Estimation of parameters in heavy-tailed distribution
when its second order tail parameter is known *†

Changryong Baek
University of North Carolina

Vladas Pipiras
CEMAT, Instituto Superior Técnico and
University of North Carolina

April 23, 2009

Abstract

Estimating parameters in heavy-tailed distribution plays a central role in extreme value theory. It is well known that classical estimators based on the first order asymptotics such as the Hill, rank-based and QQ-estimators are seriously biased under finer second order regular variation framework. To reduce the bias, many authors proposed the so-called second order reduced bias estimators for both first and second order tail parameters. In this work, estimation of parameters in heavy-tailed distributions are studied under the second order regular variation framework when the second order parameter in the distribution tail is known. This is motivated in large part by a recent work by the authors showing that the second order tail parameter is known for a large class of popular random difference equations (for example, ARCH models). Several least squares estimators, generalizing rank-based and QQ-estimators, and conditional maximum likelihood estimators, based on the exact form of second order regular variation, are proposed here and their basic asymptotics are established. Several other estimators adapting existing approaches (for example, that of Feuerverger and Hall) are also studied. Numerical performance of all proposed estimators is examined through Monte Carlo simulations.

1 Introduction

Heavy tails refer to a slow, power-like decay of a tail of a distribution function. This phenomenon is observed in a wide range of applications. For example, distributions of log-returns in finance, transmitted file sizes (in packets) in telecommunications to name but a few (see, for example, Embrechts, Klüppelberg and Mikosch (1997), de Haan and Ferreira (2006), Resnick (2007) and Reiss and Thomas (2007)). In mathematical terms, a heavy-tailed distribution of a random variable \( X \) is supposed to have a regularly varying tail in the sense that

\[
F(x) = P(X > x) = L(x)x^{-\alpha}, \quad \alpha > 0,
\]

where \( L(x) \) is a slowly varying function at infinity (see, for example, Bingham, Goldie and Teugels (1989)). The parameter \( \alpha \) is called a tail exponent. From an estimation perspective, (1.1) is typically replaced by

\[
F(x) = P(X > x) = c_1 x^{-\alpha} + o(x^{-\alpha}), \quad \alpha > 0, \ c_1 > 0,
\]  

†Keywords and phrases: extreme value theory, tail exponent estimation, bias reduction, heavy tails, least squares estimator, random difference equations.
as \( x \to \infty \), that is, by modelling the slowly varying function \( L \) in (1.1) as a constant \( c_1 \). Note that (1.2) implies \( \log F(x) \approx \log c_1 - \alpha \log x \) for large \( x \). If given \( n \) i.i.d. observations \( X_i, i = 1, \ldots, n \), of \( X \), consider their order statistics \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \). Since \( F(X_{(n-i+1)}) \approx i/n \), the above suggests that
\[
\log \left( \frac{i}{n} \right) \approx \log c_1 - \alpha \log X_{(n-i+1)} \tag{1.3}
\]
and hence that \( \alpha \), in particular, can be estimated through least squares of \( \log(i/n) \) on \( \log X_{(n-i+1)} \), \( i = 1, \ldots, k \), where \( k \) denotes a threshold. The resulting estimator \( \hat{\alpha}_{RK} \) is referred to here as the rank-based estimator (though there is no widely accepted terminology). It is popular in applied literature (see, for example, Gabaix and Ibragimov (2007) and references therein) and has been considered more rigorously in, for example, Csörgő and Viharos (1997). The relation (1.3) also provides a simple way to detect heavy tails, namely, by considering a log-log plot of empirical distribution function against order statistics.

It is well known that (1.2) yields
\[
F_{\max}(1 - y^{-1}) = (c_1 y)^{1/\alpha} + o(y^{1/\alpha}), \tag{1.4}
\]
as \( y \to \infty \), where \( F_{\max}(z) = \inf \{ x : F(x) \geq z \} \) is an inverse of the distribution function \( F \). Hence, one similarly expects that
\[
\log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log \left( \frac{i}{n} \right) \tag{1.5}
\]
and that \( \alpha \), in particular, can be estimated through least squares of \( \log X_{(n-i+1)} \) on \( \log(i/n) \), \( i = 1, \ldots, k \). The resulting estimator \( \hat{\alpha}_{QQ} \) is called the QQ-estimator and has been studied in detail by Kratz and Resnick (1996).

The rank-based and QQ estimators are not most efficient (supposing the strict power-tail or strict Pareto distribution) and more efficient estimators for tail exponent have been studied by many authors. One such popular estimator, the Hill estimator after Hill (1975) is defined as
\[
\hat{\alpha}_H = \frac{1}{k} \sum_{i=1}^{k} (\log X_{(n-i+1)} - \log X_{(n-k)}) = \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{i}{n} \right) \log \left( \frac{i}{n} \right), \tag{1.6}
\]
where \( k \), as above, is the number of upper order statistics. The log-spacings of order statistics \( U_i = \log X_{(n-i+1)} - \log X_{(n-i)} \) are approximately (exactly for strict Pareto) i.i.d. exponential variables with mean \( 1/\alpha \) and (1.6) can be seen as its (conditional) maximum likelihood estimator. Some further relationships among QQ, Hill and the so-called kernel Hill estimators can be found in Beirlant, Vynckier and Teugels (1996), Aban and Meerschaert (2004) and others.

The above estimators work well and are designed for distributions that are strict or very close to those of strict Pareto. It is well known, however, in both theory and practice, that estimation can be seriously biased when distribution deviates from that of strict Pareto (see, for example, Martins, Gomes and Neves (1999)). A common way to quantify and to model such deviations is through the so-called second order regular variation (see, for example, de Haan and Ferreira (2006)). The distribution tail \( F(x) = P(X > x) \) is second order regularly varying with first order tail parameter \( \alpha > 0 \) and second order tail parameter \( \rho < 0 \) (denoted as \( F \in 2RV(-\alpha, \rho) \)) if there is a suitable function \( g(x) \) such that, for any \( \alpha > 0 \),
\[
\lim_{x \to \infty} \frac{x^\alpha F(x) - (ax)^\alpha F(ax)}{g(x)} = \frac{a^\rho - 1}{\rho}. \tag{1.7}
\]
(The relation (1.1) is referred to as first order regular variation.) From a practical (modelling) perspective, the condition (1.7) is often replaced by

$$F(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha + \rho} + o(x^{-\alpha + \rho}), \quad \text{as } x \to \infty,$$

(1.8)

where $c_1 > 0$ and $\rho < 0$. The collection of distributions (1.8) is known as the class of Hall and Welsh (1985). This class and more generally (1.7) also naturally appear when proving asymptotic normality of the above estimators. These asymptotic normality results already indicate the presence of bias through non-zero mean in the limiting normal distributions.

In order to reduce bias, various authors proposed bias reduced estimators under the assumption (1.8). For example, Feuerverger and Hall (1999) and Beirlant, Dierckx, Guillou and Stáricá (2002) approximated log-spacings of order statistics by normalized exponential distribution and derived estimators based on the maximum likelihood or regression with exponential responses. The generalized jackknife estimators accommodating bias are studied by Gomes, Martins and Neves (2000) and Gomes and Martins (2002), and asymptotically best linear unbiased estimator is proposed by Gomes, Figueiredo and Mendonça (2005). A nice extensive review of this research direction can be found in Reiss and Thomas (2007), Chapter 6.

Bias reduced estimators discussed above assume (1.8) with unknown second order tail parameter $\rho$ (and unknown $\alpha$, $c_1$ and $c_2$). In this work, we are interested in estimation methods when the second order tail parameter $\rho$ is known. Note first that we may suppose without loss of generality that $\rho = -1$, that is,

$$F(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha - 1} + o(x^{-\alpha - 1}), \quad \text{as } x \to \infty.$$  

(1.9)

Indeed, if the distribution tail of $X$ follows (1.8) with known $\rho < 0$, then one may consider instead the variable $X^{-\rho}$ which follows (1.9) (with $\alpha$ replaced by $-\alpha/\rho$, in particular).

Considering the specification (1.9), we are interested in possible ways to estimate unknown parameters in (1.9), basic properties of resulting estimators and comparison of possible estimators. A range of estimators is obviously possible adapting (taking $\rho = -1$ in) available estimation methods that suppose unknown $\rho$, for example, maximum likelihood estimator of Feuerverger and Hall (1999), and will be considered below. On the other hand, the specification (1.9) also suggests other simple estimators. For example, note that (1.9) implies that

$$\log \left( \frac{i}{n} \right) \approx \log c_1 - \alpha \log X_{(n-i+1)} + \frac{c_2}{c_1} X_{(n-i+1)}^{-1},$$

(1.10)

In particular, the tail exponent $\alpha$ can be estimated by regressing $\log(i/n)$ on $(1, \log X_{(n-i+1)}, X_{(n-i+1)}^{-1})$, generalizing the rank-based estimator based on (1.3). The analogue of the QQ-estimator can also be introduced, and is considered below.

**Remark 1.1** Note that, for the second order tail parameter, we use the convention found in, for example, Smith (1987), Feuerverger and Hall (1999), Embrechts et al. (1997), p. 341. An alternative popular specification (see the papers by Beirlant, Gomes and others) is given by

$$F(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha(1-\rho^*)} + o(x^{-\alpha(1-\rho^*)}) = c_1 x^{-\alpha(1 + c_2/c_1 x^{\alpha \rho^*} + o(x^{\alpha \rho^*}))},$$

(1.11)

The connection between (1.8) and (1.11) is

$$\rho^* = \frac{\rho}{\alpha},$$

(1.12)
In the specification (1.11), we therefore assume that $\rho^* = -1/\alpha$. In particular, note that, with this specification, $\rho^*$ is unknown! On another hand, estimation methods presented below for known $\rho = -1$ can also be adapted to the case of known $\rho^*$, say $\rho^* = -1$. This is briefly discussed in Section 2.7 below.

The assumption of known $\rho$ is largely motivated by our recent work on second order properties of tails of the so-called random difference equations or RDEs for short (see Baek, Pipiras, Wendt and Abry (2009)). For example, the widely studied ARCH models are RDEs. In one dimension, a stationary solution $X$ of RDE satisfies the relation $X = AX + B$, where $(A, B)$ is a vector of variables that are independent of $X$. Since a celebrated result of Kesten (1973), it is well-known that, under mild assumptions, RDE $X$ has a power tail with the tail exponent $\alpha$ determined by the variable $A$, namely, $EA^\alpha = 1$. In Theorem 3.1 of Baek et al. (2009), it is shown that, under mild assumptions on RDE $X$,

$$\int_x^\infty (P(X > u) - P(AX > u))du \sim cx^{-\alpha}, \quad c > 0,$$

as $x \to \infty$. This can be considered as a weaker from of 2RV (1.7) in that (1.13) suggests (note from above that $EA^\alpha = 1$)

$$\lim_{x \to \infty} \frac{x^\alpha P(X > x) - x^\alpha(EA^\alpha)^{-1}P(X > A^{-1}x))}{x^{-1}} = c\alpha,$$

which can be viewed as (1.7) at random $a = A^{-1}$ (and, in fact, $\rho = -1$ as suggested by the term $x^{-1}$ in the denominator in (1.14)). Moreover as a consequence of (1.13), it is shown in Proposition 3.1 of Baek et al. (2009) that, under mild assumptions, RDE $X$ can have the form (1.8), the only form for bias reduction used in practice, only when $\rho = -1$ and also $c_2 < 0$.

Existence of a second order term in distribution tail does not, by itself, imply that significant estimation bias is present from a practical perspective (that is, for example, a second order term could be “too far” in the tail from a practical perspective). In the case of RDEs, however, the estimation bias of, for example, Hill estimator appears extremely large for larger values of tail exponents $\alpha$ (say $\alpha = 5$ or 10). Taking the second order term with $\rho = -1$ in (1.9) into account leads to a satisfactory bias correction. A preliminary simulation study about this is reported in Baek et al. (2009). In this work, we greatly expand on that study by looking at more estimators, proving some of their properties, and carrying out a more detailed simulation study.

Considering known $\rho$ should also not be surprising from the following angle. In the specification (1.11), it has been common to pay particular attention to the special case supposing known $\rho^* = -1$ (corresponding to unknown $\rho = -\alpha$ under the framework (1.9)). For example, a number of distributions such as symmetric $\alpha$-stable ($\alpha \in (1, 2)$) and Fréchet have $\rho^* = -1$, and several estimators are designed specially for this case (see, for example, Gomes and Martins (2002), Gomes, Figueiredo and Mendonça (2005) and Section 2.7 below). Let us also add that popular generalized Pareto distributions (GPD) and generalized extreme value distributions (GEVD) also have a known second order tail parameter $\rho = -1$, as in (1.9).

The structure of the paper is as follows. In Section 2, we gather a number of possible estimators under the model (1.9). In Section 3, some of their properties are proved, focusing on estimators of the tail exponent $\alpha$. In Section 4, we provide a simulation study comparing the proposed estimators. Conclusions can be found in Section 5.
2 Estimation methods

In this section, we gather a number of possible estimators of parameters in the framework (1.9). Several estimators are based on least squares methods (Sections 2.1-2.3), generalized jackknife (Section 2.4) and others are maximum likelihood estimators (Sections 2.5-2.6). The estimators of Section 2.1, 2.2 and 2.6 depend particularly on the specific form (1.9) and hence could be considered new. The estimators of Section 2.4 and 2.5, on the other hand, are rather adaptations of estimators available in the case when \( \rho \) is unknown. Finally, in Section 2.7, we briefly discuss estimation in the framework (1.11) assuming known \( \rho^* = -1 \).

2.1 Rank-based, least squares estimators

From (1.9), observe that

\[
\bar{F}(x) = P(X > x) = c_1 x^{-\alpha} + c_2 x^{-\alpha - 1} + o(x^{-1}) = c_1 x^{-\alpha} \left(1 + \frac{c_2}{c_1} x^{-1}\right) + o(x^{-1}),
\]

as \( x \to \infty \). By taking the logarithm, as \( x \to \infty \),

\[
\log(\bar{F}(x)) \approx \log c_1 - \alpha \log x + \log \left(1 + \frac{c_2}{c_1} x^{-1}\right) \approx \log c_1 - \alpha \log x + \frac{c_2}{c_1} x^{-1}.
\]

Therefore, one expects that parameters \( \alpha, c_1 \) and \( c_2 \) could be estimated by a linear regression of the logarithm of empirical distribution tail of \( X \) on \((1, \log x, 1/x)\) that is,

\[
\arg\min_{\beta_0, \alpha, \beta_1} \sum_{i=1}^{k} \left(\log\left(\frac{i}{n}\right) - \beta_0 + \alpha \log \left(X_{n-i+1}\right) - \beta_1 \left(X_{n-i+1}\right)\right)^2,
\]

where \( \beta_0 = \log c_1 \) and \( \beta_1 = c_2/c_1 \). This approach generalizes the least squares tail estimator based on the first order asymptotics (1.3). We denote the corresponding tail exponent estimator as \( \hat{\alpha}_{RK2} \).

Equivalently, problem (2.3) can be written as

\[
\arg\min_{\beta_0, \alpha, \beta_1} \sum_{i=1}^{k} \left(\log\left(\frac{i}{k}\right) - \beta_0 + \alpha \log \left(X_{n-i+1}/X_{n-k}\right) - \beta_1 \left(X_{n-i+1}/X_{n-k}\right)\right)^2,
\]

where \( \beta_0 = \log c_1 - \log(k/n) - \alpha \log X_{n-k} \), \( \beta_1 = c_2/(c_1 X_{n-k}) \). The tail exponent estimator, in particular, can be expressed as

\[
\hat{\alpha}_{RK2} = -\frac{A - B}{C - D},
\]

where

\[
A = \left(1 - \frac{1}{k} \sum_{i=1}^{k} \left(X_{n-k}/X_{n-i+1}\right)^2\right) - \left(1 - \frac{1}{k} \sum_{i=1}^{k} X_{n-k}/X_{n-i+1}\right)^2 \times \left(1 - \frac{1}{k} \sum_{i=1}^{k} \log\left(X_{n-i+1}/X_{n-k}\right) \log\left(\frac{i}{k}\right) - \frac{1}{k^2} \sum_{i=1}^{k} \log\left(X_{n-k}/X_{n-i+1}\right) \sum_{i=1}^{k} \log\left(\frac{i}{k}\right)\right),
\]

\[
B = \left(1 - \frac{1}{k} \sum_{i=1}^{k} \left(X_{n-k}/X_{n-i+1}\right) \log\left(X_{n-i+1}/X_{n-k}\right) - \frac{1}{k^2} \sum_{i=1}^{k} X_{n-k}/X_{n-i+1}\right) \times \left(1 - \frac{1}{k} \sum_{i=1}^{k} \log\left(X_{n-i+1}/X_{n-k}\right) \log\left(\frac{i}{k}\right) - \frac{1}{k^2} \sum_{i=1}^{k} \log\left(X_{n-k}/X_{n-i+1}\right) \sum_{i=1}^{k} \log\left(\frac{i}{k}\right)\right).
\]
Remark 2.1 Applying continuity correction for empirical distribution, namely replacing \( \log((i - 0.5)/k) \) by \( \log((i - 0.5)/k) \) in (2.4) improves estimation in smaller samples. In the context of (1.3), similar correction is well-known and is justified recently in Gabaix and Ibragimov (2007).

2.2 Analogues of the QQ-estimator

The estimators below can be viewed as generalizations of the QQ-estimators based on (1.5). Observe from (2.1) that an inverse function of \( F(x) \) satisfies

\[
F^{-1}\left(1 - \frac{1}{y}\right) = (c_1 y)^{1/\alpha} \left(1 + \frac{c_2}{\alpha c_1} (c_1 y)^{-1/\alpha}\right) + o(1), \quad y \to \infty
\]  

(2.10)

(this can be seen by making the change of variables, \( x = 1 - c_1 y^\alpha - c_2 y^{\alpha+1} \) in (1.9) and using the approximation \((1 + z)^{-\gamma} = 1 - \gamma z + o(z)\), as \( z \to 0 \)). By replacing \( 1/y \) by \( i/n \) and taking the logarithm in (2.10) gives

\[
\log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log \left(\frac{i}{n}\right) + \log \left(1 + \frac{c_2}{\alpha c_1} (c_1 \frac{n}{i})^{-1/\alpha}\right).
\]

This suggests nonlinear least squares estimation

\[
\arg\min_{\beta_0, \beta_1, \alpha} \sum_{i=1}^{k} \left(\log X_{(n-i+1)} - \beta_0 + \frac{1}{\alpha} \log \left(\frac{i}{n}\right) - \log \left(1 + \beta_1 \left(\frac{i}{n}\right)^{1/\alpha}\right)\right)^2,
\]  

(2.11)

where \( \beta_0 = \log c_1/\alpha \) and \( \beta_1 = c_2/(\alpha c_1^{1+1/\alpha}) \). The nonlinear minimization (2.11) can be reduced to that over \( \beta_1 \) and \( \alpha \). We will denote the tail exponent estimator based on (2.11) as \( \hat{\alpha}_{\text{QQn}} \).

Furthermore, observe from (2.10) that

\[
F^{-1}\left(1 - \frac{1}{y}\right) - \frac{c_2}{\alpha c_1} = (c_1 y)^{1/\alpha} + o(1).
\]  

(2.12)

Taking the logarithm gives

\[
\log \left(\frac{X_{(n-i+1)}}{c_2/\alpha c_1}\right) \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log \left(\frac{i}{n}\right).
\]  

(2.13)
and approximating the left-hand side yields

\[ \log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log \left( \frac{i}{n} \right) + \frac{c_2}{\alpha c_1} \frac{1}{X_{(n-i+1)}}. \]  

Relation (2.14) suggests linear regression estimators through

\[ \text{argmin} \sum_{i=1}^{k} \left( \log X_{(n-i+1)} - \beta_0 + \frac{1}{\alpha} \log \left( \frac{i}{n} \right) - \beta_1 \frac{1}{X_{(n-i+1)}} \right)^2, \]

where \( \beta_0 = \log c_1/\alpha, \beta_1 = c_2/(\alpha c_1) \), or equivalently

\[ \text{argmin} \sum_{i=1}^{k} \left( \log \left( \frac{X_{(n-i+1)}}{X_{(n-k)}} \right) - \beta_0 + \frac{1}{\alpha} \log \left( \frac{i}{k} \right) - \beta_1 \frac{X_{(n-k)}}{X_{(n-i+1)}} \right)^2, \]

where \( \beta_0 = \log c_1/\alpha - \log X_{(n-k)} - \log(k/n)/\alpha, \beta_1 = c_2/(\alpha c_1 X_{(n-k)}) \). This is, in fact, the inverse regression of (2.4) by reversing the roles of log order statistics and empirical distribution. We denote the corresponding tail exponent estimator as \( \hat{\alpha}_{QQ2} \). Similar calculation as for (2.5) (or just reversing the roles of \( \log X_{(n-i+1)} \) and \( \log(i/k) \)) gives an explicit form of \( \hat{\alpha}_{QQ2} \) as

\[ \hat{\alpha}_{QQ2}^{-1} = -\frac{A - B}{E - F}, \]  

where \( A \) and \( B \) are given in (2.6) and (2.7), and

\[ E = \left( \frac{1}{k} \sum_{i=1}^{k} \log^2 \left( \frac{i}{k} \right) - \left( \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{i}{k} \right) \right)^2 \right) \left( \frac{1}{k} \sum_{i=1}^{k} \left( \frac{X_{(n-k)}}{X_{(n-i+1)}} \right)^2 - \left( \frac{1}{k} \sum_{i=1}^{k} \frac{X_{(n-k)}}{X_{(n-i+1)}} \right)^2 \right), \]

\[ F = \left( \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{i}{k} \right) - \frac{1}{k^2} \sum_{i=1}^{k} \frac{X_{(n-k)}}{X_{(n-i+1)}} \sum_{i=1}^{k} \log \left( \frac{i}{k} \right) \right)^2. \]

### 2.3 Generalized least squares methods

In the first order regular variation framework, Aban and Meerschaert (2004) examined the generalized least squares estimator based on (1.5). We briefly comment here on similar ideas, for the second order framework (1.9). Recall that

\[ -\log(1 - F(X_{(n-i+1)})) = \frac{e_1}{n} + \frac{e_2}{n-1} + \ldots + \frac{e_{n+1-i}}{i} =: Y_{(n-i+1)}, \]

where \( e_i \) are i.i.d. exponential random variables with mean 1. Based on the assumption (1.9), this leads to

\[ -\log c_1 + \alpha \log X_{(n-i+1)} - \frac{c_2}{c_1} \frac{1}{X_{(n-i+1)}} \]

approximately having the same distribution as \( Y_{(n-i+1)} \). The expectation and covariance of \( Y_{(n-i+1)} \) are calculated as

\[ \nu_i := EY_{(n-i+1)} = \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{i}, \]

7
This suggests generalized least squares estimator for known \( \beta_1 = c_2/(\alpha c_1) \) as

\[
\arg\min_{\beta_0, \alpha} \left( \log \mathbf{X} - \beta_1 \mathbf{1}/\mathbf{X} - \beta_0 - \frac{1}{\alpha} \mathbf{E}\mathbf{Y} \right)^T \alpha^2 \Sigma^{-1} \left( \log \mathbf{X} - \beta_1 \mathbf{1}/\mathbf{X} - \beta_0 - \frac{1}{\alpha} \mathbf{E}\mathbf{Y} \right),
\]

where \( \log \mathbf{X} = (\log X_n \log X_{n-1} \ldots \log X_{(n-k+1)})^T \), \( \mathbf{1}/\mathbf{X} = (X_n^{-1} X_{n-1}^{-1} \ldots X_{(n-k+1)}^{-1})^T \), \( \mathbf{E}\mathbf{Y} = (\nu_1, \nu_2, \ldots, \nu_k) \) and \( \beta_0 = \log c_1/\alpha \). The corresponding tail exponent estimator \( \hat{\alpha}_{GLS} \) can be expressed as in equation (2.9) of Aban and Meerschaert (2004),

\[
\hat{\alpha}_{GLS}^{-1} = \frac{1}{k} \sum_{i=1}^{k} \left( \log X_{(n-i+1)} - X_{(n-k+1)} - \beta_1 \left( X_{(n-i+1)}^{-1} - X_{(n-k+1)}^{-1} \right) \right).
\]

(2.19)

This estimator has nice properties according to Aban and Meerschaert (2004). However, in practice, \( \beta_1 \) needs to be estimated. One possible approach is to use \( \beta_1 \) from QQ2 estimator in Section 2.2, which is a GLS estimator with identity covariance matrix.

**Remark 2.2** Assuming \( \beta_1 \) is unknown in (2.18) yields inconsistent estimators of tail exponent \( \alpha \). The same happens if one tried the generalized least squares regression

\[
\arg\min_{\beta_0, \alpha, \beta_1} (-\mathbf{E}(\mathbf{Y}) - \beta_0 + \alpha \log \mathbf{X} - \beta_1 \mathbf{1}/\mathbf{X})^T \Sigma^{-1} (-\mathbf{E}(\mathbf{Y}) - \beta_0 + \alpha \log \mathbf{X} - \beta_1 \mathbf{1}/\mathbf{X}),
\]

(2.20)

based on (2.3). The reason for inconsistency is that, for example, (2.3) is not a standard regression problem in that, for example, response variables are constant. (Moreover, there is nothing special here about the second order framework: no consistency would be possible through (2.20) with \( \beta_1 = 0 \) in the first order framework.)

### 2.4 Generalized jackknife estimators

We introduce here several generalized jackknife estimators accommodating bias. Following Gomes, Martins and Neves (2000), consider two biased estimators \( \hat{\alpha}^{(1)}(k) \) and \( \hat{\alpha}^{(2)}(k) \) of \( \alpha \) such that

\[
\mathbf{E}\hat{\alpha}^{(1)}(k) = \alpha + \phi(\alpha)d_1(k), \quad \mathbf{E}\hat{\alpha}^{(2)}(k) = \alpha + \phi(\alpha)d_2(k).
\]

The generalized jackknife estimator of \( \alpha \) based on \( \hat{\alpha}^{(1)}(k) \) and \( \hat{\alpha}^{(2)}(k) \) is defined as

\[
\hat{\alpha}^{(G)}(\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)}) = \frac{\hat{\alpha}^{(1)} - q_k \hat{\alpha}^{(2)}}{1 - q_k},
\]

(2.21)

where the weight \( q_k \) is given by

\[
q_k = \frac{\mathbf{BIAS}(\hat{\alpha}^{(1)})}{\mathbf{BIAS}(\hat{\alpha}^{(2)})}.
\]

(2.22)

For example, consider the following three biased estimators of tail exponent,

\[
\hat{\alpha}^{(1)} = \hat{\alpha}_H, \quad \hat{\alpha}^{(2)} = \frac{2}{M_2 \hat{\alpha}_H}, \quad \hat{\alpha}^{(3)} = \sqrt{\frac{2}{M_2}}.
\]

8
where $\hat{\alpha}_H$ is the Hill estimator in (1.6) and $M_2$ is given by

$$M_2 = \frac{1}{k} \sum_{i=1}^{k} \log^2 \left( \frac{X_{n-i+1}}{X_{n-k}} \right).$$

Under the second order framework, asymptotic behavior of these three estimators is characterized by

$$\sqrt{k} (\hat{\alpha}^{(1)} - \alpha) \xrightarrow{d} \mathcal{N} \left( -\frac{\lambda \alpha^2}{1 - \rho^*}, \alpha^2 \right),$$

$$\sqrt{k} (\hat{\alpha}^{(2)} - \alpha) \xrightarrow{d} \mathcal{N} \left( -\frac{\lambda \alpha^2}{(1 - \rho^*)^2}, 2\alpha^2 \right),$$

$$\sqrt{k} (\hat{\alpha}^{(3)} - \alpha) \xrightarrow{d} \mathcal{N} \left( -\frac{\lambda \alpha^2 (2 - \rho^*)}{2(1 - \rho^*)^2}, \frac{5}{4} \alpha^2 \right),$$

where $\lambda$ appears in (3.2) below (see, for example, Gomes, Martins and Neves (2000)). Hence, the generalized jackknife estimator based on $(\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)})$ is given by

$$\hat{\alpha}^{(G)}(\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)}) = \frac{\hat{\alpha}^{(1)} - (1 - \rho^*) \hat{\alpha}^{(2)}}{\rho^*}$$

(2.23)

and is the same as an asymptotically unbiased estimator in Peng (1998). In our context, $\rho^* = -1/\alpha$ is unknown. Estimating $\rho^*$ by $-1/\hat{\alpha}_H$ gives the generalized jackknife estimator of Peng,

$$\hat{\alpha}_P = -\hat{\alpha}_H^2 + \left( \frac{1}{\hat{\alpha}_H} + 1 \right) \frac{2}{M_2}.$$  

(2.24)

Similarly, the generalized jackknife estimator based on $\hat{\alpha}^{(2)}$ and $\hat{\alpha}^{(3)}$ is given by

$$\hat{\alpha}^{(G)}(\hat{\alpha}^{(2)}, \hat{\alpha}^{(3)}) = \frac{(2 - \rho^*) \hat{\alpha}^{(2)} - 2 \hat{\alpha}^{(3)}}{-\rho^*}.$$  

(2.25)

By replacing $\rho^*$ by $-1/\hat{\alpha}_H$, this yields a jackknife estimator

$$\hat{\alpha}_{JK} = \left( 2 + \frac{1}{\hat{\alpha}_H} \right) \frac{2}{M_2} - 2 \hat{\alpha}_H \sqrt{\frac{2}{M_2}}.$$  

(2.26)

**Remark 2.3** Gomes, Martins and Neves (2000) derive generalized jackknife estimators for $\gamma = 1/\alpha$. For example, replacing $\rho^* = -1/\hat{\alpha}_H$ in Peng’s jackknife estimator gives

$$\hat{\gamma}_P = \frac{1}{\hat{\alpha}^{(2)}} + \frac{\hat{\alpha}^{(1)}}{\hat{\alpha}^{(2)}} - 1.$$  

In terms of $\alpha$, this gives

$$\frac{1}{\hat{\gamma}_P} = \frac{\hat{\alpha}^{(2)}}{1 + \hat{\alpha}^{(1)} - \hat{\alpha}^{(2)}} = \frac{2}{M_2 \hat{\alpha}_H (1 + \hat{\alpha}_H) - 2},$$

which is different from $\hat{\gamma}_P$ in (2.24). It can be easily checked, however, that the asymptotic normality results (Section 3) for $\hat{\alpha}_P$ and $1/\hat{\gamma}_P$ are the same.
2.5 Approximate normalized log-spacings

Feuerverger and Hall (1999) proposed parameter estimators based on normalized log-spacings of order statistics and their approximations by a normalized Exponential distribution. Under the assumption (1.9), consider the normalized log-spacings

\[ U_i = i(\log X_{(n-i+1)} - \log X_{(n-i)}) \]

and set \( \delta(x) = -\alpha^{-1}c_1^{-\alpha^{-1}+1}c_2x^{-1/\alpha} = Dx^{1/\alpha} \). Then, one expects that

\[ U_i \approx \alpha^{-1}(1 + \delta(i/n))Z_i \approx \alpha^{-1}\exp(\delta(i/n))Z_i, \quad (2.27) \]

where \( Z_i \) are independent Exponential random variables with mean 1. This suggests the maximum likelihood estimator of \( \alpha \) based on maximizing

\[
L(D, \alpha) = \sum_{i=1}^{k} \left\{ \log \alpha - D \left( \frac{i}{n} \right)^{\alpha^{-1}} - \alpha U_i \exp \left( -D \left( \frac{i}{n} \right)^{\alpha^{-1}} \right) \right\}.
\]

We denote the corresponding estimator as \( \hat{\alpha}_{FH} \). A related estimator based on regression is the following. Observe from (2.27) that taking logarithm gives

\[
\log U_i \approx -\log \alpha + \delta(i/n) + \log Z_i.
\]

\[ =: -\log \alpha - E(\log(Z_i)) + \delta(i/n) + u_i = \theta + \delta(i/n) + u_i, \]

where \( \theta = -\log \alpha - E(\log(Z_i)) \) and \( u_i \) are i.i.d. random variable with mean zero and variance \( \sigma_i^2 \).

The mean of \( \log Z_i \) is known as Euler’s constant (\( E(\log(Z_i)) = .5772 \ldots \)) and \( \sigma_i^2 = \text{Var}(\log(Z_i)) = \pi^2/6 = 1.644934 \). This leads to the nonlinear regression estimator \( \hat{\alpha}_{FHn} \).

2.6 Conditional maximum likelihood estimators

We derive here conditional maximum likelihood estimators supposing that the distribution tail behaves exactly as

\[
\bar{F}(x) = c_1x^{-\alpha} + c_2x^{-\alpha-1}, \quad x > u, \quad (2.28)
\]

for fixed known threshold \( u \), or equivalently, the corresponding density is given as

\[
f(x) = x^{-\alpha-1}(\alpha c_1 + (\alpha + 1)c_2x^{-1}), \quad x > u,
\]

where

\[
\alpha c_1 + (\alpha + 1)c_2x^{-1} > 0 \quad \text{for all } x > u. \quad (2.29)
\]

Several approaches are possible and are explained next.

First approach:

For a given order statistics \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \), suppose that \( k \) upper order observations are above the threshold \( u \). The joint density of \( k \) upper order statistics (see, for example, Embrechts et al. (1997), p. 185) is

\[
f(X_{(n-k+1)} = x_k, \ldots, X_{(n)} = x_1) = \frac{n!}{(n-k)!}F(x_k)^{n-k} f(x_k)f(x_{k-1})\ldots f(x_1)
\]
\[
= n! \frac{1}{(n-k)!} (1 - c_1 x_k^{-\alpha} - c_2 x_k^{-\alpha-1})^{n-k} \prod_{i=1}^{k} x_i^{-\alpha-1} (\alpha c_1 + (\alpha + 1)c_2 x_i^{-1}).
\] (2.30)

The corresponding maximum likelihood estimators \(\hat{\alpha}_{ML}, \hat{c}_{1ML}, \hat{c}_{2ML}\) are obtained by minimizing negative log-likelihood, namely,

\[
\arg\min_{\alpha,c_1,c_2} \left( \alpha \sum_{i=1}^{k} \log x_i - \sum_{i=1}^{k} \log(\alpha c_1 + (\alpha + 1)c_2 x_i^{-1}) - (n-k) \log(1 - c_1 x_k^{-\alpha} - c_2 x_k^{-\alpha-1}) \right),
\] (2.31)

subject to (2.29), where we denote \(x_i = X_{(n-i+1)}\) for notational simplicity. This is equivalent to finding solutions to

\[
(n-k) \frac{c_1 x_k + c_2}{x_k^{\alpha+1} - c_1 x_k - c_2} - \sum_{i=1}^{k} \log x_i + \sum_{i=1}^{k} \frac{c_1 x_i + c_2}{\alpha c_1 x_i + (\alpha + 1)c_2} = 0,
\] (2.32)

\[
(n-k) \frac{-x_k}{x_k^{\alpha+1} - c_1 x_k - c_2} + \sum_{i=1}^{k} \frac{\alpha x_i}{\alpha c_1 x_i + (\alpha + 1)c_2} = 0,
\] (2.33)

\[
(n-k) \frac{-1}{x_k^{\alpha+1} - c_1 x_k - c_2} + \sum_{i=1}^{k} \frac{\alpha + 1}{\alpha c_1 x_i + (\alpha + 1)c_2} = 0,
\] (2.34)

subject to the condition (2.29).

Substituting (2.33) and (2.34) into (2.32), we get that

\[
\sum_{i=1}^{k} \frac{1}{\alpha + c_2/(c_1 x_i + c_2)} = \sum_{i=1}^{k} \log \left( \frac{x_i}{x_k} \right).
\] (2.35)

Observe also that adding \(c_1 \times (2.33)\) and \(c_2 \times (2.34)\) gives

\[
c_1 x_k + c_2 = \frac{k}{n} x_k^{\alpha+1}
\] (2.36)

and relations (2.33), (2.34), (2.35) and (2.36) give

\[
\frac{c_1}{\alpha} \frac{n}{x_k^{\alpha}} + \frac{c_2}{\alpha + 1} \frac{n}{x_k^{\alpha+1}} = \sum_{i=1}^{k} \log \left( \frac{x_i}{x_k} \right).
\] (2.37)

Solving linear equations (2.36) and (2.37) gives a closed-form expression for \(c_1\) and \(c_2\) in terms of \(\alpha\),

\[
c_1 = \frac{k}{n} x_k^{\alpha} \left( \frac{\alpha + 1}{\alpha_H} - 1 \right), \quad c_2 = \frac{k}{n} x_k^{\alpha+1} \left( 1 - \alpha \left( \frac{\alpha + 1}{\alpha_H} - 1 \right) \right),
\] (2.38)

where

\[
\tilde{\alpha}_H^{-1} = \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{x_i}{x_k} \right) = \frac{k-1}{k} \tilde{\alpha}_H^{-1}
\]

is a version of the original Hill estimator. Finally, \(\hat{\alpha}_{ML}\) is the solution of nonlinear equation,

\[
\sum_{i=1}^{k} \left( \frac{1}{\alpha + w_i(\alpha)} - \frac{1}{\hat{\alpha}_{ML}} \right) = 0,
\] (2.39)
where weights are given by
\[ w_i(\alpha) = \frac{x_k(\alpha + 1)(\tilde{\alpha}_H - \alpha)}{\alpha(\alpha + 1 - \tilde{\alpha}_H)x_i + x_k(\alpha + 1)(\tilde{\alpha}_H - \alpha)}. \]

**Remark 2.4** Observe from (2.35) and (2.36) that if \( c_2 = 0 \), then \( \hat{\alpha}_{ML} \) leads to conditional maximum likelihood estimator of Hill (1975) for the first order asymptotics,
\[ \hat{\alpha}_{ML} = \tilde{\alpha}_H, \quad \hat{c}_{1,ML} = \frac{k}{n} \frac{\tilde{\alpha}_H}{x_k}. \]

In the case of RDEs, for example, one expects that \( c_1 > 0 \) and \( c_2 < 0 \). This leads to further restrictions on the solutions of (2.39). Note that the inequality (2.29) and (2.38) imply that
\[ 1 < \alpha(((\alpha + 1)/\tilde{\alpha}_H - 1) < \alpha + 1. \tag{2.40} \]
This is equivalent to
\[ \tilde{\alpha}_H < \alpha < \tilde{\alpha}_H - \frac{1}{2} + \sqrt{\frac{\alpha^2}{4} + 1}. \tag{2.41} \]

From computational perspective, by using relation (2.36), maximum likelihood estimator \( \hat{\alpha}_{ML} \) is obtained by minimizing the negative log-likelihood
\[ \text{argmin}_\alpha \left( \alpha \sum_{i=1}^{k} \log x_i - \sum_{i=1}^{k} \log(\alpha c_1 + (\alpha + 1)c_2 x_i^{-1}) \right), \]
where \( c_1 \) and \( c_2 \) are functions of \( \alpha \) as in (2.38), subject to restriction (2.41).

*Second approach:*

Alternatively, note that for a given order statistics \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \), the joint distribution of \( (X_{(n-k+1)}, \ldots, X_{(n)}) \) given \( X_{(n-k)} = u \) is the same as the joint distribution of order statistics \( Y_{(1)}, \ldots, Y_{(k)} \) of i.i.d. random variables from a distribution
\[ F_u(y) = P(X \leq y | X > u) = \frac{F(y) - F(u)}{1 - F(u)}, \quad y \geq u \tag{2.42} \]
(see, for example, Lemma 3.4.1 of de Haan and Ferreira (2006)). Under (2.28) conditional density function becomes
\[ f_u(y) = y^{-\alpha - 2} u^\alpha (\alpha c y + (\alpha + 1)(1 - c)u), \tag{2.43} \]
where
\[ c = \frac{c_1}{c_1 + c_2 u^{-1}} \tag{2.44} \]
and
\[ \alpha c y + (\alpha + 1)(1 - c)u > 0, \quad y \geq u. \tag{2.45} \]
Therefore, maximum likelihood estimators \( (\hat{\alpha}_{ML2}, \hat{c}_{ML2}) \) are given by minimizing negative log-likelihood, namely,
\[ l(\alpha, c) = -\log k! + \sum_{i=1}^{k} (\alpha + 2) \log y_i - \alpha \log u - \log(\alpha c y_i + (\alpha + 1)(1 - c)u)), \tag{2.46} \]
subject to (2.45), where we denote $y_i = Y_{(i)} = X_{(n-k+i)}$ for notational simplicity. By solving $\partial l/\partial \alpha = 0$ and $\partial l/\partial c = 0$, the maximum likelihood estimators are given as the solution of

$$
\sum_{i=1}^{k} \frac{1}{\alpha + (1-c)u/(y_i c + (1-c)u)} = \sum_{i=1}^{k} \log \left( \frac{y_i}{u} \right),
$$

(2.47)

c = \frac{\alpha(\alpha + 1 - \alpha \hat{H})}{\hat{\alpha} \hat{H}}.

(2.48)

**Remark 2.5** The only difference between the two maximum likelihood estimators for $\alpha$ is what version of Hill estimator is used. First approach uses a version of original Hill estimator $\tilde{\alpha}_H$, while the second approach uses unbiased Hill estimator $\hat{\alpha}_H$. Note also that for the constant estimators, the first approach gives explicit estimators for two constant parameters $c_1$ and $c_2$, whereas the second approach is parameterized only by one constant $c$. However, replacing $\tilde{\alpha}_H$ by $\hat{\alpha}_H$ in $\hat{c}_{1ML}$ and $\hat{c}_{2ML}$ and plugging them into (2.44) with $u = X_{(n-k)}$ gives $\hat{c}_{ML2}$.

**Remark 2.6** Suppose that $c_1 > 0$ and $c_2 < 0$. Then, (2.44) and (2.45) with $y = u$ imply that

$$1 < c < \alpha + 1,$$

which leads further to

$$\hat{\alpha}_H < \alpha < \hat{\alpha}_H - \frac{1}{2} + \sqrt{\frac{\hat{\alpha}_H^2}{4} + 1}.$$  

(2.49)

This is exactly of the form (2.40) replacing $\tilde{\alpha}_H$ by $\hat{\alpha}_H$. In practice, maximum likelihood estimator is obtained by minimizing

$$\arg\min_{\alpha} \left( \alpha \sum_{i=1}^{k} \log(y_i/u) - \sum_{i=1}^{k} \log(\alpha c + (\alpha + 1)(1-c)(y_i/u)^{-1}) \right),$$

subject to (2.49).

### 2.7 Discussions on estimation for known $\rho^*$

We discuss here briefly estimation methods in the framework (1.11) where $\rho^*$ is known. Without loss of generality, we may suppose that $\rho^* = -1$. Maximum likelihood estimators of Sections 2.5 and 2.6, and the generalized jackknife estimators (for example, Peng’s estimator in (2.23)) can be adapted analogously to this case. The least squares regression estimators can be defined as outlined next.

As in (2.10), the inverse function of $F(x)$ can be written as

$$F^{-1} \left( 1 - \frac{1}{y} \right) = (c_1 y)^{1/\alpha} \left( 1 + \frac{c_2}{\alpha c_1} (c_1 y)^{-1}(1 + o(1)) \right), \quad y \to \infty.$$  

(2.50)

Then, the corresponding QQ-estimator can be based on

$$\log X_{(n-i+1)} \approx \frac{1}{\alpha} \log c_1 - \frac{1}{\alpha} \log \left( \frac{i}{n} \right) + \frac{c_2}{\alpha c_1^2} \left( \frac{i}{n} \right).$$  

(2.51)

To see how rank-based, least squares estimator can be defined, note that

$$F(x) = c_1 x^{-\alpha} \left( 1 + \frac{c_2}{c_1} x^{-\alpha} + o(x^{-\alpha}) \right) = c_1 x^{-\alpha} \left( 1 + \frac{c_2}{c_1^2} (c_1 x^{-\alpha}) + o(x^{-\alpha}) \right)$$

13
\[ \approx c_1 x^{-\alpha} \left( 1 + \frac{c_2}{c_1} (F(x)) \right). \]  

Taking the logarithm in (2.52) yields

\[ \log \left( \frac{i}{n} \right) \approx \log c_1 - \alpha \log X_{(n-i+1)} + \frac{c_2}{c_1} \left( \frac{i}{n} \right). \]  

3 Theoretical properties of estimators

We examine here basic theoretical properties of proposed estimators. More specifically, we focus on estimators \( \hat{\alpha}_{RK2}, \hat{\alpha}_{QQ2}, \hat{\alpha}_{JK}, \hat{\alpha}_P \) (Section 3.1) and \( \hat{\alpha}_{ML2}, \hat{\alpha}_{FH} \) (Section 3.2) of tail exponent \( \alpha \), and their asymptotic normality. The proofs are similar to what can be found in the literature, and are only outlined. The other estimators are not considered for shortness sake and for being close relatives of the considered estimators.

The asymptotic normality results will be established under the second order condition (1.7) with parameter \( \rho = -1 \), which is more convenient to write here as

\[ \lim_{x \to \infty} \frac{F(xa)}{F(x)} - a^{-\alpha} \frac{G(x)}{\alpha} = a^{-\alpha} a^{-1} - 1^{-1/\alpha^2}, \]

where \( \alpha > 0 \) and \( G \) is (ultimately) a positive or negative function with \( \lim_{x \to \infty} G(x) = 0 \). As common for similar results in related literature, we shall consider \( k = k(n) \to \infty \), \( k/n \to 0 \) as \( n \to \infty \), such that

\[ \lim_{n \to \infty} \sqrt{k} G \left( \frac{n}{k} \right) = \lambda < \infty. \]

3.1 Least squares based and generalized jackknife estimators

We state here asymptotic normality of the rank-based, QQ and generalized jackknife estimators.

**Theorem 3.1** Let \( \hat{\alpha} \) denote one of the estimators \( \hat{\alpha}_{RK2}, \hat{\alpha}_{QQ2}, \hat{\alpha}_{JK} \) and \( \hat{\alpha}_P \). Under the assumptions (3.1)-(3.2) above, we have

\[ \sqrt{k} (\hat{\alpha} - \alpha) \overset{d}{\to} \mathcal{N}(0, \sigma^2), \]

where the respective asymptotic variances \( \sigma^2 \) are given by

\[ \sigma^2_{RK2} = \sigma^2_{QQ2} = 2\alpha^2(\alpha + 1)^2, \]  

\[ \sigma^2_{JK} = \sigma^2_P = \alpha^2((\alpha + 1)^2 + \alpha^2). \]

**Remark 3.1** All proposed estimators are aimed to reduce bias in the second order framework. Therefore, it should not be surprising that the mean is zero in the limiting normal distribution (3.3).

**Remark 3.2** Note that the asymptotic normality result (3.3) implies consistency. However, consistency can also be established under milder assumptions of first order asymptotics. The idea is to use Potter’s bound to have lower and upper bounds for each individual term in considered estimators, and apply Central Limit Theorem (see, for example, Lemma 3.2.3 in de Haan and Ferreira (2006), p. 71) for Renyi’s representation of order statistics.
The proof of Theorem 3.1 is standard and will be mostly outlined. We consider only the estimator \( \hat{a}_{RK2} \) (others can be dealt with in analogous way). The main idea is to establish joint normality of each individual term entering into estimator \( \hat{a}_{RK2} \) and to apply the delta method. Let \( \gamma = 1/\alpha \) to simplify the notation and further denote \( Z_{(n-i+1)} = X_{(n-i+1)}/X_{(n-k)} \),

\[
a = \frac{1}{k} \sum_{i=1}^{k} Z_{(n-i+1)}^{-2}, \quad b = \frac{1}{k} \sum_{i=1}^{k} Z_{(n-i+1)}^{-1}, \quad c = \frac{1}{k} \sum_{i=1}^{k} \left( \log(i/k) - \frac{\log(i/k)}{n} \right) \log Z_{(n-i+1)},
\]

\[
d = \frac{1}{k} \sum_{i=1}^{k} Z_{(n-i+1)} \log Z_{(n-i+1)}, \quad e = \frac{1}{k} \sum_{i=1}^{k} \log Z_{(n-i+1)},
\]

\[
f = \frac{1}{k} \sum_{i=1}^{k} \left( \log(i/k) - \frac{\log(i/k)}{n} \right) Z_{(n-i+1)}^{-1}, \quad g = \frac{1}{k} \sum_{i=1}^{k} (\log Z_{(n-i+1)})^2,
\]

where \( \log(i/k) = \sum_{i=1}^{k} \log \left( \frac{k}{i} \right) / k \). Then, the rank-based estimator can be rewritten as

\[
\hat{a}_{RK2} = \frac{(d-be) - (a-b^2)c}{(g-c^2)(a-b^2) - (d-be)^2} =: H(h).
\]

For asymptotic normality of \( h = (a, b, \ldots, g) \), observe first that under the second order condition (3.1), we can represent log-spacings of order statistics by tail empirical quantile process (see, for example, Theorem 2.4.8 and top of p. 76 of de Haan and Ferreira (2006)). For each \( \epsilon > 0 \),

\[
\log Z_{(n-[ks]+1)} = -\gamma \log s + \frac{\gamma}{\sqrt{k}} s^{-1} B_n^0(s) + G_0 \left( \frac{n}{k} \right) \left( \frac{s^\gamma - 1}{-\gamma} + o_P(1)s^{-1/2-\epsilon} \right), \tag{3.6}
\]

where \( G_0(s) \sim G(s) \) and \( B_n^0(s) = B_n(s) - sB_n(1) \) is a Brownian bridge and \( o_P(1) \) term tends to zero in probability uniformly for \( 0 < s \leq 1 \). By using (3.6), it follows that

\[
a = \frac{1}{2\gamma + 1} - \frac{2\gamma}{\sqrt{k}} \int_0^1 s^{2\gamma-1} B_n^0(s) ds - 2G_0 \left( \frac{n}{k} \right) \int_0^1 s^{2\gamma} \left( \frac{s^{\gamma} - 1}{-\gamma} + o_P(1)s^{-1/2-\epsilon} \right) ds
\]

and hence, as \( k \to \infty \), using (3.2),

\[
\sqrt{k} \left( a - \frac{1}{2\gamma + 1} \right) = -2\gamma \int_0^1 s^{2\gamma-1} B_n^0(s) ds - 2\lambda \int_0^1 s^{2\gamma} \left( \frac{s^{\gamma} - 1}{-\gamma} \right) ds + o_P(1).
\]

Similar expansions for the rest of the terms are

\[
\sqrt{k} \left( b - \frac{1}{\gamma + 1} \right) = -\gamma \int_0^1 s^{\gamma-1} B_n^0(s) ds - \lambda \int_0^1 s^{\gamma} \left( \frac{s^{\gamma} - 1}{-\gamma} \right) ds + o_P(1),
\]

\[
\sqrt{k} (c - (-\gamma)) = \gamma \int_0^1 s^{-1}(1 + \log s) B_n^0(s) ds + \lambda \int_0^1 (1 + \log s) \left( \frac{s^{\gamma} - 1}{-\gamma} \right) ds + o_P(1),
\]

\[
\sqrt{k} \left( d - \frac{\gamma}{(\gamma + 1)^2} \right) = \gamma \int_0^1 s^{\gamma-1}(1 + \gamma \log s) B_n^0(s) ds + \lambda \int_0^1 s^{\gamma}(1 + \gamma \log s) \left( \frac{s^{\gamma} - 1}{-\gamma} \right) ds + o_P(1),
\]

\[
\sqrt{k} (e - \gamma) = \gamma \int_0^1 s^{-1} B_n^0(s) ds + \lambda \int_0^1 \left( \frac{s^{\gamma} - 1}{-\gamma} \right) ds + o_P(1),
\]

15
\[ \sqrt{k}(f - 2\gamma^2) = -2\gamma^2 \int_0^1 s^{-1} \log s \, B_n^0(s) ds - 2\lambda \int_0^1 \log s \left( \frac{s^\gamma - 1}{-\gamma} \right) ds + o_p(1), \]

\[ \sqrt{k}(g - 2\gamma^2) = -2\gamma^2 \int_0^1 s^{-1} \log s B_n^0(s) ds - 2 \gamma \int_0^1 \log s \left( \frac{s^\gamma - 1}{-\gamma} \right) ds + o_p(1), \]

by using the approximation \( e^{-x} \approx 1 - x \) as \( x \to 0 \) and

\[ (B_n^0(s))^2 / \sqrt{k} \to 0, \quad B_n^0(s) G_0(n/k) \to 0, \quad \sqrt{k}(G_0(n/k))^2 \to 0, \]

as \( k \to \infty, \, k/n \to 0. \)

Since \( B_n^0 \) is a Gaussian process, the vector \((a, b, \ldots, g)\) is asymptotically multivariate normal, and \( E B_n^0(s) = 0 \) implies that the asymptotic mean comes from the integration related to \( \lambda \). For example, for \( c \), the change of variables \( y = -\log s \) gives

\[ \lambda \int_0^1 (1 + \log s) \left( \frac{s^\gamma - 1}{-\gamma} \right) ds = \frac{\lambda}{\gamma} \int_0^\infty (1 - y)(1 - e^{-\gamma y}) e^{-y} dy = \frac{-\lambda}{(\gamma + 1)^2}. \]

Note also that the limiting covariance function only involves the Brownian bridge. For example, the limiting covariance of \( a \) and \( b \) becomes

\[ E \left( -2\gamma \int_0^1 s^{2\gamma - 1} B_n^0(s) ds \right) \left( -\gamma \int_0^1 u^{\gamma - 1} B_n^0(u) du \right) = 2\gamma^2 \int_0^1 \int_0^1 s^{2\gamma - 1} u^{\gamma - 1} (s \wedge u - su) ds du = \frac{2\gamma^2}{(\gamma + 1)(2\gamma + 1)(3\gamma + 1)}. \]

After tedious calculations not reported here, we obtain that

\[ \sqrt{k} \left( \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{bmatrix} - \begin{bmatrix} \frac{1}{\gamma + 1} \\ \frac{\gamma}{\gamma + 1} \\ -\gamma \\ \frac{-\gamma}{\gamma + 1} \\ \frac{-\gamma}{\gamma + 1} \\ \frac{-\gamma}{\gamma + 1} \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(\mu', \Sigma), \quad (3.7) \]

where

\[ \mu' = \left( \begin{array}{cc} \frac{-2\lambda}{(3\gamma + 1)(2\gamma + 1)} & \frac{-\lambda}{(3\gamma + 1)(2\gamma + 1)} \\ \frac{-\lambda}{(6\gamma + 2)(\gamma + 1)} & \frac{-\lambda}{(6\gamma + 2)(\gamma + 1)} \\ \frac{-\lambda}{(6\gamma + 2)(\gamma + 1)} & \frac{-\lambda}{(6\gamma + 2)(\gamma + 1)} \\ \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} & \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} \\ \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} & \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} \\ \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} & \frac{-\lambda}{(9\gamma + 3)(\gamma + 1)} \end{array} \right), \]

and the covariance matrix \( \Sigma \) is given by
Theorem 3.2

Under the assumptions (3.1)-(3.2) above, we have

\[ \sigma^2 = \frac{4\gamma^2}{(2\gamma + 1)^2(4\gamma + 1)} \]

where

\[ H = \frac{2\gamma^2}{(2\gamma + 1)^2} \]

Hence, we have

\[ \frac{\partial H}{\partial h} (h_0) = \left( \begin{array}{c} 0, \frac{(\gamma + 1)^2(2\gamma + 1)}{\gamma^4}, \frac{-(\gamma + 1)^2(2\gamma + 1)}{\gamma^4}, \frac{-(\gamma + 1)^2(2\gamma + 1)}{\gamma^5}, \frac{(\gamma + 1)(2\gamma^2 + 4\gamma + 1)}{\gamma^5} \end{array} \right) \]

Algebraic calculations give

\[ \sqrt{k} (H(h) - H(h_0)) \overset{d}{\rightarrow} N\left( \frac{\partial H}{\partial h} (h_0) \mu, \frac{\partial H}{\partial h} (h_0) \Sigma \frac{\partial H}{\partial h} (h_0)' \right) \]

Finally, applying the delta method to the function \( H(h) \) at the point value

\[ h_0 = \left( \frac{1}{2\gamma + 1}, \frac{1}{\gamma + 1}, -\gamma, \frac{\gamma}{(\gamma + 1)^2}, \frac{\gamma}{(\gamma + 1)^2}, 2\gamma^2 \right) \]

gives

\[ \sqrt{k} (H(h) - H(h_0)) \overset{d}{\rightarrow} N(0, 2\alpha^2(\alpha + 1)^2) \]

3.2 Maximum likelihood estimators

We state and outline the proof for the asymptotic normality of maximum likelihood estimators \( \hat{\alpha}_{ML2} \) and \( \hat{\alpha}_{FH} \).

Theorem 3.2 Under the assumption (2.28), we have

\[ \sqrt{k} (\hat{\alpha}_{ML2} - \alpha) \overset{d}{\rightarrow} N(0, \sigma^2_{ML2}(u)), \]

where \( \sigma^2_{ML2}(u) \) is given in (3.14) below, and satisfies

\[ \lim_{u \to \infty} \sigma^2_{ML2}(u) = \alpha^2(\alpha + 1)^2 =: \sigma^2_{ML2}. \]

Under the assumptions (3.1)-(3.2) above, we have

\[ \sqrt{k} (\hat{\alpha}_{FH} - \alpha) \overset{d}{\rightarrow} N(0, \sigma^2_{FH}), \]

where

\[ \sigma^2_{FH} = \alpha^2(\alpha + 1)^2. \]
The proof for \( \hat{\alpha}_{ML2} \) is based on the standard asymptotic normality result for maximum likelihood estimator. From the likelihood function \( l \) in (2.46), observe that

\[
\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{k} \frac{(c y_i + (1 - c) u)^2}{(\alpha c y_i + (\alpha + 1)(1 - c) u)^2},
\]

(3.11)

\[
\frac{\partial^2 l}{\partial \alpha \partial c} = \sum_{i=1}^{k} \frac{y_i u}{(\alpha c y_i + (\alpha + 1)(1 - c) u)^2},
\]

(3.12)

\[
\frac{\partial^2 l}{\partial c^2} = -\sum_{i=1}^{k} \frac{(\alpha y_i - (\alpha + 1) u)^2}{(\alpha c y_i + (\alpha + 1)(1 - c) u)^2}.
\]

(3.13)

Tedious calculations for information matrix lead to (3.8) with

\[
\sigma_{ML2}^2(u) = \frac{\alpha^2 L - 2\alpha(\alpha + 1)M + (\alpha + 1)^2 N}{(c^2 L + 2c(1 - c)M + (1 - c)^2 N)(\alpha^2 L - 2\alpha(\alpha + 1)M + (\alpha + 1)^2 N) - M^2},
\]

(3.14)

where

\[
L = 2F_1(1, \alpha, 1 + \alpha, (1 + \alpha)(c - 1)/\alpha c)/(\alpha^2 c),
\]

\[
M = 2F_1(1, 1 + \alpha, 2 + \alpha, (1 + \alpha)(c - 1)/\alpha c)/(\alpha(\alpha + 1)c),
\]

\[
N = 2F_1(1, 2 + \alpha, 3 + \alpha, (1 + \alpha)(c - 1)/\alpha c)/(\alpha(\alpha + 2)c)
\]

and \( 2F_1(a, b, c, z) = \Gamma(c)(\Gamma(b)\Gamma(c-b))^{-1} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt \) denotes a hypergeometric function. (Note that the dependence on \( u \) in (3.14) is through \( c \).) Finally, as \( u \to \infty \), we have \( c \to 1 \), and all hypergeometric functions converging to 1. This yields (3.9).

Theoretical properties of \( \hat{\alpha}_{FH} \) with known second order tail parameter \( \rho^* \) are studied in Theorem 2.1 of Gomes and Martins (2002). Here, we briefly argue that the same asymptotic result holds for known \( \rho = -1 \) or unknown \( \rho^* = -1/\alpha \). Denote \( \gamma = 1/\alpha \) for notational simplicity, and suppose that the upper \( k \)-order normalized log-spacings exactly follow

\[
U_i \overset{d}{=} \gamma \exp(D(i/n)^\gamma) Z_i,
\]

as in relation (2.27). Then, log-likelihood becomes

\[
l(\gamma, D) = \sum_{i=1}^{k} \left\{ -\log \gamma - D \left( \frac{i}{n} \right)^\gamma - \frac{U_i}{\gamma} \exp \left( -D \left( \frac{i}{n} \right)^\gamma \right) \right\}.
\]

From the asymptotic normality of maximum likelihood estimator,

\[
\sqrt{k} \left( \frac{\hat{\alpha}_{FH} - \gamma}{\hat{\sigma}_{FH}(\gamma)} \right) \overset{d}{\to} N(0, 1),
\]

where

\[
\hat{\sigma}_{FH}^2(\gamma) = \frac{1}{k} \Var \left( \frac{\partial l}{\partial D} \right) \left\{ \frac{1}{k} \Var \left( \frac{\partial l}{\partial D} \right) - \left( \frac{1}{k} E \left( -\frac{\partial^2 l}{\partial D^2} \right) \right)^2 \right\}^{\gamma},
\]

\[
\frac{1}{k} \Var \left( \frac{\partial l}{\partial D} \right) = \left( \frac{k}{n} \right)^{2\gamma} \frac{1}{k} \sum_{i=1}^{k} \left( \frac{i}{k} \right)^{2\gamma}.
\]

18
Figure 1: Comparison of theoretical variance.

\[
\frac{1}{k} \text{Var}\left(\frac{\partial l}{\partial \gamma}\right) = \frac{1}{k} \sum_{i=1}^{k} \gamma^2 \left(\frac{1}{\gamma^2} + \frac{D}{\gamma} \left(\frac{i}{n}\right)^\gamma \log\left(\frac{i}{n}\right)\right)^2,
\]

\[
E\left(-\frac{\partial^2 l}{\partial \gamma \partial D}\right) = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{i}{n}\right)^\gamma \left(\frac{1}{\gamma} + \frac{D}{\gamma} \log\left(\frac{i}{n}\right)\right).
\]

Using \(k = k(n) \to \infty, k/n \to 0\) as \(n \to \infty\), we have \(\sum_{i=1}^{k} \left(\frac{i}{n}\right)^\gamma \log\left(\frac{i}{n}\right) / k \to 0\). Hence, by factoring out \((k/n)^2\),

\[
\hat{\sigma}_{FH}^2(\gamma) \to \frac{\int_0^1 x^{2\gamma} dx}{\gamma^{-2} \int_0^1 x^{2\gamma} dx - \left(\int_0^1 x^\gamma dx\right)^2} = (\gamma + 1)^2,
\]

which gives (3.10).

**Remark 3.3** Figure 1 compares theoretical variances of tail exponent estimators. The following ordering takes place:

\[
\hat{\sigma}_{ML2}^2 = \hat{\sigma}_{FH}^2 < \hat{\sigma}_{JK}^2 = \hat{\sigma}_{P}^2 < \hat{\sigma}_{RK2}^2 = \hat{\sigma}_{QQ2}^2.
\]

**Remark 3.4** Just comparing variances is not the right way to measure performance of estimators. Bias can arise by introducing third order framework, and it can seriously affect asymptotic mean squared error. For example, under the simpler second order framework (3.1)-(3.2), the Hill estimator behaves as

\[
\sqrt{k}(\hat{\alpha}_H - \alpha) \xrightarrow{d} N\left(-\frac{\lambda \alpha^3}{\alpha + 1}, \alpha^2\right),
\]

while the QQ-estimator behaves as

\[
\sqrt{k}(\hat{\alpha}_{QQ} - \alpha) \xrightarrow{d} N\left(-\frac{\lambda \alpha^4}{(\alpha + 1)^2}, 2\alpha^2\right).
\]

That is the QQ-estimator has twice larger variance than the Hill estimator though it has smaller bias. The respective asymptotic mean squared errors (AMSE) become

\[
AMSE(\hat{\alpha}_H) = \frac{1}{k} \frac{\lambda^2 \alpha^6}{(\alpha + 1)^2} + \alpha^2 \frac{1}{k}, \quad AMSE(\hat{\alpha}_{QQ}) = \frac{1}{k} \frac{\lambda^2 \alpha^8}{(\alpha + 1)^4} + 2\alpha^2 \frac{1}{k},
\]
simple algebra gives that if 
$$\lambda > \sqrt{\frac{(\alpha + 1)^4}{\alpha^4(2\alpha + 1)^2}},$$
then AMSE of the QQ-estimator is smaller than that of the Hill estimator. Similar observation is also made in Feuerverger and Hall (1999) by comparing FHn and FH estimators. FHn estimator produces 64% greater variance than $\hat{\alpha}_{FH}$, but numerical studies indicate that $\hat{\alpha}_{FHn}$ is sometimes less biased. Compared to $\hat{\alpha}_{FHn}$, for example, which is based on nonlinear regression, our RK2 or QQ2 estimator has 22% greater variance. However, they are based on linear regression that is easy to implement and are free of initial estimators to solve nonlinear regression.

4 Simulation study

In this section, we present a simulation study examining performance of proposed estimators on several models, and discuss other issues. We consider two models for distribution of $X$: beta prime distributions (Section 4.1), and ARCH(1) or autoregressive conditionally heteroscedastic models of order 1 (Section 4.2). These are known examples of random difference equations and have also been considered in Baek et al. (2009). Beta prime distributions have a closed form with the tail satisfying (1.9). The same behavior (1.9) is expected for ARCH(1) models by the results of Baek et al. (2009).

We shall examine seven estimators $\hat{\alpha}_{RK2}$, $\hat{\alpha}_{QQ2}$, $\hat{\alpha}_{GLS}$, $\hat{\alpha}_{P}$, $\hat{\alpha}_{JK}$, $\hat{\alpha}_{FH}$, $\hat{\alpha}_{ML}$ of heavy tail exponent $\alpha$. All simulations are based on 1,000 realizations. In some realizations, when applying FH method, estimates were very unstable (see below for further discussion on FH method). Therefore, we choose to present here robust measures of performance. We focus on median of tail exponent estimators and median absolute error given by

$$\text{MAE}(\hat{\alpha}(k)) = |\hat{\alpha}(i, k) - \alpha|,$$

where $k$ denotes the number of upper order statistics and $i$ represents realization. Another obvious measure of interest is median absolute deviation given by the median of $|\hat{\alpha}(i, k) - \text{median}(\hat{\alpha}(i, k))|$. We do not report it for shortness sake though we find that its ordering among estimators corresponds to that from theoretical analysis (3.15) (with the exception of FH which is unstable). Finally, several other issues are discussed in Section 4.3. We contrast our estimation methods in correctly specified models to those that assume unknown $\rho$, and also examine our estimation methods on misspecified models.

4.1 Results for beta prime distribution

Beta prime distribution $\beta(a, b)$, also known as beta distribution of the second kind, is given by its density

$$\frac{1}{B(a, b)} x^{a-1} (1 + x)^{-a-b} 1_{\{x > 0\}}, \quad (4.1)$$

where $a > 0$, $b > 0$ are two parameters. It is related to a standard beta distribution $B(a, b)$ through $\beta(a, b) \overset{d}{=} B(b, a)^{-1} - 1$. Observe that the tail of beta prime distribution behaves as

$$F(x) = P(X > x) \overset{d}{=} \frac{1}{B(a, b)} \int_x^\infty u^{a-1} (1 + u)^{-a-b} du = \frac{x^{-b}}{B(a, b)b} - \frac{(a + b)x^{-b-1}}{B(a, b)(b + 1)} + o(x^{-b-1}), \quad (4.2)$$
as $x \to \infty$, and hence satisfies the relation (1.9) with

$$\alpha = b, \quad c_1 = \frac{1}{B(a, b)}, \quad c_2 = -\frac{a + b}{B(a, b)(b + 1)}.$$

Figure 2 presents simulation results for $\beta(9, 5)$ (hence $\alpha = 5$) distribution when the sample size is $N = 5,000$. Median plot suggests that least squares-based ($\hat{\alpha}_{RK2}, \hat{\alpha}_{QQ2}, \hat{\alpha}_{GLS}$) and generalized jackknife ($\hat{\alpha}_P, \hat{\alpha}_{JK}$) estimators perform well. In particular, $\hat{\alpha}_P, \hat{\alpha}_{JK}$ work well for moderate numbers of order statistics used, while $\hat{\alpha}_{QQ2}, \hat{\alpha}_{GLS}$ have the smallest bias over a relatively large $k$ used and $\hat{\alpha}_{RK2}$ yields stable estimates regardless of (reasonable) $k$ used. By comparing MAE, simulation study shows that maximum likelihood, generalized jackknife and least squares estimators work well for a small, moderate and larger number of order statistics $k$ used, respectively. In Figure 3, we also present similar plots for the same distribution when the sample size is $N = 500$. Analogues conclusions can be drawn in this case as well. By considering stability of median plot and ease of implementation, we would prefer the estimator $\hat{\alpha}_{RK2}$ in these simulations. Note also that $\hat{\alpha}_{FH}$ estimator is worst in this setting because of unstable numerical optimization.

**Remark 4.1** Numerical instabilities associated with FH method are well known. See, for example, p. 770 in Feuerverger and Hall (1999), p. 187 in Beirlant et al. (1999) or p. 7 in Gomes.
and Martins (2002). More specifically, the instabilities result from a flat likelihood surface which can yield both larger (true) maximum likelihood values for $\alpha$ and local (as opposed to global) maximum values. The problems are especially pronounced when $\rho$ or $\rho^*$ is small (or $\alpha$ is large in our context). These numerical instabilities can and have been somewhat addressed. For example, Gomes and Martins (2002) use an approximate maximum likelihood which yields an explicit solution. Beirlant et al. (1999) use least squares method in the spirit of FHn method described in Section 2.5. Both methods (Gomes and Martins (2002), and Beirlant et al. (1999)) require a plug-in estimate for $\rho$. In our context, using the Hill estimator as the plug-in leads to satisfactory estimators in our simulations. In particular, we find that the estimator based on Gomes and Martins (2002) performs similarly to generalized jackknife estimator, and that based on Beirlant et al. (1999) similarly to least squares estimators. Finally, let us mention that, in contrast to FH method, conditional maximum likelihood estimator is stable in numerical optimization.

4.2 Results for ARCH model

A popular ARCH(1) model is defined by

$$\xi_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \beta + \eta \xi_{t-1}^2, \quad (4.4)$$

where $\{\epsilon_t\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables and $\beta > 0$, $\eta > 0$ are parameters. If $\xi$ denotes a stationary distribution of (4.4), it is well-known that the distribution $\xi^2$ has the tail exponent $\alpha$ satisfying

$$\Gamma(\alpha + 1/2) = \sqrt{\pi}(2\sigma^2\eta)^{-\alpha}. \quad (4.5)$$

For example, if $\sigma^2 = 1$, numerical calculations yield the following tail exponents $\alpha$ for given values of $\eta$:

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>.577</th>
<th>.312</th>
<th>.214</th>
<th>.163</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

By the results proved in Baek et al. (2009), the distribution tail of $\xi^2$ is expected to satisfy (1.9) (with $c_2 < 0$ in particular).

**Remark 4.2** Note also that, by symmetry, the tail exponent of $\xi$ is

$$\alpha_{\xi} = 2\alpha \quad (4.6)$$

because

$$P(\xi > x) = 1/2P(\xi^2 > x^2) \sim c/2x^{-2\alpha}. \quad (4.7)$$

Moreover, if the tail distribution of $\xi^2$ satisfies (1.9), then that of $\xi$ satisfies

$$P(\xi > x) = c_1/2x^{-2\alpha} + c_2/2x^{-2\alpha-2} + o(x^{-2\alpha-2}) \quad (4.8)$$

and corresponding to (1.8) with $\rho = -2$.

In simulations, we have chosen parameters $\eta = .5$, $\beta = 1$ and $\sigma^2 = 1$. This yields tail exponents $\alpha = 4.73/2$ and $\alpha_{\xi} = 4.73$ for $\xi^2$ and $\xi$, respectively. The simulations are based on independent copies of $\xi^2$, not the time series data (see also remark below). Since one is more interested in the series itself, we report performance of estimator for $\alpha_{\xi}$ (which are obtained from those of $\alpha$ through (4.6)). The results are plotted in Figures 4 and 5 for sample sizes $N = 5,000$ and $N = 500$, respectively. Analogues observations can be made here as in the case of beta prime distribution. Small differences are that $\hat{\alpha}_{FH}$ performs well around $k = 200$ in Figure 4, and $\hat{\alpha}_{GLS}$ works quite well in Figure 5.
Figure 4: ARCH(1) with $\alpha_\xi = 4.73$ and sample size $N = 5,000$.

Figure 5: ARCH(1) with $\alpha_\xi = 4.73$ and sample size $N = 500$.

Remark 4.3 Simulations for ARCH(1) model above are based on independent copies of a stationary solution. (In fact, we take independent copies of $\xi_T$ in (4.4) for large $T$ but these can be considered as good approximations to the stationary solution as discussed in Appendix A of Baek et al. (2009).) We removed temporal dependence in order not to confuse the effects of temporal dependence and second order terms of distribution tails. If temporal dependence is also taken into account, then all considered estimators generally perform worse than when observations are independent. This is also briefly discussed in Section 5.2 of Baek et al. (2009).

4.3 Some other issues

We also compare here the performance of our proposed estimator to those of estimators with estimated second order tail parameter $\rho^*$. Since FH estimator is numerically unstable, we compare generalized jackknife estimators instead. Gomes and Martins (2002) suggest to estimate $\rho^*$ as

$$\hat{\rho}^* = - \frac{3(T(k_1) - 1)}{(T(k_1) - 3)},$$

(4.9)
where

\[ T(k_1) = \frac{M_1 - (M_2/2)^{1/2}}{(M_2/2)^{1/2} - (M_3/6)^{1/3}}, \quad M_j = \frac{1}{k_1} \sum_{i=1}^{k_1} \log^j \left( \frac{X_{(n-i+1)}}{X_{(n-k)}} \right), \]

with the choice of \( k_1 = \min(n - 1, \lceil 2n / \log \log n \rceil) \).

Figure 6 compares performance of our suggested generalized jackknife estimators \( \hat{\alpha}_P, \hat{\alpha}_{JK} \) and the generalized jackknife estimators based on estimated \( \rho^* \) in (2.23) and (2.25) as

\[
\hat{\alpha}_{P\rho^*} = \frac{\hat{\alpha}^{(1)} - (1 - \hat{\rho}^*)\hat{\alpha}^{(2)}}{\hat{\rho}^*}, \quad \hat{\alpha}_{GJ\rho^*} = \frac{(2 - \hat{\rho}^*)\hat{\alpha}^{(2)} - 2\hat{\alpha}^{(3)}}{-\hat{\rho}^*},
\]

respectively. Generalized jackknife estimators with estimated \( \rho^* \) certainly perform better than the Hill estimator, but not as good as our proposed jackknife estimators taking into accounts an exact second order tail parameter.

On a different side, suppose that the second order tail parameter \( \rho \) is not \(-1\). Consider, for example, the Burr model with the distribution tail

\[
\overline{F}(x) = (1 + x^{-\rho^* \alpha})^{1/\rho^*}, \quad x \geq 0, \quad \alpha > 0, \quad \rho^* < 0.
\]

Figure 6: \( \beta(9, 5) \) with sample size \( N = 5,000 \).

Figure 7: Burr(4, \(-0.5\)) with sample size \( N = 5,000 \) (\( \rho = -2 \)).
Figures 7 and 8 illustrate the performance of our proposed estimators and the generalized jackknife methods with estimated $\rho^*$ for the Burr model with $\alpha = 2$, $\rho^* = -0.5$ ($\rho = -2$) and $\alpha = 2$, $\rho = -0.25$ ($\rho = -0.5$). These plots are representatives of what to expect for $\rho < -1$ and $\rho > -1$. The generalized jackknife methods using estimated $\rho^*$ performs best among all estimators. Performance of our estimators (especially rank-based and generalized jackknife) is less sensitive to misspecifications with $\rho < -1$.

5 Conclusions

In this work, we studied various tail parameter estimators in the second order framework with known second order tail parameter. Assuming known second order parameter was largely motivated by our earlier work on second order properties of distribution tails of random difference equations. Second order term in distribution tails can significantly affect the bias of common tail estimators and this work shows how the bias can satisfactorily be removed (supposing a known second order tail parameter).

Among various estimators considered, we generally found least squares estimators performing best, especially the rank-based estimator. They consistently show smallest bias over a large range of upper order statistics considered, and are easy to implement. These estimators would also lead to more reliable confidence intervals for larger number of upper order statistics, despite their asymptotic variance being the largest in theory (among the estimators considered). In the next order of our preference go generalized jackknife and conditional maximum likelihood estimators. The other, FH maximum likelihood estimator has nice theoretical properties but generally suffers from numerical instabilities in practice.

Acknowledgment

We thank an anonymous referee whose suggestions and comments led to a substantial improvement of the paper. The second author is supported in part by the NSF grant DMS-0505628.

References


Changryong Baek and Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Hanes Hall,
Chapel Hill, NC 27599, USA
crbaek@email.unc.edu, pipiras@email.unc.edu