Adapting extreme value statistics to financial time series: dealing with bias and serial dependence

Laurens DE HAAN[(1)]
ldehaan@ese.eur.nl

Cécile MERCADIER[(2)]
mercadier@math.univ-lyon1.fr

Chen ZHOU[(1,3)]
zhou@ese.eur.nl

[(1)]Dept. of Economics,
Erasmus university,
P.O. Box 1738,
3000 DR Rotterdam,
The Netherlands.

[(2)]Université de Lyon,
CNRS, Université Lyon 1,
Institut Camille Jordan,
43 blvd du 11 novembre 1918,
69622 Villeurbanne, France.

[(3)]Economic and Research Div.
De Nederlandsche Bank,
P.O. Box 98,
1000AB Amsterdam,
The Netherlands.

Abstract

We handle two major issues in applying extreme value analysis to financial time series, bias and serial dependence, jointly. This is achieved by studying bias correction method when observations exhibit weakly serial dependence, namely the $\beta-$mixing series. For estimating the extreme value index, we propose an asymptotically unbiased estimator and prove its asymptotic normality under the $\beta-$mixing condition. The bias correction procedure and the dependence structure have an interactive impact on the asymptotic variance of the estimator. Then, we construct an asymptotically unbiased estimator of high quantiles. Simulations show that finite sample performance of the estimators reflects their theoretical properties. We apply the new method to estimate the Value-at-Risk of the daily return on the Dow Jones Industrial Average Index.

Keywords: Hill estimator; bias correction; β-mixing condition; tail empirical quantile function

AMS 2000 Subject Classification: Primary 62G32; secondary 60G70

1 Introduction

In financial risk management, a key concern is on modeling and evaluating potential losses occurring with extremely low probabilities, i.e. tail risks. For example, the Basel committee on banking supervision suggests regulators to require banks holding adequate capital against the tail risk of bank assets measured by the Value-at-Risk (VaR). The VaR refers to high quantile of the loss distribution with an extremely low tail probability.[1] Estimating such risk measures thus relies on modeling the tail region of distribution functions of asset values. To serve such a purpose, statistical tools stemming from Extreme Value Theory (EVT) are obvious candidates. By investigating data in an

[1]In the revised Basel II accord and the subsequent Basel III accord, the VaR measures for risks on both trading and banking books must be calculated at a 99.9% level.

intermediate region close to the tail, extreme value statistics employs models to extrapolate intermediate properties to the tail region. Although such an attractive feature of extreme value statistics makes it a popular tool for evaluating tail events in many scientific fields such as meteorology and engineering, it has not yet emerged into as a dominating tool in financial risk management. This is potentially due to some crucial critiques on applying EVT to financial data; see, e.g. Diebold et al. [2000]. The critiques are mainly on two issues: the difficulty in selecting the intermediate region in estimation and the validity of the maintained assumptions in EVT for financial data. This paper tries to deal with the two critiques simultaneously and provide adapted EVT methods that overcome the two issues jointly.

We start with explaining the problem on selecting the intermediate region in estimation. Extreme value statistics usually use only observations in an intermediate region. This has been achieved by selecting the highest (or lowest when dealing with lower tail) $k = k(n)$ observations in a sample with size n. The problem on selecting k is sometimes referred to as "selecting the cutoff point". Theoretically, the statistical properties of EVT-based estimators are established for k such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. For a finite sample application, such a condition can not be verified. Thus, in practice, it is necessary to solve the problem on the choice of the number of high observations used in estimation, or selecting the "cutoff point". For financial practitioner, two difficulties arise: firstly, there is no automatical procedure on the choice of k; secondly, the performance of the EVT estimators is rather sensitive to this choice. More specifically, there is a bias-variance tradeoff: with low level of k, the estimation variance is at a high level which may not be acceptable for application; by increasing k, i.e. using progressively more data, the variance is reduced, but at the cost of an increasing bias, because the EVT models are assumed to hold only in the tail region.

Recent developments in extreme value statistics provide two solutions. One solution considers establishing an automatic procedure to select the optimal cutoff point that balances the bias and variance by a bootstrapping method, see, e.g. Danielsson et al. [2001]. The other solution considers correcting the bias that exists in EVT estimators and consequently allowing a higher and less sensitive level of k, see, e.g. Gomes et al. [2008]. Compared to the bootstrapping method, the bias correction method allows eventually a larger value of k and consequently a lower level of estimation variance without suffering from the bias issue. Moreover, it leads to a more flexible choice of k. For application, the bias correction method is thus preferred for handling the problem on selecting the cutoff point.

The other criticism on applying extreme value statistics to financial data is on the fact that most existing EVT methods require independent and identically distributed (i.i.d.) observations whereas financial time series exhibits obvious serial dependence feature such as volatility clustering. This issue has been addressed in works dealing with weakly serial dependence, see, e.g. Hsing [1991] and Drees [2000]. The main message from these studies is that usual EVT methods are still valid under weakly serial dependence, although the asymptotic variance of estimators may differ from that in the i.i.d. case.

With the aforementioned developments in extreme value statistics, it seems that one can handel the two major critiques on EVT methods, and the EVT tools should be ready for applications to financial data. However, the literature addressing these two issues are mutually exclusive: in the bias correction literature, it is always assumed that the observations form an i.i.d. sample; in the literature on dealing with serial dependence, the choice of k is assumed to be sufficiently low such that there is no asymptotic bias. Therefore, it is still an open question whether we can apply the bias correction technique to datasets that exhibit weakly serial dependence. The answer to such a question can ultimately clear the two hurdles that prevent the application of EVT-type tools. This is what we tend to address in this paper.

We consider bias correction procedure on estimating the extreme value index and high quantiles for $\beta-$mixing stationary time series with common heavy-tailed distribution. The bias term stems from the approximation of the tail region of distribution functions. In EVT, a second order condition is often imposed to characterize such an approximation. Such a condition is almost indispensable for establishing asymptotic properties of estimators. To correct the bias, one needs to estimate the second order scale function, the function A in (3) below. The existing literature is restricted to the case $A(t) = Ct^{\rho}$ with constants $C \neq 0$ and $\rho < 0$. The estimation of C requires extra conditions. Instead we estimate the function A in a non-parametric way which makes the analysis and application smoother. We prove that the asymptotically unbiased estimators possess usual statistical properties such as asymptotic normality. The bias correction procedure and the serial dependence structure have an interactive impact on the asymptotic variances of the estimators. We conduct simulations to show the performance of the asymptotically unbiased estimators under weakly serial dependence. The method is applied to estimate VaR of the daily returns on a stock market index.

The asymptotically unbiased estimators we obtain have the following advantages. Firstly, it allows serial dependence in the observations. Secondly, one may apply the unbiased estimator with a higher value of k, which reduces the asymptotic variance and ultimately the estimation error thanks to the bias correction feature. Thirdly, the theoretical range of potential choices of k in our unbiased estimators is larger than the original estimators. This makes the choice of k less crucial. All these features become apparent in simulations and application.

The paper is organized as follows. Under a simplified model without serial dependence, Section 2 presents the bias correction idea for the Hill estimator. Section 3 presents the general model with serial dependence, particularly, the regulatory conditions we are dealing with. Section 4 defines the asymptotically unbiased estimators of the extreme value index and quantiles. In addition, we state the main theorems on the asymptotic normality of these two estimators. The asymptotic variances of the unbiased estimators are computed in Section 5 for particular models. Section 6 demonstrates finite sample performance of the asymptotically unbiased estimators based on simulations. An application to estimate the VaR of daily returns on the Dow Jones Industrial Average Index is given in Section 7. All proofs are postponed to Appendix A.

2 The idea of bias correction under independence

For the sake of simplicity, we first introduce our bias correction idea under the assumption of independent and identically distributed (i.i.d.) observations in this section. We will show later that our bias correction procedure also works for $\beta-$mixing series.

2.1 The origin of bias

Let $\{X_1, X_2, \ldots\}$ be an i.i.d. sequence of random variables with a common distribution function F. We assume that the distribution function F belongs to the domain of attraction. We present the domain of attraction condition with respect to the quantile function $U := (1/1-F)^{\leftarrow}$, where $^{\leftarrow}$ denotes the left-continuous inverse function. That is, there exists a positive function $a(t)$ and a real number γ such that

$$\lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad \text{for} \;\; x \geq 1.$$

The parameter γ is called the extreme value index, which largely determines the tail behavior of F. We focus on the case $\gamma > 0$ in which case the domain of attraction condition is equivalent to

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \qquad x > 0\,. \tag{1}$$

The distribution function F with a positive γ is also referred as a heavy-tailed distribution. For a heavy-tailed distribution, the relation (1) governs how a high quantile, say $U(tx)$, can be extrapolated from an intermediate quantile $U(t)$. Clearly, estimating the extreme value index γ is a major step in estimating high quantiles.

In the heavy-tailed case, Hill [1975] proposes the following estimator of the parameter γ,

$$\hat{\gamma}_k := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}\;, \tag{2}$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ are the order statistics and k is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

To obtain the asymptotic normality of the Hill estimator (and most other estimators in extreme value statistics), it is necessary to quantify the speed of convergence in (1). We thus assume a second-order condition on the function U as follows. Suppose that there exist a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho \leq 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}\;,$$

for all $x > 0$. It is equivalent to

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^\rho - 1}{\rho} \; ; \tag{3}$$

see, for instance de Haan and Ferreira [2006, Proof of Theorem 3.2.5]. Compared to the second order condition, condition (1) is sometimes called the first order condition.

The estimator $\hat{\gamma}_k$ is consistent under the first order condition. Under the second order condition (3), the asymptotic normality can be established for i.i.d. observations as

$$\sqrt{k_\lambda}\left(\hat{\gamma}_{k_\lambda} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \; , \tag{4}$$

if the intermediate sequence k_λ satisfies $\lim_{n\to\infty} \sqrt{k_\lambda} A(n/k_\lambda) = \lambda$. This condition imposes an upper bound on the speed at which k_λ goes to infinity. The asymptotic bias for the Hill estimator is consequently given by the term $\frac{\lambda}{1-\rho}$.

To obtain an asymptotically unbiased estimator, we will first estimate the bias term and then subtract that from $\hat{\gamma}_k$. The asymptotically unbiased estimator is then given as $\hat{\gamma}_k - \widehat{\text{Bias}_k}$, where

$$\text{Bias}_k := \frac{A(n/k)}{1-\rho} \; . \tag{5}$$

A formal definition of the asymptotically unbiased estimator is given in equation (11) below.

One important consequence of the bias correction method is on the choice of the intermediate sequence k. In our asymptotically unbiased estimator, the limit of $\sqrt{k}A(n/k)$ does not need to be finite as that of k_λ, see below the condition (6). Thus it allows for larger value of k. In other words, we can reduce the level of estimation variance without suffering from the bias issue. Moreover, our asymptotically unbiased estimator allows a larger range of k. The "cutoff point" is thus solved thanks to the less sensitivity with respect to k. We shall demonstrate this showing that the estimates stay at a stable level for a much wider range of the k values in simulations.

2.2 Estimating the bias term

The estimation of the bias term requires estimating the second-order parameter ρ and the second-order scale function, $A(n/k)$, appearing in the condition (3). The parameter ρ controls the speed of convergence of most γ estimators. In the following we restrict the study to the case $\rho < 0$. In the literature of bias correction, in order to establish the asymptotic property of estimators of ρ, it is necessary to chose a higher intermediate sequence $k_\rho = k_\rho(n)$ such that $k_\rho \to \infty$, $k_\rho/n \to 0$ and

$$\sqrt{k_\rho}\, A(n/k_\rho) \to \infty \; , \tag{6}$$

as $n \to \infty$, see e.g. Gomes et al. [2002]. This provides a lower bound to the speed at which k_ρ goes to infinity. Furthermore, the asymptotic normality of all estimators of ρ is achieved under the assumption of i.i.d. observation and a third-order condition. The latter is given as follows. Suppose that there exist a positive or negative function B with $\lim_{t\to\infty} B(t) = 0$ and a real number $\rho' \leq 0$ such that

$$\lim_{t\to\infty} \frac{1}{B(t)} \left\{ \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} - \frac{x^\rho - 1}{\rho} \right\} = \frac{1}{\rho'} \left\{ \frac{x^{\rho+\rho'} - 1}{\rho + \rho'} - \frac{x^\rho - 1}{\rho} \right\} , \tag{7}$$

for all $x > 0$. If the observations are i.i.d., the asymptotic normality of all existing estimators of ρ, including that of the one we use in (10) below, holds under the condition (7) and with a sequence k_ρ such that as $n \to \infty$, $k_\rho \to \infty$, $k_\rho/n \to 0$ and

$$\sqrt{k_\rho} A(n/k_\rho) \to \infty, \sqrt{k_\rho} A^2(n/k_\rho) \to \lambda_1, \sqrt{k_\rho} A(n/k_\rho) B(n/k_\rho) \to \lambda_2 , \tag{8}$$

where λ_1 and λ_2 are both finite constants; see for instance Gomes et al. [2002] and Ciuperca and Mercadier [2010]. Here, since we are going to deal with $\beta-$mixing series, we need to re-establish the asymptotic property of the ρ estimator. The details are left to Appendix A.

In order to avoid extra bias stemming from the third order condition, the k sequence we use for the asymptotically unbiased estimator of the extreme value index is of a lower order, compared to the k_ρ sequence. More specifically, we use a sequence k_n such that as $n \to \infty$, $k_n \to \infty$, $k_n/k_\rho \to 0$ and

$$\sqrt{k_n} A(n/k_n) \to \infty, \sqrt{k_n} A^2(n/k_n) \to 0, \sqrt{k_n} A(n/k_n) B(n/k_n) \to 0 . \tag{9}$$

Comparing our asymptotically unbiased estimator with the original Hill estimator, the k sequences used for estimation are at different level. The conditions on k_n and k_λ imply that $k_n/k_\lambda \to +\infty$ as $n \to \infty$. Since the asymptotic variance of both the asymptotically unbiased estimator and the original Hill estimator is of an order $1/k$, using a sequence k_n increasing faster than k_λ leads to a lower asymptotic variance of our asymptotic unbiased estimator compared to that of the original Hill estimator.

In addition, the k sequence used for the asymptotically unbiased estimator is more flexible in the following sense. The condition that $\sqrt{k_\lambda} A(n/k_\lambda) \to \lambda$ restricts the level of k_λ as $k_\lambda = O\left(n^{\frac{2\rho}{2\rho-1}}\right)$, whereas condition (9) implies that $k_n = O\left(n^\tau\right)$ for any $\tau \in \left(\frac{2\rho}{2\rho-1}, \frac{2(\rho+\max(\rho,\rho'))}{2(\rho+\max(\rho,\rho'))+1}\right)$.

3 The serial dependence conditions

In this section, we present the serial dependence conditions on the time series we are going to deal with. The serial dependence structure follows from the so-called $\beta-$mixing conditions. The $\beta-$mixing conditions have been introduced by Rootzén [1995], Drees [2000, 2003] and Rootzén [2009] as follows. Let $\{X_1, X_2, \ldots\}$ be a stationary time series with common distribution function F. Let $\mathcal{B}_i^j$ denote the σ-algebra generated by $X_i, \ldots, X_j$. The sequence is said to be β-mixing or absolutely regular if

$$\beta(m) := \sup_{\ell \geq 1} \mathbb{E} \left(\sup_{E \in \mathcal{B}_{\ell+m+1}^{\infty}} \left| \mathbb{P}(E|\mathcal{B}_1^{\ell}) - \mathbb{P}(E) \right| \right) \to 0$$

as $m \to \infty$. The constants $\beta(m)$ are called the β-mixing constants of the sequence.

The asymptotic normality of the original Hill estimator has been established for β-mixing sequences in Drees [2000, 2003] with some mild extra conditions. With a sequence k_λ such that $\sqrt{k_\lambda} A(n/k_\lambda) \to \lambda$ as $n \to \infty$, it is proved that

$$\sqrt{k_\lambda} \left(\hat{\gamma}_{k_\lambda} - \gamma \right) \xrightarrow{d} N \left(\frac{\lambda}{1-\rho}, \sigma^2 \right) ,$$

where σ^2 is equal to γ^2 under independence but is more complicated otherwise. The extra conditions for establishing the asymptotic normality of the Hill estimator are the following list of *regulatory conditions*. Suppose there exist a constant $\varepsilon > 0$ and a sequence $\ell = \ell_n$ such that as $n \to \infty$,

(a) $\frac{\beta(\ell)}{\ell} n + \ell k^{-1/2} \log^2 k \to 0$,

(b) $\frac{n}{\ell k} \mathrm{Cov} \left(\sum_{i=1}^{\ell} \mathbf{1}_{\{X_i > F^{-1}(1-kx/n)\}}, \sum_{i=1}^{\ell} \mathbf{1}_{\{X_i > F^{-1}(1-ky/n)\}} \right) \to r(x,y)$, for any $0 \leq x, y \leq 1+\varepsilon$,

(c) for some constant C,

$$\frac{n}{\ell k} \mathbb{E} \left[\left(\sum_{i=1}^{\ell} \mathbf{1}_{\{F^{-1}(1-ky/n) < X_i \leq F^{-1}(1-kx/n)\}} \right)^4 \right] \leq C(y-x) ,$$

for any $0 \leq x < y \leq 1+\varepsilon$ and $n \in \mathbb{N}$.

The regulatory conditions (a)-(c) can be substituted by simpler sufficient versions, see Proposition 2.1 in Drees [2002]. In Section 5 below, we provide examples of time series models that satisfy these assumptions.

We intend to correct the bias while allowing the observations to follow the $\beta-$mixing condition and the regulatory conditions. Since the asymptotic bias of the original Hill estimator under serial dependence has the same form as in (5), we can construct an asymptotic unbiased estimator for β-mixing sequences with exactly the same form as in the independence case. Nevertheless, due to the serial dependence, the asymptotic property of the estimator has to be reestablished. This is what we do in the next section.

4 Main Results

We start by introducing the estimator of the second-order parameter. Then we state our main results on the asymptotic properties of the asymptotic unbiased estimators of the extreme value index and high quantiles.

4.1 Estimating the second-order parameter

Similar to Gomes et al. [2002], we introduce the following notations. For any positive number α, denote

$$M_k^{(\alpha)} := \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right)^{\alpha} \ ,$$

$$R_k^{(\alpha)} := \frac{M_k^{(\alpha)} - \Gamma(\alpha+1)\left(M_k^{(1)}\right)^{\alpha}}{M_k^{(2)} - 2\left(M_k^{(1)}\right)^{2}} \ ,$$

$$S_k^{(\alpha)} := \frac{\alpha(\alpha+1)^2\Gamma^2(\alpha)}{4\Gamma(2\alpha)}\frac{R_k^{(2\alpha)}}{\left(R_k^{(\alpha+1)}\right)^{2}} \ ,$$

$$s^{(\alpha)}(\rho) := \frac{\rho^2\left(1-(1-\rho)^{2\alpha} - 2\alpha\rho(1-\rho)^{2\alpha-1}\right)}{\left(1-(1-\rho)^{\alpha+1} - (\alpha+1)\rho(1-\rho)^{\alpha}\right)^2} \ .$$

Then the estimator of the second-order parameter ρ is defined as

$$\hat{\rho}_k^{(\alpha)} := \left(s^{(\alpha)}\right)^{\leftarrow}(S_k^{(\alpha)}) \ . \tag{10}$$

4.2 Asymptotically unbiased estimator of the extreme value index

We now write explicitly the asymptotically unbiased estimator of the extreme value index. Let k_n and k_ρ satisfying (9) and (8) be the number of observations selected for estimating γ and ρ respectively. For some positive real number α, we define the asymptotically unbiased estimator as

$$\hat{\gamma}_{k_n,k_\rho,\alpha} := \hat{\gamma}_{k_n} - \frac{M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}} \ , \tag{11}$$

where $\hat{\gamma}_{k_n}$ denotes the original Hill estimator as in (2). We remark that it is asymptotically equivalent to its positive modification

$$\tilde{\gamma}_{k_n,k_\rho,\alpha} := \hat{\gamma}_{k_n}\exp\left(-\frac{M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2}{2\hat{\gamma}_{k_n}^2\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}}\right) \ ,$$

which might be preferred in applications since we are dealing with positive γ.

The following theorem shows the asymptotic normality of our asymptotically unbiased estimator for $\beta-$mixing series. The consistency of the estimator could be obtained under the second-order condition without requiring the third-order condition.

Theorem 4.1. *Suppose that $\{X_1, X_2, \ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third-order condition (7) with parameters $\rho < 0$ and $\rho' < 0$. Suppose the two intermediate sequences k_ρ and k_n satisfy* *the conditions in (8) and (9) respectively. Suppose that the regulatory conditions hold with the intermediate sequence k_n. Then,*

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma\right)\xrightarrow{d} N\left(0,\sigma^2\right) ,$$

where

$$\sigma^2 := \frac{\gamma^2}{\rho^2}\left((2-\rho)^2 c_{1,1}+(1-\rho)^2 c_{2,2}+2(2-\rho)(\rho-1)c_{1,2}\right) ,$$

with

$$c_{i,j} := \iint_{[0,1]^2}(-\log s)^{i-1}(-\log t)^{j-1}\left\{\frac{r(s,t)}{st}-\frac{r(s,1)}{s}-\frac{r(1,t)}{t}+r(1,1)\right\} ds\, dt ,$$

and $r(s,t)$ defined in the regulatory condition (b).

We remark that our estimator is also valid as an asymptotically unbiased estimator of the extreme value index when the observations are i.i.d.. In that case, the result is simplified to

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma\right)\xrightarrow{d} N\left(0,\frac{\gamma^2}{\rho^2}\left\{\rho^2+(1-\rho)^2\right\}\right) .$$

4.3 Asymptotically unbiased estimator of high quantiles

We consider the estimation of high quantiles. High quantile refers to the quantile at a probability level $(1-p)$, where the tail probability $p=p_n$ depends on the sample size n: as $n\to\infty$, $p_n=O(1/n)$. The goal is to estimate the quantile $x(p)=U(1/p)$. In extreme case such that $np_n<1$, it is not possible to have non-parametric estimate of such a quantile.

We derive an asymptotically unbiased estimator of such high quantile from a variant of the high quantile estimator in Weissman [1978] as follows:

$$\hat{x}_{k_n,k_\rho,\alpha}(p) := X_{n-k_n,n}\left(\frac{k_n}{np}\right)^{\hat{\gamma}_{k_n,k_\rho,\alpha}} .$$

The following theorem gives its asymptotic normality.

Theorem 4.2. *Suppose that $\{X_1,X_2,\ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third-order condition (7) with parameters $\rho<0$ and $\rho'<0$. Suppose the two intermediate sequences k_ρ and k_n satisfy the conditions in (8) and (9) respectively. Assume in addition that $n\to\infty$, $np_n/k_n\to 0$ and $\log(np_n)/\sqrt{k_n}\to 0$. Suppose that the regulatory conditions hold with k_n. Then*

$$\frac{\sqrt{k_n}}{\log(k_n/(np_n))}\left(\frac{\hat{x}_{k_n,k_\rho,\alpha}(p_n)}{x(p_n)}-1\right)\xrightarrow{d} N\left(0,\sigma^2\right) ,$$

with σ^2 as defined in Theorem 4.1.

5 Examples

In our framework, we model the serial dependence by the $\beta-$mixing condition and the extra regulatory conditions. In this section, we give a few examples that satisfy those conditions. The fact that these examples satisfy the conditions are documented in the referred literature as follows.

- Rootzén [1995], Drees [2003], Rootzén [2009]: the k-dependent process and the autoregressive (AR) process: AR(1);
- Resnick and Stărică [1997], Drees [2002]: the AR(p) processes and the infinite moving averages (MA) processes;
- Hsing [1991], Rootzén [1995], Drees [2002], Rootzén [2009]: the finite MA processes;
- Drees [2002, 2003]: the autoregressive conditional heteroskedasticity process: ARCH(1);
- Stărică [1999], Drees [2000]: the generalized autoregressive conditional heteroskedasticity (GARCH) processes.

We review some simple cases of these processes and provide the comparison of the asymptotic variances under dependence to that under independence, and to that of the original Hill estimator under serial dependence.

5.1 Autoregressive model

Consider first the stationary solution of the following AR(1) equation

$$X_i = \theta X_{i-1} + Z_i \ , \tag{12}$$

for some $\theta \in (0,1)$ and i.i.d. random variables Z_i. The distribution function of the innovation is denoted by F_Z. Assume that F_Z admits a positive Lebesgue density that satisfies the Lipschitz condition of order 1 (Billingsley [1979], pp. 418). Suppose that as $x \to \infty$,

$$1 - F_Z(x) \sim p x^{-1/\gamma} \ell(x) \, \text{and} \, F_Z(-x) \sim q x^{-1/\gamma} \ell(x) \tag{13}$$

for some slowly varying function ℓ and $p = 1 - q \in (0,1)$. Then from Section 3.2 of Drees [2002] we get that $1 - F(x) \sim d_\theta \, (1 - F_Z(x))$ as $x \to \infty$, where $d_\theta = p(1 - \theta^{1/\gamma})^{-1}$. Furthermore, the regulatory conditions hold with

$$r(x, y) = x \wedge y + \sum_{m=1}^{\infty} \{c_m(x, y) + c_m(y, x)\},$$

where $c_m(x, y) = p^{-1}(x \wedge y\theta^{m/\gamma})$.

Let us denote by $\sigma^2(\theta, p)$ the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$. First, we compare the asymptotic variance under model (12) with that under independence by calculating the ratio $\sigma^2(\theta, p)/\sigma^2(0, p)$. Second, we compare $\sigma^2(\theta, p)$ with the asymptotic variance of the original Hill estimator under serial dependence, σ_H^2, when using the same k sequence. From Drees [2000], we get that under serial dependence $\sqrt{k}(\hat{\gamma}_H - \gamma)$ converges to a normal distribution with asymptotic variance $\sigma_H^2 = \gamma^2 r(1, 1)$. The two ratios are given as follows. Set $\kappa = \theta^{1/\gamma}$. Then,

$$\frac{\sigma^2(\theta,p)}{\sigma^2(0,p)} = 1 + \frac{2\kappa}{p(1-\kappa)} + \frac{2\rho(1-\rho)}{1-2\rho(1-\rho)} \frac{\kappa \log \kappa}{p(1-\kappa)^2} \, ,$$

$$\frac{\sigma^2(\theta,p)}{\sigma_H^2} = \frac{1}{\rho^2} \left(1 - 2\rho(1-\rho) + 2\rho(1-\rho) \frac{\kappa \log \kappa}{p(1-\kappa)^2 + 2\kappa(1-\kappa)} \right) \, .$$

In the first row of Figure 1, we plot these ratios against κ for different values of the parameters p and ρ. From Figure 1a, we note that the variation of the first ratio is mainly due to that of κ and p. The parameter ρ plays a relative minor role here. We further give a numerical illustration with $\gamma = 1$ and $\rho = -1$. With i.i.d. observations, the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$ is 5. Instead, if the observations follow the AR(1) model with $\theta = 0.4$ and $p = 0.5$, then the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$ is close to 26. Hence, overlooking the serial dependence may severely underestimate the range of confidence intervals.

Differently, we observe from Figure 1b that the variation of the second ratio is mainly due to that of ρ, but not to the variation of other parameters. It is important to remark that although this ratio is greater than one, the asymptotically unbiased estimator does not necessarily have a higher asymptotic variance. The current comparison is conducted with assuming the same k value, while the k value used in our asymptotically unbiased estimator can be at a much higher level than that used for the Hill estimator. Theoretically the conditions on k_n and k_λ guarantees that $k_n/k_\lambda \to +\infty$. Thus the asymptotic variance of the asymptotically unbiased estimator is at a lower level. Practically, for $\rho = -1$, the ratio is in between 6 and 8. Thus, once we use a k_n that is seven times higher than the k_λ used for the original Hill estimator, we will get an estimator with lower asymptotic variance. Together with the fact that the asymptotic bias is zero, we end up with a better performance in terms of lower asymptotic mean squared error. Such a feature will show up in the simulation studies in Section 6 below.

5.2 Moving average model

Consider now the stationary solution of the MA(1) equation

$$X_i = \theta Z_{i-1} + Z_i \, , \tag{14}$$

where the innovation Z satisfies the same conditions as in the AR(1) model in the previous subsection. We calculate the two ratios when comparing the asymptotic variance of the asymptotically unbiased estimator under serial dependence to that under independence, and that of the original

Hill estimator under dependence as follows. Set again $\kappa = \theta^{1/\gamma}$. Then,

$$\begin{aligned}
\frac{\sigma^2(\theta,p)}{\sigma^2(0,p)} &= 1 + \frac{2\kappa}{p(1+\kappa)} + \frac{2\rho(1-\rho)}{1-2\rho(1-\rho)}\frac{\kappa\log\kappa}{p(1+\kappa)}\ , \\
\frac{\sigma^2(\theta,p)}{\sigma_H^2} &= \frac{1}{\rho^2}\left(1-2\rho(1-\rho)+2\rho(1-\rho)\frac{\kappa\log\kappa}{p(1+\kappa)+2\kappa}\right)\ .
\end{aligned}$$

In the second row of Figure 1, we plot the variations of these ratios with respect to κ for different values of ρ and p. The general feature is comparable to that observed from the first row. A notable difference between Figures 1a and 1c is that although the ratios are both increasing in κ and p, their convexities with respect to κ are different in the two models: we observe a concave (resp. convex) relation in κ under the MA(1) (resp. AR(1)) model.

5.3 Generalized autoregressive conditional heteroskedasticity model

Consider the stationary solution to the following recursive system of equations

$$\begin{cases} X_t &= \varepsilon_t\sigma_t, \\ \sigma_t^2 &= \lambda_0 + \lambda_1 X_{t-1}^2 + \lambda_2\sigma_{t-1}^2, \end{cases}$$

where ε_t are i.i.d. innovations with zero mean and unit variance. The stationary solution of this GARCH(1,1) model, X_t, follows a heavy-tailed distribution, even if the innovations ε_t are normally distributed, see Kesten [1973] and Goldie [1991]. The extreme value index of the GARCH(1,1) model can be derived from the Kesten theorm on stochastic difference equations, see Kesten [1973]. Nevertheless, the calculation is not explicit.

In addition, the stationary GARCH(1,1) series satisfies the $\beta-$mixing condition and the regulatory conditions, see Stărică [1999] and Drees [2000]. Thus, it can be considered as an example for which we can apply the asymptotically unbiased estimators. Since it is difficult to explicitly calculate the r function and consequently the asymptotic variance, we opt to use simulations to show the performance of the asymptotically unbiased estimator under the GARCH model.

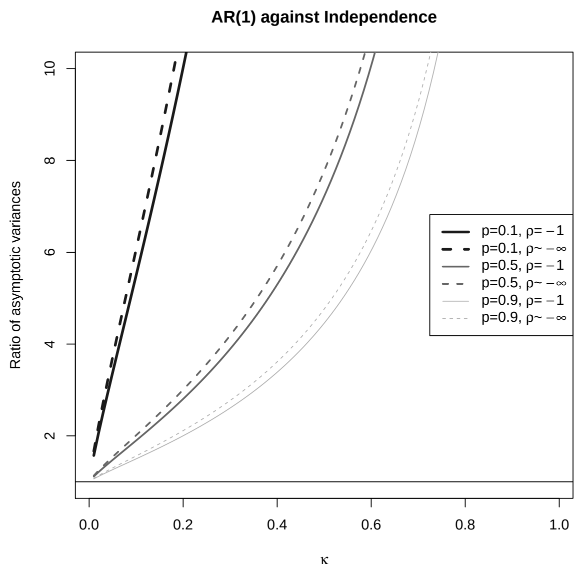
6 Simulations

The simulations are set up as follows. We consider four models on the serial dependence of the simulated samples. Suppose Z follows the distribution F_Z given by

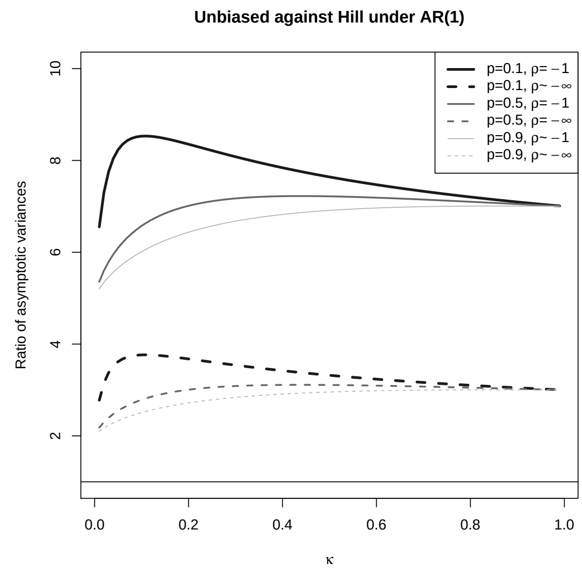
$$F_Z(x) = \begin{cases} (1-p)(1-\tilde{F}(-x)) & \text{if } x < 0, \\ 1-p+p\tilde{F}(x) & \text{if } x > 0, \end{cases}$$

where $\tilde{F}$ is the standard Fréchet distribution function: $\tilde{F}(x) = \exp(-1/x)$ for $x > 0$, and $p = 0.75$. Then F_Z belongs to the domain of attraction with extreme value index 1. We construct three time

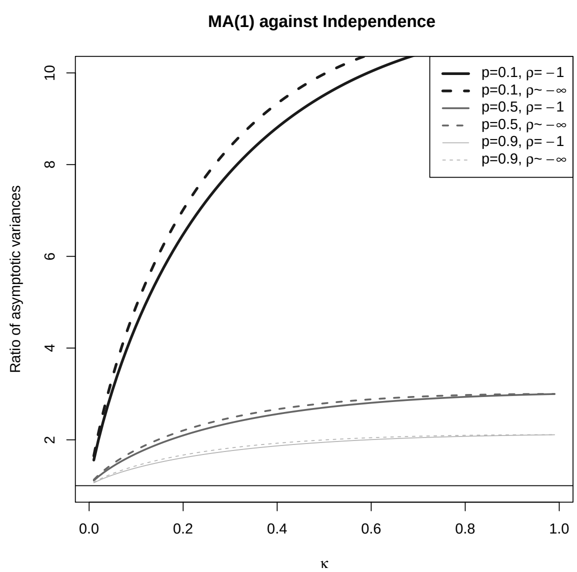
Figure 1: Ratios between asymptotic variances



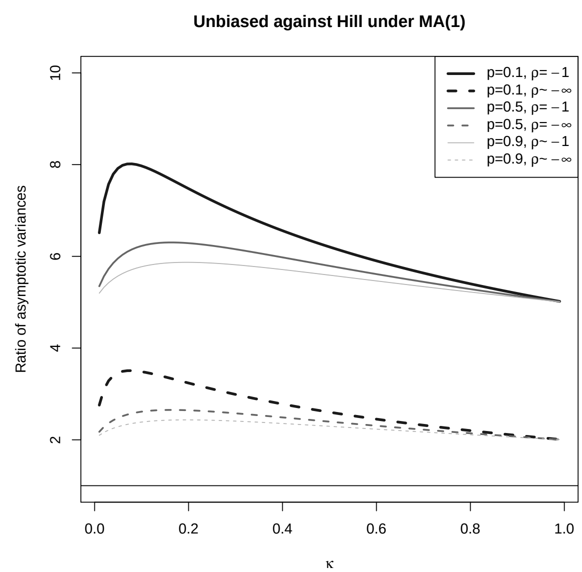
(a) Comparison with the i.i.d. case: AR(1)



(b) Comparison with the Hill estimator under AR(1)



(c) Comparison with the i.i.d. case: MA(1)



(d) Comparison with the Hill estimator under MA(1)

Note: Figure 1a shows the ratio between the asymptotic variance of the asymptotically unbiased estimator under the AR(1) model and that under the i.i.d. case. Figure 1b shows the ratio between the asymptotic variance of the asymptotically unbiased estimator and that of the original Hill estimator, under the AR(1) model. Figure 1c and 1d show the corresponding ratios when the serial dependence is modeled by the MA(1) model.

series models based on i.i.d. observations Z_t as follows:

(Model 1) Independence: $X_t = Z_t$ (can be regarded as MA(1) with $\theta = 0$),

(Model 2) AR(1): X_t given by (12) with $\theta = 0.3$,

(Model 3) MA(1): X_t given by (14) with $\theta = 0.3$.

In all three models, the theoretical value of γ is 1. In addition, we construct a GARCH(1,1) model as in Subsection 5.3. We remark that although the heavy-tailed feature of the GARCH(1,1) model does not depend on that of the innovations, there are empirical evidence supporting using heavy-tailed innovations for financial time series, see, e.g. McNeil and Frey [2000] and Sun and Zhou [2013]. To be consistent with the empirical literature, we use the Student-t distribution with degree of freedom 6 as noise distribution.[2] Other parameters in the simulated GARCH(1,1) model are also chosen as close to what has been estimated from financial time series, see Sun and Zhou [2013].

(Model 4) GARCH(1,1): X_t given as in Subsection 5.3 with $\lambda_0 = 0.5$, $\lambda_1 = 0.07$, $\lambda_2 = 0.91$ and with initial value $X_0 = 0$ and $\sigma_0^2 = 25$.

Following Kesten [1973] we calcluate the extreme value index of the series in Model 4 at 0.214.

Next, for each model, we simulate $N = 500$ samples with sample size $n = 1000$ each. We estimate $\hat{\gamma}_k$ for each value of $k = 10, 11, \ldots, 600$ using the following methods:

(Method 1) the Hill estimator (2);

(Method 2) the asymptotically unbiased estimator (11) with $k_\rho = \sup\{k \le n^{0.98}, \hat{\rho}_k \text{ exists}^3\}$.

Then we calculate the average absolute bias (ABias) and the mean square error (MSE) by

$$\text{ABias}_k = \frac{1}{N}\sum_{j=1}^{N} |\hat{\gamma}_k^j - \gamma| \, ,$$

and

$$\text{MSE}_k = \frac{1}{N}\sum_{j=1}^{N} (\hat{\gamma}_k^j - \gamma)^2 \, .$$

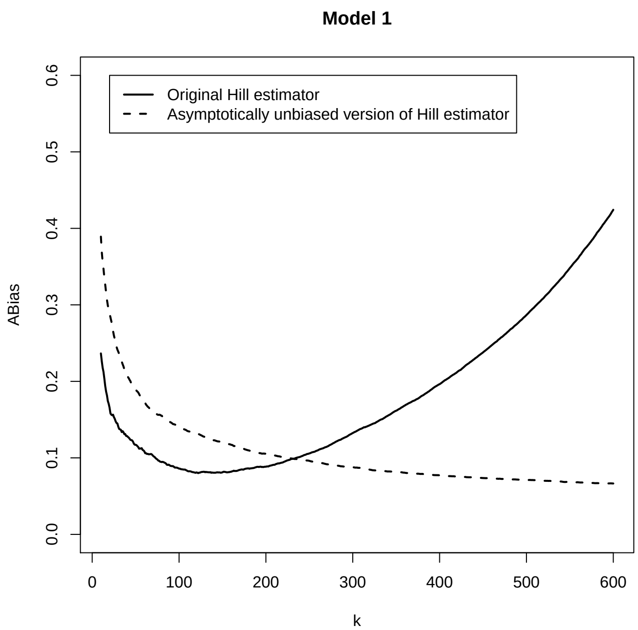
The ABias and MSE for the two estimation methods are given in Figure 2– 5 with each figure corresponding to one model. We observe that even with a rather high level of k, our asymptotically unbiased estimator does not suffer from a large bias, at least for the first three models. In Model 4, the bias term increases with respect to k, but still stays at a much lower level compared to that of the original Hill estimator. As a consequence, the MSE of our asymptotically unbiased estimator

[2] In order to get a unit variance, we simply divide the standard Student-t distribution by $\sqrt{3/2}$.

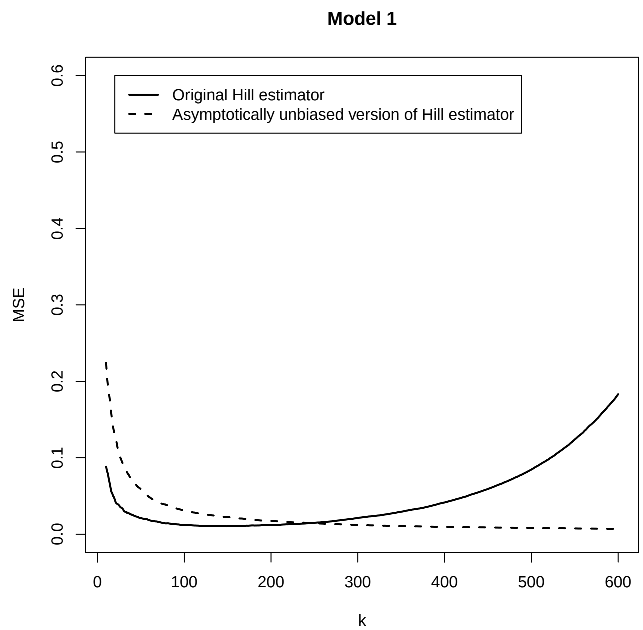
[3] The choice of $n^{0.98}$ follows from the suggestion in Cai et al. [2013]. In addition, we estimate the ρ parameter by (10) with $\alpha = 2$. Thus, the estimate exists if and only if the value $S_k^{(2)}$ lies in the interval $[2/3, 3/4]$.

decreases with respect to k in the first three models, and remains at a lower level than that of the original Hill estimator. Thus, our asymptotically unbiased estimator provides not only a relatively more accurate estimate, but also allows for a flexible choice of k, even if serial dependence is present. This helps to clear the two major hurdles for applying extreme value statistics to financial time series as we intend to.

Figure 2: Absolute bias and MSE for Model 1.



(a) ABias under Model 1.



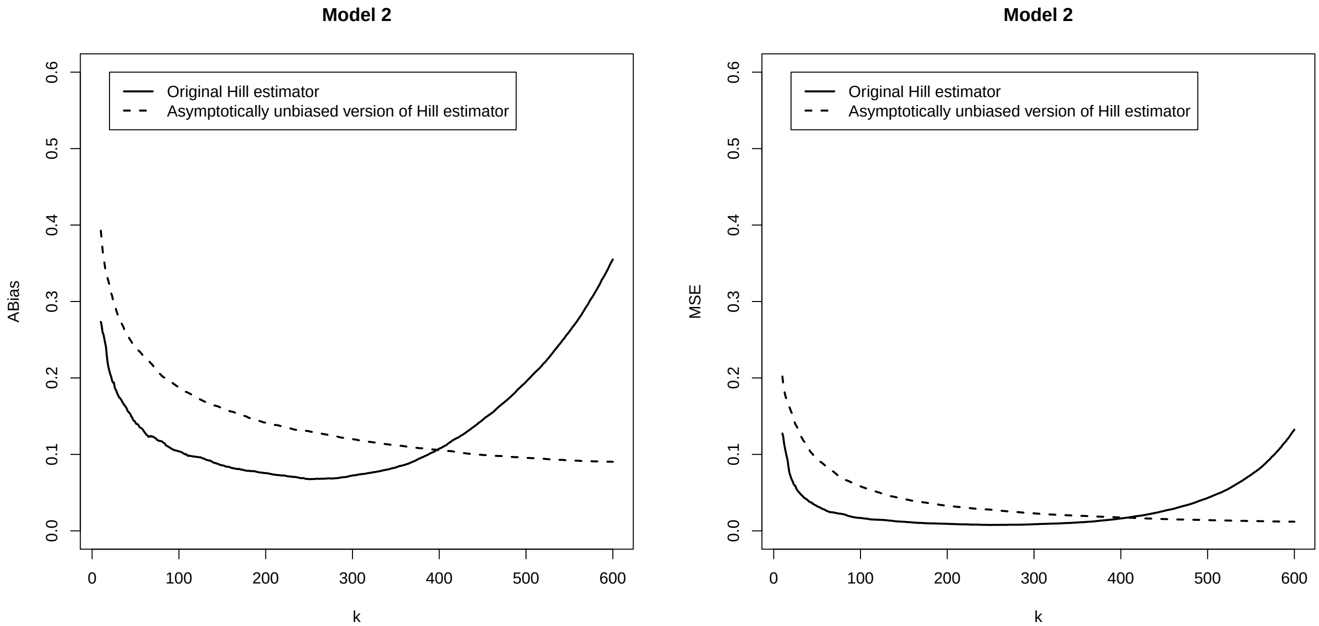
(b) MSE under Model 1.

7 Application

We apply the asymptotically unbiased estimators on the extreme value index and high quantiles to evaluate the downside tail risk in the Dow Jones Industrial Average index. We collect the daily index from 1980 to 2010 and compute the daily loss returns. The indices and loss returns are presented in Figure 6a and 6b. From the figures, we observe that although the loss return series can be regarded as stationary, there is evidence of serial dependence such as volatility clustering. Therefore, one should not treat the series as i.i.d. observations. The serial dependence has to be accounted for when performing extreme value analysis.

Our goal is to estimate the Value-at-Risk of the return series at 99.9% level, which corresponds to a high quantile with tail probability 0.1%. From 8088 daily observations, a non-parametric estimate can be obtained by taking the eighth highest order statistic. We thus get 7.16% as the empirical estimate. Next, we apply both the original Hill estimator and the asymptotically unbiased estimator

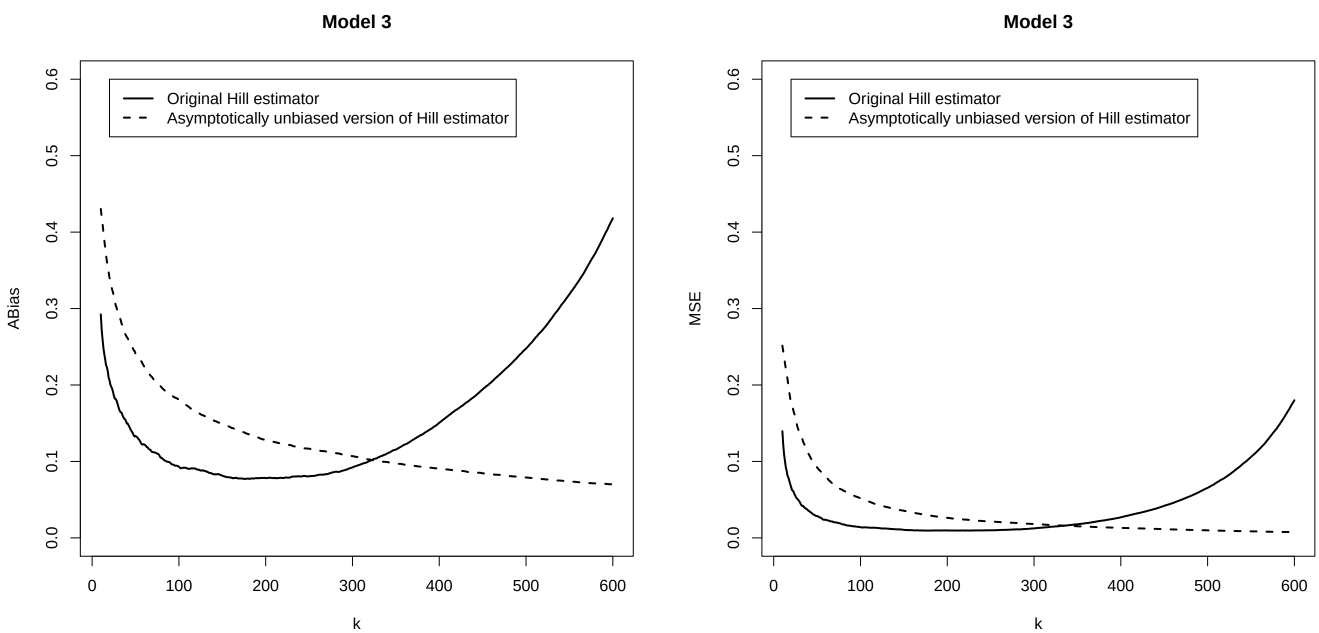
Figure 3: Absolute bias and MSE for Model 2.



(a) ABias under Model 2.

(b) MSE under Model 2.

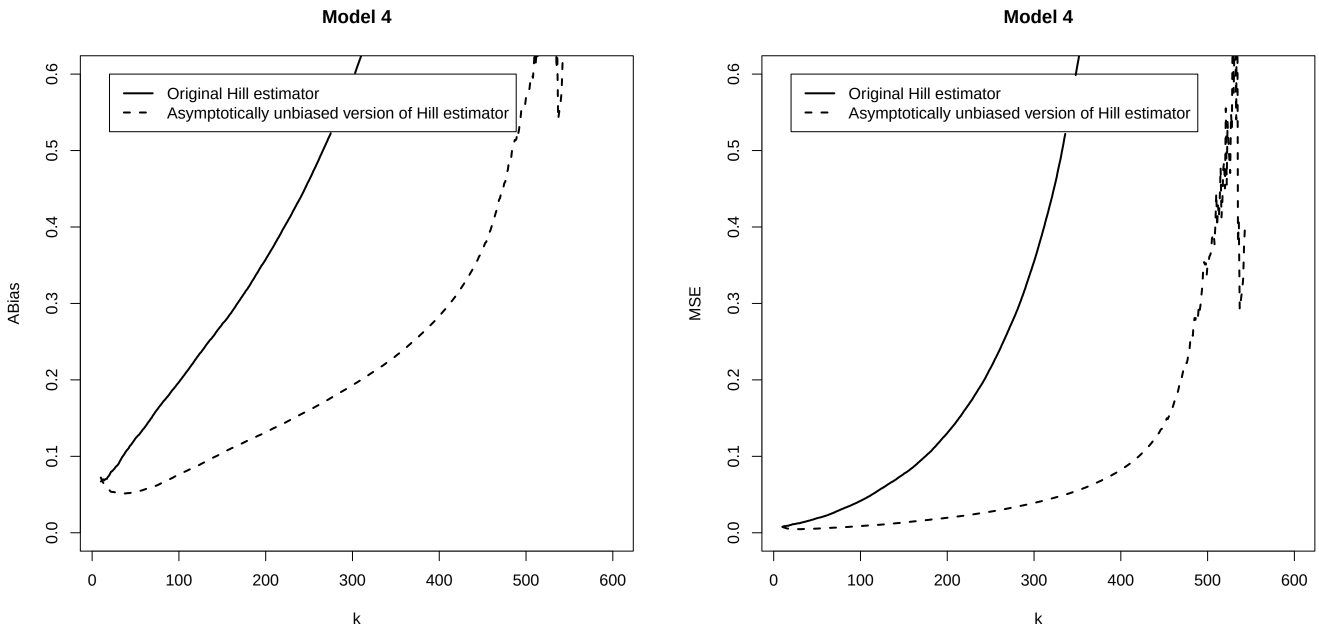
Figure 4: Absolute bias and MSE for Model 3.



(a) ABias under Model 3.

(b) MSE under Model 3.

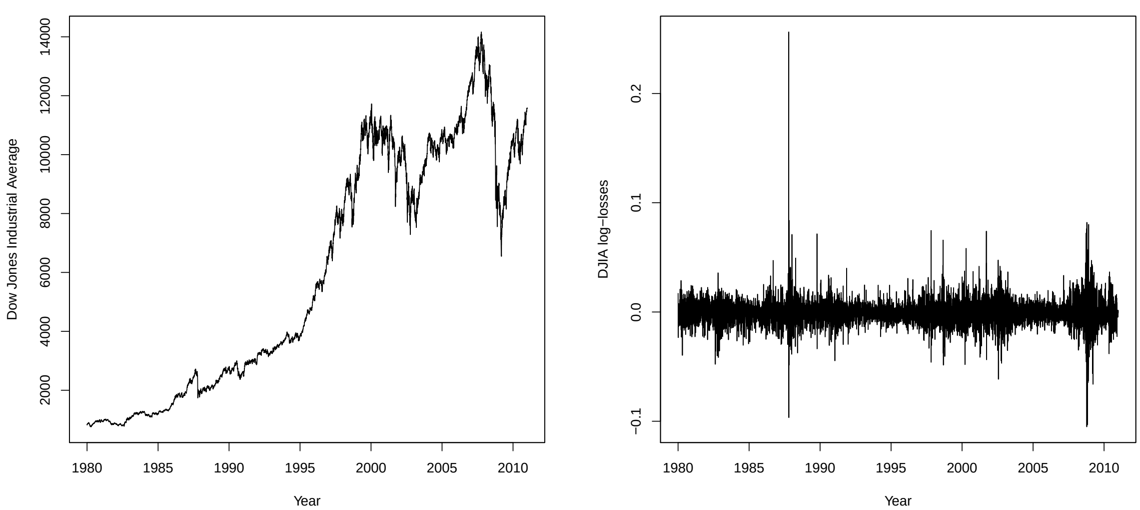
Figure 5: Absolute bias and MSE for Model 4.



(a) ABias under Model 4.

(b) MSE under Model 4.

Figure 6: Historical representation of the real data set.



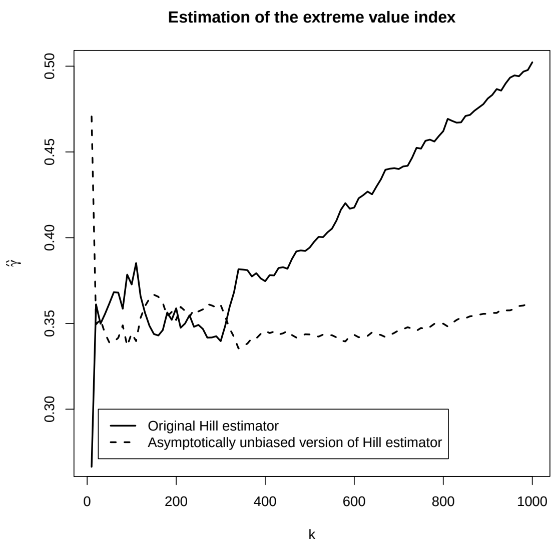
(a) Dow Jones Industrial Average (DJIA).

(b) Loss returns of DJIA.

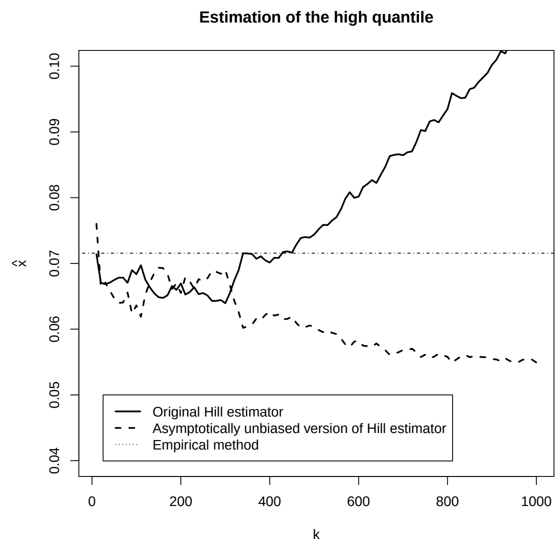
Note: The figures show the daily prices and loss returns of the DJIA index from 1980 to 2010.

to estimate the extreme value index of the loss return series. Then, we use these estimates as input in the estimation of the VaR. We plot the estimates of the extreme value index against different choices of k in Figure 7a, and those of the Value-at-Risk in Figure 7b.

Figure 7: Application to the DJIA index.



(a) Estimating the extreme value index



(b) Estimating the VaR at 99.9% level

Note: The figures present the estimates of the extreme value index and the VaR at 99.9% level for the loss returns of the DJIA index with varying choice of k. The returns are calculated from daily prices in the period from 1980 and 2010. The estimation uses the asymptotically unbiased estimators of the extreme value index and high quantile provided in Section 4.1 and 4.2, respectively.

From the two figures, we observe that the estimates using the bias correction technique stays stable for a larger range of k values. In contrast, the estimates based on the original Hill estimator suffers from a large bias starting from $k \geq 400$. Thus, when applying the original EVT estimators, it is possible to choose k only around 250, which corresponds to 3% of the total sample. With our asymptotically unbiased estimators, we can take $k = 1000$ and obtain an estimated extreme value index at 0.361 with an estimated VaR at 5.49%. Note that our result is lower than the empirical estimate.

A – Appendix – Proofs

The asymptotically unbiased estimator of the extreme value index is constructed based on the moments

$$M_k^{(\alpha)} := \frac{1}{k} \sum_{i=1}^{k} (\log X_{n-i+1,n} - \log X_{n-k,n})^{\alpha} \ ,$$

defined in Subsection 4.1. One can write these statistics as functionals of the tail quantile process $\left\{Q_n(t) := X_{n-[kt],n}\right\}_{t\in[0,1]}$ as follows

$$M_k^{(\alpha)} = \int_0^1 \left(\log \frac{Q_n(t)}{Q_n(1)}\right)^\alpha dt \, .$$

Therefore, to derive the asymptotic property of the asymptotically unbiased estimator, we first establish those of the tail quantile process and the moments. We first show that the tail quantile process can be approximated by a Gaussian process as in the following proposition.

Proposition A.1. *Suppose that $\{X_1, X_2, \ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third-order condition (7) with parameters $\rho < 0$ and $\rho' < 0$. Suppose an intermediate sequence k satisfies that as $n \to \infty$, $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k)B(n/k) = O(1)$. In addition, assume that the regulatory conditions holds. Then, for a given $\varepsilon > 0$, under a Skorohod construction, there exists two functions $\tilde{A} \sim A$ and $\tilde{B} = O(B)$, where A and B are the second and third order scale functions in (7), and a centered Gaussian process $\{e(t)\}_{t\in[0,1]}$ with covariance function r defined as in the regulatory condition (b), such that, as $n \to \infty$,*

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon} \left| \sqrt{k}\left(\log\left(\frac{Q_n(t)}{U(n/k)}\right) + \gamma \log(t)\right) - \gamma t^{-1} e(t) \right.$$
$$\left. -\sqrt{k}\tilde{A}(n/k)\frac{t^{-\rho}-1}{\rho} - \sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\frac{t^{-\rho-\rho'}}{\rho+\rho'} \right| \to 0 \text{ a.s.}$$

Proof of Proposition A.1. By writing $X_i = U(Y_i)$ where each Y_i follows a standard Pareto distribution, we obtain that $\{Y_1, Y_2, \cdots\}$ is a stationary $\beta-$mixing series satisfying the regulatory conditions. This is a direct consequence of $Y_i = 1/(1 - F(X_i))$. We write $Q_n(t) = X_{n-[kt],n} = U(Y_{n-[kt],n})$ and focus first on the asymptotic property of the process $\left\{Y_{n-[kt],n}\right\}_{t\in[0,1]}$. By verifying the conditions in Drees [2003, Theorem 2.1], we get that under a Skorohod construction, there exists a centered Gaussian process $\{e(t)\}_{t\in[0,1]}$ with covariance function r defined in the regulatory condition (b), such that for $\varepsilon > 0$, as $n \to \infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon} \left| \sqrt{k}\left(t\frac{Y_{n-[kt],n}}{n/k} - 1\right) - t^{-1}e(t) \right| \to 0 \text{ a.s..} \tag{15}$$

Next, we present an inequality on the U function based on the third-order condition (7). Under the third-order condition, there exists two functions $\tilde{A} \sim A$ and $\tilde{B} = O(B)$, such that for any $\delta > 0$, there exists some positive number $t_0 = t_0(\varepsilon)$ such that for all $t \geq t_0$ and $tx \geq t_0$,

$$\left| \frac{\frac{\log U(tx) - \log U(t) - \gamma \log x}{\tilde{A}(t)} - \frac{x^\rho - 1}{\rho}}{\tilde{B}(t)} - \frac{x^{\rho+\rho'}}{\rho+\rho'} \right| \leq \delta x^{\rho+\rho'} \max(x^\delta, x^{-\delta}) \, . \tag{16}$$

This inequality is a direct consequence of applying de Haan and Ferreira [2006, Theorem B.3.10] to the function $f(t) = \log U(t) - \gamma \log t$.

We combine the asymptotic property of $\left\{Y_{n-[kt],n}\right\}_{t\in[0,1]}$ in (15) with the inequality (16) as follows. Taking $t = n/k$ and $tx = Y_{n-[kt],n}$ in (16), we get that given any $0 < \delta < -\rho - \rho'$, for sufficiently large $n > n_0(\delta)$, with probability 1,

$$\left|\log Q_n(t) - \log U(n/k) - \gamma \log\left(\frac{k}{n}Y_{n-[kt],n}\right) - \tilde{A}(n/k)\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho}-1}{\rho}\right.$$
$$\left. -\tilde{A}(n/k)\tilde{B}(n/k)\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'}}{\rho+\rho'}\right| \leq \delta\tilde{A}(n/k)\tilde{B}(n/k)\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'+\delta} . \tag{17}$$

By applying (15), we bound the four terms in (17) that contain $\frac{k_n}{n}Y_{n-[k_nt],n}$ as

$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\log\left(\frac{k}{n}Y_{n-[kt],n}\right)+\log t\right)-t^{-1}e(t)\right| \to 0 \text{ a.s.},$$
$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho}-1}{\rho}-\frac{t^{-\rho}-1}{\rho}\right)-t^{-\rho-1}e(t)\right| = o_{\mathbb{P}}(t^{-\rho}) \to 0 \text{ a.s.},$$
$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'}-t^{-\rho-\rho'}\right)-(\rho+\rho')\left(t^{-\rho-\rho'-1}e(t)\right)\right| = o_{\mathbb{P}}\left(t^{-\rho-\rho'}\right) \to 0 \text{ a.s.},$$
$$t^{1/2+\varepsilon}\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'+\delta} = O_{\mathbb{P}}(t^{1/2-\rho-\rho'+\varepsilon-\delta}) = O_{\mathbb{P}}(1) .$$

When taking $n \to \infty$, with the facts that $\sup_{t\in(0,1]} t^{1/2+\varepsilon}t^{-1}|e(t)| = O_{\mathbb{P}}(1)$, $\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k) = O(1)$ and $\tilde{A}(n/k), \tilde{B}(n/k) \to 0$, the proposition is proved due to the free choice of δ. □

By applying Proposition A.1, we get the asymptotic property of the moments $M_k^{(\alpha)}$ as follows.

Corollary A.2. *Assume that the conditions in Proposition A.1 hold. Then, as* $n \to \infty$

$$\sqrt{k}\left(M_k^{(\alpha)} - \gamma^{\alpha}\Gamma(\alpha+1)\right) - \alpha\gamma^{\alpha}P_1^{(\alpha)} - \sqrt{k}\tilde{A}(n/k)\gamma^{\alpha-1}\frac{\Gamma(\alpha+1)}{\rho}\left(\frac{1}{(1-\rho)^{\alpha}}-1\right)$$
$$-\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\gamma^{\alpha-1}\frac{\Gamma(\alpha+1)}{\rho+\rho'}\left(\frac{1}{(1-\rho-\rho')^{\alpha}}-1\right)$$
$$-\sqrt{k}\tilde{A}(n/k)^2\gamma^{\alpha-2}\frac{\Gamma(\alpha+1)}{2\rho^2}\left(\frac{1}{(1-2\rho)^{\alpha}}-\frac{2}{(1-\rho)^{\alpha}}+1\right) \to 0 \ a.s.,$$

where $P_1^{(\alpha)}$ *are normally distributed random variables with mean zero. In addition*

$$\mathrm{Cov}(P_1^{(\alpha)}, P_1^{(\tilde{\alpha})}) = \iint_{[0,1]^2}(-\log s)^{\alpha-1}(-\log t)^{\tilde{\alpha}-1}\left\{\frac{r(s,t)}{st}-\frac{r(s,1)}{s}-\frac{r(1,t)}{t}+r(1,1)\right\} ds\, dt ,$$

with the covariance function r *defined as in regulatory condition (b).*

Proof of Corollary A.2. Recall that

$$M_k^{(\alpha)} = \int_0^1 \left(\log \frac{Q_n(t)}{U(n/k)} - \log \frac{Q_n(1)}{U(n/k)}\right)^\alpha dt\ .$$

From Proposition A.1, we get that as $n \to \infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon} \left| \sqrt{k}\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right) + \gamma(-\log t)\right) - \gamma(t^{-1}e(t) - e(1)) - \sqrt{k}\tilde{A}(n/k)\frac{t^{-\rho}-1}{\rho} \right.$$
$$\left. -\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\frac{t^{-\rho-\rho'}-1}{\rho+\rho'}\right| \to 0 \text{ a.s.}\ .$$

The second order expansion $(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2}x^2 + o(x^2)$ yields that, as $n \to \infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon} \left| \sqrt{k}\left(\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right)\right)^\alpha + \gamma^\alpha(-\log t)^\alpha\right) - \alpha\gamma^\alpha(-\log t)^{\alpha-1}(t^{-1}e(t) - e(1)) \right.$$
$$- \sqrt{k}\tilde{A}(n/k)\alpha\gamma^{\alpha-1}(-\log t)^{\alpha-1}\frac{t^{-\rho}-1}{\rho} - \sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\alpha\gamma^{\alpha-1}(-\log t)^{\alpha-1}\frac{t^{-\rho-\rho'}-1}{\rho+\rho'}$$
$$\left. -\sqrt{k}\tilde{A}^2(n/k)\frac{\alpha(\alpha-1)}{2}\gamma^{\alpha-2}(-\log t)^{\alpha-2}\left(\frac{t^{-\rho}-1}{\rho}\right)^2\right| \to 0 \text{ a.s.}.$$

Some terms are omitted here because $\sup_{t\in(0,1]} t^{1/2+\varepsilon}t^{-1}|e(t)| = O_{\mathbb{P}}(1)$ and $\tilde{A}(n/k) \to 0$ as $n \to \infty$.

By taking $\varepsilon < 1/2$, we can then take the integral of $\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right)\right)^\alpha$ on $(0,1]$ and use the fact that $\int_0^1 (-\log t)^{a-1}t^{-b}dt = \dfrac{\Gamma(a)}{(1-b)^a}$ for $b < 1$ to obtain the result in the corollary. The random term is $P_1^{(\alpha)} = \int_0^1 (-\log t)^{\alpha-1}(t^{-1}e(t) - e(1))\, dt$. The covariance can be calculated from there. □

Next, we handle the asymptotic property of the estimator of the second-order parameter ρ. The estimator of ρ is based on a different k sequence, k_ρ, satisfying (8). Because k_ρ satisfies the condition in Proposition A.1, we get the asymptotic properties of the moments $M_{k_\rho}^{(\alpha)}$ as in Corollary A.2. Then, following the same lines as in the proof of Gomes et al. [2002, Theorem 2.2], we get the following proposition.

Proposition A.3. *Suppose that $\{X_1, X_2, \ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third-order condition (7) with parameters $\rho < 0, \rho' < 0$. Suppose an intermediate sequence k_ρ satisfies (8). In addition, assume that the regulatory conditions holds. Then, for the ρ estimator defined in (10) and as $n \to \infty$*

$$\sqrt{k_\rho}\tilde{A}(n/k_\rho)\left(\hat{\rho}_{k_\rho}^{(\alpha)} - \rho\right)$$

is asymptotically normally distributed.

We remark that analogous to the result in Theorem 2.1 in Gomes et al. [2002], the consistency of the ρ estimator for β-mixing time series can be proved under only the second-order condition (3) and weaker conditions on k_ρ.

Finally, we can use the tools built in Corollary A.2 and PropositionA.3 to prove our main results.

Proof of Theorem 4.1. From Corollary A.2, with k_n satisfying (9), the Hill estimator has the following expansion

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n}-\gamma\right)-\gamma P_1^{(1)}-\sqrt{k_n}\tilde{A}(n/k_n)\frac{1}{1-\rho}\to 0 \text{ a.s.}$$

which leads to

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n}^2-\gamma^2\right)-2\gamma^2 P_1^{(1)}-\sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma}{1-\rho}\to 0 \text{ a.s.}\ .$$

Together with the asymptotic property of $M_{k_n}^{(2)}$ obtained again from Corollary A.2, it implies that

$$\sqrt{k_n}\left(M_{k_n}^{(2)}-2\hat{\gamma}_{k_n}^2\right)-2\gamma^2(P_1^{(2)}-2P_1^{(1)})-\sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma\rho}{(1-\rho)^2}\to 0 \text{ a.s.}\ .$$

Thus, the asymptotic unbiased estimator has the following expansion, almost surely as $n\to\infty$,

$$\begin{aligned}
&\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma\right)\\
=&\sqrt{k_n}\left(\hat{\gamma}_{k_n}-\gamma\right)-\frac{1}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}}\sqrt{k_n}\left(M_{k_n}^{(2)}-2\hat{\gamma}_{k_n}^2\right)\\
=&\gamma P_1^{(1)}+\sqrt{k_n}\tilde{A}(n/k_n)\frac{1}{1-\rho}-\frac{1}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}}\left(2\gamma^2(P_1^{(2)}-2P_1^{(1)})+\sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma\rho}{(1-\rho)^2}\right)\\
=&\gamma P_1^{(1)}-\frac{\gamma(1-\hat{\rho}_{k_\rho}^{(\alpha)})}{\hat{\rho}_{k_\rho}^{(\alpha)}}\left(P_1^{(2)}-2P_1^{(1)}\right)+\sqrt{k_n}\tilde{A}(n/k_n)\frac{\rho}{(1-\rho)^2}\left(\frac{1-\rho}{\rho}-\frac{1-\hat{\rho}_{k_\rho}^{(\alpha)}}{\hat{\rho}_{k_\rho}^{(\alpha)}}\right).
\end{aligned}\tag{18}$$

In the last step we use the fact that $\hat{\gamma}_{k_n}\to\gamma$ a.s., as $n\to\infty$. Further, the relation $k_n/k_\rho\to 0$ implies that $\frac{\sqrt{k_n}\tilde{A}(n/k_n)}{\sqrt{k_\rho}\tilde{A}(n/k_\rho)}\to 0$, as $n\to\infty$. Thus, according to Proposition A.3 and Cramér's Delta method, we get that as $n\to\infty$,

$$\sqrt{k_n}\tilde{A}(n/k_n)\frac{\rho}{(1-\rho)^2}\left(\frac{1-\rho}{\rho}-\frac{1-\hat{\rho}_{k_\rho}^{(\alpha)}}{\hat{\rho}_{k_\rho}^{(\alpha)}}\right)\xrightarrow{\mathbb{P}}0.$$

Together with the consistency of $\hat{\rho}_{k_\rho}^{(\alpha)}$, the expansion (18) implies that as $n\to\infty$,

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma\right)\xrightarrow{\mathbb{P}}\frac{\gamma}{\rho}\left(P_1^{(1)}(2-\rho)+P_1^{(2)}(\rho-1)\right).$$

The theorem is proved by using the covariance structure of $\left(P_1^{(1)},P_1^{(2)}\right)$ given in Corollary A.2. □

Proof of Theorem 4.2. The proof follows the same lines in the proof of Theorem 4.3.8 in de Haan and Ferreira [2006]. Only the asymptotic limit of the estimator of the extreme value index contributes to that of the quantile estimator. The theorem thus follows. □

Acknowledgments

Laurens de Haan's research was partially supported by ENES project /PTDC/MAT/112770/2009. Cécile Mercadier and Chen Zhou's research was partially supported by the "EUR fellowship" grant.

References

P. Billingsley. *Probability and measure.* John Wiley & Sons, 1979.

J.-J. Cai, L. de Haan, and C. Zhou. Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201, 2013.

G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13: 55–87, 2010.

J. Danielsson, L. de Haan, L. Peng, and C. G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*, 76(2):226–248, 2001.

L. de Haan and A. Ferreira. *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. New York, NY: Springer., 2006.

F.X. Diebold, T. Schuermann, and J.D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance, The*, 1(2):30–35, 2000.

H. Drees. Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.*, 10 (4):1274–1301, 2000.

H. Drees. Tail empirical processes under mixing conditions. In *Empirical process techniques for dependent data*, pages 325–342. Birkhäuser Boston, Boston, MA, 2002.

H. Drees. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4): 617–657, 2003.

C. M Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1(1): 126–166, 1991.

I.M. Gomes, L. de Haan, and L.H. Rodrigues. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):31–52, 2008.

M. I. Gomes, L. de Haan, and L. Peng. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5(4):387–414 (2003), 2002.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5): 1163–1174, 1975.

T. Hsing. On tail index estimation using dependent data. *Ann. Statist.*, 19(3):1547–1569, 1991.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973.

A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300, 2000.

S. Resnick and C. Stărică. Asymptotic behavior of Hill's estimator for autoregressive data. *Comm. Statist. Stochastic Models*, 13(4):703–721, 1997.

H. Rootzén. The tail empirical process for stationary sequence. Preprint, Chalmers Univ. Gothenburg, 1995.

H. Rootzén. Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.*, 119(2):468–490, 2009.

C. Stărică. On the tail empirical process of solutions of stochastic difference equations. Preprint, Chalmers Univ. Gothenburg, 1999. URL www.math.chalmers.se/-starica/resume/aarch.ps.gz.

P. Sun and C. Zhou. Diagnosing the distribution of garch innovations. *Unpublished manuscript*, 2013.

I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73(364):812–815, 1978.