Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 2867 – 2886
journal of statistical planning and inference
www.elsevier.com/locate/jspi

# A LAN based Neyman smooth test for Pareto distributions

Michael Falk[a], Armelle Guillou[b,*], Gwladys Toulemonde[c]

[a] *Institute of Mathematics, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany*
[b] *IRMA, Université Louis Pasteur, 7 rue René Descartes, F-67084 Strasbourg Cedex, France*
[c] *LSTA, Université Pierre et Marie Curie, Boîte 158, 175 rue du chevaleret, F-75013 Paris, France*

Received 4 May 2007; received in revised form 25 October 2007; accepted 31 October 2007
Available online 12 November 2007

**Abstract**

The Pareto distribution is found in a large number of real world situations and is also a well-known model for extreme events. In the spirit of Neyman [1937. Smooth tests for goodness of fit. Skand. Aktuarietidskr. 20, 149–199] and Thomas and Pierce [1979. Neyman's smooth goodness-of-fit test when the hypothesis is composite. J. Amer. Statist. Assoc. 74, 441–445], we propose a smooth goodness of fit test for the Pareto distribution family which is motivated by LeCam's theory of local asymptotic normality (LAN). We establish the behavior of the associated test statistic firstly under the null hypothesis that the sample follows a Pareto distribution and secondly under local alternatives using the LAN framework. Finally, simulations are provided in order to study the finite sample behavior of the test statistic.


*MSC:* 62F03; 62F05

*Keywords:* Goodness of fit test; Neyman smooth test; LeCam's theory; LAN; Pareto distributions; Power

## 1. Introduction

The Pareto distribution is a well-known model which was originally introduced to describe the distribution of income. It has been applied in many fields like in insurance to model claims (Benktander, 1970), in climatology–hydrology (Katz et al., 2002) to describe the occurrence of extreme weather, and also in economy (Fisk, 1961), in finance (Danielsson and de Vries, 1997) or in hydrogeology (Gustafson and Fransson, 2005) among others. For general overviews of the role of the Pareto distribution in many other fields, we refer to Arnold (1983) who studied extensively this distribution. Moreover, with a specific parameterization, the Pareto distribution can be considered as a particular case of the generalized Pareto distribution which is an essential model in the study of exceedances over a high threshold in the extreme value framework (Reiss and Thomas, 2007).

If we use a statistical analysis that involves fitting a parametric model, it is always advisable to check the adequacy of the model. For this, a goodness of fit test could be used to decide whether a sample of independent random variables (rvs) $X_1, \ldots, X_n$ is distributed according to a Pareto distribution. To keep a link with extreme value theory, we use the

* Corresponding author. Tel.: +33 3 90 24 01 99; fax: +33 3 90 24 03 28.
*E-mail address:* guillou@math.u-strasbg.fr (A. Guillou).

0378-3758/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.10.007

following parameterization of the Pareto distribution density:

$$f(x,\beta)=\frac{1}{\sigma}\left(1+\frac{\xi x}{\sigma}\right)^{-1/\xi-1} \quad \text{where } \beta^t=(\sigma,\xi)\in\Theta=(0,\infty)\times(0,\infty) \text{ and } x\in[0,\infty).$$

Clearly, we will have to estimate the parameters of the Pareto distribution since in most of the applications, they are unknown. To this aim, different techniques have been proposed (see Arnold and Press, 1989; Castillo and Hadi, 1997; Davison, 1984; Hosking and Wallis, 1987; Malik, 1966; Smith, 1984). In this paper we will introduce an estimator, having the same properties as the maximum likelihood one, but which is easier to compute. This point will be discussed later on.

The problem of testing the fit of a Pareto distribution has not been studied extensively in the literature. Davison and Smith (1990), for instance, pointed out the lack of tests in the case of the generalized Pareto distribution. They used tables for testing the exponential distribution, which, however, requires very high critical values. Porter et al. (1992) presented tables of critical values for Pareto distributions with known shape parameter for test statistics based on the empirical distribution function (EDF). More recently, Choulakian and Stephens (2001) introduced goodness of fit tests for generalized Pareto distributions also based on EDF but with an unknown shape parameter. Although these test statistics, based on the EDF, are consistent against essentially all alternatives, there are numerous empirical studies (Kopecky and Pierce, 1979; Miller and Quesenberry, 1979; Rayner and Best, 1986) where smooth tests have been shown to be more powerful than the usual tests like these EDF tests or the Pearson's $\chi^2$ test. The latter results caused renewed interest in smooth tests.

Historically, Neyman (1937) introduced these smooth goodness of fit tests for testing uniformity. Many generalizations of them have been proposed, see for instance Rayner and Best (1989, 1990) and Thomas and Pierce (1979). The basic idea behind these tests is to embed the null density into, say a $J$-dimensional exponential family of the form

$$g_J(x,\theta,\beta)=f(x,\beta)\exp\left\{\sum_{s=1}^{J}\theta_s\overline{F}^s(x,\beta)-K(\theta)\right\} \tag{1}$$

and then to construct an asymptotically optimal test for the parametric testing problem. Here $F(.,\beta)$ is the df of the Pareto distribution, $\overline{F}(.,\beta)$ the corresponding survival function, whereas $K(\theta)$ is a normalizing function ensuring integration to 1. It is easy to check that $K(\theta)$ only depends on $\theta$ and is given by

$$K(\theta)=\log\left\{\int_0^1\exp\left(\sum_{s=1}^{J}\theta_s t^s\right)\mathrm{d}t\right\}.$$

From (1), we can deduce that the goodness of fit test for the Pareto distribution can be reduced to the following test: $\mathrm{H}_0:\theta=0$ where $\theta=(\theta_1,\ldots,\theta_J)^t$ versus $\mathrm{H}_1:\theta\neq 0$.

In Section 3, we study the behavior under $\mathrm{H}_0$ of the test statistic which is defined in Proposition 5. To this aim, we need to use an estimator of $\beta$. Various methods have been proposed to estimate the parameters of the Pareto distribution. In hydrology for instance, the method of probability weighted moments (Hosking and Wallis, 1987) is widely applied. The maximum likelihood estimator (MLE) is also quite common. It is known to have good asymptotic properties, but it is typically hard to compute. In this paper, we introduce an estimator, $\widehat{\beta}_n$, having the same asymptotic properties as the MLE but which is easier to compute. Using this estimator, the limiting distribution of the test statistic is obtained under $\mathrm{H}_0$ using Pfanzagl's (1994) general results on parametric statistics. In Section 4 the behavior under local alternatives $\mathrm{H}_{1,n}$ is studied. They are defined as

$$\mathrm{H}_{1,n}:\theta=\theta_n=\frac{\delta}{\sqrt{n}}(1+\mathrm{o}(1)), \tag{2}$$

where $\delta=(\delta_1,\ldots,\delta_J)^t\neq 0$ with $\delta_i\in\mathbb{R}$, for all $i=1,\ldots,J$.

This investigation is used in the local asymptotic normality (LAN) framework of LeCam (1960). Here the notation $\delta(1+\mathrm{o}(1))$ has to be interpreted as a vector $(\delta_1(1+r_1(n)),\ldots,\delta_J(1+r_J(n)))^t$, with functions $r_s(n)$ converging to 0 as $n\to\infty$. The fifth section of this paper is devoted to simulations in order to study the finite sample behavior of our test statistic. Finally the last section deals with concluding remarks. The details of the proofs are given in the appendix.

## 2. Notations and definitions

We start this section by setting some notations. We denote by $\{\mathbb{P}_{H_0,\beta};\ \beta\in\Theta\}$ the Pareto family of probability measures with density $f$. First, define $\ell^{(.)}(.,\beta)$ and $\ell^{(\mu\nu)}(.,\beta)$ as follows:

$$\ell^{(.)}(.,\beta)=\begin{pmatrix}\ell^{(1)}(.,\beta)\\ \ell^{(2)}(.,\beta)\end{pmatrix}=\begin{pmatrix}\dfrac{\partial\log f(.,\beta)}{\partial\sigma}\\ \dfrac{\partial\log f(.,\beta)}{\partial\xi}\end{pmatrix},$$

$$\ell^{(\mu\nu)}(.,\beta)=\frac{\partial^2\log f(.,\beta)}{\partial\beta[\mu]\partial\beta[\nu]}\quad\text{for } \mu,\nu=1,2 \text{ with } \beta[1]=\sigma \text{ and } \beta[2]=\xi.$$

Direct computations lead to

$$\ell^{(.)}(x,\beta)=\begin{pmatrix}-\dfrac{1}{\sigma}+(1+\xi)\dfrac{x}{\sigma^2}\left(1+\dfrac{\xi x}{\sigma}\right)^{-1}\\ \dfrac{1}{\xi^2}\log\left(1+\dfrac{\xi x}{\sigma}\right)-\dfrac{1+\xi}{\xi\sigma}x\left(1+\dfrac{\xi x}{\sigma}\right)^{-1}\end{pmatrix}\tag{3}$$

and

$$\ell^{(11)}(x,\beta)=\frac{1}{\sigma^2}-2\frac{1+\xi}{\sigma^3}x\left(1+\frac{\xi x}{\sigma}\right)^{-1}+\frac{1+\xi}{\sigma^4}\xi x^2\left(1+\frac{\xi x}{\sigma}\right)^{-2},\tag{4}$$

$$\ell^{(12)}(x,\beta)=\frac{1}{\sigma^2}x\left(1+\frac{\xi x}{\sigma}\right)^{-1}-\frac{1+\xi}{\sigma^3}x^2\left(1+\frac{\xi x}{\sigma}\right)^{-2},\tag{5}$$

$$\ell^{(22)}(x,\beta)=-\frac{2}{\xi^3}\log\left(1+\frac{\xi x}{\sigma}\right)+\frac{2}{\xi^2\sigma}x\left(1+\frac{\xi x}{\sigma}\right)^{-1}+\frac{1+\xi}{\xi\sigma^2}x^2\left(1+\frac{\xi x}{\sigma}\right)^{-2}.\tag{6}$$

Further define the Fisher information matrix

$$\begin{aligned}I_{\beta\beta}&=\mathbb{E}_{H_0,\beta}\begin{pmatrix}(\ell^{(1)}(X,\beta))^2 & \ell^{(1)}(X,\beta)\ell^{(2)}(X,\beta)\\ \ell^{(1)}(X,\beta)\ell^{(2)}(X,\beta) & (\ell^{(2)}(X,\beta))^2\end{pmatrix}\\ &=\begin{pmatrix}1/\{\sigma^2(1+2\xi)\} & 1/\{\sigma(1+\xi)(1+2\xi)\}\\ 1/\{\sigma(1+\xi)(1+2\xi)\} & 2/\{(1+\xi)(1+2\xi)\}\end{pmatrix}\end{aligned}$$

which leads to

$$I_{\beta\beta}^{-1}=\begin{pmatrix}2\sigma^2(1+\xi) & -\sigma(1+\xi)\\ -\sigma(1+\xi) & (1+\xi)^2\end{pmatrix}.$$

Now, the methodology used in order to study the behavior of the test statistic under $H_0$ is based on Pfanzagl's (1994) results, while, under $H_{1,n}$, the LAN theory is applied. To this aim, it seems useful to recall some definitions, in particular concerning the locally uniformly weak convergence, which is a fundamental notion in our method. Also, as already mentioned, we have to estimate the parameter of the Pareto distribution. Therefore, the definition of a $\sqrt{n}$-consistent locally uniformly estimator will be also recalled below. Most of the definitions and notations below are adopted from Pfanzagl (1994).

**Definition 1.** Let $X_1,\dots,X_n$ be iid rvs from an arbitrary df depending on some parameter $\zeta\in\mathscr{S}$, say $F_\zeta$. Set $Y_n=H(X_1,\dots,X_n;\zeta)$, where the function $H$ is $\mathbb{R}^q$-valued and let $Q_{n,\zeta}$ be the law of $Y_n$. The sequence $Y_n$, $n\in\mathbb{N}$, converges weakly to $Q_\zeta$ locally uniformly on $\mathscr{S}$ if for every $\zeta_0\in\mathscr{S}$, there exists a neighborhood $V(\zeta_0)$ of $\zeta_0\in\mathscr{S}$ such that

$$\lim_{n\to\infty}\sup_{\zeta\in V(\zeta_0)}|Q_{n,\zeta}(h)-Q_\zeta(h)|=0\quad\text{for every } h\in\mathscr{C},$$

where $\mathscr{C}$ denotes the class of all bounded and continuous functions $h:\mathbb{R}^q\mapsto\mathbb{R}$.

**Definition 2.** Denote by $\mathbb{P}_\zeta$ the probability under $F_\zeta$. A random quantity $Q(X_1,\ldots,X_n;\zeta)$ is $\sqrt{n}$-consistent locally uniformly on $\mathscr{S}$ for $Q(\zeta)$ if for every $\zeta_0$, there exists a neighborhood $V(\zeta_0)$ of $\zeta_0\in\mathscr{S}$ such that for every $\varepsilon>0$, there exists $M_\varepsilon>0$ such that

$$\sup_{\zeta\in V(\zeta_0)} \mathbb{P}_\zeta(\sqrt{n}\|Q(X_1,\ldots,X_n;\zeta)-Q(\zeta)\|>M_\varepsilon)<\varepsilon \quad \text{for } n\in\mathbb{N},$$

where $\|\cdot\|$ is the usual Euclidean norm.

Remark also that a random quantity $Q(X_1,\ldots,X_n;\zeta)$ is $\sqrt{n}$-consistent locally uniformly on $\mathscr{S}$ for $Q(\zeta)$ if the quantity $\sqrt{n}(Q(X_1,\ldots,X_n;\zeta)-Q(\zeta))$ converges weakly locally uniformly on $\mathscr{S}$.

**Definition 3.** A random quantity $Q(X_1,\ldots,X_n;\zeta)$ converges stochastically locally uniformly on $\mathscr{S}$ to $\mathscr{Q}(\zeta)$ if for every $\zeta_0\in\mathscr{S}$, there exists a neighborhood $V(\zeta_0)$ of $\zeta_0$ such that

$$\lim_{n\to\infty}\sup_{\zeta\in V(\zeta_0)} \mathbb{P}_\zeta(\|Q(X_1,\ldots,X_n;\zeta)-\mathscr{Q}(\zeta)\|>\varepsilon)=0 \quad \text{for } \varepsilon>0.$$

Now, in order to use the LAN theory, we have to compute the loglikelihood ratio given by

$$\begin{aligned} L_n &= \log\prod_{i=1}^n \frac{g_J(X_i,(\delta/\sqrt{n})(1+o(1)),\beta)}{f(X_i,\beta)} \\ &= \sum_{i=1}^n\left[\log\left(g_J\left(X_i,\frac{\delta}{\sqrt{n}}(1+o(1)),\beta\right)\right)-\log(f(X_i,\beta))\right]. \end{aligned}$$

By (1) and (2), $L_n$ may be rewritten as

$$\begin{aligned} &\sum_{i=1}^n\sum_{s=1}^J \frac{\delta_s}{\sqrt{n}}(1+r_s(n))\overline{F}^s(X_i,\beta)-nK\left(\frac{\delta}{\sqrt{n}}(1+o(1))\right) \\ &\quad=\sum_{s=1}^J\delta_s(1+r_s(n))\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\overline{F}^s(X_i,\beta)-\frac{1}{s+1}\right)+\sqrt{n}\sum_{s=1}^J\frac{\delta_s}{s+1}(1+r_s(n))-nK\left(\frac{\delta}{\sqrt{n}}(1+o(1))\right). \end{aligned}$$

Using a Taylor expansion, we obtain that

$$\begin{aligned} L_n &= \sum_{s=1}^J\delta_s(1+r_s(n))\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\overline{F}^s(X_i,\beta)-\frac{1}{s+1}\right)+\sqrt{n}\sum_{s=1}^J\frac{\delta_s}{s+1}(1+r_s(n)) \\ &\quad -n\left[\frac{(\delta(1+o(1)))^t}{\sqrt{n}}\left.\frac{\partial K(\theta)}{\partial\theta}\right|_{\theta=0}+\frac{1}{2}\frac{1}{\sqrt{n}}(\delta(1+o(1)))^t\left.\frac{\partial^2K(\theta)}{\partial\theta\partial\theta}\right|_{\theta=0}\frac{1}{\sqrt{n}}\delta(1+o(1))\right. \\ &\quad \left.+o\left(\left\|\frac{\delta}{\sqrt{n}}(1+o(1))\right\|^2\right)\right]. \end{aligned}$$

Denote by $Z_n(\beta)$ the column vector

$$\left.\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\overline{F}^s(X_i,\beta)-\frac{1}{s+1}\right]\right|_{s=1,\ldots,J},$$

which is bounded in probability, by the Central Limit Theorem.

Note that the matrix

$$I:=\left.\frac{\partial^2K(\theta)}{\partial\theta\partial\theta}\right|_{\theta=0}$$

corresponds to the variance of $Z_n(\beta)$ (see (14)).

Since for $s = 1, \dots, J$, the functions $r_s(n)$ converge to 0 and

$$\frac{\partial K(\theta)}{\partial \theta}\bigg|_{\theta=0} = \frac{1}{s+1}\bigg|_{s=1,\dots,J},$$

we obtain

$$L_n = \delta^t Z_n(\beta) - \tfrac{1}{2}\delta^t I \delta + o_{\mathbb{P}_{H_0,\beta}}(1). \tag{7}$$

The central sequence $Z_n(\beta)$ is significant in LAN theory (see Strasser, 1985, Chapter 13). Consequently, the statistic for testing $H_0$ against $H_{1,n}$ should be based on $Z_n(\beta)$. In the next section we consider the null hypothesis.

## 3. The behavior of our test statistic under $H_0$

We assume in this section that the sample $X_1, \dots, X_n$ comes from a Pareto distribution with unknown parameter $\beta = (\sigma, \xi)^t \in \Theta$. Since the MLE is difficult to compute, we propose to use an estimator $\widehat{\beta}_n$ of $\beta$ having the same asymptotic properties as the MLE but which is easier to compute. This estimator $\widehat{\beta}_n$ will be called a two-step estimator. Indeed, the first step consists in computing an initial $\sqrt{n}$-consistent estimator (locally uniformly) $\widetilde{\beta}_n$ whereas the second one consists in the achieved construction of the estimator $\widehat{\beta}_n$ defined as

$$\widehat{\beta}_n = \widetilde{\beta}_n + I^{-1}_{\widetilde{\beta}_n\widetilde{\beta}_n} \frac{1}{n}\sum_{i=1}^{n} \ell^{(\cdot)}(X_i, \widetilde{\beta}_n). \tag{8}$$

The estimator $I^{-1}_{\widetilde{\beta}_n\widetilde{\beta}_n}$ is obtained by the plug-in method which consists in replacing the parameter $\beta$ by the estimator $\widetilde{\beta}_n$ in the matrix $I^{-1}_{\beta\beta}$. Indeed, a local uniform version of the continuous mapping theorem can be deduced by straightforward arguments from the usual one (see van der Vaart, 1998, 2.3). The estimator $\widetilde{\beta}_n$ is consistent for $\beta$, locally uniformly on $\Theta$, therefore applying the local uniform version of the continuous mapping theorem, we deduce that the estimator $I^{-1}_{\widetilde{\beta}_n\widetilde{\beta}_n}$ is consistent for $I^{-1}_{\beta\beta}$, locally uniformly on $\Theta$.

The following proposition is a key result to establish the behavior of the test statistic defined in Proposition 5. It states a local uniform weak convergence of the two-step estimator. Here $\rightsquigarrow$ denotes weak convergence.

**Proposition 4.** *Under our assumption* $H_0$ *and using* $\widehat{\beta}_n$ *as defined by* (8), *we have*

$$\sqrt{n}(\widehat{\beta}_n - \beta) - I^{-1}_{\beta\beta}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^{(\cdot)}(X_i, \beta) \to 0 \quad \textit{locally uniformly on } \Theta, \tag{9}$$

$$\sqrt{n}(\widehat{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, I^{-1}_{\beta\beta}) \quad \textit{locally uniformly on } \Theta. \tag{10}$$

We can now construct our test as follows.

**Proposition 5.** *Let* $Z_n(\widehat{\beta}_n)$ *and* $\Sigma_J(\beta)$ *be defined as follows*:

$$Z_n(\widehat{\beta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(1 - F(X_i, \widehat{\beta}_n))^s - \frac{1}{s+1}\right]\bigg|_{s=1,\dots,J},$$

$$\Sigma_J(\beta) = \left[\frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\xi)(uv+\xi+(u+1)(v+1))}{(v+\xi+1)(u+\xi+1)(u+1)^2(v+1)^2}\right]\bigg|_{u,v=1,\dots,J}.$$

*Under the assumption* $H_0 : X_1, \dots, X_n$ *are iid rvs according to a Pareto distribution, the statistic* $Z_n(\widehat{\beta}_n)$ *converges weakly to the normal distribution* $\mathcal{N}(0, \Sigma_J(\beta))$ *with mean vector* 0 *and covariance matrix* $\Sigma_J(\beta)$, *and the test statistic* $\Psi_J^2 := Z_n^t(\widehat{\beta}_n)\Sigma_J^{-1}(\widehat{\beta}_n)Z_n(\widehat{\beta}_n)$ *converges weakly to a chi-square distribution with J degrees of freedom as* $n \to \infty$. *Therefore, the null hypothesis will be rejected for large values of* $\Psi_J^2$.

**Remark 6.** When we deal with the null hypothesis $\theta=0$ against $\theta \neq 0$, there exists three tests asymptotically equivalent: the Wald test, the score test and the likelihood ratio test. It is possible to show that our test statistic corresponds to the score test which is commonly used in the theory of smooth tests of goodness of fit. Indeed the use of the score test in the Neyman one appears to be the most convenient, essentially because it does not require the computation of the MLE of the $J$-dimensional parameter $\theta$ unlike the two other tests. Moreover the parameter $\theta$ has no physical interpretation. For further discussion we refer to Rayner and Best (1989).

As an illustration, we propose to use the moment estimator (see Hosking and Wallis, 1987, Section 3.2) in order to construct our new estimator $\widehat{\beta}_n$ of $\beta$. This moment estimator $\widetilde{\beta}_n = (\widetilde{\sigma}_n, \widetilde{\xi}_n)^t$ is defined as follows:

$$\widetilde{\sigma}_n = \frac{1}{2}\overline{X}\left(\frac{\overline{X}^2}{\overline{X^2}-\overline{X}^2}+1\right), \quad \widetilde{\xi}_n = \frac{1}{2}\left(1-\frac{\overline{X}^2}{\overline{X^2}-\overline{X}^2}\right), \tag{11}$$

where $\overline{X} = (1/n)\sum_{i=1}^n X_i$ and $\overline{X^2} = (1/n)\sum_{i=1}^n X_i^2$.

The following proposition proves that $\widetilde{\beta}_n$ is a $\sqrt{n}$-consistent estimator (locally uniformly) of $\beta$.

**Proposition 7.** *Let $X_1, \ldots, X_n$ be iid rvs according to a Pareto distribution with parameter $\beta$. The moment estimator $\widetilde{\beta}_n$ defined in* (11) *is $\sqrt{n}$-consistent for $\beta = (\sigma, \xi)^t$, locally uniformly on $\mathscr{B} \subset \Theta$ where $\mathscr{B} = (0, \infty) \times (0, \frac{1}{4})$.*

Note that the condition $\xi \in (0, \frac{1}{4})$ seems to be restrictive, but it is only due to the fact that we use here the moment estimator. A weaker condition could be obtained for other $\sqrt{n}$-consistent estimators, for instance the probability weighted moments estimator (see Hosking and Wallis, 1987, Section 3.3).

In the simulations in Section 5 we use this moment estimator as the initial one. Before that, we investigate in the next section the behavior of the test statistic $\Psi_J^2$ under the alternatives $\mathrm{H}_{1,n}$.

## 4. The behavior of our test statistic under $\mathrm{H}_{1,n}$

In order to study the behavior of our test statistic under $\mathrm{H}_{1,n}$ some preliminary results are required.

### 4.1. Preliminary results

We denote by $\overset{\mathbb{P}_{\mathrm{H}_0,\beta}}{\longrightarrow}$ (respectively, by $\overset{\mathbb{P}_{\mathrm{H}_{1,n}}}{\longrightarrow}$) the convergence in probability when $\mathrm{H}_0$ (respectively, $\mathrm{H}_{1,n}$) holds. In what follows the rvs $X_1, \ldots, X_n$ are iid with either a Pareto distribution (in case we are under $\mathrm{H}_0$) or a common density $g_J$ as defined in (1) (in case we are under the alternatives $\mathrm{H}_{1,n}$).

The following proposition will be a crucial tool. It is a consequence of the concept of contiguity (see LeCam, 1960).

**Proposition 8.** *For any statistic $T_n = T(X_1, \ldots, X_n; \beta)$, the convergence $T_n \overset{\mathbb{P}_{\mathrm{H}_0,\beta}}{\rightarrow} 0$ holds if and only if $T_n \overset{\mathbb{P}_{\mathrm{H}_{1,n}}}{\rightarrow} 0$.*

The next two propositions will be useful to establish the asymptotic behavior of our test statistic $\Psi_J^2$ under $\mathrm{H}_{1,n}$.

**Proposition 9.** *With the previous notations, combined with*

$$I := \left.\frac{uv}{(u+v+1)(u+1)(v+1)}\right|_{u,v=1,\ldots,J} \tag{12}$$

*and with*

$$I_\beta := \left.\left(\frac{-u}{(u+1+\xi)(u+1)\sigma} \; \frac{-u}{(u+1+\xi)(u+1)^2}\right)\right|_{u=1,\ldots,J}$$

*we have under* $\mathrm{H}_{1,n}$

$$\begin{pmatrix} Z_n(\beta) \\ \sqrt{n}(\widehat{\beta}_n-\beta) \end{pmatrix} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} I\delta \\ I_{\beta\beta}^{-1} I_\beta^t \delta \end{pmatrix}, \begin{pmatrix} I & I_\beta I_{\beta\beta}^{-1} \\ I_{\beta\beta}^{-1} I_\beta^t & I_{\beta\beta}^{-1} \end{pmatrix}\right].$$

Now, an application of the delta method leads to the following result.

**Proposition 10.** *Under* $\mathrm{H}_{1,n}$, *we have the following weak convergence*:

$$\sqrt{n}\left(1-\frac{\widehat{\xi}_n \sigma}{\widehat{\sigma}_n \xi}\right) \rightsquigarrow \mathcal{N}\left(-\delta^t \left.\frac{u^2(1+\xi)(1+2\xi)}{\xi(u+1+\xi)(u+1)^2}\right|_{u=1,\ldots,J}, \frac{(1+\xi)^2(1+2\xi)}{\xi^2}\right).$$

Now, in order to judge the quality of our test, we have to compute its power, which is, by definition, the probability that it will correctly lead to the rejection of a false null hypothesis. In other words, it is the ability of a test to detect an effect, if the effect actually exists.

### *4.2. Power of the test of the Pareto distribution*

In order to evaluate the power of our test statistic, what we have to do is to study the asymptotic behavior of the test statistic $\Psi_J^2$ under our local alternatives $\mathrm{H}_{1,n}$. This is the aim of our next proposition.

**Proposition 11.** *Under the assumption* $\mathrm{H}_{1,n}: \theta=\theta_n=(\delta/\sqrt{n})(1+\mathrm{o}(1))$ *where* $\delta$ *is defined as in* (2), *the statistic* $Z_n(\widehat{\beta}_n)$ *converges weakly to the normal distribution* $\mathcal{N}(\Sigma_J(\beta)\delta, \Sigma_J(\beta))$, *and* $\Psi_J^2 = Z_n^t(\widehat{\beta}_n)\Sigma_J^{-1}(\widehat{\beta}_n)Z_n(\widehat{\beta}_n)$ *converges weakly to a non-central chi-square distribution with J degrees of freedom and noncentrality parameter* $\delta^t\Sigma_J(\beta)\delta$.

Note that the noncentrality parameter partly depends on the distance between the alternatives and the null hypothesis.

## 5. Simulations

To study the finite sample behavior of the test statistic $\Psi_J^2$, we generated 1000 samples of iid observations from several distributions with sample sizes $n=20, 50, 100, 1000$ and $\delta_1=\cdots=\delta_J=\delta\in\{0, 0.1, 2, 5\}$.

We set $\xi=0.1$ and $\sigma=1$. In all our simulations, the estimator $\widehat{\beta}_n$ is constructed with the moment estimator $\widetilde{\beta}_n$ (defined in (11)) as the initial one. In case where $\delta\neq 0$, four different values of $J$ have been taken in order to see the impact of this factor. The solid line corresponds to $J=1$, the dashed one is used for $J=2$, the dotted one for $J=3$ and 4 is represented by the dotted-dashed line.

Fig. 1 exemplifies the numerous simulations. It shows quantile plots of 1000 independent realizations of the asymptotic $p$-value $p=1-\chi_J^2(\Psi_J^2)$ corresponding to the test statistic $\Psi_J^2$, where $\chi_J^2$ denotes the df of the chi-square distribution with $J$ degrees of freedom. The 1000 $p$-values were ordered on the $Y$-axis, $p_{1:1000}\leqslant\cdots\leqslant p_{1000:1000}$ and the points $(i/1001, p_{i:1000})$, $i\leqslant 1000$, were plotted. The horizontal line is drawn at the 5%-level, since a realization of the $p$-value below 5% usually leads to a rejection of the null hypothesis.

We observe that the power increases with $J$, with a noticeable difference for a large $\delta$ (typically $\delta=5$) but not for a small one close to 0 (corresponding to the null hypothesis). Nevertheless the problem of selecting an optimal value for this parameter is outside the scope of this paper. This problem has already been studied in the literature, see for instance Thomas and Pierce (1979) or Kopecky and Pierce (1979) for some specific values as an empirical basis, whereas Inglot et al. (1997) recommended a selection rule.

The $p$-value is (asymptotically as the sample size increases) uniformly distributed on (0, 1) under the null hypothesis $\delta=0$. The quantile plot should in this case be close to the straight line connecting the points (0, 0) and (1, 1). The corresponding plots in Fig. 1 show that this is actually true, with a small tendency to reject the null hypothesis in case of a very small sample size ($n=20$). As expected, the power of the test increases with $\delta$. This is visualized by the quantile plots of the $p$-value which tend to be below the 5% line as $\delta$ increases.

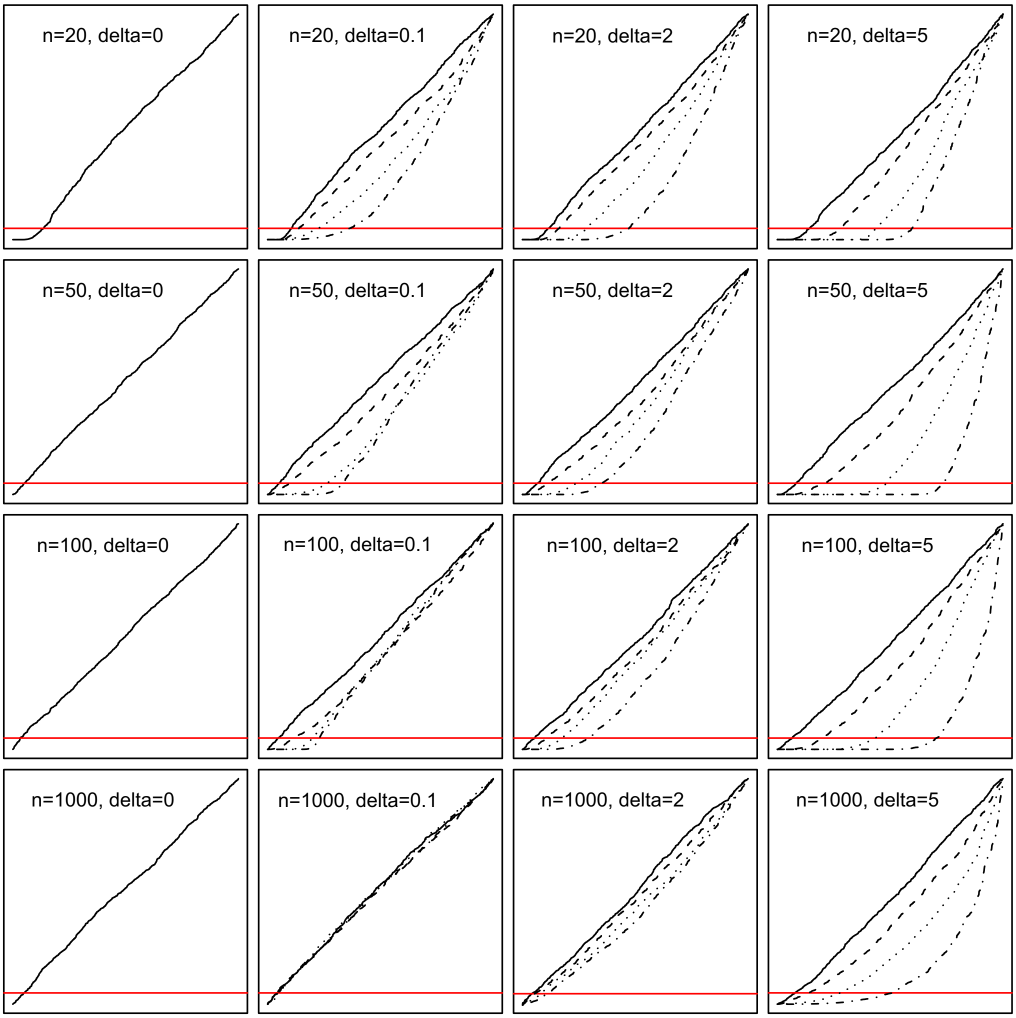


Fig. 1. Quantile plots of 1000 $p$-values of the test statistic $\Psi_J^2$. Here $\xi = 0.1$, $\sigma = 1$ and $\delta_1 = \cdots = \delta_J = \delta \in \{0, 0.1, 2, 5\}$. The solid line corresponds to $J = 1$, the dashed one is used for $J = 2$, the dotted one for $J = 3$ and 4 is represented by the dotted-dashed line. The horizontal line is always drawn at the 5% level and the $X$- and $Y$-axes from 0 to 1.

## 6. Concluding remarks

In this paper a smooth goodness of fit test for the Pareto distribution family has been proposed. It is motivated by Le Cam's LAN theory. Since the Pareto distribution corresponds to the heavy tailed generalized Pareto distribution, it is an essential model for extreme events. A classical approach in the extreme framework is the peaks over threshold (POT) method. It relies on the fact that exceedances over high thresholds in an iid sample follow approximately a generalized Pareto distribution for a large class of underlying distributions. Therefore an adaptation of our test to the case of "approximately Pareto distribution" would be useful and will lead to further research. Equally, according to an idea of Davison and Smith (1990), a further application would be the use of this test to select the threshold in the POT approach.

## Acknowledgment

The authors are grateful to the referees for their constructive comments from which the paper has benefitted a lot.

## Appendix

**Proof of Proposition 4.** This proof is based on Theorem 7.5.9. in Pfanzagl (1994). It can be divided into two parts.

*First part*: We have to check five conditions which are denoted in Pfanzagl's (1994) book by conditions 7.5.1, 7.5.2, 7.5.3, (i) and (ii) in Theorem 7.5.5. The first one consists in checking that $\mathbb{E}_{H_0,\beta}(\ell^{(\cdot)}(X,\beta))=0$ with $\beta\in\Theta$. To satisfy the second one we have to verify that

$$\mathbb{E}_{H_0,\beta}(\ell^{(\mu\nu)}(X,\beta))+\mathbb{E}_{H_0,\beta}(\ell^{(\mu)}(X,\beta)\ell^{(\nu)}(X,\beta))=0 \quad \text{for } \mu,\nu=1,2 \text{ and } \beta\in\Theta.$$

The third condition requires that the functions $\ell^{(\nu)}(.,\beta)$ for $\nu=1,2$ and $\beta\in\Theta$, are linearly $\mathbb{P}_{H_0,\beta}$-independent.

It is easy to see that these three conditions 7.5.1–7.5.3 are satisfied.

It remains for the first part to show that the following two conditions (i) and (ii) are satisfied for every $\beta_0\in\Theta$.

*Condition* (i) *in Theorem* 7.5.5 *of* Pfanzagl (1994): We have to check that there exists a neighborhood $U(\sigma_0,\xi_0)$ of $\beta_0$ such that

$$\beta\mapsto\ell^{(\mu\nu)}(x,\beta) \text{ is continuous on } U(\sigma_0,\xi_0) \quad \text{for } \mu,\nu=1,2 \text{ for every } x$$

and $\sup_{\beta_1\in U(\sigma_0,\xi_0)}|\ell^{(\mu\nu)}(.,\beta_1)|$ is integrable with respect to the Pareto distribution $(\beta)$ locally uniformly at $\beta_0$.

Note that in order to establish the latter point, it is sufficient to prove that

$$\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)}\mathbb{E}_{H_0,\beta}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}|\ell^{(\mu\nu)}(X,\beta_1)|^{1+\gamma}\right)<\infty$$

for some $\gamma>0$, see Pfanzagl (1994, p. 215).

In view of (4), (5) and (6), it is enough to consider the case of $\ell^{(22)}(X,\beta_1)$. We have

$$\begin{aligned}&\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)}\mathbb{E}_{H_0,\beta}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}|\ell^{(22)}(X,\beta_1)|^{1+\gamma}\right)\\&\quad=\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)}\mathbb{E}_{H_0,\beta}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\left|\frac{\partial^2}{\partial\xi_1\partial\xi_1}\log f(X,\beta_1)\right|^{1+\gamma}\right).\end{aligned}$$

With the change of variable $X=(\sigma/\xi)(U^{-\xi}-1)$, where $U$ is a uniformly distributed rv on (0, 1), it is equivalent to study

$$\begin{aligned}&\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)}\mathbb{E}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\left|\frac{\partial^2}{\partial\xi_1\partial\xi_1}\log f\left(\frac{\sigma}{\xi}(U^{-\xi}-1),\beta_1\right)\right|^{1+\gamma}\right)\\&\quad=\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)}\mathbb{E}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\left|-\frac{2}{\xi_1^3}\log\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right.\right.\\&\qquad\left.\left.+\frac{2}{\xi_1^2\sigma_1}\frac{(\sigma/\xi)(U^{-\xi}-1)}{1-(\xi_1\sigma/\sigma_1\xi)(1-U^{-\xi})}+\frac{1+\xi_1}{\xi_1\sigma_1^2}\frac{(\sigma/\xi)^2(1-U^{-\xi})^2}{(1-(\xi_1\sigma/\sigma_1\xi)(1-U^{-\xi}))^2}\right|^{1+\gamma}\right).\end{aligned}$$

If we take $\gamma = 1$, we only need to consider the terms $\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\{(1-U^{-\xi})^4\}/\{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^4\}$, $\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\log^2\{1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi})\}$ and $\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\{\log\{1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi})\}\times\{(1-U^{-\xi})^2\}/\{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^2\}\}$.

Just like $\sigma$ and $\xi$, the parameters $\sigma_1$ and $\xi_1$ are in the neighborhood of $\sigma_0$ and $\xi_0$. Consequently, we can assume that there exists $\varepsilon$ such that $\varepsilon\in(0,1)$ and $(\xi_1\sigma)/(\sigma_1\xi)\in(1-\varepsilon,1+\varepsilon)$.

Concerning the first term, we have, with $\varepsilon$ fixed, that

$$\begin{aligned} C_1 &= \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\left\{\frac{1-U^{-\xi}}{1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi})}\right\}^4\right) \\ &\leqslant \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\left\{\frac{1-U^{-\xi}}{1-(1-\varepsilon)(1-U^{-\xi})}\right\}^4\right). \end{aligned}$$

For $u\in(0,1)$, let $h_{1,\varepsilon}(u):=\{1-u^{-\xi}\}/\{1-(1-\varepsilon)(1-u^{-\xi})\}$. Since $\xi$ is positive, $h_{1,\varepsilon}(.)$ is a negative and increasing function.

Since $\varepsilon\in(0,1)$, we have

$$\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\left\{\frac{1-U^{-\xi}}{1-(1-\varepsilon)(1-U^{-\xi})}\right\}^4\right) \leqslant h_{1,\varepsilon}^4(0) = \frac{1}{(1-\varepsilon)^4} < \infty.$$

Concerning the second term,

$$\begin{aligned} C_2 &= \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \log^2\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right) \\ &\leqslant \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}(\log^2(1-(1+\varepsilon)(1-U^{-\xi}))) \\ &= \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}([-\xi\log U+\log(1+\varepsilon(-U^{\xi}+1))]^2) \\ &\leqslant \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}([-\xi\log U+\log(1+\varepsilon)]^2) \\ &\leqslant \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} (2\xi^2+2\xi\log(1+\varepsilon))+\log^2(1+\varepsilon) \\ &< \infty. \end{aligned}$$

Concerning the third term,

$$\begin{aligned} C_3 &= \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)}\left\{\log\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right\}\left\{\frac{(1-U^{-\xi})^2}{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^2}\right\}\right) \\ &\leqslant \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \log\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right\}\right. \\ &\qquad \left.\times\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \frac{(1-U^{-\xi})^2}{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^2}\right\}\right). \end{aligned}$$

By applying the Cauchy–Schwarz inequality, we obtain that

$$
\begin{aligned}
C_3 \leqslant & \sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \left( \sqrt{\mathbb{E}\left(\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \log\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right\}^2\right)} \right.\\
& \left. \times \sqrt{\mathbb{E}\left(\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \frac{(1-U^{-\xi})^2}{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^2}\right\}^2\right)}\right)\\
\leqslant & \sqrt{\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \log\left(1-\frac{\xi_1\sigma}{\sigma_1\xi}(1-U^{-\xi})\right)\right\}^2\right)}\\
& \times \sqrt{\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}\left(\left\{\sup_{(\sigma_1,\xi_1)\in U(\sigma_0,\xi_0)} \frac{(1-U^{-\xi})^2}{(1-(\xi_1\sigma/(\sigma_1\xi))(1-U^{-\xi}))^2}\right\}^2\right)}.
\end{aligned}
$$

Consequently, since $C_1$ and $C_2$ are finite, $C_3$ is also finite.

*Condition* (ii) *in Theorem* 7.5.5 *of* Pfanzagl (1994): For $v=1,2$, the family of functions $x\mapsto[\ell^{(v)}(x,\beta)]^2$ must be integrable with respect to the Pareto distribution with parameter $\beta$, locally uniformly at $\beta_0$, i.e. we have to check that for some $\gamma>0$,

$$\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}_{\mathrm{H}_0,\beta}(|\ell^{(v)}(X,\beta)|^{2(1+\gamma)})<\infty.$$

Let $\gamma=1/2$. We have

$$[\ell^{(1)}(X,\beta)]^3=-\frac{1}{\sigma^3}+\frac{(1+\xi)^3}{\sigma^6}X^3\left(1+\frac{\xi X}{\sigma}\right)^{-3}-\frac{3(1+\xi)^2}{\sigma^5}X^2\left(1+\frac{\xi X}{\sigma}\right)^{-2}+\frac{3(1+\xi)}{\sigma^4}X\left(1+\frac{\xi X}{\sigma}\right)^{-1}.$$

With $X=(\sigma/\xi)(U^{-\xi}-1)$, $U\sim\mathrm{U}(0,1)$, we obtain that $[\ell^{(1)}(X,\beta)]^3$ is also equal to

$$-\frac{1}{\sigma^3}-\frac{(1+\xi)^3}{\xi^3\sigma^3}(U^{\xi}-1)^3-\frac{3(1+\xi)^2}{\xi^2\sigma^3}(U^{\xi}-1)^2-\frac{3(1+\xi)}{\xi\sigma^3}(U^{\xi}-1).$$

Similarly, we have

$$
\begin{aligned}
[\ell^{(2)}(X,\beta)]^3=&\frac{1}{\xi^6}\log^3\left(1+\frac{\xi X}{\sigma}\right)-\frac{(1+\xi)^3}{\sigma^3\xi^3}X^3\left(1+\frac{\xi X}{\sigma}\right)^{-3}+\frac{3(1+\xi)^2}{\sigma^2\xi^4}X^2\left(1+\frac{\xi X}{\sigma}\right)^{-2}\log\left(1+\frac{\xi X}{\sigma}\right)\\
&-\frac{3(1+\xi)}{\sigma\xi^5}X\left(1+\frac{\xi X}{\sigma}\right)^{-1}\log^2\left(1+\frac{\xi X}{\sigma}\right)
\end{aligned}
$$

which is, by the same change of variable, equal to

$$-\frac{1}{\xi^3}\log^3(U)+\frac{(1+\xi)^3}{\xi^6}(U^{\xi}-1)^3-\frac{3(1+\xi)^2}{\xi^5}(U^{\xi}-1)^2\log(U)+\frac{3(1+\xi)}{\xi^4}(U^{\xi}-1)\log^2(U).$$

Using Hölder's inequalities, we get

$$
\begin{aligned}
&\mathbb{E}((U^{\xi}-1)^2\log(U))\leqslant(\mathbb{E}(U^{\xi}-1)^{2(3/2)})^{2/3}\times(\mathbb{E}(\log^3(U)))^{1/3},\\
&\mathbb{E}((U^{\xi}-1)\log^2(U))\leqslant(\mathbb{E}(U^{\xi}-1)^3)^{1/3}\times(\mathbb{E}(\log^{2(3/2)}(U)))^{2/3}.
\end{aligned}
$$

Since $\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}(U^{\xi}-1)^3$ and $\sup_{(\sigma,\xi)\in V(\sigma_0,\xi_0)} \mathbb{E}(\log^3(U))$ are finite, condition (ii) is fulfilled. This completes the first part of the proof of Proposition 4.

*Second part*: We make use of the particular estimator $\widehat{\beta}_n$ of $\beta$ given in Eq. (8). A direct application of Theorem 7.5.9 in Pfanzagl (1994) leads to

$$\sqrt{n}(\widehat{\beta}_n-\beta)-I_{\beta\beta}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^{(.)}(X_i,\beta)\to 0 \quad \text{locally uniformly at } \beta_0.$$

Moreover, since $(1/\sqrt{n})\sum_{i=1}^{n}\ell^{(.)}(X_i,\beta)\rightsquigarrow\mathcal{N}(0,I_{\beta\beta})$ locally uniformly at $\beta_0$ by the corresponding version of the Central Limit Theorem (see Pfanzagl, 1994, 7.7.11), we obtain

$$\sqrt{n}(\widehat{\beta}_n-\beta)\rightsquigarrow\mathcal{N}(0,I_{\beta\beta}^{-1}) \quad \text{locally uniformly at } \beta_0.$$

Since these two convergences hold for any $\beta_0$ in $\Theta$, the proof of Proposition 4 is complete. □

**Proof of Proposition 5.** Let $X_1,\dots,X_n$ be iid rvs from $F(.,\beta)$. Under $H_0$ we have $X_i=F^{-1}(1-U_i,\beta)=(\sigma/\xi)(U_i^{-\xi}-1)$, $i=1,\dots,n$, where $U_1,\dots,U_n$ are iid rvs from the uniform distribution on (0, 1). Then

$$\begin{aligned}
Z_n(\widehat{\beta}_n) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(1-F(X_i,\widehat{\beta}_n))^s-\frac{1}{s+1}\right]\Bigg|_{s=1,\dots,J}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(1-F(X_i,\beta))^s-\frac{1}{s+1}+(1-F(X_i,\widehat{\beta}_n))^s-(1-F(X_i,\beta))^s\right]\Bigg|_{s=1,\dots,J}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[U_i^s-\frac{1}{s+1}+\left(U_i^{-\xi}+1-U_i^{-\xi}-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}(1-U_i^{-\xi})\right)^{-s/\widehat{\xi}_n}-(U_i^{-\xi})^{-s/\xi}\right]\Bigg|_{s=1,\dots,J}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[U_i^s-\frac{1}{s+1}\right]\Bigg|_{s=1,\dots,J}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}-(U_i^{-\xi})^{-s/\xi}\right]\Bigg|_{s=1,\dots,J}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\widehat{\xi}_n}\right.\\
&\quad\left.-\left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}\right]\Bigg|_{s=1,\dots,J}\\
&=: A+B+C.
\end{aligned}$$

We will study the three terms separately. First, concerning $B$, we have

$$\begin{aligned}
B &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}-(U_i^{-\xi})^{-s/\xi}\right]\Bigg|_{s=1,\dots,J}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^s\left[\left(1+(U_i^{\xi}-1)\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}-1\right]\Bigg|_{s=1,\dots,J}.
\end{aligned}$$

Using a Taylor expansion with $f(x)=(1+x)^{\alpha}$ and then applying the law of large numbers, we obtain

$$\begin{aligned} B &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^s\left[\frac{s}{\xi}(1-U_i^{\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right]\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \sqrt{n}\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\frac{s}{(s+\xi+1)(s+1)}\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1). \end{aligned}$$

Concerning the term C, we have by a Taylor expansion of $f(x)=(1+x)^{\alpha}$

$$\begin{aligned} C &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\widehat{\xi}_n} - \left(U_i^{-\xi}+(1-U_i^{-\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}\right]\Bigg|_{s=1,\dots,J} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[U_i^{s\xi/\widehat{\xi}_n}\left(1+(U_i^{\xi}-1)\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\widehat{\xi}_n} - U_i^s\left(1+(U_i^{\xi}-1)\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)^{-s/\xi}\right]\Bigg|_{s=1,\dots,J} \\ &= \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^{s\xi/\widehat{\xi}_n}\left(1+\frac{s}{\widehat{\xi}_n}(1-U_i^{\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)\right. \\ &\quad \left.-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^s\left(1+\frac{s}{\xi}(1-U_i^{\xi})\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right)\right]\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[U_i^{s\xi/\widehat{\xi}_n}-U_i^s] + \sqrt{n}\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\frac{1}{n}\sum_{i=1}^{n}U_i^{\xi}\left[-\frac{s}{\widehat{\xi}_n}\exp\left(\frac{s}{\widehat{\xi}_n}\xi\log U_i\right)+\frac{s}{\xi}\exp\left(\frac{s}{\xi}\xi\log U_i\right)\right]\right. \\ &\quad \left.-\sqrt{n}\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\frac{1}{n}\sum_{i=1}^{n}\left[-\frac{s}{\widehat{\xi}_n}\exp\left(\frac{s}{\widehat{\xi}_n}\xi\log U_i\right)+\frac{s}{\xi}\exp\left(\frac{s}{\xi}\xi\log U_i\right)\right]\right]\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1). \end{aligned}$$

Using Taylor expansions for the functions $f(x)=x\exp(-x\xi\log U_i)$ and $f(x)=\exp(x)$, we obtain that

$$\begin{aligned} C &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[U_i^{s\xi/\widehat{\xi}_n}-U_i^s]\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^s(U_i^{s(\xi/\widehat{\xi}_n-1)}-1)\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^s\left[\exp\left(s\left(\frac{\xi}{\widehat{\xi}_n}-1\right)\log(U_i)\right)-1\right]\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \sqrt{n}s\left(\frac{\xi}{\widehat{\xi}_n}-1\right)\frac{1}{n}\sum_{i=1}^{n}U_i^s\log(U_i)\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\ &= \frac{s}{(s+1)^2}\frac{1}{\widehat{\xi}_n}\sqrt{n}(\widehat{\xi}_n-\xi)\Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \end{aligned}$$

by the law of large numbers.

Summarizing the above expansions we obtain

$$
\begin{aligned}
Z_n(\widehat{\beta}_n) &= \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^s - \frac{1}{s+1} \right) + \sqrt{n} \left( 1 - \frac{\widehat{\xi}_n \sigma}{\widehat{\sigma}_n \xi} \right) \frac{s}{(s+\xi+1)(s+1)} \right. \\
&\quad \left. + \frac{s}{(s+1)^2} \frac{1}{\widehat{\xi}_n} \sqrt{n}(\widehat{\xi}_n - \xi) \right] \Bigg|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&=: [A'(s) + B'(s) + C'(s)]|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1). \qquad (13)
\end{aligned}
$$

Next we compute the asymptotic variance of $Z_n(\widehat{\beta}_n)$. In what follows, the asymptotic expectation, variance and covariance will be denoted by $\mathbb{E}_a$, $\mathbb{V}\mathrm{ar}_a$ and $\mathbb{C}\mathrm{ov}_a$. We have

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}_a(A'(u), A'(v)) &= \mathbb{C}\mathrm{ov}_a \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^u - \frac{1}{u+1} \right), \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^v - \frac{1}{v+1} \right) \right) \\
&= \mathbb{C}\mathrm{ov}_a(U_i^u, U_i^v) = \frac{1}{u+v+1} - \frac{1}{(u+1)(v+1)} \\
&= \frac{uv}{(u+v+1)(u+1)(v+1)}.
\end{aligned}
$$

Note that

$$
\mathbb{C}\mathrm{ov}_a(A'(u), A'(v))|_{u,v=1,\dots,J} = \left. \frac{\partial^2 K(\theta)}{\partial\theta\partial\theta} \right|_{\theta=0} = I. \qquad (14)
$$

Similarly,

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}_a(B'(u), B'(v)) &= \mathbb{C}\mathrm{ov}_a \left( \sqrt{n} \left( 1 - \frac{\widehat{\xi}_n \sigma}{\widehat{\sigma}_n \xi} \right) \frac{u}{(u+\xi+1)(u+1)}, \sqrt{n} \left( 1 - \frac{\widehat{\xi}_n \sigma}{\widehat{\sigma}_n \xi} \right) \frac{v}{(v+\xi+1)(v+1)} \right) \\
&= \mathbb{V}\mathrm{ar}_a \left( \sqrt{n} \left( 1 - \frac{\widehat{\xi}_n \sigma}{\widehat{\sigma}_n \xi} \right) \right) \frac{uv}{(u+\xi+1)(v+\xi+1)(u+1)(v+1)} \\
&= \mathbb{V}\mathrm{ar}_a \left( \sqrt{n} \left( \frac{\xi(\widehat{\sigma}_n - \sigma) - (\widehat{\xi}_n - \xi)\sigma}{\xi \widehat{\sigma}_n} \right) \right) \frac{uv}{(u+\xi+1)(v+\xi+1)(u+1)(v+1)}.
\end{aligned}
$$

From (9) in Proposition 4, we derive that

$$
\begin{aligned}
\sqrt{n}(\widehat{\xi}_n - \xi) &= \frac{1}{\sqrt{n}} \left( \frac{1}{\xi^2}(1+\xi)^2(1+2\xi) \sum_{i=1}^n (U_i^\xi - 1) - \frac{1}{\xi}(1+\xi)^2 \sum_{i=1}^n \log(U_i) + n(1+\xi) \right) \\
&\quad + o_{\mathbb{P}_{H_0,\beta}}(1) \qquad (15)
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{n}(\widehat{\sigma}_n - \sigma) &= \frac{1}{\sqrt{n}} \left( -\frac{\sigma}{\xi^2}(1+\xi)^2(1+2\xi) \sum_{i=1}^n (U_i^\xi - 1) + \frac{\sigma}{\xi}(1+\xi) \sum_{i=1}^n \log(U_i) - 2n\sigma(1+\xi) \right) \\
&\quad + o_{\mathbb{P}_{H_0,\beta}}(1). \qquad (16)
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
\sqrt{n} \left( \frac{\xi(\widehat{\sigma}_n - \sigma) - (\widehat{\xi}_n - \xi)\sigma}{\xi \widehat{\sigma}_n} \right) &= \frac{1}{\sqrt{n}} \frac{1}{\xi \widehat{\sigma}_n} \left[ (1+\xi)^3(1+2\xi) \frac{\sigma}{\xi^2} \sum_{i=1}^n (1 - U_i^\xi) \right. \\
&\quad \left. + \frac{\sigma}{\xi}(1+\xi)(1+2\xi) \sum_{i=1}^n \log(U_i) - n\sigma(1+\xi)(1+2\xi) \right] + o_{\mathbb{P}_{H_0,\beta}}(1). \qquad (17)
\end{aligned}
$$

Direct computations lead to the following expression:

$$\mathbb{C}\mathrm{ov}_a(B'(u), B'(v)) = \frac{(1+\xi)^2(1+2\xi)}{\xi^2}\frac{uv}{(u+\xi+1)(v+\xi+1)(u+1)(v+1)}.$$

As the variance of $\sqrt{n}(\widehat{\xi}_n - \xi)$ is $(1+\xi)^2$ according to Proposition 4, we have

$$\mathbb{C}\mathrm{ov}_a(C'(u), C'(v)) = \frac{uv(1+\xi)^2}{\xi^2(u+1)^2(v+1)^2}.$$

We have now to compute the covariances between the terms $A'$, $B'$ and $C'$.

Using (15) and (17), we obtain

$$\mathbb{C}\mathrm{ov}_a\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_i^s - \frac{1}{s+1}\right), \sqrt{n}(\widehat{\xi}_n - \xi)\right) = \frac{s(1+\xi)(s-\xi)}{(s+1)^2(s+\xi+1)},$$

$$\mathbb{C}\mathrm{ov}_a\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_i^s - \frac{1}{s+1}\right), \sqrt{n}\left(1 - \frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\right) = -\frac{s^2(1+\xi)(1+2\xi)}{\xi(s+1)^2(s+\xi+1)},$$

$$\mathbb{C}\mathrm{ov}_a\left(\sqrt{n}\left(1 - \frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right), \sqrt{n}(\widehat{\xi}_n - \xi)\right) = -\frac{1}{\xi}(1+\xi)(1+2\xi).$$

It follows that

$$\mathbb{C}\mathrm{ov}_a(A'(u), B'(v)) + \mathbb{C}\mathrm{ov}_a(A'(v), B'(u)) = -\frac{u^2v(1+\xi)(1+2\xi)}{\xi(u+1)^2(u+\xi+1)(v+\xi+1)(v+1)} - \frac{v^2u(1+\xi)(1+2\xi)}{\xi(v+1)^2(v+\xi+1)(u+\xi+1)(u+1)},$$

$$\mathbb{C}\mathrm{ov}_a(A'(u), C'(v)) + \mathbb{C}\mathrm{ov}_a(A'(v), C'(u)) = \frac{uv(1+\xi)(2uv+u+v-2\xi^2-2\xi)}{\xi(u+1)^2(v+1)^2(u+\xi+1)(v+\xi+1)},$$

$$\mathbb{C}\mathrm{ov}_a(B'(u), C'(v)) + \mathbb{C}\mathrm{ov}_a(B'(v), C'(u)) = -\frac{uv(1+\xi)(1+2\xi)[(u+1)(v+\xi+1)+(v+1)(u+\xi+1)]}{\xi^2(v+\xi+1)(u+\xi+1)(u+1)^2(v+1)^2}.$$

Finally, the asymptotic variance of $Z_n(\widehat{\beta}_n)$ follows:

$$\mathbb{V}\mathrm{ar}_a(Z_n(\widehat{\beta}_n)) = \left[\frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\xi)(uv+\xi+(u+1)(v+1))}{(v+\xi+1)(u+\xi+1)(u+1)^2(v+1)^2}\right]\Bigg|_{u,v=1,\dots,J}.$$

Combining the Central Limit Theorem with the representations (13), (15) and (17), we deduce that

$$Z_n(\widehat{\beta}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_J(\beta)).$$

By (10) and continuity the proof of Proposition 5 is complete. □

**Proof of Proposition 7.** Let $\mu_r = \mathbb{E}_{\mathrm{H}_0,\beta}(X^r) < \infty$, $r = 1, 2, 3, 4$. We have

$$\mu_1 = \frac{\sigma}{1-\xi},$$

$$\mu_2 = \frac{2\sigma^2}{(1-\xi)(1-2\xi)},$$

$$\mu_3 = \frac{6\sigma^3}{(1-\xi)(1-2\xi)(1-3\xi)},$$

$$\mu_4 = \frac{24\sigma^4}{(1-\xi)(1-2\xi)(1-3\xi)(1-4\xi)}.$$

We can express the parameter $\beta$ as a function of $\mu_1$ and $\mu_2$.

To this aim, we write $\beta[v]=h_v(\mu_1,\mu_2)$ with $h_1:(0,\infty)\times(0,\infty)\mapsto(0,\infty)$ and $h_2:(0,\infty)\times(0,\infty)\mapsto(0,\infty)$. Then, we have

$$\beta[1]=\sigma=h_1(\mu_1,\mu_2)=\frac{1}{2}\mu_1\left(\frac{\mu_1^2}{\mu_2-\mu_1^2}+1\right),$$
$$\beta[2]=\xi=h_2(\mu_1,\mu_2)=\frac{1}{2}\left(1-\frac{\mu_1^2}{\mu_2-\mu_1^2}\right)$$

with $\mu_2-\mu_1^2>0$. For every $(y_1,y_2)\in(0,\infty)\times(0,\infty)$, $y\mapsto h_v(y,y_2)$ and $y\mapsto h_v(y_1,y)$ are two continuous functions.

Since by the strong law of large numbers, $(1/n)\sum_{i=1}^n X_i^r\overset{\text{a.s.}}{\to}\mathbb{E}_{\mathrm{H}_0,\beta}(X^r)=\mu_r$ for $r=1,2$, the moment estimator $\widetilde{\beta}_n[v]=h_v((1/n)\sum_{i=1}^n X_i,(1/n)\sum_{i=1}^n X_i^2)$ for $v=1,2$ defines a sequence of estimators which is strongly consistent.

The partial derivatives of $h_1$ and $h_2$ exist, they are continuous and $H(\sigma,\xi)$, defined as

$$H(\sigma,\xi)=\frac{\partial h_\mu}{\partial y_v}(y_1,y_2)\bigg|_{(y_1,y_2)=(\sigma/(1-\xi),2\sigma^2/[(1-\xi)(1-2\xi)])},\quad \mu,v=1,2$$

has rank 2. Therefore, by Pfanzagl (1994, Propositions 7.6.8 and 7.2.1) we have the locally uniformly $\sqrt{n}$-consistency of $\widetilde{\beta}_n[v]$. More precisely, we have

$$\sqrt{n}(\widetilde{\beta}_n-\beta)\rightsquigarrow\mathcal{N}(0,H(\beta)\Sigma(\beta)H(\beta)^t)\quad\text{locally uniformly on }\mathscr{B}$$

with $H(\beta)$ and $\Sigma(\beta)$ defined as follows:

$$H(\beta)=\begin{pmatrix}(3-4\xi)(1-\xi) & -\dfrac{1}{2\sigma}(1-\xi)(1-2\xi)^2\\ \dfrac{2}{\sigma}(1-\xi)^2(1-2\xi) & -\dfrac{1}{2\sigma^2}(1-\xi)^2(1-2\xi)^2\end{pmatrix},$$

$$\Sigma(\beta)=\mathbb{V}\mathrm{ar}_{\mathrm{H}_0,\beta}\begin{pmatrix}X\\X^2\end{pmatrix}=\begin{pmatrix}\mu_2-\mu_1^2 & \mu_3-\mu_1\mu_2\\ \mu_3-\mu_1\mu_2 & \mu_4-\mu_2^2\end{pmatrix}$$
$$=\begin{pmatrix}\dfrac{\sigma^2}{(1-\xi)^2(1-2\xi)} & \dfrac{4\sigma^3}{(1-\xi)^2(1-2\xi)(1-3\xi)}\\ \dfrac{4\sigma^3}{(1-\xi)^2(1-2\xi)(1-3\xi)} & \dfrac{20\sigma^4-44\xi\sigma^4}{(1-\xi)^2(1-2\xi)^2(1-3\xi)(1-4\xi)}\end{pmatrix}.$$

Now, in order to establish our results under the alternatives $\mathrm{H}_{1,n}$, we use LeCam's First and Third Lemma (see van der Vaart, 1998, 6.4 and 6.7). By $\overset{\mathrm{H}_0}{\rightsquigarrow}$ (respectively, $\overset{\mathrm{H}_{1,n}}{\rightsquigarrow}$) we denote weak convergence under $\mathrm{H}_0$ (respectively, under $\mathrm{H}_{1,n}$). □

**LeCam's First Lemma.** *Let $P_n$ and $Q_n$ be sequences of probability measures on measurable spaces $(\Omega_n,\mathscr{A}_n)$. Then the following statements are equivalent*:

(i) *$Q_n\lhd P_n$, i.e. $Q_n$ is contiguous with respect to $P_n$.*
(ii) *There exists some subsequence $m(n)$ such that if $\mathrm{d}P_{m(n)}/\mathrm{d}Q_{m(n)}\overset{Q_{m(n)}}{\rightsquigarrow}U$, then $\mathbb{P}(U>0)=1$.*
(iii) *There exists some subsequence $m(n)$ such that if $\mathrm{d}Q_{m(n)}/\mathrm{d}P_{m(n)}\overset{P_{m(n)}}{\rightsquigarrow}V$, then $\mathbb{E}(V)=1$.*
(iv) *For any statistic $T_n:\Omega_n\mapsto\mathbb{R}^k$ : if $T_n\overset{P_n}{\to}0$, then $T_n\overset{Q_n}{\to}0$.*

**LeCam's Third Lemma.** *Let $P_n$ and $Q_n$ be sequences of probability measures on measurable spaces $(\Omega_n, \mathscr{A}_n)$ and let $X_n : \Omega_n \mapsto \mathbb{R}^k$ be a sequence of random vectors.*

*Suppose that $Q_n \lhd P_n$ and*

$$\begin{pmatrix} X_n \\ \log \dfrac{\mathrm{d}Q_n}{\mathrm{d}P_n} \end{pmatrix} \overset{P_n}{\rightsquigarrow} \mathscr{N}_{k+1}\left(\begin{pmatrix} \mu \\ -\dfrac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^t & \sigma^2 \end{pmatrix}\right) \quad \textit{then } X_n \overset{Q_n}{\rightsquigarrow} \mathscr{N}_k(\mu+\tau, \Sigma).$$

**Proof of Proposition 8.** This result is a consequence of LeCam's First Lemma. By choosing $\mathrm{d}P_n = \prod_{i=1}^n f(X_i, \beta)$ (corresponding to the assumption $\mathrm{H}_0$) and $\mathrm{d}Q_n = \prod_{i=1}^n g_J(X_i, \theta_n, \beta)$ (corresponding to the alternatives $\mathrm{H}_{1,n}$), in view of (7) and (14), the Central Limit Theorem entails that $L_n$ converges weakly under $\mathrm{H}_0$ to $\mathscr{N}(-\frac{1}{2}\delta^t I\delta, \delta^t I\delta)$.

Now we can check statement (iii). Indeed, $\log V$ follows a $\mathscr{N}(-\frac{1}{2}\delta^t I\delta, \delta^t I\delta)$ distribution implying that $\mathbb{E}(V) = 1$. Moreover, in (ii), with the roles of $P_n$ and $Q_n$ interchanged, $\log U$ follows a $\mathscr{N}(-\frac{1}{2}\delta^t I\delta, \delta^t I\delta)$ distribution implying that $\mathbb{P}(U > 0) = 1$. Consequently, by (i), the sequences $P_n$ and $Q_n$ are mutually contiguous and this is denoted by $P_n \lhd\rhd Q_n$. Now by (iv), $T_n \overset{P_n}{\to} 0$ if and only if $T_n \overset{Q_n}{\to} 0$. $\square$

**Proof of Proposition 9.** By the multivariate Central Limit Theorem and by Proposition 4, we obtain that

$$\begin{pmatrix} Z_n(\beta) \\ \dfrac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \end{pmatrix} \overset{\mathrm{H}_0}{\rightsquigarrow} \mathscr{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I & I_\beta \\ I_\beta^t & I_{\beta\beta} \end{pmatrix}\right].$$

Under $\mathrm{H}_0$ the loglikelihood ratio $L_n$ converges weakly to the normal distribution $\mathscr{N}(-\frac{1}{2}\delta^t I\delta, \delta^t I\delta)$. Moreover, $L_n$ is essentially a function of $Z_n(\beta)$ (see (7)). Hence, the vector

$$\begin{pmatrix} Z_n(\beta) \\ \dfrac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \\ \log \prod_{i=1}^n \dfrac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)} \end{pmatrix}$$

is asymptotically Gaussian. Its asymptotic covariance matrix has the entries

$$\mathbb{C}\mathrm{ov}_a\left(\log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)}, Z_n(\beta)\right) = \mathbb{C}\mathrm{ov}\left(\delta^t Z_n(\beta) - \frac{1}{2}\delta^t I\delta, Z_n(\beta)\right) = \delta^t\, \mathbb{V}\mathrm{ar}(Z_n(\beta)) = \delta^t I,$$

$$\begin{aligned} \mathbb{C}\mathrm{ov}_a\left(\log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)}, \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta)\right) &= \mathbb{C}\mathrm{ov}\left(\delta^t Z_n(\beta) - \frac{1}{2}\delta^t I\delta, \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta)\right) \\ &= \delta^t\, \mathbb{C}\mathrm{ov}\left(Z_n(\beta), \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta)\right) \\ &= \delta^t I_\beta \end{aligned}$$

and, consequently, we have

$$\begin{pmatrix} Z_n(\beta) \\ \dfrac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \\ \log \prod_{i=1}^n \dfrac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)} \end{pmatrix} \overset{\mathrm{H}_0}{\rightsquigarrow} \mathscr{N}_{J+3}\left[\begin{pmatrix} 0 \\ 0 \\ -\dfrac{1}{2}\delta^t I\delta \end{pmatrix}, \begin{pmatrix} I & I_\beta & I\delta \\ I_\beta^t & I_{\beta\beta} & I_\beta^t\delta \\ \delta^t I & \delta^t I_\beta & \delta^t I\delta \end{pmatrix}\right].$$

By LeCam's Third Lemma, we obtain that

$$\begin{pmatrix} Z_n(\beta) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i,\beta) \end{pmatrix} \overset{\mathrm{H}_{1,n}}{\rightsquigarrow} \mathcal{N}_{J+2}\left[\begin{pmatrix} I\delta \\ I_\beta^t \delta \end{pmatrix}, \begin{pmatrix} I & I_\beta \\ I_\beta^t & I_{\beta\beta} \end{pmatrix}\right].$$

Now, using Proposition 4, the following asymptotic representation holds:

$$\sqrt{n}(\widehat{\beta}_n - \beta) = I_{\beta\beta}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i,\beta) + \mathrm{o}_{\mathbb{P}_{\mathrm{H}_0,\beta}}(1).$$

Also, by Proposition 8, this asymptotic representation is also true under $\mathrm{H}_{1,n}$, i.e.

$$\sqrt{n}(\widehat{\beta}_n - \beta) = I_{\beta\beta}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^{(\cdot)}(X_i,\beta) + \mathrm{o}_{\mathbb{P}_{\mathrm{H}_{1,n}}}(1).$$

Hence, the result follows. □

**Proof of Proposition 10.** This proof is based on the delta method with the function $h:\mathbb{R}^2 \mapsto \mathbb{R}$ defined as follows:

$$h\begin{pmatrix} x \\ y \end{pmatrix} = \frac{y\sigma}{x\xi}.$$

Denoting by $\dot{h}$ its gradient, we have

$$\dot{h}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y\sigma/(x^2\xi) \\ \sigma/(x\xi) \end{pmatrix}.$$

The delta method relies on the following development which is easily obtained using Proposition 9 and a Taylor expansion:

$$\sqrt{n}(h(\widehat{\beta}_n) - h(\beta)) = \dot{h}(\beta)^t\sqrt{n}(\widehat{\beta}_n - \beta) + \mathrm{o}_{\mathbb{P}_{\mathrm{H}_{1,n}}}(1). \tag{18}$$

Therefore, since we have $\sqrt{n}(\widehat{\beta}_n - \beta) \overset{\mathrm{H}_{1,n}}{\rightsquigarrow} \mathcal{N}(I_{\beta\beta}^{-1}I_\beta^t\delta, I_{\beta\beta}^{-1})$ by Proposition 9, we also get

$$\sqrt{n}(h(\widehat{\beta}_n) - h(\beta)) \overset{\mathrm{H}_{1,n}}{\rightsquigarrow} \mathcal{N}(\dot{h}(\beta)^t I_{\beta\beta}^{-1}I_\beta^t\delta, \dot{h}(\beta)^t I_{\beta\beta}^{-1}\dot{h}(\beta))$$

and therefore

$$\sqrt{n}\left(1 - \frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right) \overset{\mathrm{H}_{1,n}}{\rightsquigarrow} \mathcal{N}\left(-\delta^t \left.\frac{u^2(1+\xi)(1+2\xi)}{\xi(u+1+\xi)(u+1)^2}\right|_{u=1,\dots,J}, \frac{(1+\xi)^2(1+2\xi)}{\xi^2}\right). \qquad \square$$

**Proof of Proposition 11.** According to Proposition 8, the asymptotic representation of $Z_n(\widehat{\beta}_n)$ under $\mathrm{H}_{1,n}$ is the same as the one previously obtained, i.e.

$$\begin{aligned} Z_n(\widehat{\beta}_n) = Z_n(\beta) &+ \sqrt{n}\left(1 - \frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right) \left.\frac{s}{(s+\xi+1)(s+1)}\right|_{s=1,\dots,J} \\ &+ \sqrt{n}(\widehat{\xi}_n - \xi)\frac{1}{\widehat{\xi}_n}\left.\frac{s}{(s+1)^2}\right|_{s=1,\dots,J} + \mathrm{o}_{\mathbb{P}_{\mathrm{H}_{1,n}}}(1). \end{aligned} \tag{19}$$

By Propositions 9 and 10, we then obtain with $I$ defined in (12)

$$\mathbb{E}_{\mathrm{H}_{1,n}}(Z_n(\beta)) = I\delta = \left.\frac{uv}{(u+v+1)(u+1)(v+1)}\right|_{u,v=1,\ldots,J}\delta,$$

$$\mathbb{E}_{\mathrm{H}_{1,n}}\left(\sqrt{n}\left(1-\frac{\widehat{\xi}_n\sigma}{\widehat{\sigma}_n\xi}\right)\left.\frac{v}{(v+\xi+1)(v+1)}\right|_{v=1,\ldots,J}\right)$$

$$= -\left.\frac{u^2v(1+\xi)(1+2\xi)}{\xi(u+1+\xi)(v+1+\xi)(u+1)^2(v+1)}\right|_{u,v=1,\ldots,J}\delta,$$

$$\mathbb{E}_{\mathrm{H}_{1,n}}\left(\sqrt{n}(\widehat{\xi}_n-\xi)\left.\frac{v}{\widehat{\xi}_n(v+1)^2}\right|_{v=1,\ldots,J}\right) = \left.\frac{uv(1+\xi)(u-\xi)}{\xi(u+1+\xi)(u+1)^2(v+1)^2}\right|_{u,v=1,\ldots,J}\delta. \tag{20}$$

Direct computations lead to the following expression for the asymptotic expectation of $Z_n(\widehat{\beta}_n)$ under $\mathrm{H}_{1,n}$:

$$\mathbb{E}_{a;\mathrm{H}_{1,n}}(Z_n(\widehat{\beta}_n)) = \left[\frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\xi)(uv+\xi+(u+1)(v+1))}{(v+\xi+1)(u+\xi+1)(u+1)^2(v+1)^2}\right]\Bigg|_{u,v=1,\ldots,J}\delta$$
$$= \Sigma_J(\beta)\delta. \tag{21}$$

Using representation (18), (19) and Propositions 8 and 9, it is obvious to see that the covariance terms are unchanged under $\mathrm{H}_{1,n}$ or under $\mathrm{H}_0$. Indeed, looking at LeCam's Third Lemma, only expectation terms are affected. Consequently, we have $\mathbb{V}\mathrm{ar}_{a;\mathrm{H}_{1,n}}(Z_n(\widehat{\beta}_n)) = \Sigma_J(\beta)$. Now, from (19), we observe that the quantity $Z_n(\widehat{\beta}_n)$ can be decomposed into three terms plus a remainder one. According to Propositions 9 and 10 and Slutsky's lemma, the vector formed by these three terms is asymptotically Gaussian under $\mathrm{H}_{1,n}$. As a consequence, under $\mathrm{H}_{1,n}$, $Z_n(\widehat{\beta}_n)$ converges weakly to a $\mathcal{N}_J(\Sigma_J(\beta)\delta, \Sigma_J(\beta))$ distribution. Therefore, $Z_n^t(\widehat{\beta}_n)\Sigma_J^{-1}(\beta)Z_n(\widehat{\beta}_n) \overset{\mathrm{H}_{1,n}}{\rightsquigarrow} \chi^2_{J,\delta^t\Sigma_J(\beta)\delta}$. Then, by Proposition 9 and by continuity, the result follows. □

## References


Arnold, B.C., 1983. Pareto Distributions. International Co-operative Publishing House, Burtonsville, MD.

Arnold, B.C., Press, S.J., 1989. Bayesian estimation and prediction for Pareto data. J. Amer. Statist. Assoc. 84, 1079–1084.

Benktander, G., 1970. Schadenverteilung nach Grösse in der Nicht-Leben-Versicherung. Mitt. Verein. Schweiz. Versich.-Mathr. 70, 268–284.

Castillo, E., Hadi, A.S., 1997. Fitting the generalized Pareto distribution to data. J. Amer. Statist. Assoc. 92, 1604–1620.

Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit tests for the generalized Pareto distribution. Technometrics 43, 478–484.

Danielsson, J., de Vries, C.G., 1997. Tail index and quantile estimation with very high frequency data. J. Empir. Finance 4, 241–257.

Davison, A.C., 1984. Modelling excesses over high thresholds, with an application. In: Tiago de Oliveira, J. (Ed.), Statistical Extremes and Applications. Reidel, Dordrecht, pp. 461–482.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. J. Roy. Statist. Soc. Ser. B 52, 393–442.

Fisk, P.R., 1961. The graduation of income distributions. Econometrica 29, 171–185.

Gustafson, G., Fransson, A., 2005. The use of the Pareto distribution for fracture transmissivity assessment. Hydrogeology J. 14, 15–20.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339–349.

Inglot, T., Kallenberg, W.C.M., Ledwina, T., 1997. Data driven smooth tests for composite hypotheses. Ann. Statist. 25, 1222–1250.

Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. Adv. Water Res. 25, 1287–1304.

Kopecky, K.J., Pierce, D.A., 1979. Efficiency of smooth goodness-of-fit tests. J. Amer. Statist. Assoc. 74, 393–397.

LeCam, L., 1960. Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. Univ. California Publ. Statist. 3, 37–98.

Malik, H.J., 1966. Estimation of the parameters of the Pareto distribution. Metrika 15, 126–132.

Miller, F.L., Quesenberry, C.P., 1979. Power studies of tests for uniformity, II. Comm. Statist. B Simulation Comput. 8, 271–290.

Neyman, J., 1937. Smooth tests for goodness of fit. Skand. Aktuarietidskr. 20, 149–199.

Pfanzagl, J., 1994. Parametric Statistical Theory. Walter de Gruyter & Co., Berlin.

Porter, J.E., Coleman, J.W., Moore, A.H., 1992. Modified KS, AD, and C-VM tests for the Pareto distribution with unknown location and scale parameters. IEEE Trans. Reliability 41, 112–117.

Rayner, J.C.W., Best, D.J., 1986. Neyman-type smooth tests for location-scale families. Biometrika 73, 437–446.

Rayner, J.C.W., Best, D.J., 1989. Smooth Tests of Goodness of Fit. Oxford University Press, New York.

Rayner, J.C.W., Best, D.J., 1990. Smooth tests of goodness of fit, an overview. Internat. Statist. Rev. 58, 9–17.

Reiss, R.D., Thomas, M., 2007. Statistical Analysis of Extreme Values. third ed. Birkhäuser, Basel.

Smith, R.L., 1984. Threshold methods for sample extremes, with an application. In: Tiago de Oliveira, J. (Ed.), Statistical Extremes and Applications. Reidel, Dordrecht, pp. 621–638.

Strasser, H., 1985. Mathematical Theory of Statistics. De Gruyter, Berlin.

Thomas, D.R., Pierce, D.A., 1979. Neyman's smooth goodness-of-fit test when the hypothesis is composite. J. Amer. Statist. Assoc. 74, 441–445.

van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.