2})$ (see Lemma 4) and $\delta_N N^{1/2}\to\infty$. We first study I_{11n}, which can be written as

$$I_{11n}=\frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}-\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\right)\chi_I(\varepsilon_t)+\frac{1}{N}\sum_{t=1}^{n}\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\chi_D(\varepsilon_t)$$
$$=I_{111n}+I_{112n} \text{ where } t\in\tilde{E}-E \text{ denotes that } t\in\tilde{E} \text{ and } t\notin E.$$

By the Mean Value theorem, for some $\lambda_t\in(0,1)$ and $Z_t^*=Z_t+\lambda_t(\tilde{Z}_t-Z_t)$ we have

$$\left|\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\chi_I(\varepsilon_t)-\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\chi_I(\varepsilon_t)\right|=\frac{|k_0|/\sigma_N}{\left(1-\frac{k_0Z_t^*}{\sigma_N}\right)^2}|\tilde{Z}_t-Z_t|\chi_I(\varepsilon_t)$$

$$=\frac{1}{\left(1+\frac{Z_t^*}{q(a_n)}\right)^2}\frac{|\tilde{Z}_t-Z_t|}{q(a_n)}\chi_I(\varepsilon_t),$$

where the last equality follows from $\sigma_N = -q(a_n)k_0$ and $k_0 < 0$. Note that we can write,

$$\hat{\varepsilon}_t - \varepsilon_t = (m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t))(\hat{h}^{-1/2}(\mathbf{X}_t) - h^{-1/2}(\mathbf{X}_t))\chi_{\{\hat{h}(\mathbf{X}_t)>0\}}$$
$$+ h^{-1/2}(\mathbf{X}_t)(m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t))\left(\chi_{\{\hat{h}(\mathbf{X}_t)>0\}} - 1\right) + h^{-1/2}(\mathbf{X}_t)(m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t))$$
$$+ \left(\frac{h^{1/2}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)} - 1\right)\chi_{\{\hat{h}(\mathbf{X}_t)>0\}}\varepsilon_t + \left(\chi_{\{\hat{h}(\mathbf{X}_t)>0\}} - 1\right)\varepsilon_t. \tag{20}$$

By Lemmas 2, 3, Corollary 1 and the fact that $h(\mathbf{x})$ is uniformly bounded away from zero, we have

$$|\hat{\varepsilon}_t - \varepsilon_t| = O_p(L_{1n}) + (O_p(L_{1n}) + O_p(L_{2n}))\,|\varepsilon_t| \text{ uniformly in } \mathcal{G}. \tag{21}$$

Consequently, since $\tilde{Z}_t - Z_t = \hat{\varepsilon_t} - \varepsilon_t - (\tilde{q}(a_n) - q_n(a_n))$, we have

$$\frac{|\tilde{Z}_t - Z_t|}{q(a_n)} \leq \frac{1}{q(a_n)}\left(O_p(L_{1n}) + O_p(L_{1n} + L_{2n})|\varepsilon_t|\right) + \frac{|\tilde{q}(a_n) - q_n(a_n)|}{q(a_n)}$$
$$= \frac{1}{q(a_n)}\left(O_p(L_{1n}) + O_p(L_{1n} + L_{2n})|\varepsilon_t|\right) + O_p(N^{-1/2}) \tag{22}$$

where the last equality follows from Lemma 4. Note that in the set $E \cap \tilde{E}$ we have $Z_t > 0$ and $0 < \varepsilon_t = Z_t + q_n(a_n)$. Hence, since $\frac{q_n(a_n)}{q(a_n)} = O_p(1)$ and $\frac{L_{1n}}{q(a_n)} = o_p(N^{-1/2})$ provided that $N \propto n^{\frac{2s}{2s+d}-\delta}$ and $h_{in} \propto n^{-\frac{1}{2s+d}}$ for $i = 1, 2$, we have

$$\frac{|\tilde{Z}_t - Z_t|}{q(a_n)} = \left(\frac{Z_t}{q(a_n)} + O_p(1)\right)O_p\left(L_{1n} + L_{2n}\right) + O_p(N^{-1/2}) \text{ uniformly in } \mathcal{G}.$$

In addition, $\frac{Z_t^*}{q(a_n)} = \frac{Z_t}{q(a_n)} + \lambda_t\left(\frac{1}{q(a_n)}\left(O_p(L_{1n}) + O_p(L_{1n} + L_{2n})\varepsilon_t\right) + O_p(N^{-1/2})\right)$, and given that $\lambda_t < 1$ and provided that $h_{in} \propto n^{-\frac{1}{2s+d}}$ we can write

$$\frac{Z_t^*}{q(a_n)} = \frac{Z_t}{q(a_n)}\left(1 + O_p(L_{1n} + L_{2n})\right) + O_p(N^{-1/2}) = \frac{Z_t}{q(a_n)}\left(1 + o_p(1)\right) + o_p(1) \text{ uniformly in } \mathcal{G}.$$

Thus,

$$I_{111n} \leq \frac{1}{N}\sum_{t=1}^{n}\frac{1}{\left(1 + \frac{Z_t}{q(a_n)}(1 + o_p(1)) + o_p(1)\right)^2}\left(\left(\frac{Z_t}{q(a_n)} + O_p(1)\right)O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})\right)\chi_I(\varepsilon_t).$$

Given that $Z_t > 0$ whenever $t \in \chi_I(\varepsilon_t)$ and $\frac{x}{(1+x)^2} < 1$ for $x > 0$ we have $I_{111n} \leq (O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2}))\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t) = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$ since $\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t) = O_p(1)$. We now consider I_{112n}, which can be written as $I_{112n} = \frac{1}{N}\sum_{t=1}^{n}\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\left(\tilde{\chi}(\varepsilon_t) - \chi(\varepsilon_t)\right)\chi_D(\varepsilon_t)$. For $\delta_1, \delta_2 > 0$ we define the events $A = \left\{\omega : \frac{|\hat{\varepsilon}_t - \varepsilon_t|}{q_n(a_n)} < \delta_1\right\}$ 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 $C = \{\omega : \tilde{\chi}(\varepsilon_t) - \chi(\varepsilon_t) = 0\}$. The indicator function for an arbitrary set S is defined by χ_S. Hence, $\chi_{C^c} \leq \chi_{A^c} + \chi_{B^c}$ and

$$\begin{aligned} I_{112n} &\leq \frac{1}{N}\sum_{t=1}^{n}\left|\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\right|\chi_{A^c}\chi_D(\varepsilon_t) + \frac{1}{N}\sum_{t=1}^{n}\left|\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\right|\chi_{B^c}\chi_D(\varepsilon_t) \\ &= I_{1121n} + I_{1122n}. \end{aligned}$$

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

$$I_{1121n} < \frac{1}{N}\sum_{t=1}^{n}\left|\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\right|\frac{|\hat{\varepsilon}_t - \varepsilon_t|}{\delta_1 q_n(a_n)}\chi_D(\varepsilon_t)$$

and

$$I_{1122n} < \frac{1}{N}\sum_{t=1}^{n}\left|\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\right|\frac{|\tilde{q}(a_n) - q_n(a_n)|}{\delta_2 q_n(a_n)}\chi_D(\varepsilon_t).$$

Since $k_0 < 0$, $\sigma_N > 0$ and $\tilde{Z}_t > 0$ whenever $t \in \tilde{E} - E$ we have that $\left|\left(1 - \frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\right| < C$. From Lemma 4 we can immediately conclude that $I_{1122n} \leq \frac{1}{\delta_2}O_p(N^{-1/2})\frac{1}{N}\sum_{t=1}^{n}\chi_D(\varepsilon_t)$, and since $\frac{1}{N}\sum_{t=1}^{n}\chi_D(\varepsilon_t) = O_p(1)$ we have $I_{1122n} = O_p(N^{-1/2})$. Now, given (21), we have $\frac{|\hat{\varepsilon}_t - \varepsilon_t|}{q_n(a_n)} = \frac{|\varepsilon_t|}{q_n(a_n)}O_p(L_{1n} + L_{2n}) + o_p(N^{-1/2})$, therefore $I_{1121n} \leq O_p(L_{1n} + L_{2n})\frac{1}{N\delta_1}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t) + o_p(N^{-1/2})\frac{1}{N\delta_1}\sum_{t=1}^{n}\chi_D(\varepsilon_t)$. The second term following the inequality is $o_p(N^{-1/2})$ since $\frac{1}{N\delta_1}\sum_{t=1}^{n}\chi_D(\varepsilon_t) = O_p(1)$. For the first term, note that $\frac{|\varepsilon_t|}{q_n(a_n)} = \left|\frac{Z_t}{q_n(a_n)} + 1\right|$ and for $t \in \tilde{E} - E$, $\varepsilon_t \leq q_n(a_n)$ and consequently if $\varepsilon_t > 0$ we have $|\frac{Z_t}{q_n(a_n)}| < \frac{|\varepsilon_t|}{q_n(a_n)} + 1 < 2$. If $\varepsilon_t \leq 0$ for $t \in \tilde{E} - E$ then

$$\begin{aligned} \hat{\varepsilon}_t &= (m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t))\left(\hat{h}^{-1/2}(\mathbf{X}_t) - h^{-1/2}(\mathbf{X}_t)\right)\chi_{\{\hat{h}(\mathbf{X}_t)>0\}} + \frac{m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t)}{h^{1/2}(\mathbf{X}_t)}\left(\chi_{\{\hat{h}(\mathbf{X}_t)>0\}} - 1\right) \\ &+ (m(\mathbf{X}_t) - \hat{m}(\mathbf{X}_t))\,h^{-1/2}(\mathbf{X}_t) + \frac{h^{1/2}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)}\chi_{\{\hat{h}(\mathbf{X}_t)>0\}}\varepsilon_t \\ &= O_p(L_{1n}) + \frac{h^{1/2}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)}\chi_{\{\hat{h}(\mathbf{X}_t)>0\}}\varepsilon_t > 0. \end{aligned}$$

Since $\frac{h^{1/2}(\mathbf{X}_t)}{\hat{h}^{1/2}(\mathbf{X}_t)}\chi_{\{\hat{h}(\mathbf{X}_t)>0\}} = 1 + o_p(1)$ and $L_{1n} \to 0$, it must be that $\varepsilon_t > 0$ with probability approaching 1. Consequently, for N sufficiently large and $t \in \tilde{E} - E$ we have $|\frac{Z_t}{q_n(a_n)}| < 2$ and $I_{1121n} = O_p(L_{1n} + L_{2n}) + o_p(N^{-1/2})$. Combining the orders of I_{1121n} and I_{1122n} we have $I_{112n} = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$ and

$I_{11n} = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$. We now consider I_{12n} and note that

$$I_{12n} \leq \frac{1}{N\delta_1}\sum_{t=1}^{n}\left|\left(1-\frac{k_0 Z_t}{\sigma_N}\right)^{-1}\frac{k_0 Z_t}{\sigma_N}\right|\frac{|\hat{\varepsilon}_t - \varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) + \frac{1}{N\delta_2}\sum_{t=1}^{n}\left|\left(1-\frac{k_0 Z_t}{\sigma_N}\right)^{-1}\frac{k_0 Z_t}{\sigma_N}\right|\frac{|\tilde{q}(a_n) - q_n(a_n)|}{q_n(a_n)}\chi(\varepsilon_t).$$

By Lemma 4 and the fact that $\left|\left(1-\frac{k_0 Z_t}{\sigma_N}\right)^{-1}\frac{k_0 Z_t}{\sigma_N}\right| < C$, the second term following the inequality is $O_p(N^{-1/2})$ given that $\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$. Again, using $\frac{|\hat{\varepsilon}_t-\varepsilon_t|}{q_n(a_n)} = \frac{|\varepsilon_t|}{q_n(a_n)}O_p(L_{1n}+L_{2n}) + o_p(N^{-1/2})$ we have that the first term after the inequality is bounded by $O_p(L_{1n}+L_{2n})\frac{1}{N\delta_1}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) + o_p(N^{-1/2})\frac{1}{N\delta_1}\sum_{t=1}^{n}\chi(\varepsilon_t)$. Since $\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$ we need only investigate the order of $\frac{1}{N}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t)$. Note that $\frac{1}{N}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) \leq C\frac{1}{N}\sum_{t=1}^{n}\frac{Z_t}{q(a_n)}\chi(\varepsilon_t) + O_p(1)$ since $\frac{q(a_n)}{q_n(a_n)} = O_p(1)$, $Z_t > 0$ whenever $t \in E$ and $\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$. Now, let $Z'_t = \varepsilon_t - q(a_n)$ and note that

$$\begin{aligned}\frac{1}{N}\sum_{t=1}^{n}\frac{Z_t}{q(a_n)}\chi(\varepsilon_t) &= \frac{1}{N}\sum_{t=1}^{n}\frac{Z'_t}{q(a_n)}\chi(\varepsilon_t) - \frac{q_n(a_n)-q(a_n)}{q(a_n)}\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t)\\ &= \frac{1}{N}\sum_{t=1}^{N}\frac{Z'_t}{q(a_n)} - O_p(N^{-1/2})O_p(1) \text{ by Lemma 4.}\end{aligned}$$

If $q(a_n) \leq q_n(a_n)$ then $Z'_t > 0$ and $E\left(\left|\frac{Z'_t}{q(a_n)}\right|\right) = \frac{1}{\alpha-1} + O(\phi(q(a_n)))$ and $\frac{1}{N}\sum_{t=1}^{N}\frac{Z'_t}{q(a_n)} = O_p(1)$. If $q(a_n) > q_n(a_n)$ and N_1 denotes the number of elements in $\{\varepsilon_t\}_{t=1}^{N}$ that exceed $q(a_n)$, we have $N > N_1$ and we write $\frac{1}{N}\sum_{t=1}^{N}\left|\frac{Z'_t}{q(a_n)}\right| \leq \frac{1}{N}\sum_{t=1}^{N_1}\left|\frac{Z'_t}{q(a_n)}\right| + \frac{1}{N}\sum_{t=1}^{N-N_1}\left|\frac{Z'_t}{q(a_n)}\right|$. For the terms in the second sum on the right side of the inequality $\left|\frac{Z'_t}{q(a_n)}\right| \leq \left|\frac{\varepsilon_t}{q(a_n)}\right| + 1 \leq 2$ since in this case $\varepsilon_t \leq q(a_n)$. The Z'_t in the first sum are all positive and we have $\frac{1}{N}\sum_{t=1}^{N_1}\frac{Z'_t}{q(a_n)} = O_p\left(\frac{N_1}{N}\right)\left(\frac{1}{\alpha-1} + O\left(\phi(q(a_n))\right)\right)$. Thus, $\frac{1}{N}\sum_{t=1}^{N}\frac{Z'_t}{q(a_n)} = O_p\left(\frac{N_1}{N}\right)\left(\frac{1}{\alpha-1} + O\left(\phi(q(a_n))\right)\right) + O_p\left(\frac{N_1-N}{N}\right) = O_p(1)$ since $\frac{N_1}{N} = O_p(1)$. Thus, $\frac{1}{N}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$ and we conclude that $I_{12n} = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$. Combining the orders of I_{11n} and I_{12n} we have $\tilde{I}_{1N} - I_{1N} = \frac{1}{\delta_N}(k_0^{-1}-1)(O_p(L_{1n}+L_{2n}) + O_p(N^{-1/2})) + o_p(1)$. Since, $\delta_N N^{1/2} \to \infty$ and $\sqrt{N}L_{1n}, \sqrt{N}L_{2n} \to 0$ as $n \to \infty$ whenever $N \propto n^{\frac{2s}{2s+d}-\delta}$ for $0 < \delta$, $h_{1n} \propto n^{-\frac{1}{2s+d}}$ and $h_{2n} \propto n^{-\frac{1}{2s+d}}$ we have $\tilde{I}_{1N} - I_{1N} = o_p(1)$. We now turn to establishing that $\tilde{I}_{4N} - I_{4N} = o_p(1)$. We write

$$\begin{aligned}\tilde{I}_{4N} - I_{4N} = \frac{1}{\delta_N}\Bigg(\frac{1}{N}\sum_{t=1}^{n}\Bigg(&-\frac{1}{k_0^2}log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right) + \frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)^{-1}\frac{k_0\tilde{Z}_t}{\sigma_N}\\ &-\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0 Z_t}{\sigma_N}\right) + \frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0 Z_t}{\sigma_N}\right)^{-1}\frac{k_0 Z_t}{\sigma_N}\right)\chi(\varepsilon_t)\Bigg)\tilde{\chi}(\varepsilon_t)\end{aligned}$$

$$+ \frac{1}{N}\sum_{t=1}^{n}\left(-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\right)\right)\chi(\varepsilon_t)(\tilde{\chi}(\varepsilon_t)-\chi(\varepsilon_t))$$
$$= \frac{1}{\delta_N}(I_{41n}+I_{42n}).$$

First, note that

$$I_{41n} = -\frac{1}{k_0^2}\left(\frac{1}{N}\sum_{t=1}^{n}\left(log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)-log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\right)\chi_I(\varepsilon_t)+\frac{1}{N}\sum_{i=1}^{n}log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\chi_D(\varepsilon_t)\right)$$
$$+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)I_{11n} = -\frac{1}{k_0^2}(I_{411n}+I_{412n})+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)I_{11n}$$

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

$$\left|log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)-log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\right|\chi_I(\varepsilon_t) = \left(1+\frac{Z_t^*}{q(a_n)}\right)^{-1}\frac{|\tilde{Z}_t-Z_t|}{q(a_n)}\chi_I(\varepsilon_t).$$

Using the same arguments when studying the order of I_{111n} we immediately have $I_{411n} \leq O_p(L_{1n}+L_{2n}) + O_p(N^{-1/2})$. Given that $\chi_D(\varepsilon_t) = \tilde{\chi}(\varepsilon_t)(\tilde{\chi}(\varepsilon_t)-\chi(\varepsilon_t))$ and $\chi_A = \chi_A^2$ we have for $\delta_1, \delta_2 > 0$

$$I_{412n} \leq \frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\left(\frac{|\hat{\varepsilon}_t-\varepsilon_t|}{\delta_1 q_n(a_n)}+\frac{|\tilde{q}(a_n)-q_n(a_n)|}{\delta_2 q_n(a_n)}\right)\chi_D(\varepsilon_t)$$
$$\leq O_p(L_{1n}+L_{2n})\frac{1}{\delta_1 N}\sum_{t=1}^{n}log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t)+O_p(N^{-1/2})\frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\chi_D(\varepsilon_t).$$

Note that by the Mean Value theorem $\left|log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\right| = \left|\left(1-\frac{k_0Z_t^*}{\sigma_N}\right)^{-1}\frac{-k_0\tilde{Z}_t}{\sigma_N}\right| < -\frac{k_0}{\sigma_N}\tilde{Z}_t$ since $\tilde{Z}_t > 0$ whenever $t \in \tilde{E} - E$, $Z_t^* = \lambda_t\tilde{Z}_t > 0$ for some $0 < \lambda_t < 1$. Hence,

$$I_{412n} \leq O_p(L_{1n}+L_{2n})\frac{1}{\delta_1 N}\sum_{t=1}^{n}\frac{-k_0}{\sigma_N}\tilde{Z}_t\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t)+O_p(N^{-1/2})\frac{1}{N}\sum_{t=1}^{n}\frac{-k_0}{\sigma_N}\tilde{Z}_t\chi_D(\varepsilon_t).$$

Since $\frac{q(a_n)}{q_n(a_n)} = 1 + o_p(1)$, $\frac{|\varepsilon_t|}{q_n(a_n)} \leq \frac{|Z_t|}{q_n(a_n)} + 1$ and $\frac{\tilde{Z}_t}{q(a_n)} \leq \frac{|Z_t|}{q_n(a_n)}(1+O_p(L_{1n}+L_{2n}))+O_p(N^{-1/2})$ we have, for the first term following the inequality,

$$\frac{1}{N}\sum_{t=1}^{n}\frac{-k_0}{\sigma_N}\tilde{Z}_t\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t) \leq \frac{1}{N}\sum_{t=1}^{n}\left(\frac{|Z_t|}{q_n(a_n)}(1+O_p(L_{1n}+L_{2n}))+O_p(N^{-1/2})\right)\left(1+\frac{|Z_t|}{q_n(a_n)}\right)\chi_D(\varepsilon_t).$$

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

$$\begin{aligned}
I_{42n} &< \frac{1}{N}\sum_{t=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\right|\frac{|\hat{\varepsilon}_t-\varepsilon_t|}{\delta_1 q_n(a_n)}\chi(\varepsilon_t)\\
&+\frac{1}{N}\sum_{t=1}^{n}\left|-\frac{1}{k_0^2}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)+\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\frac{k_0Z_t}{\sigma_N}\right|\frac{|\tilde{q}(a_n)-q_n(a_n)|}{\delta_2 q_n(a_n)}\chi(\varepsilon_t)\\
&= I_{421n}+I_{422n}.
\end{aligned}$$

Since $\frac{|\hat{\varepsilon}_t-\varepsilon_t|}{q_n(a_n)} \leq \frac{|\varepsilon_t|}{q_n(a_n)}O_p(L_{1n}+L_{2n}) + o_p(N^{-1/2})$ we write

$$\begin{aligned}
I_{421n} &\leq O_p(L_{1n}+L_{2n})\frac{1}{k_0^2\delta_1 N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t)\\
&+O_p(L_{1n}+L_{2n})\left|\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\right|\frac{1}{\delta_1 N}\sum_{t=1}^{n}\left|\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_t}{\sigma_N}\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t)\\
&+o_p(N^{-1/2})\frac{1}{k_0^2\delta_1 N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\right|\chi(\varepsilon_t)\\
&+o_p(N^{-1/2})\left|\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\right|\frac{1}{\delta_1 N}\sum_{t=1}^{n}\left|\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_t}{\sigma_N}\right|\chi(\varepsilon_t) \qquad (23)
\end{aligned}$$

Since $k_0 < 0$ and $Z_t > 0$ for all $t \in E$ we have that $\left|\left(1-\frac{k_0Z_t}{\sigma_N}\right)^{-1}\right|\left|\frac{k_0Z_t}{\sigma_N}\right| < C$ and the second and fourth terms following the inequality are $O_p(L_{1n}+L_{2n})$ and $o_p(N^{-1/2})$ since $\frac{1}{N}\sum\limits_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$ (see the order of I_{12n}) and $\frac{1}{N}\sum\limits_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$. Note that for all $t \in E$, since $Z_t > 0$ and $k_0 < 0$ we have

$$\begin{aligned}
\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) &\leq \frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\frac{Z_t}{q_n(a_n)}\chi(\varepsilon_t)+\frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\chi(\varepsilon_t)\\
&= (1+o_p(1))\frac{1}{N}\sum_{t=1}^{n}log\left(1+\frac{Z_t}{q(a_n)}\right)\frac{Z_t}{q(a_n)}\chi(\varepsilon_t)+\frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0Z_t}{\sigma_N}\right)\chi(\varepsilon_t)\\
&= (1+o_p(1))T_{1N}+T_{2N}
\end{aligned}$$

by using the fact that $q(a_n)/q_n(a_n) = 1 + o_p(1)$ and $-\sigma_N/k_0 = q(a_n)$. Let $\Psi(z) = log\,(1+z)\,z$ and write $T_{1N} = -\frac{1}{N}\sum_{t=1}^{N}(\Psi(Z_t'/q(a_n)) - \Psi(Z_t/q(a_n))) + \frac{1}{N}\sum_{t=1}^{N}\Psi(Z_t'/q(a_n))$. If $q(a_n) \leq q_n(a_n)$ then $Z_t, Z_t' > 0$ and by the Mean Value Theorem, there exists $\lambda_t \in (0,1)$ and $Z_t^* = Z_t' + \lambda_t(q_n(a_n) - q(a_n)) > 0$ such that

$$\Psi(Z_t'/q(a_n)) - \Psi(Z_t/q(a_n)) = \left(\left(1+\frac{Z_t^*}{q(a_n)}\right)^{-1}\frac{Z_t^*}{q(a_n)} + log\left(1+\frac{Z_t^*}{q(a_n)}\right)\right)\frac{q_n(a_n)-q(a_n)}{q(a_n)}.$$

We observe that $\left(1+\frac{Z_t^*}{q(a_n)}\right)^{-1}\frac{Z_t^*}{q(a_n)} < C$ and $\frac{q_n(a_n)-q(a_n)}{q(a_n)} = O_p(N^{-1/2})$. Also, $log\left(1+\frac{Z_t^*}{q(a_n)}\right) \leq \frac{Z_t^*}{q(a_n)} = \frac{Z_t'}{q(a_n)} + \lambda_t\frac{q_n(a_n)-q(a_n)}{q(a_n)} = \frac{Z_t'}{q(a_n)} + o_p(1)$. Consequently, $\frac{1}{N}\sum_{t=1}^{N}(\Psi(Z_t'/q(a_n)) - \Psi(Z_t/q(a_n))) = O_p(N^{-1/2})$

since $\frac{1}{N}\sum_{t=1}^{N}\frac{Z_t'}{q(a_n)} = O_p(1)$. Now, $\frac{1}{N}\sum_{t=1}^{N}\Psi(Z_t') = O_p(1)$ since $E\left(\left|log\left(1+\frac{Z_t'}{q(a_n)}\right)\frac{Z_t'}{q(a_n)}\right|\right) = \frac{1}{(\alpha-1)^2} + \frac{1}{\alpha(\alpha-1)} + O(\phi(q(a_n)))$. Thus, combining these orders we have $T_{1N} = O_p(1)$. Let $Q(z) = log(1+z)$, then $T_{2N} = -\frac{1}{N}\sum_{t=1}^{N}(Q(Z_t'/q(a_n)) - Q(Z_t/q(a_n))) + \frac{1}{N}\sum_{t=1}^{N}Q(Z_t'/q(a_n))$. By the Mean Value Theorem, there exists $\lambda_t \in (0,1)$ and $Z_t^* = Z_t' + \lambda_t(q_n(a_n) - q(a_n)) > 0$ such that

$$Q(Z_t'/q(a_n)) - Q(Z_t/q(a_n)) = \left(1+\frac{Z_t^*}{q(a_n)}\right)^{-1}\frac{q_n(a_n)-q(a_n)}{q(a_n)} = O_p(N^{-1/2}).$$

Furthermore, $\frac{1}{N}\sum_{t=1}^{N}Q(Z_t'/q(a_n)) = O_p(1)$ since $E(|Q(Z_t'/q(a_n))|) = \frac{1}{\alpha} + o(1)$. Hence, $T_{2N} = O_p(1)$ which combines with the order of T_{1N} to give $\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0 Z_t}{\sigma_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$.

Now, consider the case where $q(a_n) > q_n(a_n)$ and consequently $N > N_1$. Then,

$$T_{1N} = -\frac{1}{N}\sum_{t=1}^{N_1}(\Psi(Z_t'/q(a_n)) - \Psi(Z_t/q(a_n))) + \frac{1}{N}\sum_{t=1}^{N_1}\Psi(Z_t'/q(a_n)) + \frac{1}{N}\sum_{t=1}^{N-N_1}\Psi(Z_t/q(a_n)).$$

The first two sums on the right side of the equality have $Z_t' > 0$ and can be treated as in the case where $q(a_n) \le q_n(a_n)$. In the last sum $q_n(a_n) < \varepsilon_t < q(a_n)$ and we have

$$\frac{1}{N}\sum_{t=1}^{N-N_1}\Psi(Z_t/q(a_n)) \le \frac{1}{N}\sum_{t=1}^{N-N_1}\left(\frac{Z_t}{q(a_n)}\right)^2 \le \frac{1}{N}\sum_{t=1}^{N-N_1}\left(\frac{\varepsilon_t}{q(a_n)}\right)^2 = O_p\left(\frac{N-N_1}{N}\right) = O_p(N^{-1/2}),$$

which combines with the order of the first two terms to give $T_{1N} = O_p(1)$. Now, write

$$T_{2N} = -\frac{1}{N}\sum_{t=1}^{N_1}(Q(Z_t'/q(a_n)) - Q(Z_t/q(a_n))) + \frac{1}{N}\sum_{t=1}^{N-N_1}Q(Z_t/q(a_n)).$$

The first term on the right side of the inequality can her treat as in the case where $q(a_n) \le q_n(a_n)$. For the second term, $\frac{1}{N}\sum_{t=1}^{N-N_1}Q(Z_t/q(a_n)) \le \frac{1}{N}\sum_{t=1}^{N-N_1}\frac{Z_t}{q(a_n)} \le O_p\left(\frac{N-N_1}{N}\right) = O_p(N^{-1/2})$. Consequently, $T_{2N} = O_p(N_1/N) + O_p(N^{-1/2})$ which combines with the order of T_{1N} to give $\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0 Z_t}{\sigma_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$. Thus, the first term on the right side of inequality in (23) is $O_p(L_{1n} + L_{2n})$. Similar arguments establish that $\frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{k_0 Z_t}{\sigma_N}\right)\chi(\varepsilon_t) = O_p(1)$ and that the order of the third term is $O_p(N^{-1/2})$.

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 $\left|\left(1-\frac{k_0 Z_t}{\sigma_N}\right)^{-1}\frac{k_0 Z_t}{\sigma_N}\right| < C$ we write $I_{422n} \le \frac{1}{\delta_2}O_p(N^{-1/2})\left(\frac{1}{k_0^2\delta_2}\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0 Z_t}{\sigma_N}\right)\right|\chi(\varepsilon_t) + C\left|\frac{1}{k_0}\left(1-\frac{1}{k_0}\right)\right|\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t)\right)$. Since we have already established that $\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{k_0 Z_t}{\sigma_N}\right)\right|\chi(\varepsilon_t) = O_p(1)$ and $\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$ we conclude that

$I_{422n} = O_p(N^{-1/2})$. Combining all orders obtained we have that $I_{41n} + I_{42n} = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$ and consequently $\tilde{I}_{4N} - I_{4N} = o_p(1)$, since $\delta_N N^{1/2} \to \infty$ as $n \to \infty$.

We now investigate the order of $\tilde{I}_{2N} - I_{2N}$. Consider arbitrary $\dot{\sigma}_N = \sigma_N(1+\delta_N\tau_1\lambda_1)$ and $\dot{k} = k_0+\delta_N\tau_2\lambda_2$ and write

$$\tilde{I}_{2N} - I_{2N} = \frac{\tau_1}{(1+\tau_1\delta_N\lambda_1)^2}\left((-2)\left(\frac{1}{\dot{k}}-1\right)\frac{1}{N}\sum_{t=1}^{N_s}\left(\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}+\frac{1}{2}\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^2\right)\right.$$
$$\left.+\ 2\left(\frac{1}{\dot{k}}-1\right)\frac{1}{N}\sum_{t=1}^{N}\left(\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_t}{\dot{\sigma}_N}+\frac{1}{2}\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^2\right)\right)$$
$$+\frac{\tau_1}{(1+\tau_1\delta_N\lambda_1)^2}\left(\frac{N_s}{N}-1\right)$$

Hence, it suffices to examine

$$\frac{1}{N}\sum_{t=1}^{N_s}\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}-\frac{1}{N}\sum_{t=1}^{N}\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}=\frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}\right.$$
$$\left.-\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\chi(\varepsilon_t)\right)\tilde{\chi}(\varepsilon_t)+\frac{1}{N}\sum_{t=1}^{n}\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\chi(\varepsilon_t)(\tilde{\chi}(\varepsilon_t)-\chi_{E_t})=I_{nl1}+I_{nl2}$$

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

$$I_{nl11} = \frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}-\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\right)\chi_I(\varepsilon_t) \text{ and}$$
$$I_{nl12} = \frac{1}{N}\sum_{t=1}^{n}\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}\chi_D(\varepsilon_t).$$

By the Mean Value theorem, there exists $Z_t^* = \tilde{Z}_t + \lambda_t(\tilde{Z}_t - Z_t)$ for $\lambda_t \in (0, 1)$ such that

$$I_{nl11} \le l\frac{1}{N}\sum_{t=1}^{n}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{l-1}q_n(a_n)\right|$$
$$\times\left(O_p(L_{1n}+L_{2n})\left(\frac{Z_t}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_I(\varepsilon_t). \tag{24}$$

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

$$\sup_{S_T}\frac{1}{N}\sum_{t=1}^{n}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{\dot{k}}{\dot{\sigma}_N}\left(\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{l-1}\frac{Z_t}{q_n(a_n)}q_n(a_n)\right|\chi_I(\varepsilon_t)\le O_p(1)\sup_{S_T}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\dot{\sigma}_N}\right|$$
$$\times\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t)\sup_{S_T}\left|\left(\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{l-1}\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-l-1}\frac{Z_t}{q_n(a_n)}\right|.$$

Given that $\delta_N \to 0$ we have for N sufficiently large $\underset{S_T}{sup}\left|\frac{\dot{k}}{k_0}\frac{\sigma_N}{\dot{\sigma}_N}\right| < C$ and $\underset{S_T}{sup}\left|\left(\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{l-1}\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-l}\right| < C$. Hence, to establish the order of the left hand side of the inequality it suffices to obtain the order of $\upsilon_n = \frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t)\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\frac{Z_t}{q_n(a_n)}\right|$. Note that,

$$\begin{aligned}
\upsilon_n &\leq C\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t)\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\left(-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right|\underset{S_T}{sup}\left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)} \\
&\leq C\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t)\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\left(-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right| \quad \text{since } \underset{S_T}{sup}\left(-\frac{\dot{k}}{\dot{\sigma}_N}\right)^{-1}\frac{1}{q_n(a_n)} < C \\
&\leq C\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t)\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}q(a_n)\left(\frac{Z_t}{q(a_n)}(1+o_p(1))+o_p(1)\right)\right)^{-1}\left(-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right| = O_p(1)
\end{aligned}$$

since $\frac{Z_t^*}{q(a_n)} = \left(\frac{Z_t}{q(a_n)}(1+o_p(1))+o_p(1)\right)$, $\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}}{\dot{\sigma}_N}q(a_n)\left(\frac{Z_t}{q(a_n)}(1+o_p(1))+o_p(1)\right)\right)^{-1}\left(-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right| < C$ and $\frac{1}{N}\sum_{t=1}^{n}\chi_I(\varepsilon_t) = O_p(1)$. Consequently, $I_{nl11} = O_p(L_{1n}+L_{2n}) + O_p(N^{-1/2})$ as all remaining terms in (24) are of the same order. Now, we write

$$I_{nl12} \leq \frac{1}{N}\sum_{t=1}^{n}\left|\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}\right|\chi_D(\varepsilon_t)\left(O_p(L_{1n}+L_{2n})\frac{1}{\delta_1}\left(\frac{|\varepsilon_t|}{q_n(a_n)}+\frac{1}{q_n(a_n)}\right)+\frac{1}{\delta_2}O_p(N^{-1/2})\right)$$

and obtain the order of $\nu_n = \frac{1}{N}\sum_{t=1}^{n}\left|\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t)$. Note that

$$\nu_n \leq \frac{1}{N}\sum_{t=1}^{n}\underset{S_T}{sup}\left|\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{l}\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t) \leq C\frac{1}{N}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi_D(\varepsilon_t) = O_p(1)$$

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

$$I_{nl2} \leq \frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\right)\left(O_p(L_{1n}+L_{2n})\frac{1}{\delta_1}\left(\frac{|\varepsilon_t|}{q_n(a_n)}+\frac{1}{q_n(a_n)}\right)+\frac{1}{\delta_2}O_p(N^{-1/2})\right)\chi(\varepsilon_t). \tag{25}$$

Letting $T_n = \frac{1}{N}\sum_{t=1}^{n}\left(\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\right)\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t)$, we note that $T_n \leq C\frac{1}{N}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$ from the study of I_{12n} given that $\delta_N \to 0$, for N sufficiently large we have $\dot{k} < 0$, $\dot{\sigma}_N > 0$ and $\left|\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\right| < C$. Consequently, $I_{nl2} = O_p(L_{1n}+L_{2n}) + O_p(N^{-1/2})$ since all other terms in inequality (25) are of the same order given $\left|\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-l}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{l}\right| < C$ and the fact that $\frac{1}{N}\sum_{t=1}^{n}\chi(\varepsilon_t) = O_p(1)$. Combining the orders of I_{nl1}

and I_{nl2} we conclude that $\tilde{I}_{2N} - I_{2N} = o_p(1)$ uniformly on S_T. Now, note that $\tilde{I}_{3N} - I_{3N} = \tilde{I}_{5N} - I_{5N}$ and

$$\tilde{I}_{3N} - I_{3N} = \frac{\tau_2}{1+\delta_N\tau_1\lambda_1}\frac{1}{N}\sum_{t=1}^{N_s}\left(-\frac{1}{\dot{k}}\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N} + \frac{1}{\dot{k}}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^2\right)$$
$$+ \frac{\tau_2}{1+\delta_N\tau_1\lambda_1}\frac{1}{N}\sum_{t=1}^{N}\left(\frac{1}{\dot{k}}\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}Z_t}{\dot{\sigma}_N} - \frac{1}{\dot{k}}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^2\right).$$

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

$$\tilde{I}_{6N} - I_{6N} = \tau_2\left(\frac{1}{N}\sum_{t=1}^{N_s}\left(\frac{2}{\dot{k}^3}log\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right) + \frac{2}{\dot{k}^3}\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)\right)\right.$$
$$- \frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)^2\Bigg)$$
$$- \frac{1}{N}\sum_{t=1}^{N}\left(\frac{2}{\dot{k}^3}log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right) + \frac{2}{\dot{k}^3}\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-1}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right.$$
$$\left.\left. - \frac{1}{\dot{k}^2}\left(\frac{1}{\dot{k}}-1\right)\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^{-2}\left(\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)^2\right)\right)$$
$$= \frac{2\tau_2}{\dot{k}^3}\left(\frac{1}{N}\sum_{t=1}^{N_s}log\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right) - \frac{1}{N}\sum_{t=1}^{N}log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right) + o_p(1) \text{ uniformly in } S_T. \quad (26)$$

The last equality follows form the arguments used above when investigating the order of $\tilde{I}_{2N} - I_{2N}$. The first term in equation (26) can be written as (excluding the constant $2\tau_2/\dot{k}^3$) $I_{61n} + I_{62n}$, where

$$I_{61n} = \frac{1}{N}\sum_{t=1}^{n}\left(log\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\chi(\varepsilon_t)\right)\tilde{\chi}(\varepsilon_t),$$
$$I_{62n} = \frac{1}{N}\sum_{t=1}^{n}log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\chi(\varepsilon_t)(\tilde{\chi}(\varepsilon_t) - \chi(\varepsilon_t)).$$

Now, $I_{62n} \leq \frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right|\chi(\varepsilon_t)\left(\frac{O_p(L_{1n}+L_{2n})}{\delta_1}\left(\frac{|\varepsilon_t|}{q_n(a_n)} + \frac{1}{q_n(a_n)}\right) + \frac{1}{\delta_2}O_p(N^{-1/2})\right)$ and we consider the order of $\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t)$. Note that $Z_t > 0$ whenever $t \in E$ 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 $\frac{1}{N}\sum_{t=1}^{n}\left|log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right|\frac{|\varepsilon_t|}{q_n(a_n)}\chi(\varepsilon_t) = O_p(1)$ which follows from the order of I_{421n}. Hence, $I_{62n} = O_p(L_{1n} + L_{2n}) + O_p(N^{-1/2})$ uniformly on S_T. We write $I_{61n} = I_{611n} + I_{612n}$ where $I_{611n} = \frac{1}{N}\sum_{t=1}^{n}\left(log\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right) - log\left(1-\frac{\dot{k}Z_t}{\dot{\sigma}_N}\right)\right)\chi_I(\varepsilon_t)$and $I_{612n} =$

$\frac{1}{N}\sum_{t=1}^{n} log\left(1-\frac{\dot{k}\tilde{Z}_t}{\dot{\sigma}_N}\right)\chi_D(\varepsilon_t)$. Then,

$$\begin{aligned}
I_{611n} &= \frac{1}{N}\sum_{t=1}^{n}\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\frac{\dot{k}}{\dot{\sigma}_N}(\tilde{Z}_t - Z_t)\chi_I(\varepsilon_t) \\
&\leq \frac{1}{N}\sum_{t=1}^{n}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\right|\left|\frac{\dot{k}}{\dot{\sigma}_N}q_n(a_n)\right|\left(O_p(L_{1n}+L_{2n})\left(\frac{Z_t}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_I(\varepsilon_t) \\
&\leq \sup_{S_T}\left|\frac{\dot{k}}{\dot{\sigma}_N}q_n(a_n)\right|\frac{1}{N}\sum_{t=1}^{n}\sup_{S_T}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\right|\left(O_p(L_{1n}+L_{2n})\left(\frac{Z_t}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_I(\varepsilon_t) \\
&= O_p(1)\frac{1}{N}\sum_{t=1}^{n}\sup_{S_T}\left|\left(1-\frac{\dot{k}Z_t^*}{\dot{\sigma}_N}\right)^{-1}\right|\left(O_p(L_{1n}+L_{2n})\left(\frac{Z_t}{q_n(a_n)}+1\right)+O_p(N^{-1/2})\right)\chi_I(\varepsilon_t) \\
&= O_p(L_{1n}+L_{2n})+O_p(N^{-1/2}) \text{ uniformly on } S_T \text{ given the order of } \upsilon_n.
\end{aligned}$$

Since $\tilde{Z}_t > 0$ whenever $t \in \tilde{E} - E$ 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 $I_{612n} = \frac{1}{N}\sum_{t=1}^{n} log\left(1-\frac{k_0\tilde{Z}_t}{\sigma_N}\right)\chi_D(\varepsilon_t)+o_p(1)$. From the order of I_{412n} we have $I_{612n} = O_p(L_{1n}+L_{2n})+O_p(N^{-1/2})$ and consequently $I_{61n} = O_p(L_{1n}+L_{2n})+O_p(N^{-1/2})$, which combined with the order of I_{62n} gives $\tilde{I}_{6N} - I_{6N} = o_p(1)$ uniformly on S_T. □

Theorem 3.

Proof. Let $\tilde{r}_N = \frac{\tilde{\sigma}_{\tilde{q}(a_n)}}{\sigma_N} = 1+\delta_N\tau_1^*$, $\tilde{k} = k_0+\delta_N\tau_2^*$ and note that

$$\begin{pmatrix} \frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_1}L_{TN}(\tau_1^*,\tau_2^*) \\ \frac{1}{\delta_N^2}\frac{\partial}{\partial\tau_2}L_{TN}(\tau_1*^*,\tau_2^*) \end{pmatrix} = \frac{1}{\delta_N N}\begin{pmatrix} \sum_{i=1}^N \frac{\partial}{\partial r_N} log\, g(\tilde{Z}_i;\tilde{r}_N\sigma_N,\tilde{k}) \\ \sum_{i=1}^N \frac{\partial}{\partial k} log\, g(\tilde{Z}_i;\tilde{r}_N\sigma_N,\tilde{k}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{27}$$

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

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

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

where $\tilde{I}_{1N}$, I_{1N}, $\tilde{I}_{4N}$, I_{4N} are as defined in Theorem 2. By a Taylor's expansion of the first order condition in (27) around $(1,k_0)$ we have

$$H_N(r_N^*,k^*)\sqrt{N}\begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} = \upsilon_N(1,k_0). \tag{28}$$

We start by investing the asymptotic properties of $v_N(1,k_0)$. Let $b_1 = -\frac{\alpha(1+\alpha)}{2+\alpha}$, $b_2 = \left(-\frac{\alpha^2(1+\alpha)}{2+\alpha} + \frac{\alpha^3}{1+\alpha}\right)$ and observe that from Theorem 2 and Lemma 4 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 5 and the fact that $N_1 - N = O_p(N^{1/2})$

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

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

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

Note that we can write

$$\begin{aligned} \frac{1}{\sqrt{N}}\left(\sum_{i=1}^N \frac{\partial}{\partial\sigma} log\, g(Z_i';\sigma_N,k_0)\sigma_N - b_\sigma\right) &= \sum_{t=1}^n N^{-1/2}\left(\frac{\partial}{\partial\sigma} log\, g(Z_t';\sigma_N,k_0)\sigma_N - b_\sigma\right)\chi_{\{\varepsilon_t > q(a_n)\}} \\ &= \sum_{t=1}^n Z_{t1} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{N}}\left(\sum_{i=1}^N \frac{\partial}{\partial k} log\, g(Z_i';\sigma_N,k_0)\sigma_N - b_k\right) &= \sum_{t=1}^n N^{-1/2}\left(\frac{\partial}{\partial k} log\, g(Z_t';\sigma_N,k_0)\sigma_N - b_k\right)\chi_{\{\varepsilon_t > q(a_n)\}} \\ &= \sum_{i=1}^n Z_{i2}. \end{aligned}$$

Also, from Lemma 4, $\sqrt{N}\frac{\tilde{q}(a_n)-q(a_n)}{q(a_n)}$ is distributed asymptotically as $\sum_{t=1}^n k_0(n(1-F(y_n)))^{-1/2}(q_{1n} - E(q_{1n})) + o_p(1) = \sum_{t=1}^n Z_{t3} + o_p(1)$ where $q_{1n} = \frac{1}{h_{3n}}\int_{-\infty}^{y_n} K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy$ and $y_n = q(a_n)(1+N^{-1/2}z)$ for arbitrary z. It can be easily verified that $E(Z_{t1}) = E(Z_{t2}) = E(Z_{t3}) = 0$. In addition,

$$\begin{aligned} V(Z_{t1}) &= N^{-1}E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_i';\sigma_N,k_0)\sigma_N - b_\sigma\right)^2 P(\{\varepsilon_t > q(a_n)\}) \\ &= n^{-1}E\left(\frac{\partial}{\partial\sigma_N} log\, g(Z_i';\sigma_N,k_0)\sigma_N - b_\sigma\right)^2 = n^{-1}\left(\frac{1}{1-2k_0} + o(1)\right) \end{aligned}$$

where the last equality follows from the results listed in section 3.1. Using similar arguments we obtain

$$V(Z_{t2}) = n^{-1}\left(\frac{2\alpha^2}{(1+\alpha)(2+\alpha)} + o(1)\right)$$

and from Lemma 4 we have that $V(Z_{t3}) = n^{-1}k_0^3F(y_n) + o(h_{3n})$. We now define the vector $\psi_n = \sum_{t=1}^{n}(Z_{t1}, Z_{t2}, Z_{t3})^T$ and for arbitrary $0 \neq \lambda \in \mathbb{R}^3$ we consider $\lambda^T\psi_n = \sum_{i=1}^{n}(\lambda_1 Z_{t1} + \lambda_2 Z_{t2} + \lambda_3 Z_{t3}) = \sum_{t=1}^{n} Z_{tn}$. From above, we have that $E(Z_{tn}) = 0$ and $V(Z_{tn}) = \sum_{l=1}^{3}\lambda_d^2 E(Z_{td}^2) + 2\sum_{1\leq d<d'\leq 3}\lambda_d\lambda_{d'}E(Z_{td}Z_{td'})$. First, we consider $E(Z_{t1}Z_{t2})$ which can be written as

$$E(Z_{t1}Z_{t2}) = \frac{1}{n}T_{1n} - \frac{N}{n^2}b_\sigma b_k$$

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

$$E(Z_{t1}Z_{t2}) = \frac{1}{n}T_{1n} - O\left(\frac{(N^{1/2}\phi(\varepsilon_{(n-N)}))^2}{n^2}\right) = \frac{1}{n}T_{1n} - n^{-2}O(1)$$

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

$$T_{1n} = -b_k - \frac{1}{k_0}\left(\frac{1}{k_0} - 1\right)^2 E\left(\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)^{-2}\left(\frac{k_0Z_t'}{\sigma_N}\right)^2\right)$$
$$- \frac{1}{k_0^2}\left(\frac{1}{k_0} - 1\right) E\left(log\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_t'}{\sigma_N}\right)\right).$$

From Smith (1987) we have that $E\left(\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)^{-2}\left(\frac{k_0Z_t'}{\sigma_N}\right)^2\right) = \frac{2}{(1+\alpha)(2+\alpha)} + O(\phi(\varepsilon_{(n-N)}))$ and $b_k = O(\phi(\varepsilon_{(n-N)}))$. From Lemma 6 we have that

$$E\left(log\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)\left(1 - \frac{k_0Z_t'}{\sigma_N}\right)^{-1}\left(\frac{k_0Z_t'}{\sigma_N}\right)\right) = -\frac{1}{\alpha} + \frac{\alpha}{(1+\alpha)^2} + O(\phi(\varepsilon_{(n-N)}))$$

which combined with the orders obtained for the other components of the expectation and the fact that $k_0 = -\alpha^{-1}$ give $E(Z_{t1}Z_{t2}) = -\frac{1}{n(k_0-1)(2k_0-1)} + \frac{1}{n}\phi(\varepsilon_{(n-N)})O(1) - O(n^{-2})$. We now turn to $E(Z_{t1}Z_{t3})$ which can be written as

$$E(Z_{t1}Z_{t3}) = T_{2n} - k_0E\left(N^{-1/2}\left(\frac{\partial}{\partial\sigma_N}log\, g(Z_i'; \sigma_N, k_0)\sigma_N\right)\chi_{\varepsilon_t>q(a_n)}\right)E(q_{1n})(n(1-F(y_n)))^{-1/2},$$

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

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

from Lemma 4 $E(q_{1n}) = F(y_n) + O(h_{3n}^{m+1}) = O(1)$ 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 $\frac{\sqrt{N}}{n} O(\phi(\varepsilon_{(n-N)}))O(1)N^{-1/2} = n^{-1}O(\phi(\varepsilon_{(n-N)}))$. We now turn to T_{2n}, the first term in the covariance expression. Since $(n(1-F(y_n)))^{-1/2}$ is asymptotically equivalent to $N^{-1/2}$, we have by the Cauchy-Schwartz inequality

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

Hence, $E(Z_{t1}Z_{t3}) = o(n^{-1})$. In a similar manner we obtain $E(Z_{t2}Z_{t3}) = o(n^{-1})$. Hence, $nV(Z_{tn}) = \lambda^T V_1 \lambda + o(1)$, 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 $\sum_{i=1}^{n} Z_{nt} \xrightarrow{d} \mathcal{N}(0, \lambda^T V_1 \lambda)$ provided that $\sum_{t=1}^{n} E(|Z_{tn}|^3) \to 0$. To verify this condition, it suffices to show that $(i)\ \sum_{i=1}^{n} E(|Z_{t1}|^3) \to 0$; $(ii)\ \sum_{i=1}^{n} E(|Z_{t2}|^3) \to 0$; $(iii)\ \sum_{i=1}^{n} E(|Z_{t3}|^3) \to 0$. (iii) was verified in Lemma 4, so we focus on (i) and (ii). For (i), note that $\sum_{t=1}^{n} E(|Z_{1t}|^3) \leq \frac{1}{\sqrt{N}} E\left(\left|(1/k_0 - 1)(1 - k_0 Z_t'/\sigma_N)^{-1} k_0 Z_t'/\sigma_N - 1\right|^3\right) \to 0$ provided $E(-(1-k_0 Z_t'/\sigma_N)^{-3}(k_0 Z_t'/\sigma_N)^3) < C$, which is easily verified by noting that $-(1-k_0 Z_t'/\sigma_N)^{-3}(k_0 Z_i'/\sigma_N)^3 < -(1-k_0 Z_t'/\sigma_N)^{-3}(1-k_0 Z_t'/\sigma_N)^3 = 1$. Lastly,

$$\sum_{i=1}^{n} E(|Z_{2t}|^3) \leq \frac{1}{\sqrt{N}} E\left(\left|-(1/k_0^2) log(1 - k_0 Z_t'/\sigma_N) + (1/k_0)(1 - 1/k_0)(1 - k_0 Z_t'/\sigma_N)^{-1} k_0 Z_t'/\sigma_N\right|^3\right) \to 0$$

provided $E\left(log(1 - k_0 Z_t'/\sigma_N)^3\right) < C$ given the bound we obtained in case (i). By FR2 and integrating by parts we have

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

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

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

It is easy to verify that $\int_1^\infty 3(log(t))^2 t^{-\alpha-1} dt = \frac{6}{\alpha^3}$ and consequently $\tau_{2n} = \frac{6}{\alpha^3} + O(\phi(\varepsilon_{(n-N)}))$ which verifies (ii). By the Cramer-Wold theorem we have that $\psi_n \xrightarrow{d} \mathcal{N}(0, V_1)$. Consequently, for any vector $\gamma \in \Re^2$ we have $\gamma^T \left(v_N(\sigma_N, k_0) - \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \gamma^T V_2 \gamma)$ where $V_2 = \begin{pmatrix} \frac{k_0^2 - 4k_0 + 2}{(2k_0-1)^2} & -\frac{1}{k_0(k_0-1)} \\ -\frac{1}{k_0(k_0-1)} & \frac{2k_0^3 - 2k_0^2 + 2k_0 - 1}{k_0^2(k_0-1)^2(2k_0-1)} \end{pmatrix}$. Again, by the Cramer-Wold theorem $\left(v_N(\sigma_N, k_0) - \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, V_2)$. Hence, given equation (28), provided that $H_N(r_N^*, k^*) \xrightarrow{p} H$ we have

$$\sqrt{N} \begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} - H^{-1}\sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} = H^{-1} \left(v_N(\sigma_N, k_0) - \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) \xrightarrow{d} N\left(0, H^{-1} V_2 H^{-1}\right).$$

To see that $H_N(r_N^*, k^*) \xrightarrow{p} H$, first observe that whenever $(\tau_1, \tau_2) \in S_T$ we have $(\tilde{r}_N, \tilde{k}) \in S_R$ and consequently $(r_N^*, k^*) \in S_R$. In addition, from Theorem 2 and the results from Smith (1987) we have $H_N(\tilde{r}_N, \tilde{k}) \xrightarrow{p} H$ uniformly on S_R. By Theorem 21.6 in Davidson (1994) we conclude that $H_N(r_N^*, k^*) \xrightarrow{p} H$. □

Theorem 4.

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

$$1 - \tilde{F}(\tilde{q}(a_n) + \hat{y}_{N,a}) = \frac{N}{n} \left(1 - \frac{\tilde{k}\hat{y}_{N,a}}{\tilde{\sigma}_{\tilde{q}_n(a_n)}} \right)^{1/\tilde{k}}. \tag{29}$$

Note that for a chosen N, equation (29) is satisfied with a distribution function $\dot{F}$ that is not necessarily $\tilde{F}$. However, given the continuity of $\tilde{F}$, there exists N satisfying the order relation $a > 1 - N/n$ for which (29) is satisfied by $\tilde{F}$. Hence, to avoid additional notation we proceed with $\tilde{F}$. We define the estimator for $q(a)$ as $\hat{q}(a) = \tilde{q}(a_n) + \hat{y}_{N,a}$. For $\sigma_n = q(a)(n(1-a))^{-1/2}$, arbitrary $0 < z$ and $V_n = -k_0\sqrt{n}/(1-a)^{1/2}$ we note that

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

In addition, from the proof of Lemma 4 we have that $lim_{n\to\infty} V_n((1-a) - (1 - F(q(a) + \sigma_n z))) = z$. Now, let $W_n = V_n((1 - \tilde{F}(q(a_n) + y_{N,a} + \sigma_n z)) - (1 - F(q(a) + \sigma_n z)))$ and note that $\frac{n(1-F(q(a)))}{V_n(1-F(q(a)+\sigma_n z))} W_n = \sqrt{n(1 - F(q(a)))} \left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)} - 1 \right) = -\frac{1}{k_0} W_n(1 + o(1))$. We first establish that

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

is asymptotically normally distributed. Without loss of generality consider $y_N = q(a_n)(Z_N - 1)$ for $0 < Z_N \to \mathcal{Z} < \infty$. Note that if $Z_N = \mathcal{Z}$, then $y_{N,a} = y_N = q(a_n)(\mathcal{Z} - 1)$. Then, $q(a) + \sigma_n z = q(a_n)\mathcal{Z}(1 + z((1-a)n)^{-1/2}) = q(a_n) Z_N$. 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 $Z_N \to \mathcal{Z}$, $k(Z_N)\phi(q(a_n)) - k(\mathcal{Z}) N^{-1/2} \frac{\mu(\alpha-\rho)}{C} \to 0$ and consequently

$$Z_N^{-1/k_0} \frac{1 - F(q(a_n)Z_N)}{1 - F(q(a_n)} = 1 + k(\mathcal{Z}) N^{-1/2} \frac{\mu(\alpha - \rho)}{C} + o(N^{-1/2}). \tag{30}$$

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

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

and using the notation in Theorem 3 and the mean value theorem gives

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

for $\sigma_N^* = \lambda_1\tilde{\sigma}_N + (1-\lambda_1)\sigma_N$ and $k_N^* = \lambda_2\tilde{k}_N + (1-\lambda_2)k_0$ and $\lambda_1, \lambda_2 \in (0,1)$. It follows from $\sigma_N = -k_0 q(a_n) = -\frac{k_0 y_N}{Z_N - 1}$ that $y_N = \frac{(1-Z_N)\sigma_N}{k_0}$ and from Theorem 3 we have

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

Hence, if $c_b^T = \begin{pmatrix} -k_0^{-1}(\mathcal{Z}^{-1}-1) & k_0^{-2} log(\mathcal{Z}) + k_0^{-2}(\mathcal{Z}^{-1}-1) \end{pmatrix}$ and $\mu_p^T = \begin{pmatrix} \frac{\mu(1-k_0)(1+2k_0\rho)}{1-k_0+k_0\rho} & \frac{\mu(1-k_0)k_0(1+\rho)}{1-k_0+k_0\rho} \end{pmatrix}$ we can write

$$c_b^T \sqrt{N} \begin{pmatrix} \tilde{r}_N - 1 \\ \tilde{k} - k_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(c_b^T \mu_p, c_b^T H^{-1} V_2 H^{-1}) \text{ and } \sqrt{N}(h(\tilde{\sigma}_N, \tilde{k}, y_N) - h(\sigma_N, k_0, y_N)) = O_p(1). \tag{31}$$

Now, we can conveniently write,

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

Note that $\frac{1-\tilde{F}(q(a_n)+y_N)}{1-\tilde{F}(\tilde{q}(a_n))} = \left(1 - \frac{\tilde{k} y_N}{\tilde{\sigma}_N}\right)^{1/\tilde{k}} \left(\frac{1-\tilde{F}(q(a_n))}{1-\tilde{F}(\tilde{q}(a_n))}\right)$ and $Z_N^{-1/k_0} = \left(1 - \frac{k_0 y_N}{\sigma_N}\right)^{-1/k_0} = exp(h(\sigma_N, k_0, y_N))$. Furthermore from equation (30), $Z_N^{1/k_0} \frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))} - 1 = N^{-1/2}\left(-k(\mathcal{Z})\frac{\mu(\alpha-\rho)}{C}\right) + o(N^{-1/2})$. Hence,

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

Now, we note that $\frac{1-\tilde{F}(q(a_n))}{1-\tilde{F}(\tilde{q}(a_n))} - 1 = -\frac{\tilde{F}(q(a_n)) - F(q(a_n))}{1-F(q(a_n))}$ and from Lemma 4 we have $\frac{\sqrt{n(1-F(q(a_n))}}{1-F(q(a_n))}(1 - \tilde{F}(q(a_n)) - (1-F(q(a_n))) \xrightarrow{d} \mathcal{N}(0,1)$ as $q(a_n) \to \infty$. In particular, using the notation adopted in Lemma 4 we have that

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

Hence,

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

and by equation (31) 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

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

Since $\sqrt{N}\left(Z_N^{1/k_0}\frac{1-F(q(a_n))}{(1-F(q(a_n)Z_N))}-1\right) \to -\frac{k(\mathcal{Z})\mu(\alpha-\rho)}{C}$ we focus on the joint distribution of the last two terms. By equation (31) we have that

$$\sqrt{N}(h(\tilde{\sigma}_N,\tilde{k},y_N)-h(\sigma_N,k_0,y_N)) = c_b^T\sqrt{N}\begin{pmatrix}\tilde{r}_N-1\\ \tilde{k}-k_0\end{pmatrix}+o_p(1) \tag{32}$$

and by Theorem 3 (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 $\sum_{t=1}^n Z_{t3}+o_p(1)$, $\sum_{t=1}^n Z_{t2}$ and $\sum_{t=1}^n Z_{t1}$. Hence, we define $\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n))}{(1-\tilde{F}(\tilde{q}(a_n)))}-1\right)=\sum_{t=1}^n Z_{t4}$, let $0\neq \mathbf{d}\in \mathbb{R}^4$,

$$\amalg_n^T = \begin{pmatrix}\sum_{t=1}^n Z_{t1} & \sum_{t=1}^n Z_{t2} & \sum_{t=1}^n Z_{t3} & \sum_{t=1}^n Z_{t4}\end{pmatrix}$$

and consider $\mathbf{d}^T\amalg_n = \sum_{t=1}^n\sum_{\delta=1}^4 Z_{t\delta}d_\delta = \sum_{t=1}^n Z_{nt}$. Note that Z_{nt} forms an iid sequence with $E(Z_{nt})=0$ and the asymptotic behavior of $\sum_{t=1}^n Z_{t1}$, $\sum_{t=1}^n Z_{t2}$ and $\sum_{t=1}^n Z_{t3}$ was studied in Theorem 3. In addition the asymptotic behavior of $\sum_{t=1}^n Z_{t4}$ was studied in Lemma 4. Recall that $E(Z_{t4}^2)=n^{-1}(F(y_n)+o(h_{3n}))$ and from Theorem 3 $E(Z_{t1}Z_{t4})=o(n^{-1})$ and $E(Z_{t2}Z_{t4})=o(n^{-1})$. Here we examine

$$\begin{aligned} E(Z_{t3}Z_{t4}) &= -\frac{k_0}{n((1-F(y_n))(1-F(q(a_n))))^{1/2}}E\left(q_{1n}\frac{1}{h_{3n}}\int_{-\infty}^{q(a_n)}K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy\right)\\ &- E(q_{1n})E\left(\frac{1}{h_{3n}}\int_{-\infty}^{q(a_n)}K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy\right). \end{aligned}$$

By Lemma 4 $E(q_{1n})-F(y_n)=O(h_{3n}^{m+1})$ and similarly we have $E\left(\frac{1}{h_{3n}}\int_{-\infty}^{q(a_n)}K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy\right)-F(q(a_n)) = O(h_{3n}^{m+1})$. Since in Lemma 4 we have $y_n=q(a_n)+\sigma_n z$, then for $\kappa(x)=h_{3n}^{-1}\int_{-\infty}^{x}K_3\left(\frac{y-\varepsilon}{h_{3n}}\right)dy$ we can write

$$E\left(q_{1n}\frac{1}{h_{3n}}\int_{-\infty}^{q(a_n)}K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy\right) = E(\kappa(q(a_n)+\sigma_n z)\kappa(q(a_n)))(\chi_{\{q(a_n)=y_n\}}+\chi_{\{q(a_n)\neq y_n\}}).$$

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

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

By FR2 $lim_{n\to\infty} \frac{F(q(a_n)+\sigma_n z)-F(q(a_n))}{1-F(q(a_n))} = 0$, hence $(1-F(q(a_n)))^{-1}E(\kappa(q(a_n)+\sigma_n z)\kappa(q(a_n))\chi_{\{q(a_n)\neq y_n\}}) = o(1)$ and $E\left(q_{1n}\frac{1}{h_{3n}}\int_{-\infty}^{q(a_n)} K_3\left(\frac{y-\varepsilon_t}{h_{3n}}\right)dy\right) = E\left(\kappa^2(q(a_n))\right) + o(1-F(q(a_n)))$. Consequently,

$$\begin{aligned} E(Z_{t3}Z_{t4}) &= -\frac{k_0}{n((1-F(y_n))(1-F(q(a_n))))^{1/2}}(E\left(\kappa^2(q(a_n))\right) + o(1-F(q(a_n)))) - F^2(q(a_n)) + O(h_{3n}^{m+1}) \\ &= -\frac{k_0}{n}(F(q(a_n)) + o(1)) \end{aligned}$$

and $V(Z_{in}) = \frac{1}{n}d^T V_3 d + o(n^{-1})$ 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 3 we have that $d^T \amalg_n \stackrel{d}{\to} \mathcal{N}(0, d^T V_3 d)$ and from the Cramer-Wold theorem $\amalg_n \stackrel{d}{\to} \mathcal{N}(0, V_3)$. Now, from equation (32)

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

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

$$\begin{aligned} \sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1\right) &= -\frac{k(\mathcal{Z})\mu(\alpha-\rho)}{C} - \left(c_b^T H_{.1}^{-1}\sum_{i=1}^n Z_{t1} + c_b^T H_{.2}^{-1}\sum_{i=1}^n Z_{t2}\right. \\ &+ \left(c_b^T H_{.1}^{-1}b_1 + c_b^T H_{.2}^{-1}b_2\right)\sum_{t=1}^n Z_{t3} \\ &+ \left. c_b^T H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right) + \sum_{i=1}^n Z_{t4} + o_p(1) \\ &= -\frac{k(\mathcal{Z})\mu(\alpha-\rho)}{C} - c_b^T H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \\ &+ \begin{pmatrix} -c_b^T H_{.1}^{-1} & -c_b^T H_{.2}^{-1} & -c_b^T H_{.1}^{-1}b_1 - c_b^T H_{.2}^{-1}b_2 & 1 \end{pmatrix}\amalg_n + o_p(1). \end{aligned}$$

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

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

and for $y_N = q(a_n)(Z_N - 1)$ with $Z_N \to \mathcal{Z}$ we immediately have

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

Lastly, since $-W_n/k_0 + o(1) = \sqrt{n(1-F(q(a)))}\left(\frac{1-\tilde{F}(q(a)+\sigma_n z)}{1-F(q(a)+\sigma_n z)} - 1\right)$ and if $\sqrt{n(1-F(q(a)))} = \sqrt{n(1-a)} \propto N^{1/2}$, that is, $n(1-a) \to \infty$ at the same rate as N, then

$$W_n \xrightarrow{d} N\left((-k_0)\left(-\frac{k(\mathcal{Z})\mu(\alpha-\rho)}{C} - c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right), k_0^2\left(c_b^T H^{-1} V_2 H^{-1} c_b + 2c_b^T \begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\right)$$

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

$$\zeta_1 \sim N\left((-k_0)\left(-\frac{k(\mathcal{Z})\mu(\alpha-\rho)}{C} - c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}\right), k_0^2\left(c_b^T H^{-1} V_2 H^{-1} c_b + 2c_b^T \begin{pmatrix} 2-k_0 \\ 1-k_0 \end{pmatrix} + 1\right)\right).$$

□

Theorem 5.

Proof. We write

$$\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a))}{q(a)/(1+k_0)} - 1 = \frac{\hat{q}(a)/(1+\tilde{k})}{q(a)/(1+k_0)} - 1 = \left(\frac{\hat{q}(a)}{q(a)} - 1\right)\left(\frac{k_0-\tilde{k}}{1+\tilde{k}}\right) + \frac{\hat{q}(a)}{q(a)} - 1 + \frac{k_0-\tilde{k}}{1+\tilde{k}}.$$

From Theorems 3 and 4 we have $\frac{\tilde{k}-k_0}{1+\tilde{k}} = O_p(N^{-1/2})$ and $\frac{\hat{q}(a)}{q(a)} - 1 = O_p(N^{-1/2})$. Hence,

$$\begin{aligned}
\sqrt{N}\left(\frac{\hat{q}(a)/(1+\tilde{k})}{q(a)/(1+k_0)} - 1\right) &= \sqrt{N}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) - \sqrt{N}\left(\frac{\tilde{k}-k_0}{1+k_0}\right) + o_p(1) \\
&= \begin{pmatrix} 1 & -(1+k_0)^{-1} \end{pmatrix}\begin{pmatrix} \sqrt{N}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \\ \sqrt{N}\left(\tilde{k}-k_0\right) \end{pmatrix} + o_p(1).
\end{aligned}$$

Hence, it suffices to obtain the joint distribution of the vector $\begin{pmatrix} \sqrt{N}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) \\ \sqrt{N}\left(\tilde{k}-k_0\right) \end{pmatrix}$. From Theorem 4 we have that $\sqrt{N}\left(\frac{\hat{q}(a)}{q(a)} - 1\right) - (-k_0)\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1\right) = o_p(1)$, where $y_N = q(a_n)(Z_N - 1)$ for $0 < Z_N \to \mathcal{Z} < \infty$. Also,

$$(-k_0)\sqrt{N}\left(\frac{1-\tilde{F}(q(a_n)+y_N)}{1-F(q(a_n)+y_N)} - 1\right) = (-k_0)\frac{-(\mathcal{Z}^\rho - 1)\mu(\alpha-\rho)}{\rho} - (-k_0)c_b^T H^{-1}\sqrt{N}\begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} + (-k_0)\eta^T \mathrm{II}_n + o_p(1),$$

where $\text{II}_n^T = \begin{pmatrix} \sum_{t=1}^n Z_{t1} & \sum_{t=1}^n Z_{t2} & \sum_{t=1}^n Z_{t3} & \sum_{t=1}^n Z_{t4} \end{pmatrix}^T \xrightarrow{d} \mathcal{N}(0, V_3)$ and the structure of $\sum_{t=1}^n Z_{tj}$ for $j = 1, \cdots, 4$ and V_3 are given in Theorem 4. From Theorem 3

$$\begin{aligned}
\sqrt{N}(\tilde{k} - k_0) - \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} &= \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \left(v_N(1, k_0) - \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) + o_p(1) \\
&= \begin{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} & \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^n Z_{t1} \\ \sum_{t=1}^n Z_{t2} \\ \sum_{t=1}^n Z_{t3} \end{pmatrix} \\
&+ o_p(1).
\end{aligned}$$

Hence, we can write,

$$\begin{pmatrix} \sqrt{N} \left(\frac{\hat{q}(a)}{q(a)} - 1 \right) \\ \sqrt{N} \left(\tilde{k} - k_0 \right) \end{pmatrix} = \begin{pmatrix} (-k_0) \frac{-(\mathcal{Z}^\rho - 1)\mu(\alpha - \rho)}{\rho} - (-k_0) c_b^T H^{-1} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \\ \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \end{pmatrix} + \begin{pmatrix} -k_0 \eta^T \\ \theta^T \end{pmatrix} \text{II}_n + o_p(1)$$

where $\theta^T = \left(\begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \quad \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad 0 \right)$. Consequently,

$$\begin{pmatrix} \sqrt{N} \left(\frac{\hat{q}(a)}{q(a)} - 1 - (-k_0) \frac{-(\mathcal{Z}^\rho - 1)\mu(\alpha - \rho)}{\rho} + (-k_0) c_b^T H^{-1} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) \\ \sqrt{N} \left(\tilde{k} - k_0 - \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right) \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, V_4\right).$$

where $V_4 = \begin{pmatrix} -k_0 \eta^T \\ \theta^T \end{pmatrix} V_3 \begin{pmatrix} -k_0 \eta^T \\ \theta^T \end{pmatrix}^T$. Thus, it follows immediately that

$$\begin{aligned}
\sqrt{N} \left(\frac{\hat{q}(a)/(1+\tilde{k})}{q(a)/(1+k_0)} - 1 \right) - \begin{pmatrix} 1 & -(1+k_0)^{-1} \end{pmatrix} \begin{pmatrix} (-k_0) \frac{-(\mathcal{Z}^\rho - 1)\mu(\alpha - \rho)}{\rho} - (-k_0) c_b^T H^{-1} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \\ \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \end{pmatrix} \\
\xrightarrow{d} N\left(0, \begin{pmatrix} 1 & -(1+k_0)^{-1} \end{pmatrix} V_4 \begin{pmatrix} 1 & -(1+k_0)^{-1} \end{pmatrix}^T \right).
\end{aligned}$$

Additional algebra, gives

$$\begin{aligned}
\sqrt{n(1-a)} \left(\frac{\hat{q}(a)/(1+\tilde{k})}{q(a)/(1+k_0)} - 1 \right) &\xrightarrow{d} N\left(k_0 \frac{(\mathcal{Z}^\rho - 1)\mu(\alpha - \rho)}{\rho} + k_0 c_b^T H^{-1} \lim_{n\to\infty} \sqrt{N} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix} \right. \\
&- \left. \frac{1}{1+k_0} \lim_{n\to\infty} \sqrt{N} \begin{pmatrix} 0 & 1 \end{pmatrix} H^{-1} \begin{pmatrix} b_\sigma \\ b_k \end{pmatrix}, \Sigma \right),
\end{aligned}$$

where $\Sigma = k_0^2 \eta^T V_3 \eta + 2 \frac{k_0}{1+k_0} \eta^T V_3 \theta + \frac{1}{(1+k_0)^2} \theta^T V_3 \theta$, with $\eta^T V_3 \eta = \left(c_b^T H^{-1} V_2 H^{-1} c_b + 2 c_b^T \begin{pmatrix} 2 - k_0 \\ 1 - k_0 \end{pmatrix} + 1 \right)$ from Theorem 4. □

Theorem 6.

Proof. a) We write

$$\frac{\hat{q}_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)}{q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)} - 1 = \frac{\hat{m}(\mathbf{x}) - m(\mathbf{x})}{m(\mathbf{x}) + h^{1/2}(\mathbf{x})q(a)} + \frac{\left(\hat{h}^{1/2}(\mathbf{x}) - h^{1/2}(\mathbf{x})\right)}{\left(\frac{m(\mathbf{x})}{q(a)} + h^{1/2}(\mathbf{x})\right)} \frac{\hat{q}(a)}{q(a)} + \frac{h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{q(a)} + h^{1/2}(\mathbf{x})\right)} \left(\frac{\hat{q}(a) - q(a)}{q(a)}\right).$$

From Lemma 2, $\sup_{\mathbf{x}\in\mathcal{G}}|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p(L_{1n})$. As $n \to \infty$, $a \to 1$ and $q(a) \to \infty$. Hence, given that $h^{1/2}(\mathbf{x})$ is bounded away from zero for fixed $\mathbf{x}$ by assumption A3 2), we have that $\frac{\hat{m}(\mathbf{x})-m(\mathbf{x})}{m(\mathbf{x})+h^{1/2}(\mathbf{x})q(a)} = o_p(L_{1n})$. Now, given A6 1) and $n(1-a) \propto N$ we have $\sqrt{n(1-a)}\frac{\hat{m}(\mathbf{x})-m(\mathbf{x})}{m(\mathbf{x})+h^{1/2}(\mathbf{x})q(a)} = o_p(1)$. From Corollary 1, $\sup_{\mathbf{x}\in\mathcal{G}}|\hat{h}^{1/2}(\mathbf{x}) - h^{1/2}(\mathbf{x})| = O_p(L_{1n} + L_{2n})$. Hence, given A6 1), $n(1-a) \propto N$ and the fact that $m(\mathbf{x})$ is bounded for fixed $\mathbf{x}$ we have $\sqrt{n(1-a)}\frac{\left(\hat{h}^{1/2}(\mathbf{x})-h^{1/2}(\mathbf{x})\right)}{\left(\frac{m(\mathbf{x})}{q(a)}+h^{1/2}(\mathbf{x})\right)} = o_p(1)$. From Theorem 4 we have $\frac{\hat{q}(a)}{q(a)} = 1 + o_p(1)$, which gives $\sqrt{n(1-a)}\frac{\left(\hat{h}^{1/2}(\mathbf{x})-h^{1/2}(\mathbf{x})\right)}{\left(\frac{m(\mathbf{x})}{q(a)}+h^{1/2}(\mathbf{x})\right)}\frac{\hat{q}(a)}{q(a)} = o_p(1)$. Lastly, since $q(a) \to \infty$ as $n \to \infty$, for fixed $\mathbf{x}$ we have $\frac{h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{q(a)}+h^{1/2}(\mathbf{x})\right)} \to 1$ and by Theorem 4 $\sqrt{n(1-a)}\left(\frac{\hat{q}(a)-q(a)}{q(a)}\right) \xrightarrow{d} \mathcal{N}(\mu_1, \Sigma_1)$, which gives the desired result.

b) We write

$$\frac{\hat{E}\left(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right)}{E\left(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right)} - 1 = \frac{\hat{m}(\mathbf{x}) - m(\mathbf{x})}{m(\mathbf{x}) + h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t > q(a))} + \frac{\hat{h}^{1/2}(\mathbf{x}) - h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{E(\varepsilon_t|\varepsilon_t>q(a))} + h^{1/2}(\mathbf{x})\right)} \frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a))}{E(\varepsilon_t|\varepsilon_t > q(a))}$$
$$+ \frac{h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{E(\varepsilon_t|\varepsilon_t>q(a))} + h^{1/2}(\mathbf{x})\right)} \left(\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a)) - E(\varepsilon_t|\varepsilon_t > q(a))}{E(\varepsilon_t|\varepsilon_t > q(a))}\right)$$

As in part a), since $m(\mathbf{x}) + h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t > q(a)) \to \infty$ as $n \to \infty$, given Lemma 2 and A6 1) and $n(1-a) \propto N$, $\sqrt{n(1-a)}\frac{\hat{m}(\mathbf{x})-m(\mathbf{x})}{m(\mathbf{x})+h^{1/2}(\mathbf{x})E(\varepsilon_t|\varepsilon_t>q(a))} = o_p(1)$. By equation (40) in Lemma 7 we can write

$$\sqrt{n(1-a)}\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a)) - E(\varepsilon_t|\varepsilon_t > q(a))}{E(\varepsilon_t|\varepsilon_t > q(a))} = \sqrt{n(1-a)}\left(\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a))}{\frac{q(a)}{1+k_0}} - 1\right)\left(\frac{E(\varepsilon_t|\varepsilon_t > q(a))}{\frac{q(a)}{1+k_0}}\right)^{-1}$$
$$- \frac{(1+k_0)\sqrt{n(1-a)}\left(\frac{C\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} + o\left(\phi(q(a))\right)\right)}{\frac{E(\varepsilon_t|\varepsilon_t>q(a))}{\frac{q(a)}{1+k_0}}}.$$

By Theorem 5, $\sqrt{n(1-a)}\left(\frac{\hat{E}(\varepsilon_t|\varepsilon_t>q(a))}{\frac{q(a)}{1+k_0}} - 1\right) \xrightarrow{d} \mathcal{N}(\mu_2, \Sigma_2)$ and by Lemma 7, $\frac{E(\varepsilon_t|\varepsilon_t>q(a))}{\frac{q(a)}{1+k_0}} = 1 + o(1)$. Since $n(1-a) \propto N$ we investigate the order of $(1+k_0)\frac{CN^{1/2}\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} + (1+k_0)N^{1/2}\phi(q(a))o(1)$. We note that

$$\frac{CN^{1/2}\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} = \frac{1}{(\rho-\alpha+1)(1-\alpha)}\frac{N^{1/2}C\phi(q(a_n))}{\alpha-\rho}\frac{(\alpha-\rho)\phi(q(a))}{\phi(q(a_n))}$$

and since by assumption $\frac{N^{1/2}C\phi(q(a_n))}{\alpha-\rho} \to \mu$ we need only investigate $\frac{\phi(q(a))}{\phi(q(a_n))}$. As in Theorem 4, without loss of generality, there exists a sequence $Z_N \to \mathcal{Z}$ such that $q(a) = q(a_n)Z_N$ and we write $\frac{\phi(q(a))}{\phi(q(a_n))} = \frac{\phi(q(a_n)Z_N)}{\phi(q(a_n))}$. Since, ϕ is regularly varying with index $\rho \le 0$, $\frac{\phi(q(a_n)Z_N)}{\phi(q(a_n))} \to \mathcal{Z}^\rho$ as $n \to \infty$ and we have $\frac{CN^{1/2}\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} \to \frac{\mu(\alpha-\rho)\mathcal{Z}^\rho}{(\rho-\alpha+1)(1-\alpha)}$. Similar arguments show that $N^{1/2}\phi(q(a)) = O(1)$. Hence, given that $k_0 = -\alpha^{-1}$ we have

$$\sqrt{n(1-a)}\frac{\hat{E}(\varepsilon_t|\varepsilon_t > q(a)) - E(\varepsilon_t|\varepsilon_t > q(a))}{E(\varepsilon_t|\varepsilon_t > q(a))} \xrightarrow{d} \mathcal{N}\left(\mu_2 - \frac{(\rho-\alpha)\mu}{\alpha(\rho-\alpha+1)}\mathcal{Z}^\rho, \Sigma_2\right). \tag{33}$$

An immediate consequence of equation (33) is that $\frac{\hat{E}(\varepsilon_t|\varepsilon_t>q(a))}{E(\varepsilon_t|\varepsilon_t>q(a))} = 1 + o_p(1)$. Furthermore, given Corollary 1, assumption A6 1) and $n(1-a) \propto N$ we have that $\sqrt{n(1-a)}\frac{\hat{h}^{1/2}(\mathbf{x})-h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{E(\varepsilon_t|\varepsilon_t>q(a))}+h^{1/2}(\mathbf{x})\right)} = o_p(1)$. Finally, since as $n \to \infty$, $\frac{h^{1/2}(\mathbf{x})}{\left(\frac{m(\mathbf{x})}{E(\varepsilon_t|\varepsilon_t>q(a))}+h^{1/2}(\mathbf{x})\right)} \to 1$, we have

$$\sqrt{n(1-a)}\left(\frac{\hat{E}\left(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right)}{E\left(Y_t|Y_t > q_{Y_t|\mathbf{X}_t=\mathbf{x}}(a)\right)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\mu_2 - \frac{(\rho-\alpha)\mu}{\alpha(\rho-\alpha+1)}\mathcal{Z}^\rho, \Sigma_2\right).$$

□

Lemma 1. *Let $w(\mathbf{X}_t - \mathbf{x}; \mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ and $g(\varepsilon) : \mathbb{R} \to \mathbb{R}$ be measurable functions and define*

$$s(\mathbf{x}) = \frac{1}{nh_n^d}\sum_{t=1}^{n} K\left(\frac{\mathbf{X}_t - \mathbf{x}}{h_n}\right)\left(\frac{X_{ti} - x_i}{h_n}\right)^{p_1}\left(\frac{X_{tj} - x_j}{h_n}\right)^{p_2}\left(\frac{X_{tl} - x_l}{h_n}\right)^{p_3} w(\mathbf{X}_t - \mathbf{x}; \mathbf{x})g(\varepsilon_t) \tag{34}$$

where K is a multivariate kernel given by $K(\mathbf{x}) = \prod_{j=1}^d \mathcal{K}(x_j)$, $h_n > 0$ is a bandwidth, for $i, j = 1, \cdots, d$ and $p_1, p_2, p_3 = 0, 1$. Assume that A1 and A2 are holding and that:

a) $E\left(|g(\varepsilon_t)|^a\right) < \infty$ for some $a > 2$;

b) $w(\mathbf{X}_t - \mathbf{x}; \mathbf{x})$ satisfies a Lipschitz condition of order 1, i.e., $|w(\mathbf{X}_t - \mathbf{x}; \mathbf{x}) - w(\mathbf{X}_t - \mathbf{x}^k; \mathbf{x}^k)| \le C\|\mathbf{x} - \mathbf{x}^k\|_E$ for some $C > 0$ and $\mathbf{x} \ne \mathbf{x}^k$ in $\mathbb{R}^d$ and $|w(\mathbf{X}_t - \mathbf{x}; \mathbf{x})| < C$ for all $x \in \mathbb{R}^d$;

c) The joint density of $\mathbf{X}_i$ and $\mathbf{X}_j$ conditional on ε_i and ε_j denoted by $f_{\mathbf{X}_i\mathbf{X}_j|\varepsilon_i\varepsilon_j}(\mathbf{X}_i, \mathbf{X}_j) < C$.

Then, for an arbitrary compact set $\mathcal{G} \subseteq \mathbb{R}^d$, we have

$$\sup_{\mathbf{x}\in\mathcal{G}}|s(\mathbf{x}) - E(s(\mathbf{x}))| = O_p\left(\left(\frac{\log n}{nh_n^d}\right)^{1/2}\right) \tag{35}$$

provided that for $a, B > 2$, $\theta > 0$, we have

$$n^{1-\frac{2}{a}-2\theta}h_n^d \to \infty \tag{36}$$

and

$$n^{(B+1.5)(\frac{1}{a}+\theta)-\frac{B}{2}+0.75+\frac{d}{2}}h_n^{-1.75d-\frac{d}{2}(d+B)}(log\, n)^{0.25+0.5(B-d)} \to 0. \tag{37}$$

Proof. See Martins-Filho et al. (2013). □

Lemma 2. *Assume that the kernel K_1 used to define $\hat{m}$ satisfies assumption A1 and assumptions A2 and A3 are holding. Assume also that the bandwidth h_{1n} used to define $\hat{m}$ satisfies equations (36) and (37). Then, if $E(|\varepsilon_t|^a) < \infty$, $E(h^{1/2}(\mathbf{X}_t)^a) < \infty$ for some $a > 2$ and condition c) in Lemma 1 is holding*

$$\sup_{\mathbf{x}\in\mathcal{G}}|\hat{m}(\mathbf{x}) - m(\mathbf{x})| = O_p\left(L_{1n}\right), \tag{38}$$

where $L_{1n} = \left(\frac{log\, n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$.

Proof. See Martins-Filho et al. (2013). □

Lemma 3. *Assume that the kernel K_2 used to define $\hat{h}$ satisfies assumption A1 and assumptions A2 and A3 are holding. Assume also that the bandwidth h_{2n} used to define $\hat{h}$ satisfies equations (36) and (37). Then, under the assumptions in Lemma 2, if $E(|\varepsilon_t^2 - 1|^a) < \infty$ and $E(h(\mathbf{X}_t)^a) < \infty$ for some $a > 2$,*

$$\sup_{\mathbf{x}\in\mathcal{G}}|\hat{h}(\mathbf{x}) - h(\mathbf{x})| = O_p\left(L_{1n} + L_{2n}\right), \tag{39}$$

where $L_{1n} = \left(\frac{log\, n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^s$ *and* $L_{2n} = \left(\frac{log\, n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^s$.

Proof. See Martins-Filho et al. (2013). □

Corollary 1. *Under the assumptions of Lemma 3,*

$$\sup_{\mathbf{x}\in\mathcal{G}}|\hat{h}^{1/2}(\mathbf{x}) - h^{1/2}(\mathbf{x})| = O_p\left(L_{1n} + L_{2n}\right) \quad \text{and} \quad \sup_{\mathbf{x}\in\mathcal{G}}|\chi_{\{\hat{h}(\mathbf{x})>0\}} - 1| = O_p\left(L_{1n} + L_{2n}\right),$$

where $L_{1n} = \left(\frac{log\, n}{nh_{1n}^d}\right)^{1/2} + h_{1n}^2$ *and* $L_{2n} = \left(\frac{log\, n}{nh_{2n}^d}\right)^{1/2} + h_{2n}^2$.

Lemma 4. *Under assumptions A1-A6 and conditions FR and FR2, if $\alpha \geq 1$ we have*

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

Proof. See Martins-Filho et al. (2013). □

Lemma 5. *Let* $a_n = 1 - \frac{N}{n}$ *and for* $i = 1, \cdots, N$ *define* $Z_i = \varepsilon_i - q_n(a_n)$ *whenever* $\varepsilon_i > q_n(a_n)$ *and for* $i = 1, \cdots, N_1$ *define* $Z_i' = \varepsilon_i - q(a_n)$ *whenever* $\varepsilon_i > q(a_n)$. *If* $\Delta_\sigma = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial \sigma} log\, g(Z_i; \sigma_N, k_0)\sigma_N - \frac{1}{N}\sum_{i=1}^{N_1} \frac{\partial}{\partial \sigma} log\, g(Z_i'; \sigma_N, k_0)\sigma_N$ *and* $\Delta_k = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial k} log\, g(Z_i; \sigma_N, k_0) - \frac{1}{N}\sum_{i=1}^{N_1} \frac{\partial}{\partial k} log\, g(Z_i'; \sigma_N, k_0)$, *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. The proof is identical to that of Lemma 3 in Martins-Filho et al. (2014) by substituting their $U_{(n-N)}$ with $\varepsilon_{(n-N)}$. □

Lemma 6. $E\left(log\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)\left(1 - \frac{k_0 Z_i'}{\sigma_N}\right)^{-1}\left(\frac{k_0 Z_i'}{\sigma_N}\right)\right) = -\frac{1}{\alpha} + \frac{\alpha}{(1+\alpha)^2} + O(\phi(\varepsilon_{(n-N)}))$

Proof. The proof is identical to that of Lemma 4 in Martins-Filho et al. (2014) by substituting their $U_{(n-N)}$ with $\varepsilon_{(n-N)}$. □

Lemma 7. *Under conditions FR1 with* $\alpha > 1$, *FR2 and* $a \in (a_n, 1)$ *with* $a_n = 1 - \frac{N}{n} \to 1$ *as* $n \to \infty$, *we have* $\frac{E(\varepsilon_t|\varepsilon_t > q(a))}{\frac{q(a)}{1+k_0}} = 1 + o(1)$.

Proof. Denote the distribution of $\varepsilon_t - q(a)$ given that $\varepsilon_t > q(a)$ evaluated at z by $F_{\varepsilon_t - q(a)|\varepsilon_t > q(a)}(z)$. Since $F_{\varepsilon_t - q(a)|\varepsilon_t > q(a)}(z) = \frac{F(z+q(a)) - F(q(a))}{1-F(q(a))}$, by FR2 we can write $F_{\varepsilon_t - q(a)|\varepsilon_t > q(a)}(z) = 1 - \frac{L(z+q(a))}{L(q(a))}\left(\frac{z+q(a)}{q(a)}\right)^{-\alpha}$. Observe that under FR1 $\alpha = -1/k_0$ and $q(a_n) = -\sigma_N/k_0$, hence $\frac{z+q(a)}{q(a)} = 1 - \frac{k_0 z}{\sigma_N - k_0(q(a) - q(a_n))} \equiv t_n(z)$ and we write $F_{\varepsilon_t - q(a)|\varepsilon_t > q(a)}(z) = 1 - \frac{L(t_n(z)q(a))}{L(q(a))} t_n(z)^{1/k_0}$. Hence,

$$\begin{aligned} E(\varepsilon_t|\varepsilon_t > q(a)) &= q(a) - \int_{q(a)}^{\infty} (\varepsilon - q(a)) d\left(\frac{L(t_n(\varepsilon - q(a))q(a))}{L(q(a))} t_n(\varepsilon - q(a))^{1/k_0}\right) \text{ and integrating by parts} \\ &= -(\varepsilon - q(a))\left(\frac{L(t_n(\varepsilon - q(a))q(a))}{L(q(a))} t_n(\varepsilon - q(a))^{1/k_0}\right)\Big|_{q(a)}^{\infty} \\ &+ \int_{q(a)}^{\infty} \frac{L(t_n(\varepsilon - q(a))q(a))}{L(q(a))} t_n(\varepsilon - q(a))^{1/k_0} d\varepsilon + q(a) \end{aligned}$$

Denoting the first term on the right side of the equality by I_1 and the second term by I_2, we observe that

$$I_1 = -\lim_{\varepsilon \to \infty} \frac{\varepsilon - q(a)}{(1 - F(q(a)))/1 - F(\varepsilon)} = -\lim_{\varepsilon \to \infty} \frac{\varepsilon - q(a)}{\left(1 - \frac{k_0(\varepsilon - q(a))}{\sigma_N - k_0(q(a) - q(a_n))}\right)^{-1/k_0}} \lim_{\varepsilon \to \infty} \frac{L(q(a)t_n(\varepsilon - q(a)))}{L(q(a))}.$$

By FR2 $\lim_{\varepsilon\to\infty} \frac{L(q(a)t_n(\varepsilon-q(a)))}{L(q(a))} = 1$, and since $-1 < k_0 < 0$ $(\alpha > 1)$, $\lim_{\varepsilon\to\infty} \frac{\varepsilon-q(a)}{\left(1-\frac{k_0(\varepsilon-q(a))}{\sigma_N-k_0(q(a)-q(a_n))}\right)^{-1/k_0}} = 0$. By FR1 and FR2, changing variables in I_2, we write

$$\begin{aligned} I_2 &= q(a)\int_1^\infty \frac{L(q(a)t)}{L(q(a))} t^{1/k_0} dt = q(a)\left(\int_1^\infty t^{1/k_0} dt + \phi(q(a))\int_1^\infty t^{1/k_0} k(t)dt + o(\phi(q(a)))\right) \\ &= -q(a)\frac{k_0}{1+k_0} + q(a)\left(\frac{C\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} + o\left(\phi(q(a))\right)\right) \end{aligned}$$

where the last equality follows from $\int_1^\infty t^{1/k_0} dt = -\frac{k_0}{1+k_0}$ and $\int_1^\infty t^{1/k_0} k(t)dt = \frac{C}{(\rho-\alpha+1)(1-\alpha)}$. Consequently,

$$E(\varepsilon_t|\varepsilon_t > q(a)) = \frac{q(a)}{1+k_0} + q(a)\left(\frac{C\phi(q(a))}{(\rho-\alpha+1)(1-\alpha)} + o\left(\phi(q(a))\right)\right). \tag{40}$$

Since as $n \to \infty$, $a \to 1$, then $q(a) \to \infty$ and $(\phi(q(a)) \to 0$, giving the desired result. □

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, 799–808.

Cai, Z., 2002. Regression quantiles for time series. Econometric Theory 18, 169–192.

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

Chen, S. X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6, 87–107.

Chen, S. X., Tang, C., 2005. Nonparametric inference of value at risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

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

Chernozhukov, V., Umantsev, L., 2001. Conditional Value-at-Risk: aspects of modelling and estimation. Empirical Economics 26, 271–292.

Danielsson, J., 2011. Financial risk forecasting. John Wiley and Sons, New York, NY.

Davidson, J., 1994. Stochastic limit theory. Oxford University Press, New York.

Davis, R., Resnick, S. I., 1984. Tail estimates motivated by Extreme Value Theory. Annals of Statistics 12, 1467–1487.

Doukhan, P., 1994. Mixing: properties and examples. Springer-Verlag, New York.

Duffie, D., Singleton, K., 2003. Credit risk: pricing, measurement and management. Princeton University Press, Princeton, NJ.

Efron, B., Tibshirani, R., 1993. An introduction to the Bootstrap. Chapman and Hall, New York.

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

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Gao, J., 2007. Nonlinear time series: nonparametric and parametric methods. Chapman and Hall, New York.

Gnedenko, B. V., 1943. Sur la distribution limite du terme d'une série aléatoire. Annals of Mathematics 44, 423–453.

Hardle, W., Tsybakov, A. B., 1997. Local polynomial estimators of the volatility function in nonparametric autoregression. Journal of Econometrics 81, 233–242.

Hill, J. B., 2014. Expected shortfall estimation and Gaussian inference for infinite variance time series. Journal of Financial Econometrics forthcoming.

Kato, K., 2012. Weighted Nadaraya-Watson estimation of conditional expected shortfall. Journal of Financial Econometrics 10, 265–291.

Leadbetter, M., Lindgren, G., Rootzen, H., 1983. Extremes and related properties of random sequences and processes. Springer Verlag, New York.

Linton, O. B., Xiao, Z., 2013. Estimation of and inference about the expected shortfall for time series with infinite variance. Econometric Theory 29, 771–807.

Martins-Filho, C., Yao, F., 2006. 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.

Martins-Filho, C., Yao, F., Torero, M., 2013. Technical supplement to Nonparametric Estimation of Conditional Value-at-risk and Expected Shortfall based on Extreme Value Theory. http://spot.colorado.edu/∼martinsc/Research.html., University of Colorado at Boulder.

Martins-Filho, C., Yao, F., Torero, M., 2014. High order conditional quantile estimation based on nonparametric models of regression. Econometric Reviews, forthcoming.

Masry, E., Fan, J., 1997. Local polynomial estimation of regression functions for mixing processes. Scandinavian Journal of Statistics 24, 1965–1979.

Masry, E., Tjstheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. Econometric Theory 11, 258–289.

McNeill, 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.

McNeill, A. J., Frey, B., Embrechts, P., 2005. Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton, NJ.

Pagan, A., Ullah, A., 1999. Nonparametric econometrics. Cambridge University Press, Cambridge, UK.

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

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

Ruppert, D., Sheather, S., Wand, M. P., 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association 90, 1257–1270.

Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115–129.

Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. Applied Stochastic Models in Business and Industry 24, 115–129.

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.

Yu, K., Jones, M. C., 1998. Local linear quantile regression. Journal of the American Statistical Association 93, 228–237.