Ann Inst Stat Math (2009) 61:579–598
DOI 10.1007/s10463-007-0155-z

Testing the tail index in autoregressive models

Jana Jurečková · Hira L. Koul · Jan Picek

Received: 8 May 2006 / Revised: 21 June 2007 / Published online: 14 September 2007

Abstract We propose a class of nonparametric tests on the Pareto tail index of the innovation distribution in the linear autoregressive model. The simulation study illustrates a good performance of the tests. Such tests have various applications in a study of flood flows, rainflow data, behavior of solids, atmospheric ozone layer and reliability analysis, in communication engineering, in stock markets and insurance.

Keywords Empirical process · Heavy tailed distribution · Feigin-Resnick estimator · Pareto tail index

1 Introduction

If we are interested in the extremal events such as the extreme intensity of the wind, the high flood levels of the rivers or extreme values of environmental indicators, then

Research of J. Jurečková and J. Picek was partly supported by Czech Republic Grant 201/05/2340, by the Research Project LC06024 and by the NSF grant DMS 0071619. Research of H. L. Koul was partly supported by the NSF grants DMS 0071619 and 0704130.

J. Jurečková (✉)
Department of Probability and Mathematical Statistics, Charles University in Prague, Sokolovská 83, 186 75 Prague 8, Czech Republic
e-mail: jurecko@karlin.mff.cuni.cz

H. L. Koul
Department of Statistics and Probability, Michigan State University, A413 Wells Hall, East Lansing, MI 48824-1027, USA
e-mail: koul@stt.msu.edu

J. Picek
Department of Applied Mathematics, Technical University in Liberec, Hálkova 6, 461 17 Liberec 1, Czech Republic
e-mail: jan.picek@tul.cz

we are rather interested in the tails of the underlying distribution than in its central part. A goodness-of-fit test does not provide us with a sufficient information on the shape of tails, because it concerns mostly the central part. It is important to decide whether the probability distribution function is light- or heavy-tailed. If we decide in favor of a heavy tail, in the next step we should study more closely the shape of the tail, and make the pertinent decisions.

Testing the hypothesis on the tail index of a heavy tailed distribution is an alternative inference to the classical point estimation, surprisingly not yet much elaborated in the literature, though the tests often work under weaker conditions than the point estimators, can be easily reconverted into the confidence sets, and have an intuitive interpretation.

In the present paper, we construct a class of tests on the tail index of the innovation distribution in the linear autoregressive model. Such tests have applications in the environmental and financial time series, among others.

Consider the AR(p) model where the observation X_t satisfies

$$X_t = \rho_1 X_{t-1} + \cdots + \rho_p X_{t-p} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_p)' \in \mathbb{R}^p$ is an unknown parameter and $\varepsilon_t,\ t = 0, \pm 1, \pm 2, \ldots,$ are independent identically distributed (i.i.d.) random variables with a heavy-tailed distribution function F satisfying

$$\lim_{x \to \infty} \frac{-\ln(1 - F(x))}{m \ln x} = 1 \tag{2}$$

for some $m > 0$. The l'Hospital rule and the von Mises condition (see Embrechts et al. (1997), Chap. 3, Theorem 3.3.7) imply that the distributions of type (2) satisfy

$$1 - F(x) = x^{-m} L(x), \quad x \geq x_0, \tag{3}$$

for some $x_0 > 0$, where L stands for a positive function, slowly varying at infinity. In the sequel, L, with or without a suffix will stand for such a function. Moreover, we shall throughout assume that the d.f. F is absolutely continuous having Lebesgue density f.

We shall assume that the time series is strictly stationary. According to Proposition 13.3.2 in Brockwell and Davis (1991, p. 537), a sufficient condition for this is that $1 - \rho_1 z - \cdots - \rho_p z^p \neq 0$ for $|z| \leq 1$. Moreover, then

$$X_t = \sum_{t=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \text{a.s.}, \tag{4}$$

with ψ_j's such that $\sum_{t=0}^{\infty} |\psi_j|^\gamma < \infty$, for $0 < \gamma < \frac{1}{m} \wedge 1$. On the other hand, from Theorem 3.3.7 of Embrechts et al. (1997), we obtain that under (3), for any finite n,

$$\max_{1 \leq t \leq nN} |\varepsilon_t| = O_p(N^{\frac{1}{m}} L_1(N)), \quad \text{as } N \to \infty.$$

This and (4) in turn imply

$$\max_{1\le t\le nN}|X_t| = O_p(N^{\frac{1}{m}}L_1(N)), \quad \max_{1\le t\le nN}\|\mathbf{Y}_t\| = O_p(N^{\frac{1}{m}}L_1(N)), \tag{5}$$

where $\mathbf{Y}_{t-1} := (X_{t-1},\dots,X_{t-p})'$, $t = 0, \pm 1, \dots$.

Let $m_0 > 0$ be a fixed number. The problem of interest is to test either of the following two hypotheses:

$\mathbf{H}_0$: F is of type (3), concentrated on the positive half-axis, satisfying

$$x^{m_0}(1-F(x)) \ge 1, \quad \text{for} \quad x > x_0,$$
$$\text{with hypothetical } \ 0 < m_0 \le 2 \ \text{ and for some } x_0 \ge 0,$$

against the alternative

$\mathbf{K}_0$: F is of type (3), concentrated on the positive half-axis, and $\overline{\lim}_{x\to\infty} x^{m_0}(1-F(x)) < 1$,

or

$\mathbf{H}_1$: F is of type (3) satisfying

$$x^{m_0}(1-F(x)) \ge 1, \ \text{ for } \ x > x_0, \ \text{ with hypothetical } m_0 > 2, \text{ for some } x_0 \ge 0,$$

against the alternative

$\mathbf{K}_1$: F is of type (3) satisfying $\overline{\lim}_{x\to\infty} x^{m_0}(1-F(x)) < 1$, with $m_0 > 2$.

In either of these cases we wish to test the hypothesis that the right tail of F is the same or heavier than that of the Pareto distribution with index m_0 against the alternative that the right tail of F is lighter. The reason for distinguishing between $\mathbf{H}_0$ and $\mathbf{H}_1$ is that one needs to use different estimators of autoregressive parameters in these two cases, which in turn impose different conditions on the model.

Due to (3), the problem of identifying the tails is semiparametric in its nature, involving a nuisance slowly varying function, hence the hypothesis $\mathbf{H}_1$ is the set of all distribution functions either of the form $1 - F(x) = x^{-m_0}L(x)$ where $L(x)$ runs over all slowly varying functions such that $L(x) \ge 1$ for $x > x_0$, or of the form $1 - F(x) = x^{-m}L(x)$ with $m < m_0$ and with a positive slowly varying function L. $\mathbf{H}_0$ is an analogous set of distributions concentrated on the positive half-axis.

Numerous authors have considered the problem of testing the Gumbel hypothesis, $\mathbf{H}_g$: $m = \infty$ against $m < \infty$: we refer to Castillo et al. (1989), Galambos (1982), Gomes (1989), Gomes and Alpuim (1986), Hasofer and Wang (1992), Hosking (1984), van Montfort and Gomes (1985), van Montfort and Witter (1985), Stephens (1977), Tiago de Oliveira (1984). Others, as Falk (1995a,b), Marohn (1994, 1998a,b), studied testing the Gumbel hypothesis in the frame of the local asymptotic normality (LAN). Such tests, when they reject the Gumbel hypothesis in favor of alternative $m < \infty$, do not provide any information on the heaviness of the tail of the distribution, while just that is just the information needed in practical applications.

The proposed tests of $\mathbf{H}_0$ and $\mathbf{H}_1$ are based on the extremes of segments of the residual empirical process. Tests on the Pareto index for the i.i.d. model were constructed by Fialová et al. (2004), Jurečková, and Picek (2001) and by Picek and Jurečková (2001). Tests on the tail index of errors in the linear regression model, based on the extreme regression quantiles, were proposed by Jurečková (1999). However, tests based on the residuals seem to be preferable, and we expect a similar phenomenon to hold in the linear AR time series.

Because the proposed tests are based on the residual empirical process of the AR series, in the next section we first analyze the asymptotic behavior of such processes. The tests and their asymptotic null (normal) distributions are given in Sect. 3, while Sect. 4 deals with their consistency. Section 5 is a numerical illustration.

2 Residual empirical process

Let $n,\ N$ be positive integers and let $\widehat{\boldsymbol{\rho}}_N$ be an estimator of $\boldsymbol{\rho}$ in (1) based on the data set $X_{1-p}, X_{2-p}, \dots, X_0, X_1, \dots, X_{nN}$; the estimators will be considered later.

Let

$$\widehat{\varepsilon}_t := X_t - \widehat{\boldsymbol{\rho}}_N' \mathbf{Y}_{t-1}, \quad t = 1-p, 2-p, \dots, nN, \tag{6}$$

with $\mathbf{Y}_{t-1}$ given in (5).

Now group these residuals in N groups, each of size n, so that the residuals in the tth group are $\widehat{\varepsilon}_{(t-1)n-p+1}, \dots, \widehat{\varepsilon}_{tn-p}$. Do a similar decomposition of the errors $\{\varepsilon_t\}$. Let

$$\widehat{\varepsilon}^t_{(n)} := \max_{1\le i\le n} \widehat{\varepsilon}_{(t-1)n-p+i}, \quad \varepsilon^t_{(n)} := \max_{1\le i\le n} \varepsilon_{(t-1)n-p+i}, \qquad t = 1, 2, \dots, N. \tag{7}$$

The empirical distribution function F_N^* of the maximal errors $\{\varepsilon^t_{(n)},\ t = 1, \dots, N\}$ is approximated by the empirical distribution function $\widehat{F}_N^*$ of the maximal residuals $\{\widehat{\varepsilon}^t_{(n)},\ t = 1, \dots, N\}$, where

$$F_N^*(x) := N^{-1} \sum_{t=1}^{N} I[\varepsilon^t_{(n)} \le x], \quad \widehat{F}_N^*(x) := N^{-1} \sum_{t=1}^{N} I[\widehat{\varepsilon}^t_{(n)} \le x], \qquad x \in \mathbb{R}. \tag{8}$$

Put

$$a^{(1)}_{N,m} := (nN^{1-\delta})^{1/m}, \quad 0 < \delta < 1 \tag{9}$$

or alternatively

$$a^{(2)}_{N,m} := \Big(nN(\ln N)^{-2+\eta}\Big)^{1/m}, \quad 0 < \eta < 1. \tag{10}$$

The effect of the choice of parameters δ and η is discussed in Sect. 5.

We shall first show that

$$|\widehat{F}_N^*(a_{N,m}) - F_N^*(a_{N,m})| = o_p(1) \quad \text{as } N \to \infty \text{ and for a fixed } n, \tag{11}$$

under (3) and with an appropriate estimate $\widehat{\boldsymbol{\rho}}_N$ of $\boldsymbol{\rho}$ in (6), provided m is the true value of the tail index.

To prove (11), we need to find an estimate $\widehat{\boldsymbol{\rho}}_N$ such that, for some sequence of positive numbers $d_N \to \infty$,

$$d_N(\widehat{\boldsymbol{\rho}}_N - \boldsymbol{\rho}) = O_p(1). \tag{12}$$

Possible choices of the estimator are discussed in Sect. 2.1.

If (12) is true, we can write

$$\widehat{\varepsilon}_t = \varepsilon_t - (\widehat{\boldsymbol{\rho}}_N - \boldsymbol{\rho})'\mathbf{Y}_{t-1} = \varepsilon_t - d_N^{-1}\, d_N(\widehat{\boldsymbol{\rho}}_N - \boldsymbol{\rho})'\mathbf{Y}_{t-1}.$$

In view of (5) and (12), for any $\kappa > 0$, there is a $C < \infty$ and an $N_\kappa < \infty$, such that for all $N > N_\kappa$,

$$P\left(\|d_N(\widehat{\boldsymbol{\rho}}_N - \boldsymbol{\rho})\| \le C\right) \ge 1 - \kappa, \quad P\left(\max_{1\le t\le nN} \|\mathbf{Y}_t\| \le CN^{1/m}L_1(N)\right) \ge 1 - \kappa. \tag{13}$$

For $\forall\, \mathbf{u} \in \mathbb{R}^p$, let

$$F_N^*(a_{N,m}^{(\nu)}, \mathbf{u}) = N^{-1} \sum_{t=1}^{N} \prod_{i=1}^{n} I\left[\varepsilon_{(t-1)n-p+i} \le a_{N,m}^{(\nu)} + d_N^{-1}\mathbf{u}'\mathbf{Y}_{(t-1)n-p+i}\right], \tag{14}$$

$\nu = 1, 2,$ and consider the behavior of the sequences

$$D_N^{(\nu)} := \sup_{\|\mathbf{u}\|\le C} \left|F_N^*(a_{N,m}^{(\nu)}, \mathbf{u}) - F_N^*(a_{N,m}^{(\nu)})\right|, \quad \nu = 1, 2. \tag{15}$$

For any d.f. G on $\mathbb{R}$, let $G^{-1}(u) := \inf\{x : G(x) \ge u\}$, $0 \le u \le 1$. Let

$$\xi_N := F^{-1}\left(1 - \tfrac{1}{N}\right) = N^{1/m}L_1(N) \tag{16}$$

be the "population extreme" of F, and let

$$\begin{aligned} \Delta_N^{(\nu)} &:= F\left(a_{N,m}^{(\nu)} + d_N^{-1}C^2\,\xi_N\right) - F\left(a_{N,m}^{(\nu)} - d_N^{-1}C^2\,\xi_N\right), \quad \nu = 1, 2,\\ A_N &:= \left\{\max_{1\le t\le nN} \|\mathbf{Y}_t\| \le C\,\xi_N\right\}, \end{aligned} \tag{17}$$

with C as in (13). Moreover, let $U_{Ni}(x) := F_{Ni}(x) - F(x)$, $i = 1, \ldots, n$, where $F_{Ni}(x)$ is the empirical d.f. of $\{\varepsilon_{(t-1)n-p+i},\ 1 \le t \le N\}$, $i = 1, \ldots, n$. Using the fact that indicators are monotonic functions of their arguments and bounded by 1, and the identity

$$\prod_{i=1}^{M} a_i - \prod_{i=1}^{M} b_i = \sum_{i=1}^{M}[a_i - b_i] \prod_{k=1}^{i-1} a_k \prod_{k=i+1}^{M} b_k,$$

that is valid for any positive integer M and for any real numbers $\{a_i,\ b_i\}_{i=1}^M$, we obtain that an upper bound for $D_N^{(\nu)}$ on the set A_N,

$$
\begin{aligned}
D_N^{(\nu)} &\leq N^{-1}\sum_{t=1}^{N}\sum_{i=1}^{n} I\left[a_{N,m}^{(\nu)} - d_N^{-1}C^2\,\xi_N \leq \varepsilon_{(t-1)n-p+i} \leq a_{N,m}^{(\nu)} + d_N^{-1}C^2\,\xi_N\right]\\
&= \sum_{i=1}^{n}\left[U_{Ni}(a_{N,m}^{(\nu)} + d_N^{-1}C^2\,\xi_N) - U_{Ni}(a_{N,m}^{(\nu)} - d_N^{-1}C^2\,\xi_N)\right] + n\,\Delta_N^{(\nu)}\\
&= D_{N1}^{(\nu)} + n\,\Delta_N^{(\nu)}\ \text{(say)},\quad \nu = 1, 2.
\end{aligned}
\tag{18}
$$

Using the embedding theorem of Komlós et al. (1975) for empirical process $U_{Ni}(x)$, we find that there are sequences of independent Brownian bridges $\mathcal{B}_{Ni},\ i = 1, \ldots, n$, $N = 1, 2, \ldots,$ such that

$$
\begin{aligned}
D_{N1}^{(\nu)} &= N^{-1/2}\sum_{i=1}^{n}\left[\mathcal{B}_{Ni}(F(a_{N,m}^{(\nu)} + d_N^{-1}C^2\,\xi_N)) - \mathcal{B}_{Ni}(F(a_{N,m}^{(\nu)} - d_N^{-1}C^2\,\xi_N))\right]\\
&\quad + O(N^{-1}\ln N),\qquad \text{a.s., } \nu = 1, 2
\end{aligned}
\tag{19}
$$

and the differences of the Brownian bridges in (19) are of the orders $\left(\Delta_N^{(\nu)}\right)^{1/2}$, $\nu = 1, 2$. On the other hand, from (3), (9) and (10) follow approximations for density f of F:

$$
\begin{aligned}
f(a_{N,m}^{(1)}) &\approx m(nN)^{-(1-\delta)\left(1+\frac{1}{m}\right)}L(N)\\
f(a_{N,m}^{(2)}) &\approx m(nN)^{-\left(1+\frac{1}{m}\right)}(\ln N)^{\left(1+\frac{1}{m}\right)(2-\eta)}L(N)
\end{aligned}
\tag{20}
$$

where $a_{N,m} \approx b_N$ means that $a_{N,m}/b_N \to 1$ as $N \to \infty$. Hence, (16) and (17) imply that

$$
\begin{aligned}
\Delta_N^{(1)} &\approx C^2 Km\, d_N^{-1}\,\xi_N\ f(a_{N,m}^{(1)})\ \approx\ C^2 Km\, N^{-\left(1-\delta\frac{m+1}{m}\right)}L(N)\,d_N^{-1},\\
\Delta_N^{(2)} &\approx C^2 Km\, d_N^{-1}\,\xi_N\ f(a_{N,m}^{(2)})\ \approx\ C^2 Km\, N^{-1}(\ln N)^{\frac{m+1}{m}(2-\eta)}L(N)\,d_N^{-1}
\end{aligned}
\tag{21}
$$

and, regarding (6), we obtain

$$
\begin{aligned}
D_{N1}^{(1)} &= O_p\left(N^{-1+\frac{\delta}{2}\left(1+\frac{1}{m}\right)}L^{\frac{1}{2}}(N)\,d_N^{-\frac{1}{2}}\right),\\
D_{N1}^{(2)} &= O_p\left(N^{-1}(\ln N)^{(1-\frac{\eta}{2})(1+\frac{1}{m})}L^{\frac{1}{2}}(N)\,d_N^{-\frac{1}{2}}\right).
\end{aligned}
\tag{22}
$$

Finally, (15), (18), (21) and (22) together would lead to the orders for $D_N^{(\nu)}$:

$$\begin{aligned} D_N^{(1)} &= O_p\left(N^{-1-\frac{\delta}{2}\left(1+\frac{1}{m}\right)} L^{\frac{1}{2}}(N)\, d_N^{-\frac{1}{2}}\right) + O\left(N^{-1+\delta\left(1+\frac{1}{m}\right)} L(N)\, d_N^{-1}\right), \\ D_N^{(2)} &= O_p\left(N^{-1}(\ln N)^{\left(1-\frac{\eta}{2}\right)\left(1+\frac{1}{m}\right)} L^{\frac{1}{2}}(N)\, d_N^{-\frac{1}{2}}\right) \\ &\quad + O\left(N^{-1}(\ln N)^{(2-\eta)\left(1+\frac{1}{m}\right)} L(N)\, d_N^{-1}\right). \end{aligned} \tag{23}$$

In the next subsection, we shall consider more closely the possible choice of estimator $\widehat{\boldsymbol{\rho}}_N$ and the associated choice of d_N in (12), in order to find the rate of convergence in (11).

2.1 Estimators of autoregression coefficients

The choice of estimator $\widehat{\boldsymbol{\rho}}_N$ heavily depends on our hypothetical value m_0 of the tail index. Generally, we should distinguish two cases for the hypothetical distribution of innovations:

(i) Heavy-tailed distribution satisfying (3) with $0 < m_0 \leq 2$;
(ii) distribution satisfying (3) with $m_0 > 2$.

ad (i): For distributions of the first group we find the linear programming estimator of $\boldsymbol{\rho}$, proposed by Feigin and Resnick (1994), as the most convenient.
It is defined as

$$\begin{aligned} \widehat{\boldsymbol{\rho}}_{LP} &:= \underset{u \in \mathcal{D}_N}{\operatorname{argmax}} \sum_{j=1}^{p} u_j, \\ \mathcal{D}_N &:= \left\{ \mathbf{u} \in \mathbb{R}^p : X_t \geq \sum_{j=1}^{p} u_j X_{t-j},\ t = 1, \ldots, nN \right\}. \end{aligned} \tag{24}$$

Feigin and Resnick considered a stationary autoregressive process with positive innovations, whose distribution satisfies the conditions

$$\sum_{j=1}^{p} \rho_j < \infty, \tag{25}$$

$$\lim_{s\to\infty} \frac{1-F(sx)}{1-F(s)} = x^{-\alpha}, \quad \text{for all } x > 0 \text{ and for some } \alpha > 0, \tag{26}$$

$$I\!E(\varepsilon_t^{-\beta}) = \int_0^\infty u^{-\beta} \mathrm{d}F(u) < \infty, \quad \text{for some } \beta > \alpha. \tag{27}$$

Distributions of type (3) satisfy (26) with $\alpha = m$. As examples of distributions satisfying condition (27), Feigin and Resnick mention positive stable densities; (27)

is satisfied, e.g. by the inverse normal distribution, the Fréchet distribution and the Pareto distribution with

$$1 - F(x) = \begin{cases} x^{-m} & \text{for } x > 1 \\ 1 & \text{otherwise.} \end{cases}$$

Under these conditions, Feigin and Resnick proved that $\widehat{\boldsymbol{\rho}}_{LP}$ satisfies (12) with

$$d_N := F^{-1}\left(1 - \frac{1}{nN}\right) = (nN)^{\frac{1}{m}} L(N) = O(N^{\frac{1}{m}} L(N)). \tag{28}$$

If the autoregressive process satisfies the Feigin and Resnick conditions and the residuals in (6) are calculated with respect to $\widehat{\boldsymbol{\rho}}_{LP}$, then the bounds in (22) are of the respective orders

$$\begin{aligned} D_N^{(1)} &= O_p\left(N^{-1+\frac{\delta}{2}-\frac{1-\delta}{2m}} L^{1/2}(N)\right) + O\left(N^{-1+\delta-\frac{1-\delta}{m}} L(N)\right), \\ D_N^{(2)} &= O_p\left(N^{-\left(1+\frac{1}{2m}\right)} (\ln N)^{\left(1-\frac{\eta}{2}\right)\left(1+\frac{1}{m}\right)} L^{1/2}(N)\right) \\ &\quad + O\left(N^{-\left(1+\frac{1}{m}\right)} (\ln N)^{(2-\eta)\left(1+\frac{1}{m}\right)} L(N)\right). \end{aligned}$$

Hence, as $N \to \infty$,

$$\begin{aligned} &N^{1-\frac{\delta}{2}} D_N^{(1)} = o_p(1), \quad \forall\, \delta < 2/(m+2), \\ &N (\ln N)^{-1+\frac{\eta}{2}} D_N^{(2)} = o_p(1). \end{aligned} \tag{29}$$

ad (ii): If F belongs to the second group, then we need not to restrict ourselves to positive innovations. The most convenient estimators of $\boldsymbol{\rho}$ for distributions with $m_0 > 2$ are either GM-estimators or GR-estimators; we refer to Koul (2002) for their description and profound study. These estimators are $\sqrt{N}$-consistent, and cover the popular Huber estimator; the distribution can be extended over all real line and (11) applies for $\nu = 1, 2$.

3 Construction of the tests

Our procedures are based on the dataset of observations $X_{1-p}, X_{2-p}, \ldots, X_0$, $X_1, \ldots X_{nN}$. If we want to test $\mathbf{H}_0$ with $0 < m_0 \leq 2$, then we calculate the residuals with respect to the linear programming estimator $\widehat{\boldsymbol{\rho}}_{LP}$, defined in (24). If we want to test $\mathbf{H}_1$ with $m_0 > 2$, then we calculate the residuals with respect to GM- or GR-estimators (see Koul 2002). Let $a_{N,m_0}^{(1)},\ a_{N,m_0}^{(2)}$ be as in (9), (10), respectively; calculate $\widehat{F}_N^*(a_{N,m_0}^{(\nu)}),\ \nu = 1, 2$, as per the definition (8).

We propose two tests for both $\mathbf{H}_0$ against $\mathbf{K}_0$ and $\mathbf{H}_1$ against $\mathbf{K}_1$, respectively, corresponding to $a_{N,m_0}^{(1)},\ a_{N,m_0}^{(2)}$, respectively. The first test is based on the same threshold $a^{(1)}_{N,m_0}$ as the test for i.i.d. observations proposed by Jurečková, and Picek (2001). The higher value $a^{(2)}_{N,m_0}$ in the second test is likely to reduce the probability of error of the first kind, though it leads to a slower convergence to the asymptotic null distribution.

Test (1): The test of $\mathbf{H}_0$ against $\mathbf{K}_0$ and of $\mathbf{H}_1$ against $\mathbf{K}_1$, respectively, rejects the hypothesis provided

$$\begin{aligned}
&\textbf{either } 1-\widehat{F}^*_N(a^{(1)}_{N,m_0})=0,\\
&\textbf{or } \ 1-\widehat{F}^*_N(a^{(1)}_{N,m_0})>0\\
&\text{and } \ N^{\delta/2}\Big[-\ln(1-\widehat{F}^*_N(a^{(1)}_{N,m_0}))-(1-\delta)\ln N\Big]\geq \Phi^{-1}(1-\alpha),
\end{aligned} \tag{30}$$

where Φ is the standard normal distribution function and $\alpha\in(0,1)$ is the asymptotic significance level.

Test (2): The test of $\mathbf{H}_0$ against $\mathbf{K}_0$ and of $\mathbf{H}_1$ against $\mathbf{K}_1$, respectively, rejects the hypothesis provided

$$\begin{aligned}
&\textbf{either } 1-\widehat{F}^*_N(a^{(2)}_{N,m_0})=0,\\
&\textbf{or } 1-\widehat{F}^*_N(a^{(2)}_{N,m_0})>0, \quad \text{and}\\
&(\ln N)^{1-\frac{\eta}{2}}\Big[-\ln(1-\widehat{F}^*_N(a^{(2)}_{N,m_0}))-\ln N+(2-\eta)\ln\ln N\Big]\geq \Phi^{-1}(1-\alpha).
\end{aligned} \tag{31}$$

The following theorems show that the test criteria (30) and (31) have asymptotically standard normal distributions under the exact Pareto tail corresponding to $1-F(x)=x^{-m_0}$, for $x>x_0$.

Theorem 3.1 *Consider the stationary autoregressive process* (1).
(I) *Assume that the process* (1) *satisfies the condition* (25) *and that the innovation distribution function F is absolutely continuous and of type* (3) *with tail index m_0, $0<m_0\leq 2$, concentrated on the positive half-axis and strictly increasing on the set $\{x:\ F(x)>0\}$. Let $\widehat{F}^*_N(a^{(1)}_{N,m_0})$ be the empirical distribution function of extreme residuals of N segments of length n, defined in* (8)*, where the residuals are calculated with respect to $\widehat{\boldsymbol{\rho}}_{LP}$ defined in* (24). *Then, the following hold.*

(i) *For every distribution P satisfying $\mathbf{H}_0$,*

$$\lim_{N\to\infty} P\left(0<\widehat{F}^*_N(a^{(1)}_{N,m_0})<1\right)=1. \tag{32}$$

(ii) *If $1-F(x)=x^{-m_0}$, $\forall x>x_0$, then*

$$\lim_{N\to\infty} P\left\{N^{\delta/2}\Big[-\ln(1-\widehat{F}^*_N(a^{(1)}_{N,m_0}))-(1-\delta)\ln N\Big]\leq x\right\}=\Phi(x),\ \ \forall x\in\mathbb{R}. \tag{33}$$

Hence,

$$\lim_{N\to\infty} P\left\{N^{\delta/2}\left[-\ln(1-\widehat{F}_N^*(a_{N,m_0}^{(1)}))-(1-\delta)\ln N\right]\geq \Phi^{-1}(1-\alpha)\right\}=\alpha. \tag{34}$$

(iii) *The test is asymptotically unbiased for the family of F satisfying* (3) *with $m < m_0$ and for the family of F satisfying* (3) *with $m = m_0$ and with $\underline{\lim}_{x\to\infty} L(x) \geq 1$. More precisely, then*

$$\overline{\lim}_{N\to\infty} P\left\{N^{\delta/2}\left[-\ln(1-\widehat{F}_N^*(a_{N,m_0}^{(1)}))-(1-\delta)\ln N\right]\geq \Phi^{-1}(1-\alpha)\right\}\leq\alpha.$$

(II) *Let F be of type* (3) *with tail index m_0, $m_0 > 2$, with a continuous and positive density on $\mathbb{R}$. Let $\widehat{F}_N^*(a_{N,m_0}^{(1)})$ be the empirical distribution function of extreme residuals of N segments of length n, defined in* (8)*, where the residuals are calculated with respect to a $\sqrt{N}$-consistent estimator of $\boldsymbol{\rho}$. Then the propositions (i)–(iii) remain true.*

Proof We shall prove the part (I) of the Theorem; part (II) is quite analogous.
(i) Convergence (32) was proved in Jurečková, and Picek (2001) for $F_N^*(a_{N,m_0}^{(1)})$ instead of $\widehat{F}_N^*(a_{N,m_0}^{(1)})$. Because $d_N(\widehat{\boldsymbol{\rho}}_N-\boldsymbol{\rho})$ takes on positive and negative values, not all residuals are less than $a_{N,m_0}^{(1)}$ or all greater than $a_{N,m_0}^{(1)}$, with probability tending to 1.
(ii) It follows from (29) that

$$\begin{aligned}
&N^{\frac{\delta}{2}}\left[-\ln(1-\widehat{F}_N^*(a_{N,m_0}^{(1)}))+\ln(1-F_N^*(a_{N,m_0}^{(1)}))\right]\\
&=N^{\frac{\delta}{2}}\left\{-\ln\left[\frac{1-\widehat{F}_N^*(a_{N,m_0}^{(1)})}{1-F_N^*(a_{N,m_0}^{(1)})}-1+1\right]\right\}\\
&=N^{\frac{\delta}{2}}\left(1-F_N^*(a_{N,m_0}^{(1)})\right)^{-1}\left[\widehat{F}_N^*(a_{N,m_0}^{(1)})-F_N^*(a_{N,m_0}^{(1)})\right]+O_p\left(N^{-\frac{\delta}{2}}\right)\\
&=N^{\frac{\delta}{2}}\frac{1-F^*(a_{N,m_0}^{(1)})}{1-F_N^*(a_{N,m_0}^{(1)})}\left(1-F^*(a_{N,m_0}^{(1)})\right)^{-1}o_p\left(N^{-1+\frac{\delta}{2}}\right)+O_p\left(N^{-\frac{\delta}{2}}\right)\\
&=o_p\left(N^{\frac{\delta}{2}+1-\delta-1+\frac{\delta}{2}}\right)=o_p(1),
\end{aligned} \tag{35}$$

while by Theorem 2.1 in Jurečková, and Picek (2001),

$$\lim_{N\to\infty} P\left\{N^{\frac{\delta}{2}}\left[-\ln(1-F_N^*(a_{N,m_0}^{(1)}))-(1-\delta)\ln N\right]\leq x\right\}=\Phi(x),\quad x\in\mathbb{R}. \tag{36}$$

Hence (33) follows from (35) and from (36).
(iii) If F satisfies (3) either with $m = m_0$ and with $\underline{\lim}_{x\to\infty} L(x) \geq 1$, or with $m < m_0$, then F is ultimately stochastically larger than the distribution with exact Pareto tail; hence the limiting probability of the event in (34) should not exceed α. If $m > m_0$, then it is shown in Sect. 4 that test rejects the hypothesis with probability tending to one, hence $> \alpha$ for $N > N_0$, and the test is asymptotically unbiased. The consistency of the test is studied in Sect. 4, and the numerical illustration is given in Sect. 5. □

The following theorem describes the properties of the second test with the threshold $a^{(2)}_{N,m_0}$.

Theorem 3.2 (I) *Let $\widehat{F}^*_N(a^{(2)}_{N,m_0})$ be the empirical distribution function of extreme residuals of N segments of length n, defined in* (8)*, where the residuals are calculated with respect to $\widehat{\boldsymbol{\rho}}_{LP}$ defined in* (24)*. Then, under the conditions of Part* (I) *of Theorem* 3.1*, the following hold.*

(i) *For every distribution P satisfying $\mathbf{H}_0$,*

$$\lim_{N\to\infty} P\left(0 < \widehat{F}^*_N(a^{(2)}_{N,m_0}) < 1\right) = 1. \tag{37}$$

(ii) *If $1 - F(x) = x^{-m_0}$ for $\forall x > x_0$, then for $\forall x \in \mathbb{R}$,*

$$\lim_{N\to\infty} P\left\{(\ln N)^{1-\frac{\eta}{2}}\left[-\ln(1-\widehat{F}^*_N(a^{(2)}_{N,m_0})) - \ln N + (2-\eta)\ln\ln N\right] \le x\right\}$$
$$= \Phi(x), \tag{38}$$

hence

$$\lim_{N\to\infty} P\left\{(\ln N)^{1-\frac{\eta}{2}}\left[-\ln(1-\widehat{F}^*_N(a^{(2)}_{N,m_0})) - \ln N + (2-\eta)\ln\ln N\right]\right.$$
$$\left.\ge \Phi^{-1}(1-\alpha)\right\} = \alpha.$$

(iii) *The test is asymptotically unbiased for the family of F satisfying* (3) *either with $m < m_0$ or with $m = m_0$ and with $\underline{\lim}_{x\to\infty} L(x) \ge 1$. More precisely, then*

$$\overline{\lim}_{N\to\infty} P\left\{N^{1-\frac{\eta}{2}}\left[-\ln(1-\widehat{F}^*_N(a^{(2)}_{N,m_0})) - \ln N + (2-\eta)\ln\ln N\right]\right.$$
$$\left.\ge \Phi^{-1}(1-\alpha)\right\} \le \alpha.$$

(II) *Let F be of type* (3) *with tail index $m_0,\ m_0 > 2$, with a continuous and positive density on $\mathbb{R}$. Let $\widehat{F}^*_N(a^{(2)}_{N,m_0})$ be the empirical distribution function of extreme residuals of N segments of length n, defined in* (8)*, where the residuals are calculated with respect to a $\sqrt{N}$-consistent estimator of $\boldsymbol{\rho}$. Then the propositions* (i)–(iii) *remain true.*

Proof To prove (i), we should first prove the convergence (37) for $F^*_N(a^{(2)}_{N,m_0})$: Denote

$$\varepsilon_{(N)} = \max_{1\le t\le N}\{\varepsilon^t_{(n)}\}$$

where $\varepsilon^t_{(n)}$ is defined in (7), $t = 1,\dots,N$. If F is of type (3) with $m = m_0$ and $L(x) \ge 1$ for $x > x_0$, then

$$P\left(\varepsilon_{(N)} < a^{(2)}_{N,m_0}\right) = (1-(nN)^{-1}(\ln N)^{2-\eta}L(N))^{nN} \approx \exp\left\{-(\ln N)^{(2-\eta)L(N)}\right\} \to 0,$$

as $N \to \infty$. Similarly, if F is ultimately heavier than Pareto with $m < m_0$, then

$$P\left(\varepsilon_{(N)} < a^{(2)}_{N,m_0}\right) = \left[1 - \left(nN(\ln N)^{2-\eta}\right)^{-\frac{m}{m_0}} L(N)\right]^{Nn} \to 0, \quad \text{as } N \to \infty.$$

The fact that $F^*_N(a_{N,m_0}) > 0$ with probability tending to 1 for the whole hypothesis is obvious. Hence, $P\left(0 < F^*_N(a^{(2)}_{N,m_0}) < 1\right) \to 1$, and finally we obtain (38) with the aid of (12).

Analogously as in the proof of Theorem 3.1, we conclude that with probability tending to 1 not all residuals $\widehat{\varepsilon}^t_{(n)}$ are less than $a^{(2)}_{N,m_0}$.
(ii) By the embedding theorem of Komlós et al. (1975), there is a sequence of independent Brownian bridges $\mathcal{B}_N$, $N = 1, 2, \ldots,$ such that

$$\begin{aligned} F^*_N(a^{(2)}_{N,m_0}) - F^*(a^{(2)}_{N,m_0}) &= N^{-\frac{1}{2}} \mathcal{B}_N(1 - F^*(a^{(2)}_{N,m_0})) + O\left(N^{-1}\ln N\right) \\ &= O_p\left(N^{-1}(\ln N)^{1-\frac{\eta}{2}} L^{\frac{1}{2}}(N)\right) + O\left(N^{-1}\ln N\right). \end{aligned} \tag{39}$$

If $m = m_0$ and $L(x) \geq 1$ for $x > x_0$, then it follows from (29) and from (39) that,

$$\begin{aligned} &(\ln N)^{1-\frac{\eta}{2}}\left[-\ln(1 - \widehat{F}^*_N(a^{(2)}_{N,m_0})) + \ln(1 - F^*_N(a^{(2)}_{N,m_0}))\right] \\ &= (\ln N)^{1-\frac{\eta}{2}}\left\{-\ln\left[\frac{1 - \widehat{F}^*_N(a^{(2)}_{N,m_0})}{1 - F^*_N(a^{(2)}_{N,m_0})} - 1 + 1\right]\right\} \\ &= (\ln N)^{1-\frac{\eta}{2}}(1 - F^*_N(a^{(2)}_{N,m_0}))^{-1}\left[\widehat{F}^*_N(a^{(2)}_{N,m_0}) - F^*_N(a^{(2)}_{N,m_0})\right] \\ &\quad + O_p\left((\ln N)^{-1+\frac{\eta}{2}}\right) \\ &= (\ln N)^{1-\frac{\eta}{2}} \frac{1 - F^*(a^{(2)}_{N,m_0})}{1 - F^*_N(a^{(2)}_{N,m_0})}\left(1 - F^*(a^{(2)}_{N,m_0})\right)^{-1} o_p\left(N^{-1}(\ln N)^{1-\frac{\eta}{2}}\right) \\ &\quad + O_p\left((\ln N)^{-1+\frac{\eta}{2}}\right) \\ &= o_p\left((\ln N)^{1-\frac{\eta}{2}-2+\eta+1-\frac{\eta}{2}}(L(N))^{-1}\right) + O_p\left((\ln N)^{-1+\frac{\eta}{2}}\right) = o_p(1). \end{aligned}$$

Moreover,

$$\begin{aligned} &(\ln N)^{1-\frac{\eta}{2}}\left[-\ln(1 - F^*_N(a^{(2)}_{N,m_0})) - \ln N + (2-\eta)\ln\ln N\right] \\ &= (\ln N)^{1-\frac{\eta}{2}}\left[-\ln(1 - F^*_N(a^{(2)}_{N,m_0})) + \ln(1 - F^*(a^{(2)}_{N,m_0}))\right] \\ &= (\ln N)^{1-\frac{\eta}{2}}\left\{-\ln\left[\frac{1 - F^*_N(a^{(2)}_{N,m_0})}{1 - F^*(a^{(2)}_{N,m_0})} - 1 + 1\right]\right\} \\ &= (\ln N)^{1-\frac{\eta}{2}} \frac{F^*_N(a^{(2)}_{N,m_0}) - F^*(a^{(2)}_{N,m_0})}{1 - F^*(a^{(2)}_{N,m_0})} + O_p\left((\ln N)^{-1+\eta}\right) \end{aligned} \tag{40}$$

If F has exactly the Pareto tail with $m = m_0$, then

$$(\ln N)^{1-\frac{\eta}{2}}(1 - F^*(a^{(2)}_{N,m_0}))^{-1} N^{-\frac{1}{2}} \mathcal{B}_N(1 - F^*(a^{(2)}_{N,m_0})) \to_d \mathcal{N}(0, 1) \quad \text{as } N \to \infty \tag{41}$$

while the left-hand side of (40) is ultimately stochastically smaller than that of (41) if $m = m_0$ and $L(x) \geq 1$ for $x > x_0$. The rest of the proof is analogous to the proof of Theorem 2.1 in Jurečková, and Picek (2001). □

4 Consistency considerations

Let $1 - F(x) = x^{-m} L(x)$, for $x \geq x_0 > 0$, with $m > m_0$, let $F^*(x) = F^N(x)$ be the joint distribution function of the maximal innovations $\{\varepsilon^t_{(n)},\ t = 1, \ldots, N\}$, and F^*_N their empirical distribution function. Then

$$1 - F^*(a^{(1)}_{N,m_0}) \leq n \left(a^{(1)}_{N,m_0}\right)^{-m} L(N) = N^{-(1-\delta)\frac{m}{m_0}} L_1(N)$$

and $N(1 - F^*_N(a^{(1)}_{N,m_0}))$ has the binomial distribution $\mathcal{B}(N, p_N)$ with

$$p_N = 1 - F^*(a^{(1)}_{N,m_0}) \leq N^{-(1-\delta)\frac{m}{m_0}} L_1(N).$$

Hence, $E[N^{1-\delta}(1 - F^*_N(a^{(1)}_{N,m_0}))] = O\left(N^{-(1-\delta)\left(\frac{m}{m_0}-1\right)} L_1(N)\right)$ and

$$\begin{aligned} &N^{\frac{\delta}{2}}\left(-\ln(1 - F^*_N(a^{(1)}_{N,m_0})) - (1-\delta)\ln N\right) \\ &\quad = O_p\left(N^{\frac{\delta}{2}} \ln N(1-\delta)\left(\frac{m}{m_0} - 1\right) L_2(N)\right) \to \infty. \end{aligned} \tag{42}$$

Thus, for the white noise sequence with the null autoregression we would reject the hypotheses $\mathbf{H}_0,\ \mathbf{H}_1$ with probability tending to 1. An analogous statement holds for $1 - F^*_N(\lambda a^{(1)}_{N,m_0})$ with any fixed $\lambda,\ 0 < \lambda < 1$.

Let now F be concentrated on the positive half-axis and let $\widehat{F}^*_N$ be the empirical distribution function of the maximal residuals $\{\widehat{\varepsilon}^t_{(n)},\ t = 1, \ldots, N\}$ with respect to linear programming estimator $\widehat{\boldsymbol{\rho}}_{LP}$ defined in (24). Then it follows from (24) to (28) and from (5) that

$$\begin{aligned} &\varepsilon_t - \mathbf{Y}'_{t-1}(\widehat{\boldsymbol{\rho}}_{LP} - \boldsymbol{\rho}) \geq 0, \quad t = 1, \ldots, nN, \\ &\max_{1 \leq t \leq nN} |\mathbf{Y}'_t(\widehat{\boldsymbol{\rho}}_{LP} - \boldsymbol{\rho})| = O_p\left(L(N)\right) = o_p(a_{N,m_0}), \\ &0 \leq \widehat{\varepsilon}^t_{(n)} = \max_{1 \leq i \leq n} \widehat{\varepsilon}_{(t-1)n-p+i} \leq \max_{1 \leq i \leq n} \varepsilon_{(t-1)n-p+i} + O_p\left(L(N)\right),\ t = 1, 2, \cdots, N. \end{aligned}$$

This in turn implies that, given an $0 < \eta < 1$, there exists N_0 such that for $N > N_0$, under $P_m,\ m > m_0$,

$$P\left\{1 - \widehat{F}_N^*(a_{N,m_0}^{(1)}) \le 1 - F_N^*(\lambda a_{N,m_0}^{(1)})\right\} \ge 1 - \eta,$$

for any fixed $\lambda,\ 0 < \lambda < 1$, and this together with (42) and the following remark implies that we reject the hypothesis $\mathbf{H}_0$ with probability at least $1-\eta$ for $N > N_0$.

Analogous consideration applies to the hypothesis $\mathbf{H}_1$ with hypothetical $m_0 > 2$, when the innovation distribution, possibly extended on the whole $\mathbb{R}^1$, satisfies $1 - F(x) = x^{-m}L(x)$ with $m > m_0$ and with $L(\cdot)$ slowly varying at infinity; then $\boldsymbol{\rho}$ is estimated either by GM- or by GR-estimator.

5 Simulation study

We study the performance of the test on the Pareto tail index of the innovation distribution in the autoregressive model on the following three simulated time series:

(A) $X_t = 0.05X_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, Nn,$
(B) $X_t = 0.9X_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, Nn,$
(C) $X_t = 0.6X_{t-1} - 0.3X_{t-2} + 0.2X_{t-3} + \varepsilon_t, \quad t = 1, 2, \ldots, Nn,$

with the following innovation distributions:

Pareto $\quad F(x) = 1 - \left(\frac{1}{x}\right)^m, \quad x \ge 1.$

Inverse normal $\quad F(x) = \begin{cases} 2\left(1 - \Phi\left(\frac{1}{\sqrt{x}}\right)\right) & x > 0 \\ 0 & x \le 0 \end{cases}$

Student $\quad f(x) = \frac{1}{\sqrt{m}\mathrm{B}\left(\frac{1}{2}, \frac{m}{2}\right)}\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}, \quad x \in \mathbb{R}.$

For each of these cases, the time series were simulated of the lengths $nN = 200$ and $1,000$. (The initial values for the time series were obtained as the last values of auxiliary simulated time series of length 500 with the same autoregression coefficients and innovation distribution and initial values 0.)
The computation procedure for each of the above innovation distributions and time series was as follows:

(1) the autoregressive time series $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{2n}, \ldots, X_{Nn}$ was generated;
(2) ρ was estimated by $\hat{\rho}$ (either Feigin and Resnick or Huber M-estimators);
(3) residuals $\hat{\varepsilon}_t := X_t - \hat{\rho}_N' \mathbf{Y}_{t-1},\ t = 1-p, 2-p, \ldots, nN$ were computed;
(4) the maxima $\widehat{\varepsilon}_n^{(1)}, \ldots, \widehat{\varepsilon}_n^{(N)}$ of the segments were found and the corresponding empirical distribution function $\hat{F}_N^*$ calculated;
(5) a decision was made about $\mathbf{H}_0$ or $\mathbf{H}_1$, respectively, based on $F_N^*(a_{N,m_0})$, with $a_{N,m_0}^{(1)} = \left(nN^{1-\delta}\right)^{\frac{1}{m_0}}$ and $a_{N,m}^{(2)} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$, respectively.
(6) The step (5) was repeated for various values m_0, δ, η;
(7) the steps (1)–(6) were repeated $1,000$ times.

Table 1 Numbers of rejections of the null hypothesis among 1,000 tests at level $\alpha = 0.05$ for $a_{N,m}^{(1)} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$ and some selected values of m_0; $N = 50$, $n = 4$, $\delta = 0.1$

Distribution of white noise	Time series	$m_0 = 0.25$	$m_0 = 0.4$	$m_0 = 0.5$	$m_0 = 0.6$	$m_0 = 0.75$
Pareto $m = 0.5$	A	987	675	244	38	0
	B	987	675	244	38	0
	C	987	675	244	38	0
Inverse normal	A	990	736	320	79	1
	B	990	736	320	79	1
	C	990	736	320	79	1
		$m_0 = 0.5$	$m_0 = 0.8$	$m_0 = 0.9$	$m_0 = 1.0$	$m_0 = 1.2$
Pareto $m = 1$	A	992	674	439	246	35
	B	992	674	441	246	36
	C	992	674	439	245	36
		$m_0 = 2.0$	$m_0 = 2.5$	$m_0 = 2.75$	$m_0 = 3.0$	$m_0 = 3.5$
Student $m = 3$	A	867	569	402	255	66
	B	865	565	398	254	63
	C	866	564	403	251	69

Table 1 gives numbers of rejections of $\mathbf{H}_0$ or $\mathbf{H}_1$, respectively, among 1,000 tests at level $\alpha = 0.05$ for some selected values m_0 and under $a_{N,m}^{(1)} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$; $\delta = 0.1,\ n = 4, N = 50$. Similarly, Table 2 gives numbers of rejections under $a_{N,m}^{(2)} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $\eta = 0.1,\ n = 4, N = 50$.

Remark 1 Notice that the Pareto and Student distributions with tail index m satisfy $\mathbf{H}_0$ (or $\mathbf{H}_1$) for $m_0 = m + \varepsilon,\ \forall \varepsilon > 0$; the inverse normal distribution satisfies $\mathbf{H}_0$ (or $\mathbf{H}_1$) for $m_0 = 0.5 + \varepsilon,\ \forall \varepsilon > 0$.

Remark 2 We see that the performance of the test practically depends only on the white noise and not on the structure of AR series (the time series A, B, C were generated with the same innovations values under a given distribution).

The frequencies of rejection of $\mathbf{H}_0$ or $\mathbf{H}_1$ under the Pareto distribution are plotted against m_0 in Figs. 1 and 2. Because of the small difference in behavior between the series A, B, C, the illustrations are made only for B.

The following Tables 3 and 4 and Figs. 3, 4 illustrate the influence of the choice of δ on the frequency of rejections of the null hypothesis under some fixed values of m_0. The shape of the graph under m_0 close the true m is rather typical. The situation is similar for the choice of η.

Table 2 Numbers of rejections of the null hypothesis among 1,000 cases at level $\alpha = 0.05$ for $a^{(2)}_{N,m} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $N = 50, n = 4,\ \eta = 0.1$

Distribution of white noise	Time series	$m_0 = 0.3$	$m_0 = 0.4$	$m_0 = 0.5$	$m_0 = 0.52$	$m_0 = 0.6$
Pareto $m = 0.5$	A	1000	995	83	16	0
	B	1000	995	83	16	0
	C	1000	995	83	16	0
Inverse normal	A	1000	1000	363	158	1
	B	1000	1000	363	158	1
	C	1000	1000	363	158	1
		$m_0 = 0.8$	$m_0 = 0.9$	$m_0 = 1.0$	$m_0 = 1.02$	$m_0 = 1.1$
Pareto $m = 1$	A	995	645	84	37	1
	B	995	645	84	38	1
	C	995	645	84	38	1
		$m_0 = 2.5$	$m_0 = 2.8$	$m_0 = 3.00$	$m_0 = 3.05$	$m_0 = 3.5$
Student $m = 3$	A	982	684	283	186	7
	B	983	680	282	193	4
	C	982	685	281	187	5

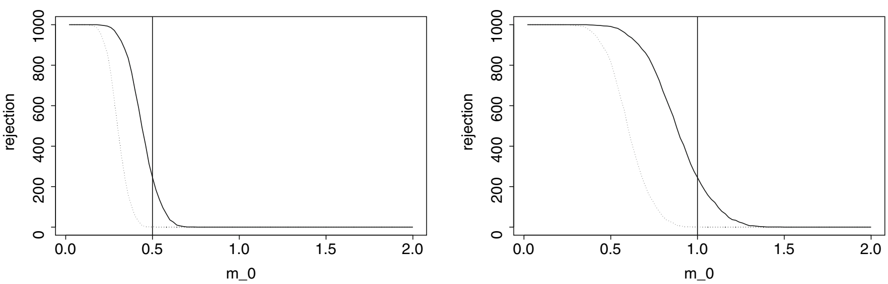
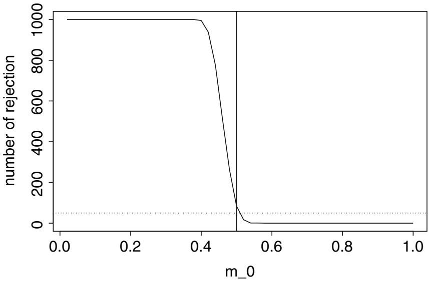


Fig. 1 Number of rejections of H_0 ($\alpha = 0.05$) plotted against m_0 for $X_t = 0.9X_{t-1} + \varepsilon_t$ and $a^{(1)}_{N,m} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1, \dots, nN$ have the Pareto distribution with $m = 0.5$ (*left*) and $m = 1$ (*right*); $N = 50, n = 4,\ \delta = 0.1$ (*solid*), $\delta = 0.5$ (*dotted*)

6 Application to the Czech daily maximum temperatures

The tests described above are applied to a 40-years dataset of daily maximum temperatures measured at three meteorological stations in Czech Republic, over the period 1961–2000. The names and coordinates of the three stations are as follows:

Praha-Ruzyně: $50°06'N,\ 14°15'E$, altitude 364 m above sea level;
Liberec: $50°46'N,\ 15°01'E$, altitude 398 m above sea level;
Brno-Tuřany: $49°09'N,\ 16°42'E$, altitude 241 m above sea level.



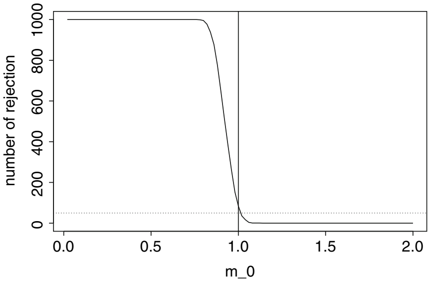


Fig. 2 Number of rejections of H_0 ($\alpha = 0.05$) plotted against m_0 for $X_t = 0.9X_{t-1} + \varepsilon_t$ and $a^{(2)}_{N,m} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1,\ldots,nN$ have the Pareto distribution with $m = 0.5$ (*left*) and $m = 1$ (*right*); $N = 200, n = 5,\ \eta = 0.1$

Table 3 Numbers of rejections of the null hypothesis among 1,000 cases for various δ under some m_0; $\alpha = 0.05,\ N = 50,\ n = 4$

Distribution of white noise		δ						
	m_0	0.01	0.1	0.2	0.3	0.4	0.5	0.6
Pareto $m = 1$	0.5	995	987	977	1000	1000	1000	1000
	0.9	551	422	233	401	404	272	281
	1.0	339	211	114	176	132	35	55
Pareto $m = 3$	2.5	786	694	589	812	894	925	997
	2.9	591	512	390	616	727	787	966
	3.0	553	469	333	566	681	747	943
Student $m = 5$	3.5	815	700	535	700	739	686	863
	4.0	95	48	15	30	33	29	87
	5.0	27	6	0	6	11	6	35

Table 4 Numbers of rejections of the null hypothesis among 1,000 cases for various δ under some m_0; $\alpha = 0.05,\ N = 200,\ n = 5$

Distribution of white noise		δ						
	m_0	0.01	0.1	0.2	0.3	0.4	0.5	0.6
Pareto $m = 1$	0.5	1000	998	1000	1000	1000	1000	1000
	0.9	621	440	551	504	482	630	718
	1.0	349	251	240	126	73	105	117
Pareto $m = 3$	2.5	828	706	877	903	957	1000	1000
	2.9	567	408	548	574	688	924	998
	3.0	483	327	459	487	596	879	995
Student $m = 5$	3.5	921	828	931	928	928	982	992
	4.0	56	11	3	2	0	0	0
	5.0	6	1	0	0	0	0	0

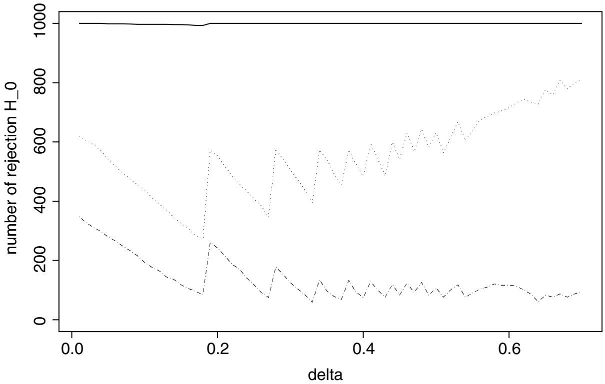


Fig. 3 Number of rejections of H_0 plotted against δ; $\alpha = 0.05,\ n = 5,\ N = 200$ and $X_t = 0.9X_{t-1} + \varepsilon_t$; $\varepsilon_t,\ t = 1, \ldots, 1,000$ have Pareto distribution ($m = 1$), $m_0 = 0.5$ (*solid*), $m_0 = 0.9$ (*dotted*), $m_0 = 1.0$ (*dashed*)

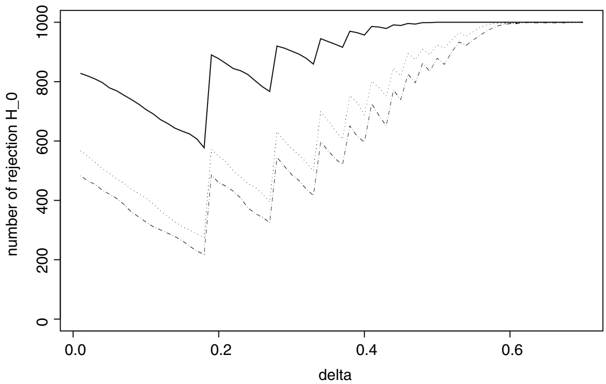


Fig. 4 Number of rejections of H_0 plotted against δ; $\alpha = 0.05,\ n = 5,\ N = 200$ and $X_t = 0.9X_{t-1} + \varepsilon_t$; $\varepsilon_t,\ t = 1, \ldots, 1,000$ have Pareto distribution ($m = 3$), $m_0 = 2.5$ (*solid*), $m_0 = 2.9$ (*dotted*), $m_0 = 3.0$ (*dashed*)

The maximum temperatures were centered and deseasonalized by subtracting the average maximum temperature computed over the 40 years. The residuals then were modeled as autoregressive series of order $p = 1$ (see Hallin et al. 1997).

Table 5 gives results of testing for all three time series for some selected values m_0 and under $a_{N,m}^{(1)} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$; $\delta = 0.1,\ n = 5$. Similarly, Table 6 gives the conclusions under $a_{N,m}^{(2)} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $\eta = 0.1,\ n = 5$.

Hence, describing the summer temperatures residuals as an autoregressive series, we could work with a heavy-tailed distribution. The results of tests indicate any small influence of the location of the station. The tests can be also used to estimate the tail index in the Hodges-Lehmann manner (see Jurečková, and Picek 2004).

Table 5 Rejection (R) and non-rejection (N) of the null hypothesis at level $\alpha = 0.05$ for $a^{(1)}_{N,m} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$ and some selected values of m_0; $n = 5$, $\delta = 0.1$

Time series	$m_0 = 3.2$	$m_0 = 3.3$	$m_0 = 3.5$	$m_0 = 3.6$	$m_0 = 3.7$
Praha	R	N	N	N	N
Liberec	R	R	N	N	N
Brno	R	R	R	R	N

Table 6 Rejection (R) and non-rejection (N) of the null hypothesis at level $\alpha = 0.05$ for $a^{(2)}_{N,m} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; and some selected values of m_0; $n = 5$, $\eta = 0.1$

Time series	$m_0 = 2.5$	$m_0 = 2.6$	$m_0 = 2.65$	$m_0 = 2.7$	$m_0 = 2.75$
Praha	R	R	N	N	N
Liberec	R	R	R	R	N
Brno	R	R	R	R	N

Acknowledgments The authors thank the referee for having read the paper carefully and for making constructive suggestions.

References

Brockwell, P. J., Davis, R. A. (1991). *Time series: theory and methods*. (2nd *ed*.). New York: Springer.

Castillo, E., Calambos, J., Sarabia, J. M. (1989). The selection of the domain of attraction of an extreme value distribution from a data set. In J. Hüsler, R.-D. Reiss (Eds.), *Extreme value theory*—Proceedings of Oberwolfach 1987. *Lecture Notes in Statistics*, vol. 51 (pp. 181–190). New York, Springer.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling of extremal events in insurance and finance.* New York: Springer.

Falk, M. (1995a). On testing the extreme value index via the POT-method. *Annals of Statistics*, *23*, 2013–2035.

Falk, M. (1995b). LAN of extreme order statistics. *Annals of Institute of Statistical Mathematics*, *47*, 693–717.

Feigin, P. D., Resnick, S. I. (1994). Limit distributions for linear programming time series estimators. *Journal of Stochastic Processes & Applications*, *51*, 135–165.

Fialová, A., Jurečková, J., Picek, J. (2004). Estimating Pareto tail index using all observations. *REVSTAT* *2*(1), 75–100.

Galambos, J. (1982). A statistical test for extreme value distribution. In B. V. Gnedenko et al. (Eds.), *Non-parametric statistical inference* (pp. 221–230). Amsterdam: North Holland.

Gomes, M. I. (1989). Comparison of extremal models through statistical choice in multidimensional background. In J. Hüsler, R.-D. Reiss (Eds.), *Extreme value theory*—Proceedings of Oberwolfach 1987. *Lecture Notes in Statistics*, vol. 51 (pp. 191–203). New York: Springer.

Gomes, M. I., Alpuim, M. T. (1986). Inference in a multivariate generalized extreme value model—asymptotic properties of two test statistics. *Scandinavian Journal of Statistics*, *13*, 291–300.

Hallin, M., Jurečková, J., Kalvová, J., Picek, J., Zahaf, T. (1997). Non-parametric tests in AR models with applications to climatic data. *Environmetrics*, *8*, 651–660.

Hasofer, A. M., Wang, Z. (1992). A test for extreme value domain of attraction. *Journal of the American Statistical Association*, *87*, 171–177.

Hosking, J. R. M. (1984). Testing whether the shape parameter is zero in generalized extreme value distribution. *Biometrika*, *71*, 367–374.

Jurečková, J. (1999). Tests of tails based on extreme regression quantiles. *Statistics & Probability Letters*, *49*, 53–61.

Jurečková, J., Picek, J. (2001). A class of tests on the tail index. *Extremes*, *4*(2), 165–183.

Jurečková, J., Picek, J. (2004). Estimates of the tail index based on nonparametric tests. In M. Hubert, G. Pison, A. Struyf, S. Van Aelst (Eds.), *Statistics for industry and technology* (pp. 141–152). Basel: Birkhäuser.

Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *32*, 111–131.

Koul, H. L. (2002). Weighted empirical processes in dynamic nonlinear models. *Lecture Notes in Statistics*, *166*.

Marohn, F. (1994). On testing exponential and Gumbel distribution. In J. Galambos et al. (Eds.), *Extreme value theory* (pp. 159–174). Dordrecht: Kluwer.

Marohn, F. (1998a). Testing the Gumbel hypothesis via the POT-method. *Extremes*, *1*, 191–213.

Marohn, F. (1998b). An adaptive efficient test for Gumbel domain of attraction. *Scandinavian Journal of Statistics*, *25*, 311–324.

van Montfort, M. A. J., Gomes, M. I. (1985). Statistical choice of extremal models for complete and censored data. *Journal of Hydrology*, *77*, 77–87.

van Montfort, M. A. J., Witter, J. V. (1985). Testing exponentially against generalized Pareto distribution. *Journal of Hydrology*, *78*, 305–315.

Picek, J., Jurečková, J. (2001). A class of tests on the tail index using the modified extreme regression quantiles. In J. Antoch, G. Dohnal (Eds.), *ROBUST'2000* (pp. 217–226). Prague: Union of Czech Mathematicians and Physicists.

Stephens, M. A. (1977). Goodness-of-fit for the extreme value distribution. *Biometrika*, *64*, 583–588.

Tiago de Oliveira, J. (1984). Univariate extremes: Statistical choice. In J. Tiago de Oliveira (Ed.), *Statistical extremes and applications* (pp. 91–107). Dodrecht: Reidel.