TESTING THE TAIL INDEX IN AUTOREGRESSIVE MODELS

Jana Jurečková, Charles University in Prague, Czech Republic

Hira L. Koul, Michigan State University, U.S.A.

Jan Picek, Technical University in Liberec, Czech Republic

Introduction

We construct a class of tests on the tail index of the innovation distribution in a stationary linear autoregressive model:

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

for some $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_p)' \in I\!R^p$,

$\varepsilon_t,\ t = 0, \pm 1, \pm 2, \ldots,$ are independent identically distributed (i.i.d.) random variables with a heavy-tailed distribution function F:

$$1 - F(x) = x^{-m} L(x),\ x \in I\!R. \tag{2}$$

Let $m_0 > 0$ be a fixed number. 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.

$\mathbf{H}_0$: F is heavy-tailed, concentrated on the positive half-axis, satisfying

$$x^{m_0}(1 - F(x)) \geq 1, \quad \forall\, x > x_0,$$

for some $x_0 \geq 0$, against the alternatives

$\mathbf{K}_0$: F is heavy-tailed, concentrated on the positive half-axis, and

$$\overline{\lim}_{x \to \infty} x^{m_0}(1 - F(x)) < 1.$$

If F is heavy tailed $(1 - F(x) = x^{-m}L(x)$ - $L(x)$ is a function, slowly varying at infinity), then F satisfies $\mathbf{H}_0$ with $m = m_0$ provided either $m = m_0$ and $L(x) \geq 1$ for $\forall x > x_0$, or $m < m_0$;

if $\overline{\lim}_{x \in R} L(x) < 1$, then F satisfies the hypothesis for $m_0 = m + \varepsilon$, $\forall \varepsilon > 0$, because $L(x)$ increases ultimately slower than any positive power of x.

The proposed tests are based on the extremes of the residual empirical process. Tests on the Pareto index for the i.i.d. model were constructed in Jurečková and Picek (2001):
A class of tests on the tail index. *Extremes*, 4:2, 165–183.

Estimators of autoregressive parameter vector

The choice of estimator $\widehat{\boldsymbol{\rho}}$ 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 $\big(1 - F(x) = x^{-m}L(x)\big)$ with $0 < m_0 \leq 2$;

(ii) Heavy-tailed distribution with $m_0 > 2$.

ad (i): For distributions of the first group we find the linear programming estimator of $\boldsymbol{\rho}$, proposed by Resnick and Feigin (1997), as the most convenient.
(Limit distributions for linear programming time series estimators. *J. Stoch. Process. & Appl.* 51, 135-165).

$$\widehat{\boldsymbol{\rho}}_{LP} := \text{argmax}_{\mathbf{u}\in\mathcal{D}_N} \sum_{j=1}^{p} u_j, \tag{3}$$

$$\mathcal{D}_N := \{\mathbf{u} := (u_1, \cdots, u_p)' \in I\!R^p : \ X_t \geq \sum_{j=1}^{p} u_j X_{t-j}, \ t = 1, \cdots nN\}.$$

Feigin and Resnick considered a stationary autoregressive process with positive innovations, whose distribution is of type $1 - F(x) = x^{-m}L(x)$.

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. These estimators are $\sqrt{N}$-consistent, and cover the popular Huber estimator; the distribution can be extended over all real line.

Construction of the tests

Let n, N be positive integers and let $\widehat{\boldsymbol{\rho}}_N$ be an estimator of $\boldsymbol{\rho}$ based on the data set X_{1-p}, $X_{2-p}, \cdots, X_0, X_1, \cdots X_{nN}$. Let

$$\widehat{\varepsilon}_t := X_t - \widehat{\boldsymbol{\rho}}'_N \mathbf{Y}_{t-1}, \quad t = 1-p, 2-p, \cdots, nN, \tag{4}$$

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

If we want to test $\mathbf{H}_0$ with $0 < m_0 \leq 2$, then we use the linear programming estimator $\widehat{\boldsymbol{\rho}}_{LP}$. If we want to test $\mathbf{H}_0$ with $m_0 > 2$, then we use GM- or GR-estimators.

Now group these residuals in N groups, each of size n, so that the residuals in the tth group are $\widehat{\varepsilon}_{(t-1)n-p+1}, \cdots, \widehat{\varepsilon}_{tn-p}$.

Let

$$\hat{\varepsilon}^t_{(n)} := \max_{1\le i\le n} \hat{\varepsilon}_{(t-1)n-p+i}, \qquad t = 1, 2, \cdots, N, \tag{5}$$

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

$$a^{(1)}_{N,m} := (nN^{1-\delta})^{\frac{1}{m}}, \qquad 0 < \delta < 1, \tag{6}$$

$$a^{(2)}_{N,m} := \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}, \quad 0 < \eta < 1. \tag{7}$$

The thresholds $a^{(1)}_{N,m}$ and $a^{(2)}_{N,m}$ lead to slightly different tests; comparing with the original $a^{(1)}_{N,m}$, used in Jurečková and Picek (2001), the new threshold $a^{(2)}_{N,m}$ seems to give better numerical results both in the linear regression and autoregression models.

The empirical distribution function $\widehat{F}_N^*$ of the maximal residuals $\{\widehat{\varepsilon}_{(n)}^t,\ t = 1, \ldots, N\}$ approximates the empirical distribution function

$$F_N^* = N^{-1} \sum_{t=1}^{N} I(\varepsilon_{(n)}^t \leq x),\ x \in \mathbb{R},$$

of the maximal errors

$$\{\varepsilon_{(n)}^t = \max_{1 \leq i \leq n} \varepsilon_{(t-1)n-p+i},,\ t = 1, \ldots, N\}.$$

If F is heavy-tailed and $\widehat{\boldsymbol{\rho}}_N$ is an appropriate estimate of $\boldsymbol{\rho}$,

$$|\widehat{F}_N^*(a_{N,m}) - F_N^*(a_{N,m})| = o_p(1), \quad \text{as} \quad N \longrightarrow \infty, \tag{8}$$

with an appropriate rate of convergence, provided m is the true value of the tail index. All limits throughout are taken as $N \longrightarrow \infty$ and for a fixed n.

We propose two tests for $\mathbf{H}_0$ against $\mathbf{K}_0$ corresponding to $a^{(1)}_{N,m_0}$, $a^{(2)}_{N,m_0}$, 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 $\mathbf{K}_0$ rejects the hypothesis provided

either $1 - \widehat{F}^*_N(a^{(1)}_{N,m_0}) = 0,$

or $1 - \widehat{F}^*_N(a^{(1)}_{N,m_0}) > 0$ and

$$N^{\delta/2}\left[-\ln(1 - \widehat{F}^*_N(a^{(1)}_{N,m_0})) - (1-\delta)\ln N\right] \geq \Phi^{-1}(1-\alpha),$$

where Φ is the standard normal distribution function.

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

either $1 - \widehat{F}_N^*(a_{N,m_0}^{(2)}) = 0,$

or $1 - \widehat{F}_N^*(a_{N,m_0}^{(2)}) > 0,$ and

$(\ln N)^{1-\frac{\eta}{2}} \left[-\ln(1 - \widehat{F}_N^*(a_{N,m_0}^{(2)})) - \ln N + (2-\eta) \ln\ln N \right] \geq \Phi^{-1}(1-\alpha).$

The test criteria have asymptotically standard normal distributions under the exact Pareto tail corresponding to $1 - F(x) = x^{-m_0}$, $\forall\, x > x_0$.

Theorem 1 *Consider the stationary autoregressive process. Assume that the process satisfies the condition* $\sum_{j=1}^{p} \rho_j < \infty$ *and that the innovation distribution function* F *is heavy-tailed 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* $\widehat{F}_N^*(a_{N,m_0}^{(1)})$ *be the empirical distribution function of extreme residuals.*

Then, the following hold:

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

$$\lim_{N\to\infty} I\!\!P \left(0 < \widehat{F}_N^*(a_{N,m_0}^{(1)}) < 1\right) = 1.$$

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

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

Hence,

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

(iii) *The test is asymptotically unbiased for the family of heavy-tailed d.f.'s* F *with* $m = m_0$ *and with* $\underline{\lim}_{x\to\infty} L(x) \ge 1$. *More precisely, then*

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

It is also asymptotically unbiased for the family with $m < m_0$.

Let F be heavy-tailed with tail index m_0, $m_0 > 2$, with a continuous and positive density on $\mathbb{R}$. Let $\widehat{F}^*_N(a^{(1)}_{N,m_0})$ be the empirical distribution function of extreme residuals, where the residuals are calculated with respect to a $\sqrt{N}$-consistent estimator of $\boldsymbol{\rho}$. Then the conclusions of (i) – (iii) above continue to hold.

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

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

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

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

$$\lim_{N\to\infty} I\!P\left\{(\ln N)^{\frac{\eta}{2}}\Big[-\ln(1-\widehat{F}_N^*(a_{N,m_0}^{(2)})) - \ln N + (2-\eta)\ln\ln N\Big] \le x\right\} = \Phi(x).$$

Hence,

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

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

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

(II) Let F *be heavy-tailed with tail index* $m_0,\ m_0 > 2$, *with a continuous and positive density on* $I\!R$. *Let* $\widehat{F}_N^*(a_{N,m_0}^{(2)})$ *be the empirical distribution function of extreme residuals, where the residuals are calculated with respect to a* $\sqrt{N}$*-consistent estimator of* $\boldsymbol{\rho}$. *Then the conclusions (i) – (iii) above continue to hold.*

Simulation study

The performance is studied on 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 white noise distributions:

Pareto: $F(x) = 1 - \left(\frac{1}{1+x}\right)^m, \quad x \geq 0$;

Burr: $F(x) = 1 - \left(\frac{1}{1+x^m}\right)^\kappa, \quad x \geq 0$;

Inverse normal: $F(x) = 2\left(1 - \Phi\left(\frac{1}{\sqrt{x}}\right)\right), \quad x > 0$

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

For each of these cases, the time series were simulated of the lengths $nN = 200$ and 1000. 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).

(1) generated the autoregressive time series;
(2) estimated ρ by $\hat{\rho}$;
(3) computed residuals $\hat{\varepsilon}_t := X_t - \hat{\rho}'_N \mathbf{Y}_{t-1}, \quad t = 1-p, 2-p, \cdots, nN$;
(4) found the maxima $\widehat{\varepsilon}_n^{(1)}, \ldots, \widehat{\varepsilon}_n^{(N)}$ of the segments and the corresponding empirical distribution function $\hat{F}_N^*$;
(5) we made a decision about $\mathbf{H}_0$;
(6) the step (5) was repeated for various values m_0, δ;
(7) the steps (1)-(6) were repeated $1\,000$ times.

Numbers of rejections of the null hypothesis for $a_{N,m}^{(1)}$,

$$\alpha = 0.05,\ N = 50, n = 4, \delta = 0.1$$

distribution of white noise	time series	m_0 0.25	0.4	0.5	0.6	0.75
Pareto	A	986	674	245	36	0
$m = 0.5$	B	986	674	245	36	0
	C	986	674	245	36	0
Burr	A	986	674	246	37	0
$m = 0.5$	B	986	674	246	37	0
$\kappa = 1$	C	986	674	246	37	0
Inverse	A	990	736	320	79	1
normal	B	990	736	320	79	1
	C	990	736	320	79	1

Numbers of rejections of the null hypothesis for $a_{N,m}^{(2)}$,

$\alpha = 0.05,\ N = 50, n = 4, \eta = 0.1$

distribution of white noise	time series	m_0				
		0.3	0.4	0.5	0.52	0.6
Pareto	A	1000	995	84	17	0
$m = 0.5$	B	1000	995	84	17	0
	C	1000	995	84	17	0
Burr	A	1000	995	107	20	0
$m = 0.5$	B	1000	995	107	20	0
$\kappa = 1$	C	1000	995	107	20	0
Inverse	A	1000	1000	363	158	1
normal	B	1000	1000	363	158	1
	C	1000	1000	363	158	1

Numbers of rejections of the null hypothesis for $a^{(1)}_{N,m}$, $\alpha = 0.05$, $N = 50, n = 4, \delta = 0.1$

distribution of white noise	time series	m_0				
		0.5	0.8	0.9	1.0	1.2
Pareto	A	991	674	438	246	36
$m = 1$	B	991	674	441	245	37
	C	991	674	439	246	37
Burr	A	991	674	442	246	37
$m = 1$	B	991	674	442	246	37
$\kappa = 1$	C	991	674	442	246	37
	m_0	2.0	2.5	2.75	3.0	3.5
Student	A	867	569	402	255	66
$m = 3$	B	865	565	398	254	63
	C	866	564	403	251	69

Numbers of rejections of the null hypothesis for $a^{(2)}_{N,m}$, $\alpha = 0.05$, $N = 50, n = 4, \eta = 0.1$

distribution of white noise	time series	m_0				
		0.8	0.9	1.0	1.02	1.1
Pareto	A	995	646	84	36	1
$m = 1$	B	995	646	84	38	1
	C	995	646	84	37	1
Burr	A	995	667	107	51	1
$m = 1$	B	995	667	107	51	1
$\kappa = 1$	C	995	667	107	51	1
	m_0	2.5	2.8	3.00	3.05	3.5
Student	A	982	684	283	186	7
$m = 3$	B	983	680	282	193	4
	C	982	685	281	187	5

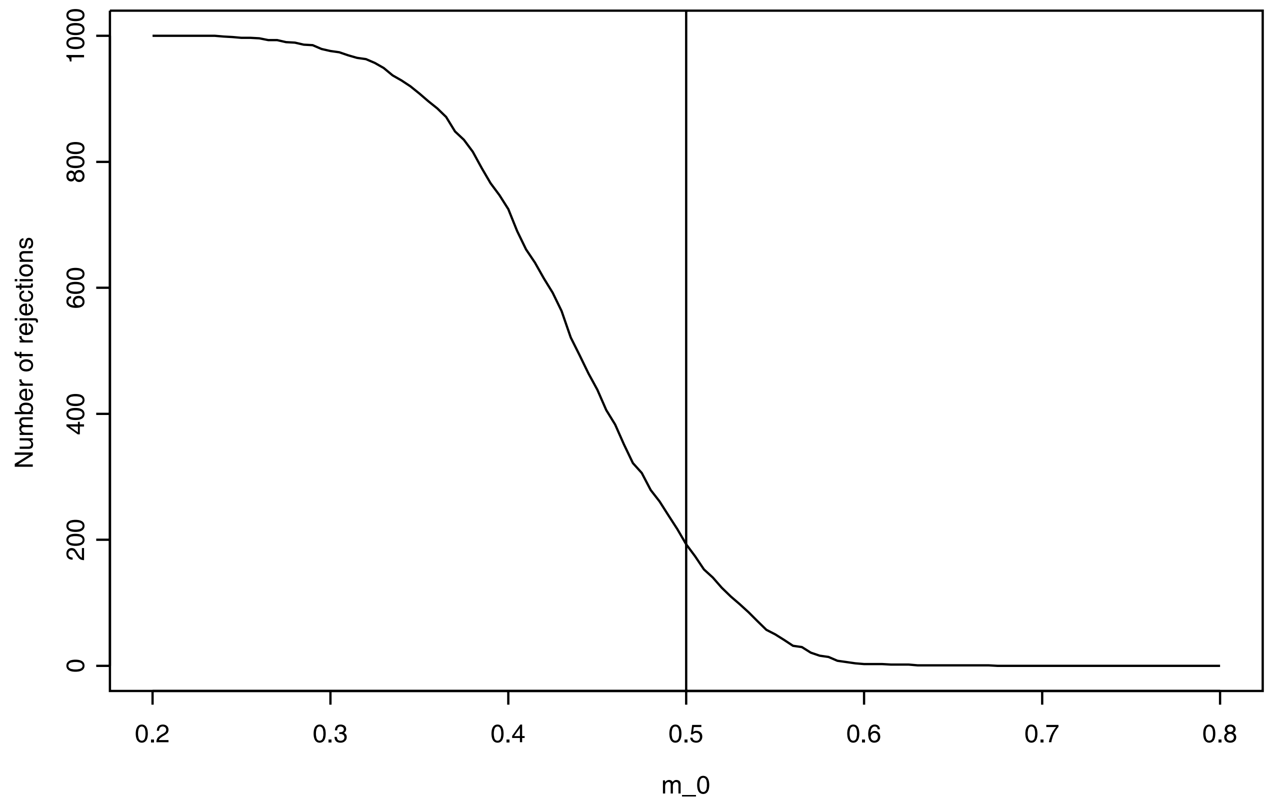

Fig.: *Number of rejections of* H_0 *(*$\alpha = 0.05$*) plotted against* m_0 *for* $X_t = 0.9X_{t-1} + \varepsilon_t$ *and* $a^{(1)}_{N,m} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1, \ldots, nN$ *have the Pareto distribution with* $m = 0.5$; $N = 50, n = 4,\ \delta = 0.1.$

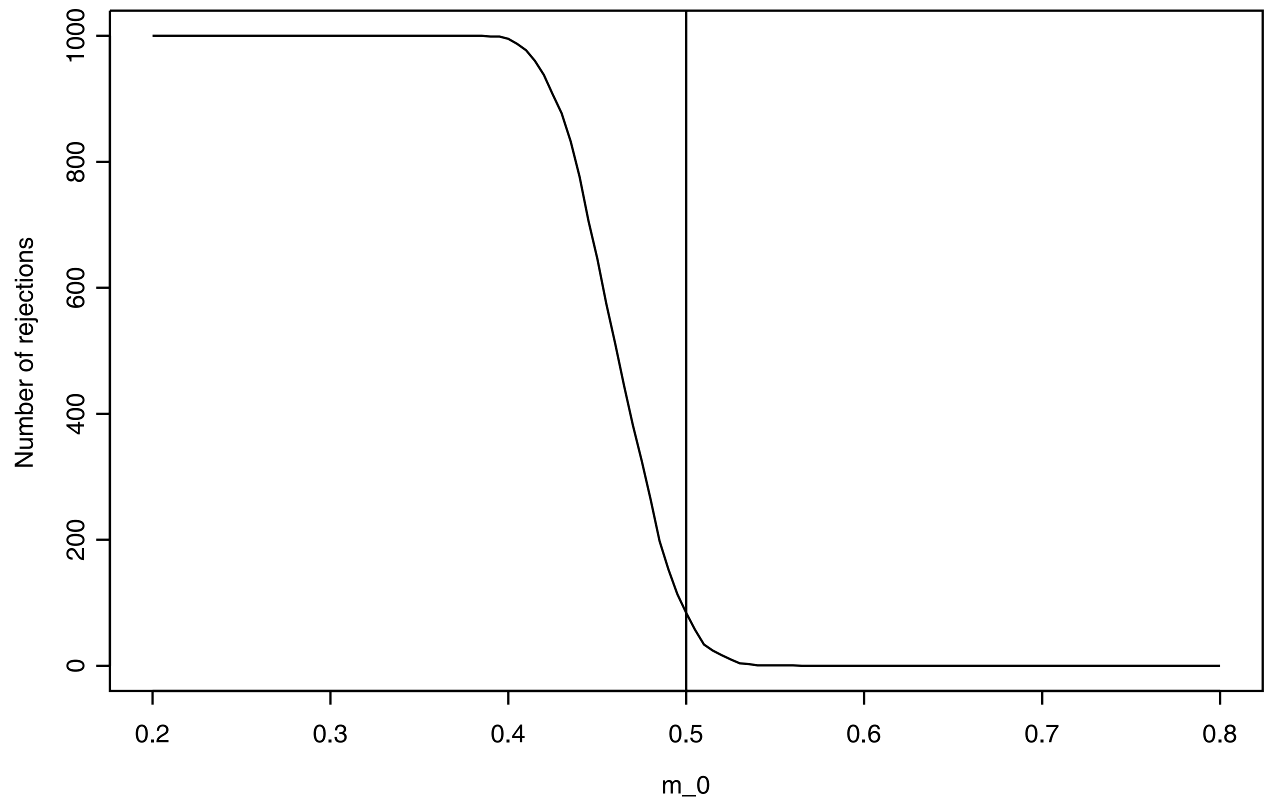

Fig.: *Number of rejections of* H_0 *(*$\alpha = 0.05$*) plotted against* m_0 *for* $X_t = 0.9X_{t-1} + \varepsilon_t$ *and* $a^{(2)}_{N,m} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1, \ldots, nN$ *have the Pareto distribution with* $m = 0.5;\ N = 200, n = 5,\ \eta = 0.1.$

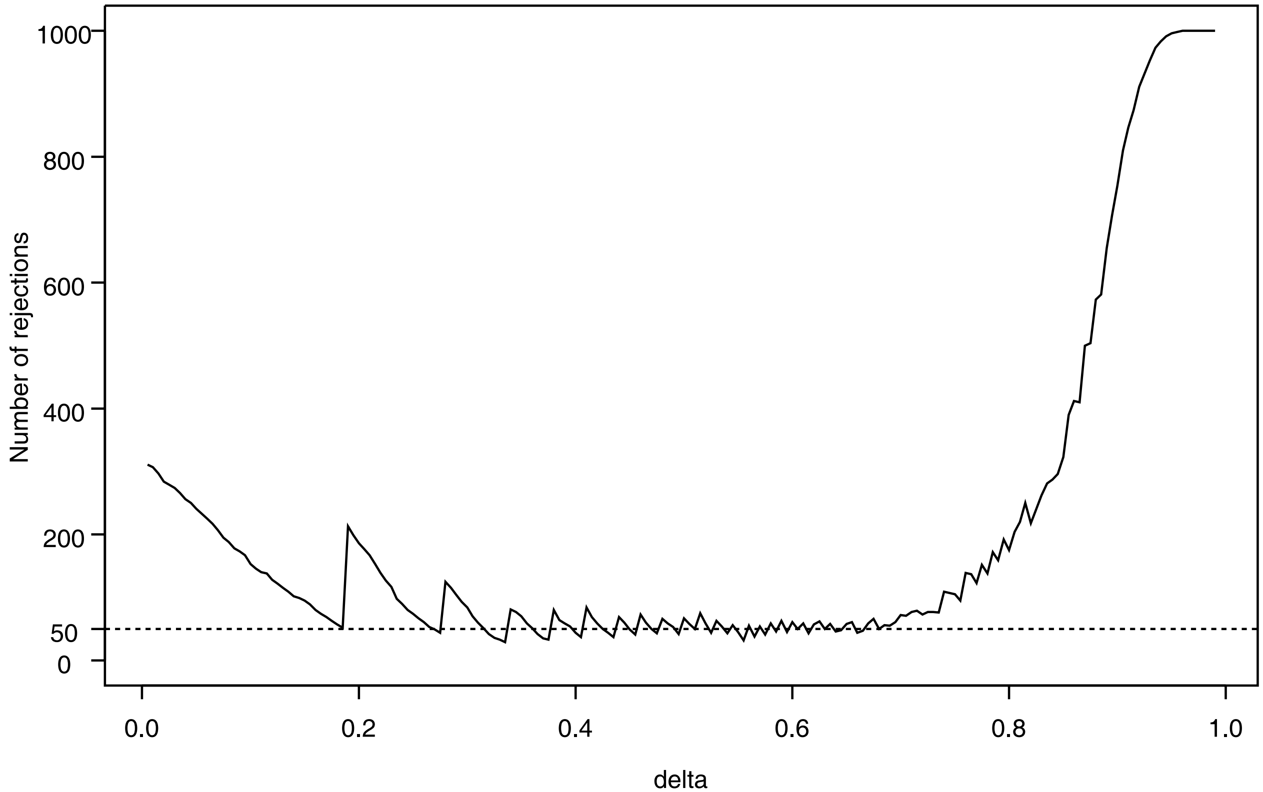

Fig. : *Number of rejections of* H_0 *(*$\alpha = 0.05$*) plotted against* δ *for* $X_t = 0.9X_{t-1} + \varepsilon_t$ *and* $a^{(1)}_{N,m} = \left(nN^{1-\delta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1, \ldots, nN$ *have the Pareto distribution with* $m = 0.5;\ N = 50, n = 4,\ m_0 = 0.51.$

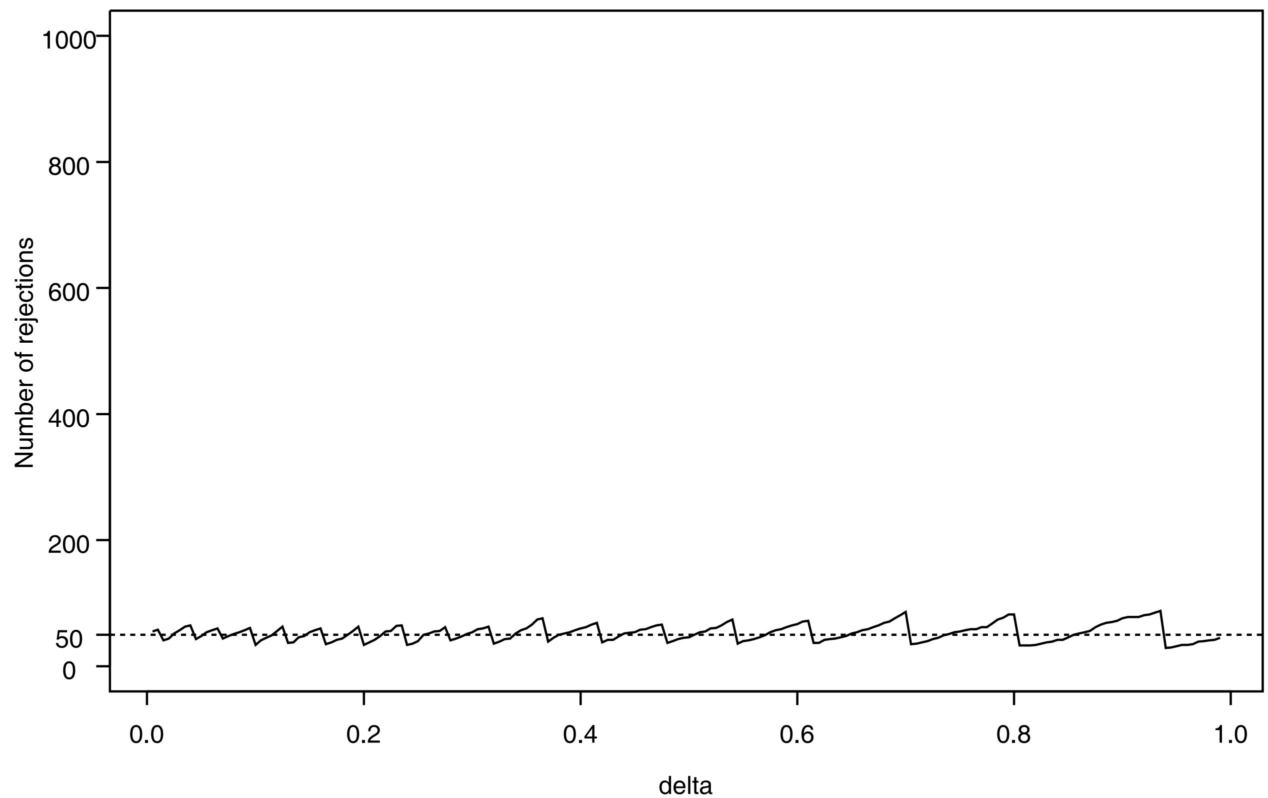

Fig.: *Number of rejections of* H_0 ($\alpha = 0.05$) *plotted against* η *for* $X_t = 0.9X_{t-1} + \varepsilon_t$ *and* $a_{N,m}^{(2)} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$; $\varepsilon_t,\ t = 1, \ldots, nN$ *have the Pareto distribution with* $m = 0.5;\ N = 200, n = 5,\ m_0 = 0.51$.

Numbers of rejections of the null hypothesis among 1000 AR series of the length 1000 for various $N,\ n$; $\alpha = 0.05,\ \delta = 0.1$.

distribution of white noise	$n,\ N$	m_0				
		0.5	0.8	0.9	1.0	1.2
Pareto $m = 1$	$n = 5,\ N = 200$	997	725	438	193	3
	$n = 10,\ N = 100$	997	745	462	221	4
	$n = 20,\ N = 50$	998	761	489	241	7
		m_0				
		2.0	2.5	2.75	3.0	3.5
Pareto $m = 3$	$n = 5,\ N = 200$	943	704	522	341	104
	$n = 10,\ N = 100$	950	723	547	355	103
	$n = 20,\ N = 50$	955	742	572	387	124

Numbers of rejections of the null hypothesis at level $\alpha = 0.05$ among 1000 AR time series, and among 1000 corresponding sequences of the white noise (WN); $N = 200,\ n = 5,\ \delta = 0.1,\ 0.5.$

distribution of white noise		δ	m_0				
			0.5	0.8	0.9	1.0	1.2
Pareto $m = 1$	AR	0.1	997	725	438	193	3
	AR	0.5	836	3	1	0	0
	WN	0.1	997	725	438	193	3
	WN	0.5	836	3	1	0	0
			2.0	2.5	2.75	3.0	3.5
Pareto $m = 3$	AR	0.1	943	704	522	341	104
	AR	0.5	310	36	4	2	0
	WN	0.1	932	623	389	204	12
	WN	0.5	169	1	0	0	0

Application to the daily maximum temperatures

The tests described above are applied to a 40-year dataset of daily maximum temperatures measured at three meteorological stations in Czech Republic, over the period of 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.

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. (1977)).

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

The same for $a^{(2)}_{N,m} = \left(nN(\ln N)^{-2+\eta}\right)^{\frac{1}{m}}$.

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