REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

Bias Reduced Peaks over Threshold Tail Estimation

Authors: Jan Beirlant *
– Department of Mathematics, LStat and LRisk, KU Leuven and
Department of Mathematical Statistics and Actuarial Science, Free State University
jan.beirlant@kuleuven.be

Gaonyalelwe Maribe
– Department of Statistics, University of Pretoria
g.maribe@up.ac.za

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement, CNRS, Université Paris-Saclay
Philippe.Naveau@lsce.ipsl.fr

Andréhette Verster
– Department of Mathematical Statistics and Actuarial Science, Free State University
verstera@ufs.ac.za

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Bias reduction in tail estimation has mainly been performed in case of Pareto-type models; see for instance Drees (1996), Peng (1998), Feuerverger and Hall (1999), Beirlant et al. (1999, 2002), Gomes and Martins (2002) and Caeiro et al. (2005, 2009). In that context, Beirlant et al. (2009) and Papastathopoulos and Tawn (2013) constructed distributional models that are based on second order rates of convergence for distributions of peaks over thresholds (POT). Such approach also allows to connect the tail and the bulk of the distribution.

  Bias reduction for all max-domains of attractions, i.e. without restricting to the Pareto-type case, received much less attention up to now. Here we extend the second-order refined POT approach started in Beirlant et al. (2009) providing a bias reduction technique for the classical generalized Pareto (GP) approximation for POTs. We consider parametric and nonparametric modelling of the second order component.

Key-Words:

- *Peaks over Threshold; Generalized Pareto distribution; Tail estimation; Mixture models.*

AMS Subject Classification:

- 62G32, 62F10, 62F15, 62J07.

*Corresponding author: Jan Beirlant, KU Leuven, Dept of Mathematics, Celestijnenlaan 200B, 3001 Heverlee,Belgium; Email: jan.beirlant@kuleuven.be

1. INTRODUCTION

Extreme value (EV) methodology starts from the assumption that the distribution of the available sample $X_1, X_2, \ldots, X_n$ belongs to the domain of attraction of a generalized extreme value distribution, i.e. there exists sequences $(b_n)_n$ and $(a_n > 0)_n$ such that as $n \to \infty$

$$\frac{\max(X_1, X_2, \ldots, X_n) - b_n}{a_n} \to_d Y_\xi, \tag{1.1}$$

where $\mathbb{P}(Y_\xi \leq y) = \exp(-(1+\xi y)^{-1/\xi})$, for some $\xi \in \mathbb{R}$ with $1 + \xi y > 0$. The parameter ξ is termed the extreme value index (EVI). It is well-known (see e.g. Beirlant et al., 2004, and de Haan and Ferreira, 2006) that (1.1) is equivalent to the existence of a positive function $t \mapsto \sigma_t$, such that

$$\mathbb{P}\left(\frac{X-t}{\sigma_t} > y | X > t\right) = \frac{\bar{F}(t + y\sigma_t)}{\bar{F}(t)} \to_{t \to x_+} \bar{H}_\xi^{GP}(y) = (1+\xi y)^{-1/\xi}, \tag{1.2}$$

where $\bar{F}(x) = \mathbb{P}(X > x)$ and x_+ denotes the endpoint of the distribution of X. The conditional distribution of $X - t$ given $X > t$ is called the peaks over threshold (POT) distribution, while $\bar{H}_\xi^{GP}$ is the survival function of the generalized Pareto distribution (GPD).

Estimation of ξ and tail quantities such as return periods is then based on fitting a GPD to the observed excesses $X - t$ given $X > t$. The main difficulty in such an EV application is the choice of the threshold t. Most often, the threshold t is chosen as one of the top data points $X_{n-k,n}$ for some $k \in \{1, 2, \ldots, n\}$ where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denotes the ordered sample. The parameters (ξ, σ) are then estimated by fitting the GPD $H_\xi^{GP}\left(\frac{y}{\sigma}\right)$ to the spacings $X_{n,n} - X_{n-k,n}, \ldots, X_{n-k+1,n} - X_{n-k,n}$.

The limit result in (1.2) requires t to be chosen as large as possible (or, equivalently, k as small as possible) for the bias in the estimation of ξ and other tail parameters to be limited. However, in order to limit the estimation variance, t should be as small as possible, i.e. the number of data points k used in the estimation should be as large as possible. Several adaptive procedures for choosing t or k have been proposed, but mainly in the Pareto-type case with $\xi > 0$, i.e. when

$$\bar{F}(x) = x^{-1/\xi}\ell(x), \tag{1.3}$$

for some slowly varying function ℓ, i.e. satisfying $\frac{\ell(yt)}{\ell(t)} \to 1$ as $t \to \infty$, for every $y > 1$. One then typically assumes a second-order specification of (1.3) of the type

$$\frac{\ell(yt)}{\ell(t)} - 1 = \delta_t\left(y^{-\beta} - 1\right), \tag{1.4}$$

where $\delta_t = \delta(t) = t^{-\beta}\tilde{\ell}(t)$, with $\beta > 0$ and $\tilde{\ell}$ slowly varying at infinity.

As an alternative, bias reduction techniques have been proposed in the Pareto-type case $\xi > 0$, among others in Feuerverger and Hall (1999), Beirlant et al. (1999, 2002) and Gomes and

Martins (2002). However while the bias is reduced, the variance is increased. In Caeiro et al. (2005, 2009) methods are proposed to limit the variance of bias-reduced estimators assuming a third-order slow variation model. These methods focus on the distribution of the log-spacings of high order statistics. Other construction methods for asymptotically unbiased estimators of $\xi > 0$ were introduced in Peng (1998) and Drees (1996).
Another approach consists of proposing penultimate limit distributions. In case $\xi > 0$, Beirlant et al. (2009) proposed an extension of the Pareto distribution (EPD) to approximate the tail probability of the POT distribution $\mathbb{P}\left(\frac{X}{t} > y | X > t\right)$ as $t \to \infty$:

$$\bar{H}^{EP}_{\xi,\delta,\rho}(y) = 1 - H^{EP}_{\xi,\delta,\rho}(y) = y^{-1/\xi}\left(1 + \delta_t\left((y^{-1/\xi})^{-\rho} - 1\right)\right),\ y > 1, \tag{1.5}$$

with δ_t satisfying $\delta_t \downarrow 0$ as $t \to \infty$ and $\rho = -\beta\xi$. In the literature, the second order parameter ρ typically is estimated externally with a different sequence of extreme order statistics than with ξ and δ, or it is given an appropriate 'canonical' value such as -1. We suppress the notation ρ from the extended distribution notation.

Fitting the extended Pareto distribution $H^{EP}_{\xi,\sigma}$ to the relative excesses $\{\frac{X_{n-j+1,n}}{X_{n-k,n}}, j = 1, \dots, k\}$ leads to estimates of ξ that are more stable as a function of k compared to the original ML estimator derived by Hill (1975)

$$\hat{\xi}^H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}},$$

which is obtained by fitting the Pareto distribution $H^{EP}_{\xi,0}$. Denoting the maximum likelihood estimators of ξ by $\hat{\xi}^{EP}_k$, it can indeed be shown under the assumption that the EP model for the excesses X/t is correct and that ρ is estimated consistently, that the asymptotic bias of $\hat{\xi}^{EP}_k$ is 0 as long as $k(k/n)^{-2\rho} \to \lambda \geq 0$ as $k, n \to \infty$, while the asymptotic bias of $\hat{\xi}^H_{k,n}$ is only 0 when $k(k/n)^{-2\rho} \to 0$. On the other hand, the asymptotic variance of $\hat{\xi}^{EP}_k$ equals $\left(\frac{1-\rho}{\rho}\right)^2 \frac{\xi^2}{k}$, where $\frac{\xi^2}{k}$ is the asymptotic variance of $\hat{\xi}^H_{k,n}$.

In case of a real-valued EVI, for the selection of an appropriate threshold or the construction of bias-reduced methods, only a few methods are available. Dupuis (1999) suggested a robust model validation mechanism to guide the threshold selection, assigning weights between 0 and 1 to each data point where a high weight means that the point should be retained since a GPD model is fitting it well. However, thresholding is required at the level of the weights and hence the method cannot be used in an unsupervised manner. Buitendag et al. (2019) present a ridge regression method to reduce the bias of the generalized Hill estimator proposed in Beirlant et al. (2005).

In this paper we concentrate on bias reduction when fitting the GPD to the distribution of POTs $X - t | X > t$ using maximum likelihood estimation. We hence extend the second-order refined POT approach based on $\bar{H}^{EP}_{\xi,\delta}$ from (1.5) to all max-domains of attraction. Here the corresponding basic second order regular variation theory can be found in Theorem 2.3.8 in de Haan and Ferreira (2006) stating that

$$\lim_{t \to x_+} \frac{\mathbb{P}(X - t > y\sigma_t | X > t) - (1 + \xi y)^{-1/\xi}}{\delta(t)} = (1 + \xi y)^{-1-1/\xi}\Psi_{\xi,\tilde{\rho}}((1 + \xi y)^{1/\xi}), \tag{1.6}$$

with $\delta(t) \to 0$ as $t \to x_+$ and $\Psi_{\xi,\tilde\rho}(x) = \frac{1}{\tilde\rho}\left(\frac{x^{\xi+\tilde\rho}-1}{\xi+\tilde\rho} - \frac{x^{\xi}-1}{\xi}\right)$ which for the cases $\xi = 0$ and $\tilde\rho = 0$ is understood to be equal to the limit as $\xi \to 0$ and $\tilde\rho \to 0$. We further allow more flexible second-order models than the ones arising from second-order regular variation theory such as in (1.6) using non-parametric modelling of the second-order component and the flexible semiparametric GP modelling introduced in Tencaliec et al. (2019). This newly proposed method can also be applied to the specific case of Pareto-type distributions.

In the next section we propose our extended GPD models, and detail the estimation methods. Some basic asymptotic results are provided in section 3. In the final section we discuss simulation results and some practical case studies.

2. TRANSFORMED AND EXTENDED GPD MODELS

In this paper we propose to approximate the POT distribution with an extended GPD model with survival function

$$(\mathcal{E}): \qquad \bar F_t^{EGP}(y) = \bar H_\xi^{GP}(\frac{y}{\sigma})\left\{1 + \delta_t B_\eta\left(\bar H_\xi^{GP}(\frac{y}{\sigma})\right)\right\},$$

where

- $\delta_t = \delta(t) \to 0$ as $t \to x_+$,
- $B_\eta(1) = 0$ and $\lim_{u\to 0} u^{1-\epsilon} B_\eta(u) = 0$ for every $0 < \epsilon < 1$,
- B_η is twice continously differentiable.

Here the parameter η represents a second order nuisance parameter. For negative δ-values one needs $\delta_t > \{\min_u (1 - \frac{d}{du}(uB_\eta(u))\}^{-1}$ to obtain a valid distribution.
Note that this model is a transformation model $G_t\left(\bar H_\xi^{GP}(\frac{y}{\sigma})\right)$ where the transformation function $G_t : (0,1) \to (0,1), u \mapsto u(1+\delta_t B_\eta(u))$ satisfies $\frac{G_t(u)}{u} \to 1$ as $t \to \infty$ for every $u \in (0,1)$ as follows from (1.2).
Also, model $(\mathcal{E})$ generalizes the EPD model (1.5) replacing the Pareto survival function $y^{-1/\xi}$ ($\xi > 0$) by the GPD survival function $\bar H_\xi^{GP}$ ($\xi \in \mathbb{R}$), and considering a general function $B_\eta(u)$.

We here detail *a parametric and non-parametric estimation procedure* for (ξ, σ) under $(\mathcal{E})$ based on excesses $Y_{j,k} = X_{n-j+1,n} - X_{n-k,n}$ $(j = 1, \ldots, k)$, while considering external estimation of the parameters in the B_η component of the model. In this we use the reparametrization (ξ, τ) with $\tau = \xi/\sigma$. Modelling the distribution of the exceedances Y with model $(\mathcal{E})$ leads to maximum likelihood estimators based on the excesses $Y_{j,k} = X_{n-j+1,n} - X_{n-k,n}$ $(j = 1, \dots, k)$:

$$(\hat{\xi}_k^E, \hat{\tau}_k^E, \hat{\delta}_k^E) \;=\; \text{argmax} \left\{ \sum_{j=1}^{k} \log\left(1 + \delta_k b_\eta((1+\tau Y_{j,k})^{-1/\xi})\right) \right.$$

$$\left. + \sum_{j=1}^{k} \log\{\frac{\tau}{\xi}(1+\tau Y_{j,k})^{-1-1/\xi}\} \right\} \tag{2.1}$$

with $b_\eta(u) = \frac{d}{du}(uB_\eta(u))$ for a given choice of B_η.
Estimates of small tail probabilities $\mathbb{P}(X > c)$ are then obtained through

$$\hat{\mathbb{P}}_k^E(X > c) = \frac{k}{n} \bar{H}^{GP}_{\hat{\xi}_k^E}\left(\frac{\hat{\tau}_k^E}{\hat{\xi}_k^E}(c - X_{n-k,n})\right)\left(1 + \hat{\delta}_k^E \hat{B}_\eta\left(\bar{H}^{GP}_{\hat{\xi}_k^E}(\frac{\hat{\tau}_k^E}{\hat{\xi}_k^E}(c - X_{n-k,n})\right)\right).$$

A general approach to choose the parameters contained in the B_η component can be to minimize the variance of the obtained estimates of ξ over $k = 2, \dots, n$. See also the simulation section 4.

A parametric approach (Ep). The second-order result (1.6) leads to the parametric choice $B_{\xi,\tilde{\rho}}(u) = \frac{u^\xi}{\tilde{\rho}}\left(\frac{u^{-\xi-\tilde{\rho}}-1}{\xi+\tilde{\rho}} - \frac{u^{-\xi}-1}{\xi}\right)$ in case $\xi + \tilde{\rho} \neq 0$ and $\xi \neq 0$.
Model $(\mathcal{E})$ allows for bias reduction in the estimation of (ξ, τ) under the assumption that the corresponding second-order model (1.6) is correct for the POTs $X - t|X > t$. Note that here the B_η component contains two parameters ξ and $\tilde{\rho}$. So in this component ξ and $\tilde{\rho}$ will be substituted with an external value.
Here

$$b_\eta(u) = u^{-\tilde{\rho}}\left(\frac{1-\tilde{\rho}}{\tilde{\rho}(\xi+\tilde{\rho})}\right) + u^\xi\left(\frac{1+\xi}{\xi(\xi+\tilde{\rho})}\right) - \frac{1}{\xi\tilde{\rho}},$$

in which the classical estimator of ξ (with $\delta_k = 0$), or an appropriate value ξ_0, is used to substitute ξ. A consistent estimator of $\tilde{\rho}$ is provided in Fraga Alves et al. (2003). Another option is to choose $(\xi_0, \tilde{\rho})$ minimizing the variance in the plot of the resulting estimates of ξ as a function of k.

A non-parametric approach $(E\bar{p})$. In practice a particular distribution probably follows laws of nature, environment or business and does not have to follow the second-order regular variation assumptions as in (1.6). A non-parametric approximation of $u \mapsto uB_\eta(u)$ can be obtained from an estimator $\hat{G}_{t_*}$ of G_{t_*}, or equivalently $\hat{G}_{k_*}$ of G_{k_*}, of the transformation $G_t(u) = u(1+\delta_t B_\eta(u))$ $(u \in (0,1))$ at some particular t_* or k_*. Indeed, using $\hat{G}^{(m)}_{k_*}(u) - u$ as an approximation of $u \mapsto \delta_{k_*} uB_\eta(u)$, and reparametrizing δ_k by δ_k/δ_{k_*}, we obtain $\hat{b}_{\eta,k_*}(u) = -1 + \frac{d}{du}\hat{G}^{(m)}_{k_*}(u)$ as an estimator of b_η.

For any t, an estimator $\hat{G}_t$ of G_t can be obtained using the Bernstein polynomial algorithm from Tencaliec et al. (2019). The Bernstein approximation of order m of a continuous distribution function G on $[0, 1]$ is given by

$$G^{(m)}(u) = \sum_{j=0}^{m} G\left(\frac{j}{m}\right)\binom{m}{j} u^j (1-u)^{m-j},\ u \in [0,1].$$

As in Babu et al. (2002) one then replaces the unknown distribution function G itself with the empirical distribution function $\hat{G}_n$ of the available data in order to obtain a smooth estimator of G:

$$\hat{G}_n^{(m)}(u) = \sum_{j=0}^{m} \hat{G}_n\left(\frac{j}{m}\right)\binom{m}{j} u^j (1-u)^{m-j}.$$

Note that G_t is the distribution function of $\bar{H}_\xi^{GP}(Y/\sigma)$. Hence, in the present application, data from G_t are only available after imputing a value for (ξ,τ). This then leads to the iterative algorithm from Tencaliec et al. (2019), which is applied to every threshold t, or every number of top k data.

(i) Set starting values $(\hat{\xi}_k^{(0)}, \hat{\tau}_k^{(0)})$. Here one can use $(\hat{\xi}_k^{ML}, \hat{\tau}_k^{ML})$ from using $G_t(u) = u$.

(ii) Iterate for $r = 0, 1, \ldots$ until the difference in loglikelihood taken in $(\hat{\xi}_k^{(r)}, \hat{\tau}_k^{(r)})$ and $(\hat{\xi}_k^{(r+1)}, \hat{\tau}_k^{(r+1)})$ is smaller than a prescribed small value:

(a) Given $(\hat{\xi}_k^{(r)}, \hat{\tau}_k^{(r)})$ construct rv's $\hat{Z}_{j,k} = \left(1 + \hat{\tau}_k^{(r)} Y_{j,k}\right)^{-1/\hat{\xi}_k^{(r)}}$

(b) Construct Bernstein approximation based on $\hat{Z}_{j,k}$ $(1 \le j \le k)$

$$\hat{G}_k^{(m)}(u) = \sum_{j=0}^{m} \hat{G}_k\left(\frac{j}{m}\right)\binom{m}{j} u^j (1-u)^{m-j}$$

with $\hat{G}_k$ the empirical distribution function of $\hat{Z}_{j,k}$

(c) Obtain new estimates $(\hat{\xi}_k^{(r+1)}, \hat{\tau}_k^{(r+1)})$ with ML:

$$\begin{aligned}(\hat{\xi}_k^{(r+1)}, \hat{\tau}_k^{(r+1)}) \;=\; \text{argmax} \Bigg\{ & \sum_{j=1}^{k} \log\{\hat{g}_k^{(m)}((1+\tau \hat{Z}_{j,k})^{-1/\xi})\} \\ & + \sum_{j=1}^{k} \log\{\frac{\tau}{\xi}(1+\tau \hat{Z}_{j,k})^{-1-1/\xi}\} \Bigg\}\end{aligned}$$

with $\hat{g}_k^{(m)}$ denoting the derivative of $\hat{G}_k^{(m)}$.

As noted in Tencaliec et al. (2019) a theoretical study of these estimates is difficult and has not been established.

Remark 1. The estimation methods described above of course can be rewritten for the specific case of Pareto-type distributions where the distribution of POTs $Y = \frac{X}{t}|X > t$ are approximated by transformed Pareto distributions. The model $(\mathcal{E})$ is then rephrased as

$$(\mathcal{E}^+): \quad \bar{F}_t^E(y) = \bar{H}_\xi^P(y)\left\{1 + \delta_t B_\eta\left(\bar{H}_\xi^P(y)\right)\right\}.$$

The likelihood estimation method, now based on the exceedances $Y_{j,k} = X_{n-j+1,n}/X_{n-k,n}$ $(j = 1, \ldots, k)$, is then adapted to

$$(\hat{\xi}_k^{E+}, \hat{\delta}_k^{E+}) = \text{argmax}\left\{\sum_{j=1}^{k} \log\left(1 + \delta_k b_\eta(Y_{j,k}^{-1/\xi})\right) + \sum_{j=1}^{k} \log\{\frac{1}{\xi}(Y_{j,k})^{-1-1/\xi}\}\right\}. \tag{2.2}$$

Note that the (Ep^+) approach using the parametric version $B_\eta(u) = u^{-\rho} - 1$ for a particular fixed $\rho < 0$ equals the EPD method from Beirlant et al. (2009), while $(E\bar{p}^+)$ is new.

Estimators of tail probabilities are then given by

$$\hat{\mathbb{P}}_k^{E+}(X > c) = \frac{k}{n} \bar{H}^P_{\hat{\xi}_k^{E+}}\left(c/X_{n-k,n}\right) \left(1 + \hat{\delta}_k^{E+} \hat{B}_\eta \left(\bar{H}^P_{\hat{\xi}_k^{E+}}(c/X_{n-k,n})\right)\right).$$

3. BASIC ASYMPTOTICS UNDER MODEL $(\mathcal{E})$

In this section we discuss the asymptotic properties of the maximum likelihood estimators solving (2.1) and (2.2). To this end, as in Beirlant et al. (2009), we develop the likelihood equations up to linear terms in δ_k since $\delta_k \to 0$ with decreasing value of k. Below we set $\bar{H}_\theta(y) = (1+\tau y)^{-1/\xi}$ when using extended GPD modelling, while $\bar{H}_\theta(y) = y^{-1/\xi}$ when using extended Pareto modelling under $\xi > 0$.

Extended Pareto POT modelling. The likelihood problem (2.2) was already considered in Beirlant et al. (2009) in case of parametric modelling for B_η. We here propose a more general treatment. The limit statements in the derivation can be obtained using the methods from Beirlant et al. (2009). Denoting the log-likelihood function in (2.2) by ℓ, the likelihood equations are given by

$$\text{(3.1)} \qquad \begin{cases} \frac{\partial}{\partial \xi}\ell &= -\frac{k}{\xi} + \frac{1}{\xi^2}\sum_{j=1}^k \log Y_{j,k} + \frac{\delta_k}{\xi^2}\sum_{j=1}^k \frac{b'_\eta(\bar{H}_\theta(Y_{j,k}))\bar{H}_\theta(Y_{j,k})\log Y_{j,k}}{1+\delta_k b_\eta(\bar{H}_\theta(Y_{j,k}))} \\ \frac{\partial}{\partial \delta_k}\ell &= \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k \sum_{j=1}^k b_\eta^2(\bar{H}_\theta(Y_{j,k})). \end{cases}$$

Extended Generalized Pareto POT modelling. The likelihood equations following from (2.1) up to linear terms in δ_k are now given by

$$\begin{cases} \frac{\partial}{\partial \xi}\ell &= -\frac{k}{\xi} + \frac{1}{\xi^2}\sum_{j=1}^k \log(1+\tau Y_{j,k}) + \frac{\delta_k}{\xi^2}\sum_{j=1}^k b'_\eta(\bar{H}_\theta(Y_{j,k}))\bar{H}_\theta(Y_{j,k})\log(1+\tau Y_{j,k}) \\ \frac{\partial}{\partial \tau}\ell &= \frac{k}{\xi\tau}\left\{-1 + (1+\xi)\frac{1}{k}\sum_{j=1}^k \frac{1}{1+\tau Y_{j,k}} \right. \\ & \qquad \left. -\frac{\delta_k}{k}\sum_{j=1}^k b'_\eta(\bar{H}_\theta(Y_{j,k}))(\tau Y_{j,k})(1+\tau Y_{j,k})^{-1-1/\xi}\right\} \\ \frac{\partial}{\partial \delta_k}\ell &= \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k \sum_{j=1}^k b_\eta^2(\bar{H}_\theta(Y_{j,k})), \end{cases}$$

from which

$$\text{(3.2)} \qquad \begin{cases} \hat{\delta}_k = \frac{\sum_{j=1}^k b_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))}{\sum_{j=1}^k b_\eta^2(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))}, \\ \frac{1}{k}\sum_{j=1}^k \log(1+\hat{\tau}_k Y_{j,k}) = \hat{\xi}_k - \frac{\hat{\delta}_k}{k}\sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))\bar{H}_{\hat{\theta}_k}(Y_{j,k})\log(1+\hat{\tau}_k Y_{j,k}), \\ \frac{1}{k}\sum_{j=1}^k \frac{1}{1+\hat{\tau}_k Y_{j,k}} = \frac{1}{1+\hat{\xi}_k} + \frac{\hat{\delta}_k}{1+\hat{\xi}_k}\left\{\frac{1}{k}\sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))\bar{H}_{\hat{\theta}_k}(Y_{j,k}) \right. \\ \qquad \left. -\frac{1}{k}\sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))\bar{H}_{\hat{\theta}_k}(Y_{j,k})\frac{1}{1+\hat{\tau}_k Y_{j,k}}\right\}. \end{cases}$$

Under the extended model we now state the asymptotic distribution of the estimators $(\hat{\xi}_k^E, \hat{\tau}_k^E)$ and $\hat{\xi}_k^{E+}$. To this end let Q denote the quantile function of F, and let $U(x) = Q(1-x^{-1})$ denote the corresponding tail quantile function. Model $(\mathcal{E})$ assumption can be rephrased in terms of U:

$$(\tilde{\mathcal{E}}): \quad \frac{\frac{U(vx)-U(v)}{\sigma_{U(v)}} - h_\xi(x)}{\delta(U(v))} \to_{v\to\infty} x^\xi B_\eta(1/x),$$

where $h_\xi(x) = (x^\xi - 1)/\xi$ and $\delta(U)$ regularly varying with index $\tilde{\rho} < 0$. Moreover in the mathematical derivations one needs the extra condition that for every $\epsilon, \nu > 0$, and v, vx sufficiently large

$$(\tilde{\mathcal{E}}_2): \quad \left| \frac{\frac{U(vx)-U(v)}{\sigma_{U(v)}} - h_\xi(x)}{\delta(U(v))} - x^\xi B_\eta(1/x) \right| \le \epsilon x^\xi |B_\eta(1/x)| \max\{x^\nu, x^{-\nu}\}.$$

Similarly, $(\mathcal{E}^+)$ is rewritten as

$$(\tilde{\mathcal{E}}^+): \quad \frac{\frac{U(vx)}{U(v)} - x^\xi}{\xi\delta(U(v)))} \to_{v\to\infty} x^\xi B_\eta(1/x).$$

The analogue of $(\tilde{\mathcal{E}}_2)$ in this specific case is given by

$$(\tilde{\mathcal{E}}_2^+): \quad \left| \frac{\frac{U(vx)}{U(v)} - x^\xi}{\xi\delta(U(v))} - x^\xi B_\eta(1/x) \right| \le \epsilon x^\xi |B_\eta(1/x)| \max\{x^\nu, x^{-\nu}\},$$

with $\delta(U)$ regularly varying with index $\rho < 0$.
Finally, in the expression of the asymptotic variances we use

$$Eb_\eta^2 = \int_0^1 b_\eta^2(u)du, \quad EB_\eta = \int_0^1 B_\eta(u)du, \quad EC_\eta = \int_0^1 u^\xi B_\eta(u)du.$$

The proof of the next theorem is outlined in the Appendix. It allows to construct confidence intervals for the estimators of ξ obtained under the extended models.
Theorem 1 *Let $k = k_n$ be a sequence such that $k, n \to \infty$ and $k/n \to 0$ such that $\sqrt{k}\delta(U(n/k)) \to \lambda \in \mathbb{R}$. Moreover assume that in* (2.1) *and* (2.2)*, B_η is substituted by a consistent estimator as $n \to \infty$. Then*

i. *when $\xi > -1/2$ with $(\tilde{\mathcal{E}}_2)$*

$$\left(\sqrt{k}(\hat{\xi}_k^E - \xi), \sqrt{k}(\frac{\hat{\tau}_k^E}{\tau} - 1)\right) \to_d \mathcal{N}_2(\mathbf{0}, \Sigma)$$

$$\Sigma = \frac{\xi^2}{D} \begin{pmatrix} \frac{1}{(1+\xi)^2(1+2\xi)} - \frac{(EC_\eta)^2}{Eb_\eta^2} & \frac{1}{\xi(1+\xi)^3} - \frac{EB_\eta EC_\eta}{\xi(1+\xi)Eb_\eta^2} \\ \frac{1}{\xi(1+\xi)^3} - \frac{EB_\eta EC_\eta}{\xi(1+\xi)Eb_\eta^2} & \frac{1}{\xi^2(1+\xi)^2}\left(1 - \frac{(EB_\eta)^2}{Eb_\eta^2}\right) \end{pmatrix}$$

where

$$D = \left(\frac{1}{(1+\xi)^2(1+2\xi)} - \frac{(EC_\eta)^2}{Eb_\eta^2}\right)\left(1 - \frac{(EB_\eta)^2}{Eb_\eta^2}\right) - \left(\frac{1}{(1+\xi)^2} - \frac{EB_\eta EC_\eta}{Eb_\eta^2}\right)^2,$$

ii. *when $\xi > 0$ with $(\tilde{\mathcal{E}}_2^+)$*

$$\left(\sqrt{k}(\hat{\xi}_k^{E+} - \xi), \sqrt{k}(\hat{\delta}_k^{E+} - \delta_k)\right) \to_d \mathcal{N}_2(\mathbf{0}, \Sigma^+),$$

$$\Sigma^+ = \frac{1}{Eb_\eta^2 - (EB_\eta)^2} \begin{pmatrix} \xi^2 Eb_\eta^2 & -\xi EB_\eta \\ -\xi EB_\eta & 1 \end{pmatrix}$$

Remark 2. The asymptotic variance of $\hat{\xi}_k^{E+}$ is larger than the asymptotic variance ξ^2 of the Hill estimator $\hat{\xi}_{k,n}^H$. Indeed,

$$\begin{aligned}(EB_\eta)^2 &= \left(\int_0^1 \log(1/u)b_\eta(u)du\right)^2 \\ &= \left(\int_0^1 (\log(1/u)-1)b_\eta(u)du\right)^2 \\ &\le \left(\int_0^1 (\log(1/u)-1)^2du\right)\left(\int_0^1 b_\eta^2(u)du\right) \\ &= (Eb_\eta^2),\end{aligned}$$

where the above inequality follows using the Cauchy-Schwarz inequality.
Similarly, one can show that

$$(EC_\eta)^2 = \xi^{-2}\left(\int_0^1 (u^\xi - \frac{1}{1+\xi})b_\eta du\right)^2 \le \frac{1}{(1+2\xi)(1+\xi)^2}(Eb_\eta^2).$$

The asymptotic variance of $\hat{\xi}_k^E$ equals

$$\frac{(1+\xi)^2}{k}\;\frac{1-(1+\xi)^2(1+2\xi)(EC_\eta)^2/(Eb_\eta^2)}{1-\frac{(1+\xi)^4(1+2\xi)}{\xi^2}(Eb_\eta^2)^{-1}[(EC_\eta)^2-2\frac{(EC_\eta)(EB_\eta)}{(1+\xi)^2}+\frac{(EB_\eta)^2}{(1+\xi)^2(1+2\xi)}]}$$

which can be shown to be larger than the asymptotic variance $(1+\xi)^2/k$ of the classical GPD maximum likelihood estimator. In the parametric case with $B_\eta(u) = \frac{u^\xi}{\tilde{\rho}}\left(\frac{u^{-\xi-\tilde{\rho}}-1}{\xi+\tilde{\rho}} - \frac{u^{-\xi}-1}{\xi}\right)$, one obtains $EB_\eta = (1+\xi)^{-1}(1-\tilde{\rho})^{-1}$, $EC_\eta = (1+\xi)^{-1}(1+2\xi)^{-1}(\xi-\tilde{\rho}+1)^{-1}$ and $Eb_\eta^2 = 2(1+2\xi)^{-1}(1-2\tilde{\rho})^{-1}(\xi-\tilde{\rho}+1)^{-1}$. It then follows that the asymptotic variance of $\hat{\xi}_k^E$ equals $\frac{(1+\xi)^2}{k}\left(\frac{1-\tilde{\rho}}{\tilde{\rho}}\right)^2$.

In case $\xi > 0$ with $B_\eta(u) = u^{-\rho}-1$, the asymptotic variance of $\hat{\xi}_k^{E+}$ is given by $\frac{\xi^2}{k}\left(\frac{1-\rho}{\rho}\right)^2$ as already found in Beirlant et al. (2009).
Finally, an asymptotic representation of $\sqrt{k}(\hat{\delta}_k^E - \delta_k)$ can be found at the end of the proof of Theorem 1 in the Appendix. □

In the case studies in the next section asymptotic confidence intervals based on Theorem 1 can be added to the analysis.

Remark 3. Since in model $(\mathcal{E})$ the B_η factor is multiplied by δ_t, the asymptotic distribution of tail estimators based on $(\mathcal{E})$ will not depend on the asymptotic distribution of the estimator of B_η. As in Beirlant et al. (2009) when using the EPD model in the Pareto-type setting, one can rely in the parametric approach on consistent estimators of the nuisance parameter η using a larger proportion k_* of the data. Alternatively, one can also consider different values of η in the parametric approach, and of (k_*, m) in the non-parametric setting, and search for values of this nuisance parameter which stabilizes the plots of the EVI estimates as a function of k using the minimum variance principle for the estimates as a function of k. Clearly one loses the asymptotic unbiasedness in Theorem 1 if B_η is not consistently estimated. For the moment no proof

is available to show that the estimators of the parameters in the second order component B_η through the minimum variance principle are consistent. Note that the estimator of $\tilde{\rho}$ presented in Fraga Alves et al. (2003) has been shown to be consistent.
As becomes clear from the simulation results, in many instances the extreme value index estimators are not very sensitive to such a misspecification, especially in the non-parametric approach leading to $E\bar{p}$ and $E\bar{p}^+$, and the proposed estimators can still outperform the classical maximum likelihood estimators based on the first order approximations of the POT distributions. □

4. SIMULATIONS AND CASE STUDIES

Simulation results and practical cases are proposed in a Shinyapp written in R:

`https://phdshinygao.shinyapps.io/ExtendedModels/`

Under *Simulations* one finds simulation results with sample sizes $n = 200$ for different distributions from each max-domain of attraction. The bias and MSE for the different estimators are plotted as a function of the number of exceedances k. Using the notation from the preceding sections one has a choice to apply the technique with $\bar{H}_\theta$ equal to the GPD, respectively the simple Pareto distribution (only when $\xi > 0$).

Sliders are provided for the following parameters:

- in case of GPD modelling: $\tilde{\rho}$ in Ep, and (k_*, m) in $E\bar{p}$ estimation,
- in case of Pareto modelling: ρ in Ep^+, and (k_*, m) in $E\bar{p}^+$ estimation.

Again one can indicate to choose these parameters so as to minimize the variance of $\hat{\xi}_k$ over $k = 2, \ldots, n$. The value of ξ in the parametric function $B_{\xi,\tilde{\rho}}$ in Ep is imputed with the classical GPD-ML estimator at the given value of k.
Also bias and RMSE plots of the corresponding tail probability estimates of $p = \mathbb{P}(X > c)$ are given, where c is chosen so that these probabilities equal $p = 0.005$ or $p = 0.003$. Here the bias, respectively RMSE, are expressed as the average, respectively the average of squared values, of $\log(p/\hat{p})$.
One can also change the vertical scale of the plots, smooth the figures by taking moving averages of a certain number of estimates. Finally one can download the figures in pdf.

While on the above link, several other distributions are used and sliders are provided for the different parameters ρ, $\tilde{\rho}$, and (k_*, m), we collect here the resulting figures for estimation of ξ and estimating 0.003 tail probabilities, when using the minimum variance principle for all parameters, in case of the following subset of models:

- *The Burr*(τ, λ) *distribution* with $\bar{F}(x) = (1 + x^\tau)^{-\lambda}$ for $x > 0$ with $\tau = 1$ and $\lambda = 2$, so that $\xi = \frac{1}{\tau\lambda} = \frac{1}{2}$ and $\rho = \tilde{\rho} = -\frac{1}{\lambda} = -\frac{1}{2}$.

- *The Fréchet(2) distribution* with $\bar{F}(x) = 1 - \exp\left(-x^{-2}\right)$ for $x > 0$, so that $\xi = \frac{1}{2}$ and $\rho = \tilde{\rho} = -1$.
- *The standard normal distribution* with $\xi = 0$ and $\tilde{\rho} = 0$.
- *The Exponential distribution* with $\bar{F}(x) = e^{-\lambda x}$ for $x > 0$, so that $\xi = 0$ and $\tilde{\rho} = 0$.
- *The Reversed Burr distribution* with $\bar{F}(x) = (1 + (1-x)^{-\tau})^{-\lambda}$ for $x < 1$ with $\tau = 5$ and $\lambda = 1$, so that $\xi = -1/(\tau\lambda) = -\frac{1}{5}$ with $\tilde{\rho} = -1/\lambda = -1$.
- *The extreme value Weibull distribution* with $\bar{F}(x) = 1 - e^{-(1-x)^{\alpha}}$ for $x < 1$ with $\alpha = 4$, so that $\xi = -\frac{1}{4}$ with $\tilde{\rho} = -1$.

We also compare the bias and RMSE results for $\hat{\xi}_k^E$ with those of the ridge regression estimator presented in Buitendag et al. (2019). This regression method is constructed on the basis of a regression model of the type

$$Y_j = \xi + b_{n,k}\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}, \; j = 1, \ldots, k,$$

where

$$Y_j = (j+1)\left(\log\frac{X_{n-j,n}\hat{\xi}^H_{j,n}}{X_{n-j-1,n}\hat{\xi}^H_{j+1,n}} - \log(1+\frac{1}{j}) + \frac{1}{j}\right), \; j = 1, \ldots, n-1.$$

In case $\xi > 0$, the results for $\hat{\xi}_k^{E+}$ are also compared with the corrected Hill method presented in Caeiro et al. (2005) and (2009), also based on regression representations of top order statistics $X_{n-j+1,n}$, and which have been shown to have asymptotic bias 0 while keeping the same asymptotic variance ξ^2/k as the Hill estimator $\hat{\xi}^H_{k,n}$ under a third-order slow variation model.

In general the minimum variance principle works well, though in some cases some improved results can be obtained by choosing specific values of the parameters ρ, $\tilde{\rho}$, and (k_*, m). This is mainly the case for the Pareto-type models when using $E\bar{p}$, such as for the Fréchet distribution. Also, in case of tail probability estimation using Ep for cases with $\xi < 0$ particular choices of the corresponding parameters lead to improvements over the minimum variance principle. Overall the Ep approach yields the best results, both in estimation of ξ and tail probabilities. The improvement over the classical GPD maximum likelihood approach is smaller for $E\bar{p}$, and in case of situations where the second order parameter $\tilde{\rho}$ equals 0 then $E\bar{p}$ basically equals the ML estimators. Note that when $\tilde{\rho} = 0$ the conditions of the main theorem are not met, in which case the GPD and the bias reductions are known to exhibit a large bias. This is typically the case when $\xi = 0$. This is also known to be the case using simple Pareto modelling when $\rho = 0$. The proposed methods compare well with the ridge regression method. One exception is the Fréchet distribution (see Figure 3) in which the ridge regression method offers exceptionally good results.

In case of simple Pareto modelling for $\xi > 0$ cases (see Figures 2 and 4) the Ep^+ and $E\bar{p}^+$ approaches yield serious improvements over the Hill estimator, with small bias for Ep^+ and $E\bar{p}^+$, while the parametric approach Ep^+ naturally exhibits the best RMSE. The results obtained with proposed methods are comparable with the CH estimator (see Figures 2 and 4).

Under *Applications* the app also offers the analysis of some case studies, some of which are discussed here in more detail. We use Belgian car insurance claim ultimates of a Belgian car insurance portfolio discussed in Albrecher et al. (2017), and lifetime data discussed in Einmahl et al. (2019). We then present estimates of ξ, σ and tail probabilities $\mathbb{P}(X > x_{n,n})$ with $x_{n,n}$ denoting the largest observation, so that the estimated probability is supposed to be close to $1/n$. An option is provided in the Shinyapp to construct asymptotic confidence intervals for ξ for the Ep and Ep^+ based estimates of ξ, on the basis of Theorem 1.

In actuarial statistics, Pareto-type modelling is customary in case of car insurance claim modelling. So here we provide both the plots of $\hat{\xi}^H_{k,n}$, Ep^+, $E\bar{p}^+$ and the CH estimator (see top left in Figure 9), as well as the GPD-ML, Ep, $E\bar{p}$ and ridge regression estimator (bottom left in Figure 9), and the corresponding tail probability estimates at the right hand side. Under the Pareto approach, confining oneselves to $\xi > 0$, the level 0.4 clearly appears for the EVI both using Ep^+ and $E\bar{p}^+$ when using the minimum variance principle. The CH estimator also shows a stable area around the value 0.5. The tail probability estimates of $\mathbb{P}(X > x_{n,n})$ are close to $1/n$ for almost all k values while the plot of the classical estimates is difficult to interpret.
With GPD based modelling two EVI levels are visible, around 0.2 and 0.4, of which the lower level is more clearly indicated when using $E\bar{p}$ with $k_* = 427$ and $m = 25$ as shown in Figure 9, bottom left. The ridge estimator is stable at the value 0.4. The corresponding tail probability estimates based on $E\bar{p}$ are also stable at the value $1/n$ for a long k range.

In Einmahl et al. (2019) the life spans are studied for Dutch males and females reaching age 92 years and higher, considering their age at death. For every year, from 1986 till 2015, the life spans of this subgroup were analyzed. The authors decided to use $k = 1500$ for every year when using the classical GPD-ML estimators, and found an EVI estimate $\hat{\xi}$ between -0.1 and -0.15 for females, while for males a value around -0.15 is common over the whole period. Here we restrict ourselves to the female data from 1986. The results of Ep with asymptotic confidence intervals as discussed in Remark 2 with $\tilde{\rho} = -0.5$ are shown in Figure 10 (left). While the classical GPD-ML estimates decrease with increasing k from 1 to 1500, the Ep estimates show a more stable plot at a negative ξ value which is rather between -0.05 and -0.1. The ridge regression method shows a similar value for $k \leq 500$. The corresponding tail probability estimates for a larger k indicate a value closer to the tail probability estimate $1/n$ based on the empirical distribution function, in contrast to the classical GPD approach.

5. CONCLUSIONS

In this contribution we have constructed bias reduced estimators of tail parameters extending the classical POT method. The bias can be modelled parametrically (for instance based on second order regular variation theory), or non-parametrically using Bernstein polynomial approximations. A basic asymptotic limit theorem is provided for the estimators of the extreme value parameters which allows to compute asymptotic confidence intervals. A shinyapp has been constructed with which the characteristics and the effectiveness of the proposed methods are illustrated through simulations and practical case studies. From this it follows that within the proposed methods it is always possible to improve upon the classical POT method both in bias and RMSE. This approach can also be used as a data analytic tool to enhance an extreme value analysis.

6. ACKNOWLEDGEMENTS

The authors want to thank the referees for their constructive suggestions.
The research of G. Maribe was supported wholly/in part by the National Research Foundation of South Africa (Grant Number 102628) and the DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (COE-Mass). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research is that of the author(s), and that the NRF accepts no liability whatsoever in this regard.
Part of P. Naveau's work was supported by the European DAMOCLES-COST-ACTION on compound events, and also benefited from French national programs, in particular FRAISE-LEFE/INSU, MELODY-ANR, and ANR-11-IDEX-0004 - 17-EURE-0006.

REFERENCES

[1] Albrecher, H., Beirlant, J. and Teugels, J., 2017. *Reinsurance: Actuarial and Statistical Aspects.* Wiley.

[2] Babu, G.J., Canty, A.J., Chaubey, Y.P., 2002. Application of Bernstein Polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, 105, 377-392.

[3] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G., 1999. Tail index estimation and an exponential regression model. *Extremes*, 2,157-180.

[4] Beirlant, J., Dierckx, G, Guillou, A. and Starica, C., 2002. On exponential representations of log-spacings of extreme order statistics. *Extremes*, 5, 257-180.

[5] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J., 2004. *Statistics of Extremes: Theory and Applications*, Wiley, UK.

[6] Beirlant, J., Dierckx, G., Matthys, G. and Guillou, A., 2005. Estimation of the extreme value index and regression on generalized quantile plots. *Bernoulli*, 11, 949-970.

[7] Beirlant, J., Joossens, E. and Segers, J., 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139(8), 2800–2815.

[8] Buitendag, S., Beirlant, J. and de Wet, T., 2019. Ridge regression estimators for the extreme value index. *Extremes*, 22, 271-292.

[9] Caeiro, F., Gomes, M.I. and Pestana, D., 2005. Direct reduction of bias of the classical Hill estimator. *Revstat*, 3(2), 113-136.

[10] Caeiro, F., Gomes, M.I. and Rodrigues, L.H., 2009. Reduced-bias tail index estimators under a third order framework. *Communications in Statistics-Theory and Methods*, 38(7), 1019-1040.

[11] Drees, H., 1996. Refined Pickands estimators with bias correction. *Communications in Statistics-Theory and Methods*, 25, 837-851.

[12] Dupuis, D., 1999. Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1, 251–261.

[13] Einmahl, J., Einmahl, J.H.J. and Haan, L. de, 2019. Limits to human life span through extreme value theory. *Journal of the American Statistical Association*, 114(527), 1075-1080.

[14] Feuerverger, A. and Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, 27, 760-781.

[15] Fraga Alves, M.I., de Haan, L. and Lin, T., 2003. Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods Statist.*, 12, 155-176.

[16] Gomes, M.I. and Martins, M.J., 2002. Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5, 5-31.

[17] Haan, L. de and Ferreira, A., 2006. *Extreme Value Theory: an Introduction*, Springer Science and Business Media, LLC, New York.

[18] Hill, B.M., 1975. A simple general approach about the tail of a distribution. *Annals of Statistics*, 3, 1163-1174.

[19] Papastathopoulos, I. and Tawn, J., 2013. Extended generalized Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143, 131–143.

[20] Peng, L., 1998. Asymptotically unbiased estimator for the extreme-value index. *Statist. Prob. Lett.*, 38, 107-115.

[21] Tencaliec, P., Favre, A.-C., Naveau, P. and Prieur, C., 2019. Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, https://doi.org/10.1002/env.2582 .

7. Appendix

In this section we provide details concerning the proof of Theorem 1.

Asymptotic distribution of $\hat{\xi}_k^{E+}$.
From (3.1) we obtain up to linear terms in δ_k that (denoting $\hat{\xi}_k$ for $\hat{\xi}_k^{E+}$)

$$\begin{cases} \hat{\delta}_k &= \frac{\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\hat{\xi}_k})}{\sum_{j=1}^k b_\eta^2(Y_{j,k}^{-1/\hat{\xi}_k})} \\ \hat{\xi}_k &= \hat{\xi}_{k,n}^H + \hat{\delta}_k B_k^{(1)}, \end{cases}$$

with $B_k^{(1)} = \frac{1}{k}\sum_{j=1}^k b'_\eta(Y_{j,k}^{-1/\hat{\xi}_k})Y_{j,k}^{-1/\hat{\xi}_k}\log Y_{j,k}$. As $k,n \to \infty$ and $k/n \to 0$ we have $B_k^{(1)} \to_p -\xi\int_0^1 b'_\eta(u)u\log u du = -\xi EB_\eta$.
Using a Taylor expansion on the numerator of the right hand side of the first equation leads to

$$\frac{1}{k}\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\hat{\xi}_k}) = \frac{1}{k}\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\xi}) - (\hat{\xi}_k - \xi)\xi^{-1}(EB_\eta)\,(1+o_p(1)),$$

so that, with $\frac{1}{k}\sum_{j=1}^k b_\eta^2(Y_{j,k}^{-1/\hat{\xi}_k}) \to_p Eb_\eta^2$, up to lower order terms

$$\hat{\delta}_k = \frac{1}{Eb_\eta^2}\frac{1}{k}\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\xi}) - (\hat{\xi}_k - \xi)\xi^{-1}\frac{EB_\eta}{Eb_\eta^2}\,(1+o_p(1)).$$

Hence, inserting this expansion into $\hat{\xi}_k = \hat{\xi}_{k,n}^H + \hat{\delta}_k B_k^{(1)}$, finally leads to

$$\begin{aligned}
\sqrt{k}(\hat{\xi}_k - \xi)(1+o_p(1)) &= \frac{Eb_\eta^2}{Eb_\eta^2 - (EB_\eta)^2}\sqrt{k}\left(\hat{\xi}_{k,n}^H - \xi\right) - \frac{\xi EB_\eta}{Eb_\eta^2 - (EB_\eta)^2}\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\xi})\right) \\
&= \frac{Eb_\eta^2}{Eb_\eta^2 - (EB_\eta)^2}\sqrt{k}\left(\hat{\xi}_{k,n}^H - \xi - \xi\delta_k EB_\eta\right) \\
&\quad - \frac{\xi EB_\eta}{Eb_\eta^2 - (EB_\eta)^2}\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\xi}) - \delta_k Eb_\eta^2\right),
\end{aligned}$$

with $\delta_k = \delta(U(n/k))$. We now show that this final expression is a linear combination of two zero centered statistics (up to the required accuracy) which is asymptotically normal with the stated asymptotic variance. To this end let $Z_{n-k,n} \le Z_{n-k+1,n} \le \ldots \le Z_{n,n}$ denote the top $k+1$ order statistics of a sample of size n from the standard Pareto distribution with distribution function $z \mapsto z^{-1}$, $z > 1$. Then from $(\tilde{\mathcal{E}}_2^+)$

$$\begin{aligned}
\hat{\xi}_{k,n}^H &= \frac{1}{k}\sum_{j=1}^k (\log U(Z_{n-j+1,n}) - \log U(Z_{n-k,n})) \\
&= \frac{1}{k}\sum_{j=1}^k \log\left\{\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^\xi \left[1 + \xi\delta(U(Z_{n-k,n}))B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)\right.\right. \\
&\qquad\qquad \left.\left. + o_p(1)|\delta(U(Z_{n-k,n}))||B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)|\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^\epsilon\right]\right\} \\
&= \xi\frac{1}{k}\sum_{j=1}^k \log\frac{Z_{n-j+1,n}}{Z_{n-k,n}} + \xi\delta(U(Z_{n-k,n}))\frac{1}{k}\sum_{j=1}^k B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right) \\
&\qquad + o_p(1)|\delta(U(Z_{n-k,n}))|\frac{1}{k}\sum_{j=1}^k |B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)|\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^\epsilon.
\end{aligned}$$

Now $\log Z_{n-j+1,n} - \log Z_{n-k,n} =_d E_{k-j+1,k}$, the $(k-j+1)$th smallest value from a standard exponential sample $E_1, \ldots, E_k$ of size k, so that $\frac{1}{k}\sum_{j=1}^k \log\frac{Z_{n-j+1,n}}{Z_{n-k,n}} =_d \frac{1}{k}\sum_{j=1}^k E_j$ and $\frac{1}{k}\sum_{j=1}^k B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right) =_d \frac{1}{k}\sum_{j=1}^k B_\eta(e^{-E_j}) =_d \frac{1}{k}\sum_{j=1}^k B_\eta(U_j)$ where $U_1, \ldots, U_k$ is a uniform (0,1) sample. Hence, since $\delta(U(Z_{n-k,n}))/\delta(U(n/k)) \to_p 1$ and $\frac{1}{k}\sum_{j=1}^k B_\eta(U_j) \to_p EB_\eta$, we have that $\hat{\xi}_{k,n}^H - \xi - \xi\delta_k EB_\eta$ is asymptotically equivalent to $\frac{1}{k}\sum_{j=1}^k \xi(E_j - 1)$ as $\sqrt{k}\delta_k \to \lambda$.

Similarly

$$
\begin{aligned}
\frac{1}{k}\sum_{j=1}^{k} b_\eta(Y_{j,k}^{-1/\xi}) &= \frac{1}{k}\sum_{j=1}^{k} b_\eta\left(\left[\frac{U\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}Z_{n-k,n}\right)}{U(Z_{n-k,n})}\right]^{-1/\xi}\right) \\
&= \frac{1}{k}\sum_{j=1}^{k} b_\eta\left(\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^{-1}\left[1+\xi\delta(U(Z_{n-k,n}))B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)\right.\right. \\
&\qquad \left.\left. +o_p(1)|\delta(U(Z_{n-k,n}))||B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)|\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^{\epsilon}\right]^{-1/\xi}\right) \\
&= \frac{1}{k}\sum_{j=1}^{k} b_\eta\left(\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^{-1}\left[1-\delta(U(Z_{n-k,n}))B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)\right.\right. \\
&\qquad \left.\left. +o_p(1)|\delta(U(Z_{n-k,n}))||B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)|\left(\frac{Z_{n-j+1,n}}{Z_{n-k,n}}\right)^{\epsilon}\right]\right) \\
&= \frac{1}{k}\sum_{j=1}^{k} b_\eta(e^{-E_j}) \\
&\quad -\delta(U(Z_{n-k,n}))\frac{1}{k}\sum_{j=1}^{k} b'_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)(1+o_p(1)).
\end{aligned}
$$

Since $\delta(U(Z_{n-k,n}))/\delta_k \to_p 1$ and $\frac{1}{k}\sum_{j=1}^{k} b'_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)B_\eta\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right)\left(\frac{Z_{n-k,n}}{Z_{n-j+1,n}}\right) \to_p -Eb_\eta^2$ it follows that $\frac{1}{k}\sum_{j=1}^{k} b_\eta(Y_{j,k}^{-1/\xi}) - \delta_k Eb_\eta^2$ is asymptotically equivalent to $\frac{1}{k}\sum_{j=1}^{k} b_\eta(e^{-E_j}) =_d \frac{1}{k}\sum_{j=1}^{k} b_\eta(U_j)$ as $\sqrt{k}\delta_k \to \lambda$, which is centered at 0 since $E(b_\eta(U)) = 0$. The results incorporating $\hat\delta_k^{E+}$ follow similarly.

Asymptotic distribution of $\hat\xi_k^E$.
This derivation follows similar lines starting from (3.2):

$$
\begin{cases}
\frac{1}{k}\sum_{j=1}^{k} b'_\eta(\bar H_{\hat\theta_k}(Y_{j,k}))\bar H_{\hat\theta_k}(Y_{j,k})\log(1+\hat\tau_k Y_{j,k}) \to_p -\xi EB_\eta, \\
\frac{1}{k}\sum_{j=1}^{k} b_\eta^2(\bar H_{\hat\theta_k}(Y_{j,k})) \to_p Eb_\eta^2, \\
\frac{1}{k}\sum_{j=1}^{k} b'_\eta(\bar H_{\hat\theta_k}(Y_{j,k}))\bar H_{\hat\theta_k}(Y_{j,k}) \to_p b_\eta(1), \\
\frac{1}{k}\sum_{j=1}^{k} b'_\eta(\bar H_{\hat\theta_k}(Y_{j,k}))\bar H_{\hat\theta_k}(Y_{j,k})\frac{1}{1+\hat\tau_k Y_{j,k}} \to_p \xi(1+\xi)EC_\eta + b_\eta(1),
\end{cases}
$$

as $k,n\to\infty$ and $k/n\to\infty$, so that the system of equations is asymptotically equivalent to

$$
\begin{cases}
\hat\delta_k = \frac{\frac{1}{k}\sum_{j=1}^{k} b_\eta(\bar H_{\hat\theta_k}(Y_{j,k}))}{Eb_\eta^2}, \\
\frac{1}{k}\sum_{j=1}^{k}\log(1+\hat\tau_k Y_{j,k}) = \hat\xi_k + \hat\xi_k\hat\delta_k EB_\eta \\
\frac{1}{k}\sum_{j=1}^{k}\frac{1}{1+\hat\tau_k Y_{j,k}} = \frac{1}{1+\hat\xi_k} - \hat\xi_k\hat\delta_k EC_\eta.
\end{cases}
$$

Using a Taylor expansion on the numerator of the right hand side of the first equation leads to

$$
\hat\delta_k Eb_\eta^2 = \frac{1}{k}\sum_{j=1}^{k} b_\eta(\bar H_\theta(Y_{j,k})) - \frac{EB_\eta}{\xi}(\hat\xi_k - \xi) + (1+\xi)EC_\eta\left(\frac{\hat\tau_k}{\tau} - 1\right).
$$

Imputing this in the second and third equation in ξ and τ, and expanding these equations linearly around the correct values (ξ,τ), while using, as $k,n\to\infty$ and $k/n\to 0$

$$\frac{1}{k}\sum_{j=1}^{k}\frac{\tau Y_{j,k}}{1+\tau Y_{j,k}}\to_p \frac{\xi}{1+\xi} \text{ and } \frac{1}{k}\sum_{j=1}^{k}\frac{\tau Y_{j,k}}{(1+\tau Y_{j,k})^2}\to_p \frac{\xi}{(1+\xi)(1+2\xi)},$$

leads to the linearized equations

$$(7.1)\quad \begin{cases} \left(\hat{\xi}_k-\xi\right)\left(-1+\frac{(EB_\eta)^2}{Eb_\eta^2}\right)+\left(\frac{\hat{\tau}_k}{\tau}-1\right)\left(\frac{\xi}{1+\xi}-\xi(1+\xi)\frac{EB_\eta\, EC_\eta}{Eb_\eta^2}\right) \\ \qquad = -\left(\frac{1}{k}\sum_{j=1}^{k}\log(1+\tau Y_{j,k})-\xi\right)+\frac{\xi EB_\eta}{Eb_\eta^2}\frac{1}{k}\sum_{j=1}^{k}b_\eta(\bar{H}_\theta(Y_{j,k})), \\ \\ \left(\hat{\xi}_k-\xi\right)\left(\frac{1}{(1+\xi)^2}-\frac{EB_\eta EC_\eta}{Eb_\eta^2}\right)+\left(\frac{\hat{\tau}_k}{\tau}-1\right)\left(-\frac{\xi}{(1+\xi)(1+2\xi)}+\xi(1+\xi)\frac{(EC_\eta)^2}{Eb_\eta^2}\right) \\ \qquad = -\left(\frac{1}{k}\sum_{j=1}^{k}\frac{1}{1+\tau Y_{j,k}}-\frac{1}{1+\xi}\right)-\frac{\xi EC_\eta}{Eb_\eta^2}\frac{1}{k}\sum_{j=1}^{k}b_\eta(\bar{H}_\theta(Y_{j,k})). \end{cases}$$

Using similar derivations as in the case $\hat{\xi}_k^{E+}$, it follows that the right hand sides in (7.1) can be rewritten as a linear combination of two zero centered statistics from which the asymptotic normality of $\left(\sqrt{k}(\hat{\xi}_k^E-\xi),\sqrt{k}(\frac{\hat{\tau}_k^E}{\tau}-1)\right)$ can be obtained, as stated in Theorem 1:

$$\begin{cases} \left(\hat{\xi}_k-\xi\right)\left(-1+\frac{(EB_\eta)^2}{Eb_\eta^2}\right)+\left(\frac{\hat{\tau}_k}{\tau}-1\right)\left(\frac{\xi}{1+\xi}-\xi(1+\xi)\frac{EB_\eta\, EC_\eta}{Eb_\eta^2}\right) \\ \qquad = -\left(\frac{1}{k}\sum_{j=1}^{k}\log(1+\tau Y_{j,k})-\xi-\xi\delta_k EB_\eta\right)+\frac{\xi EB_\eta}{Eb_\eta^2}\left(\frac{1}{k}\sum_{j=1}^{k}b_\eta(\bar{H}_\theta(Y_{j,k}))-\delta_k Eb_\eta^2\right), \\ \\ \left(\hat{\xi}_k-\xi\right)\left(\frac{1}{(1+\xi)^2}-\frac{EB_\eta EC_\eta}{Eb_\eta^2}\right)+\left(\frac{\hat{\tau}_k}{\tau}-1\right)\left(-\frac{\xi}{(1+\xi)(1+2\xi)}+\xi(1+\xi)\frac{(EC_\eta)^2}{Eb_\eta^2}\right) \\ \qquad = -\left(\frac{1}{k}\sum_{j=1}^{k}\frac{1}{1+\tau Y_{j,k}}-\frac{1}{1+\xi}+\xi\delta_k EC_\eta\right)-\frac{\xi EC_\eta}{Eb_\eta^2}\left(\frac{1}{k}\sum_{j=1}^{k}b_\eta(\bar{H}_\theta(Y_{j,k}))-\delta_k Eb_\eta^2\right). \end{cases}$$

We hence obtain the following asymptotic representation

$$\left(\hat{\xi}_k^E-\xi,\frac{\hat{\tau}_k^E}{\tau}-1\right)^t = W^{-1}\begin{pmatrix} -1 & 0 & \xi\frac{EB_\eta}{Eb_\eta^2} \\ 0 & -1 & -\xi\frac{EC_\eta}{Eb_\eta^2}\end{pmatrix}\left(U_k^{(1)},U_k^{(2)},U_k^{(3)}\right)^t$$

where

$$W=\begin{pmatrix} -1+\frac{(EB_\eta)^2}{Eb_\eta^2} & \frac{\xi}{1+\xi}-\xi(1+\xi)\frac{EB_\eta\, EC_\eta}{Eb_\eta^2} \\ \frac{1}{(1+\xi)^2}-\frac{EB_\eta EC_\eta}{Eb_\eta^2} & -\frac{\xi}{(1+\xi)(1+2\xi)}+\xi(1+\xi)\frac{(EC_\eta)^2}{Eb_\eta^2}\end{pmatrix},$$

and

$$\sqrt{k}\left(U_k^{(1)},U_k^{(2)},U_k^{(3)}\right)^t := \begin{pmatrix} \frac{1}{k}\sum_{j=1}^{k}\log(1+\tau Y_{j,k})-\xi-\xi\delta_k EB_\eta \\ \frac{1}{k}\sum_{j=1}^{k}\frac{1}{1+\tau Y_{j,k}}-\frac{1}{1+\xi}+\xi\delta_k EC_\eta \\ \frac{1}{k}\sum_{j=1}^{k}b_\eta(\bar{H}_\theta(Y_{j,k}))-\delta_k Eb_\eta^2 \end{pmatrix}$$

is asymptotically normal with variance-covariance matrix

$$\Sigma_U=\begin{pmatrix} \xi^2 & -\xi^2(1+\xi)^{-2} & \xi EB_\eta \\ \text{-}\xi^2(1+\xi)^{-2} & \xi^2(1+\xi)^{-2}(1+2\xi)^{-1} & -\xi EC_\eta \\ \xi EB_\eta & -\xi EC_\eta & Eb_\eta^2 \end{pmatrix}.$$

Concerning $\hat{\delta}_k^E$ we find the following representation:

$$(Eb_\eta^2)\sqrt{k}\left(\hat{\delta}_k^E-\delta_k\right)=\left((0\;0\;1)+(-EB_\eta/\xi\;\;(1+\xi)EC_\eta)W^{-1}\begin{pmatrix} -1 & 0 & \xi\frac{EB_\eta}{Eb_\eta^2} \\ 0 & -1 & -\xi\frac{EC_\eta}{Eb_\eta^2}\end{pmatrix}\right)\begin{pmatrix} U_k^{(1)} \\ U_k^{(2)} \\ U_k^{(3)}\end{pmatrix}.$$

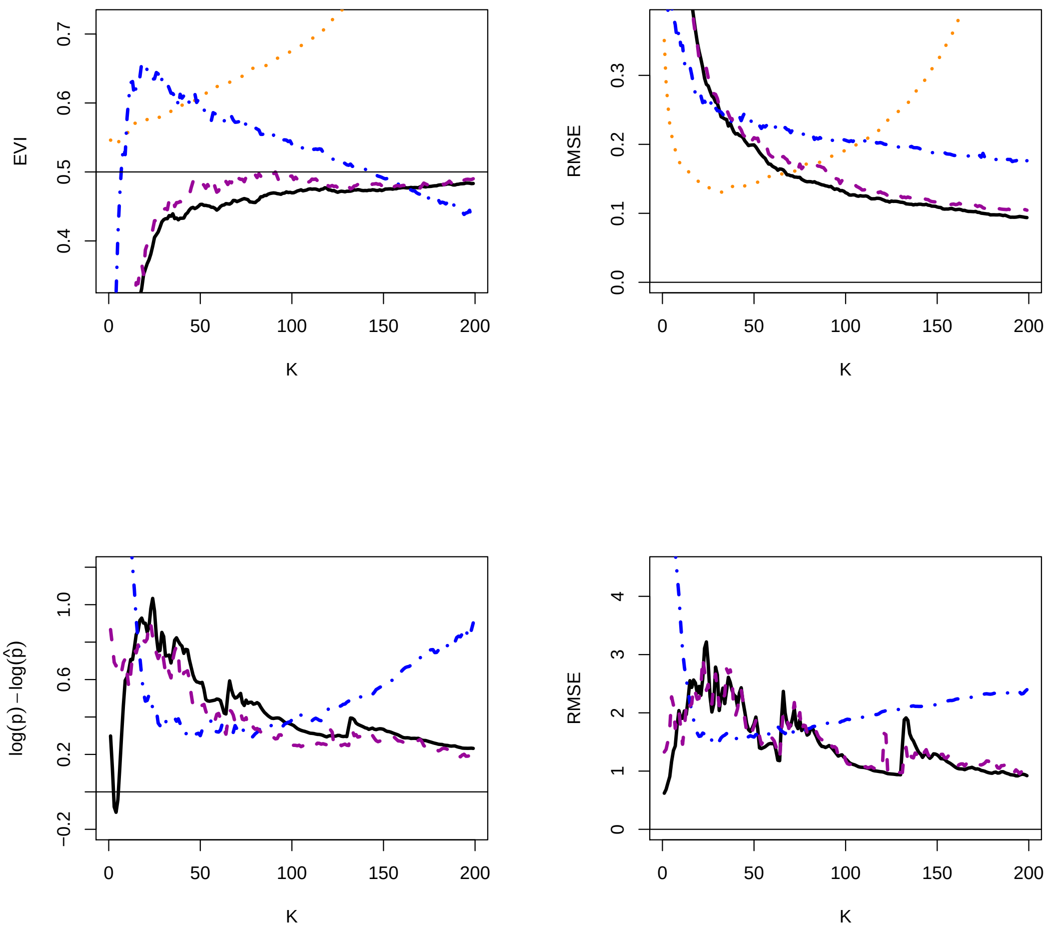

Figure 1: Burr distribution with $\xi = 0.5$ and $\rho = -0.5$. Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), Ep (dash-dotted), $E\bar{p}$ (dashed) and ridge regression estimator (dotted).

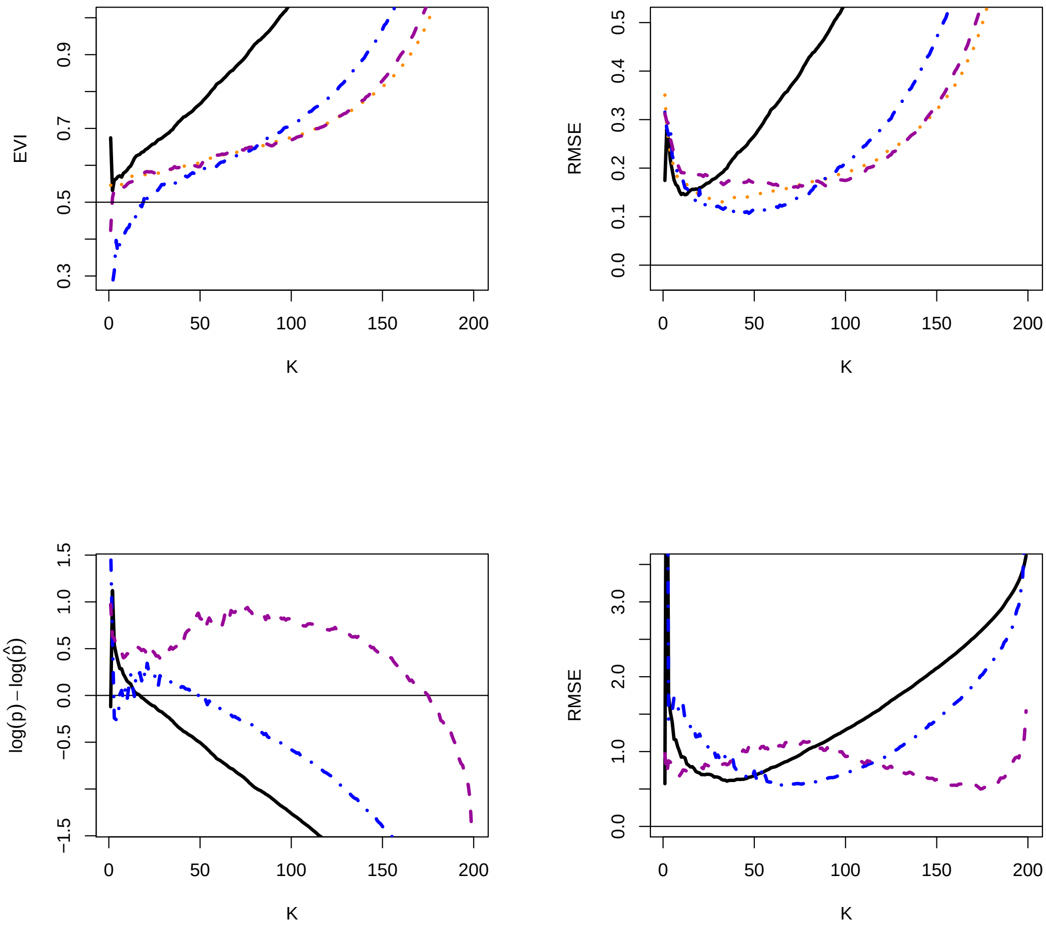

Figure 2: Burr distribution with $\xi = 0.5$ and $\rho = -0.5$. Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line), Ep^{+} (dash-dotted), $E\bar{p}^{+}$ (dashed) and corrected Hill estimator (dotted).

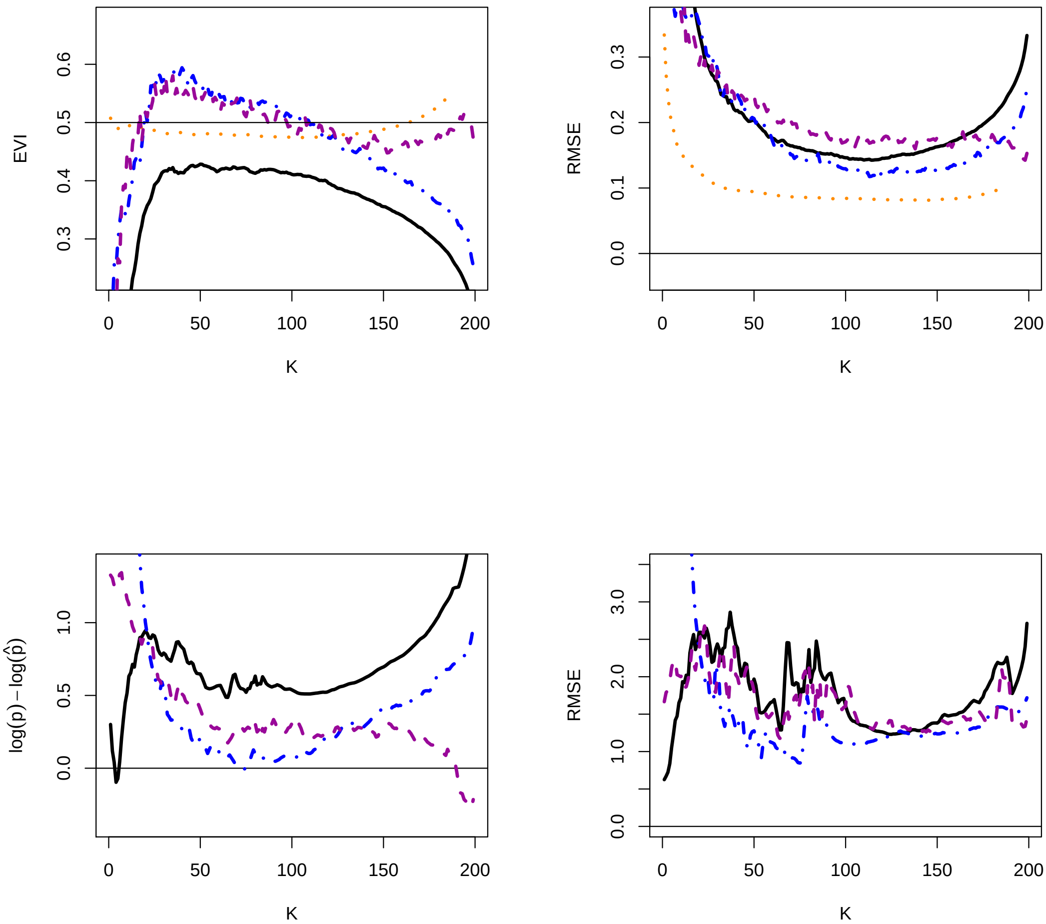

Figure 3: Fréchet distribution with $\xi = 0.5$. Estimation of ξ (top) and tail probability (bottom), bias (left), RMSE (right): GPD-ML (full line), Ep with $\rho = -2$ (dash-dotted), $E\bar{p}$ with $(k_*, m) = (190, 150)$ (dashed), and ridge regression estimator (dotted).

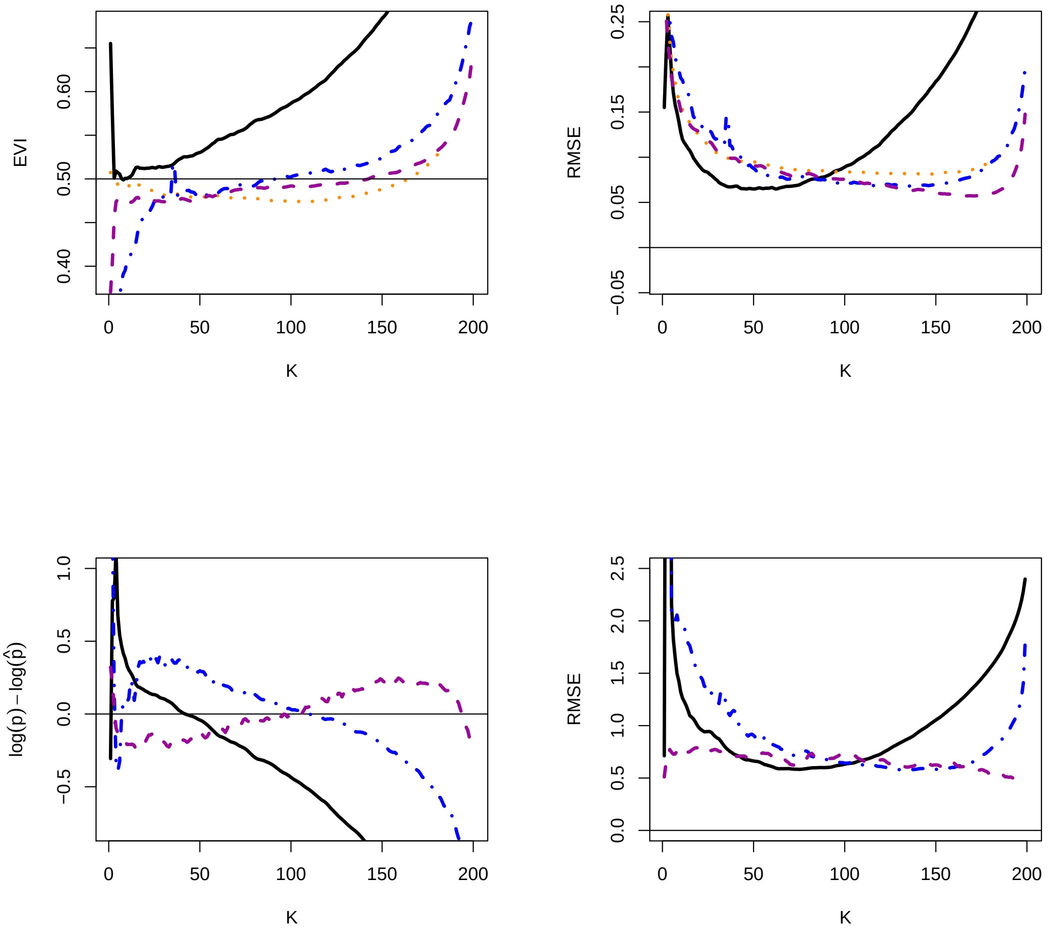

Figure 4: Fréchet distribution with $\xi = 0.5$. Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line), Ep^{+} (dash-dotted), $E\bar{p}^{+}$ (dashed) and corrected Hill estimator (dotted).

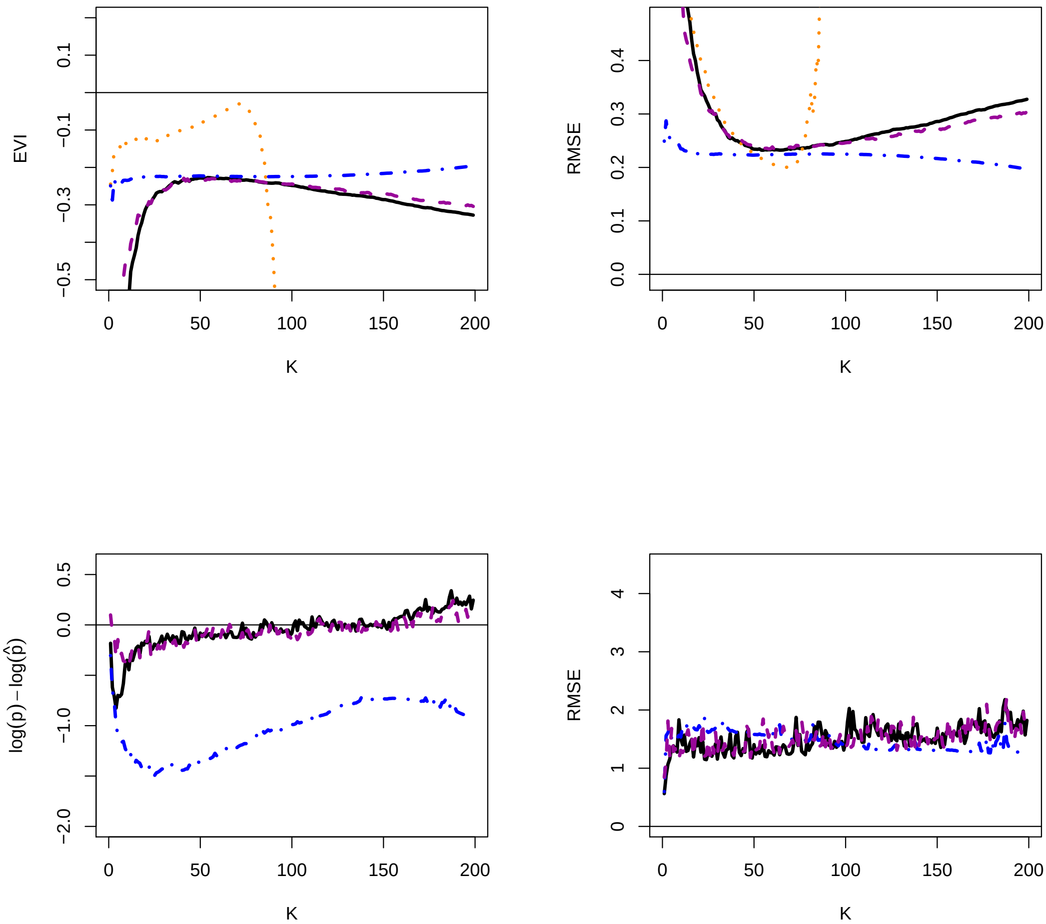

Figure 5: Standard normal distribution ($\xi = 0$ and $\tilde{\rho} = 0$). Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), Ep (dash-dotted), $E\bar{p}$ (dashed) and ridge regression estimator (dotted).

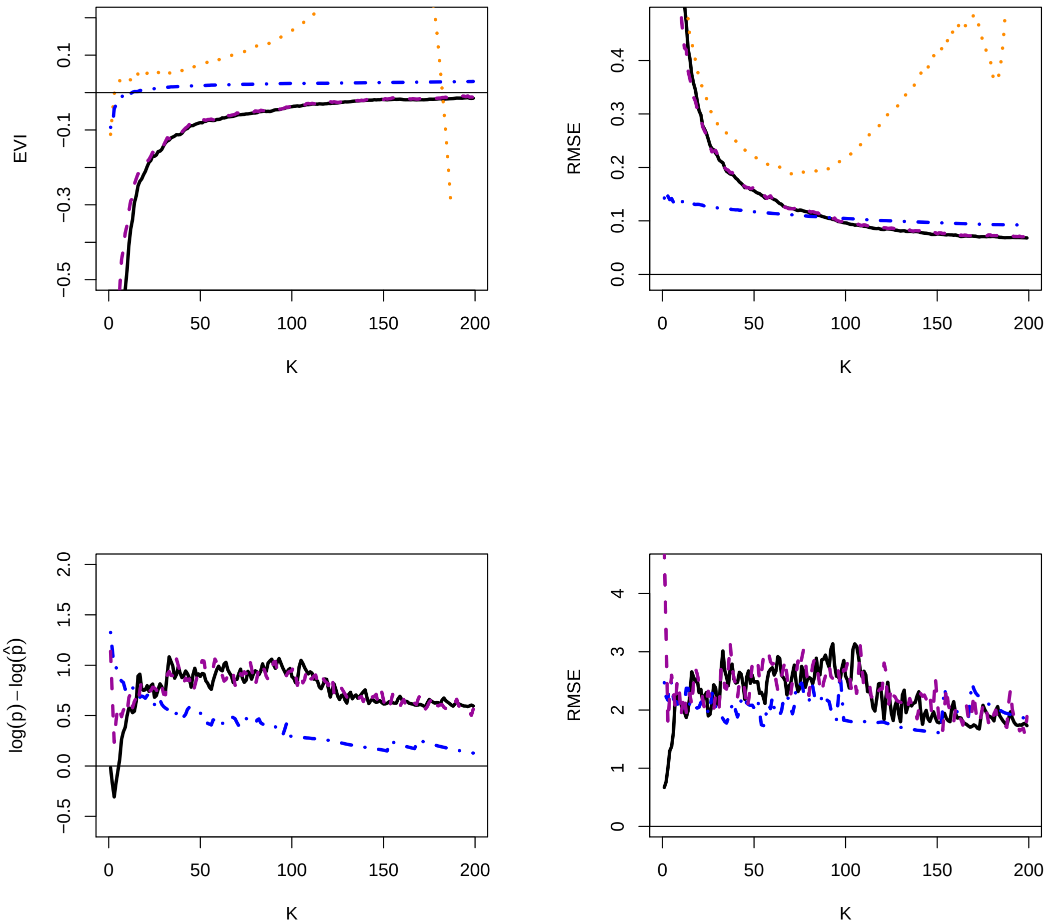

Figure 6: The exponential distribution ($\xi = 0$ and $\tilde{\rho} = 0$). Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), Ep (dash-dotted), $E\bar{p}$ (dashed) and ridge regression estimator (dotted).

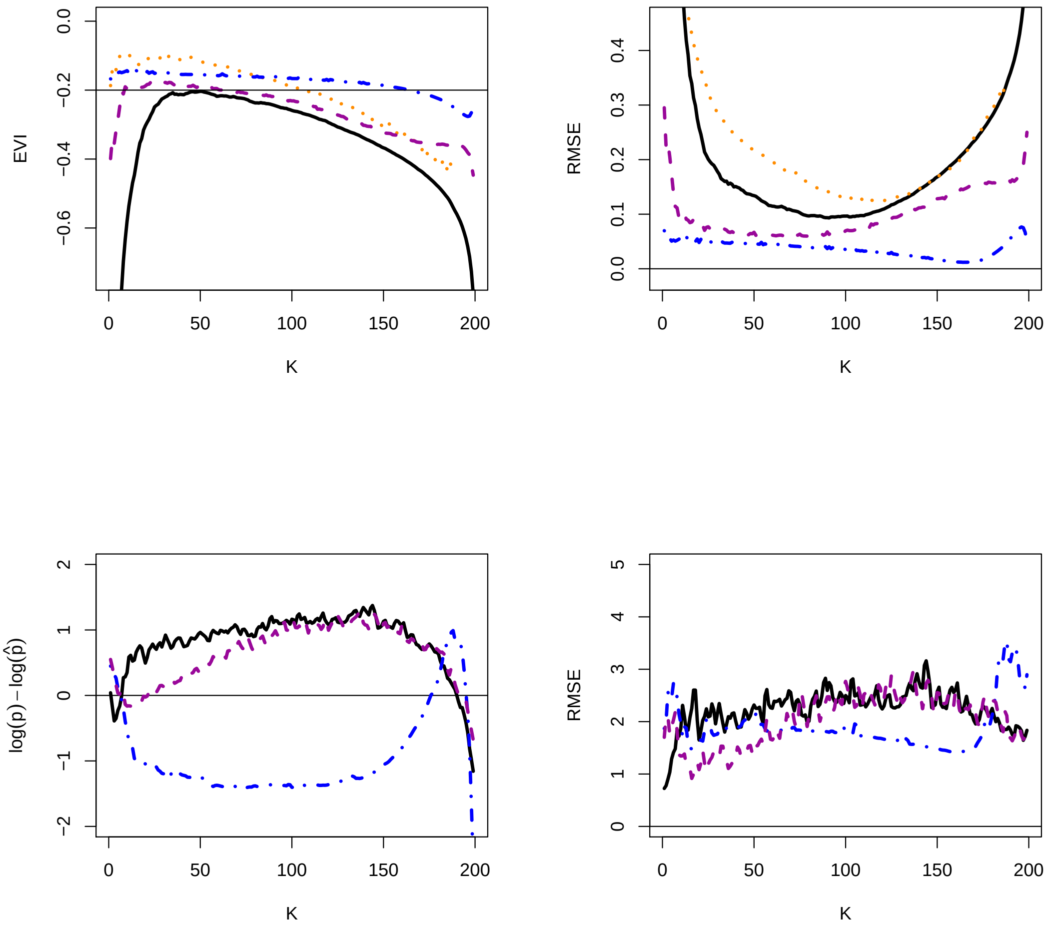

Figure 7: Reversed Burr distribution ($\xi = -0.2$ and $\tilde{\rho} = -1$). Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), Ep (dash-dotted), $E\bar{p}$ (dashed) and ridge regression estimator (dotted).

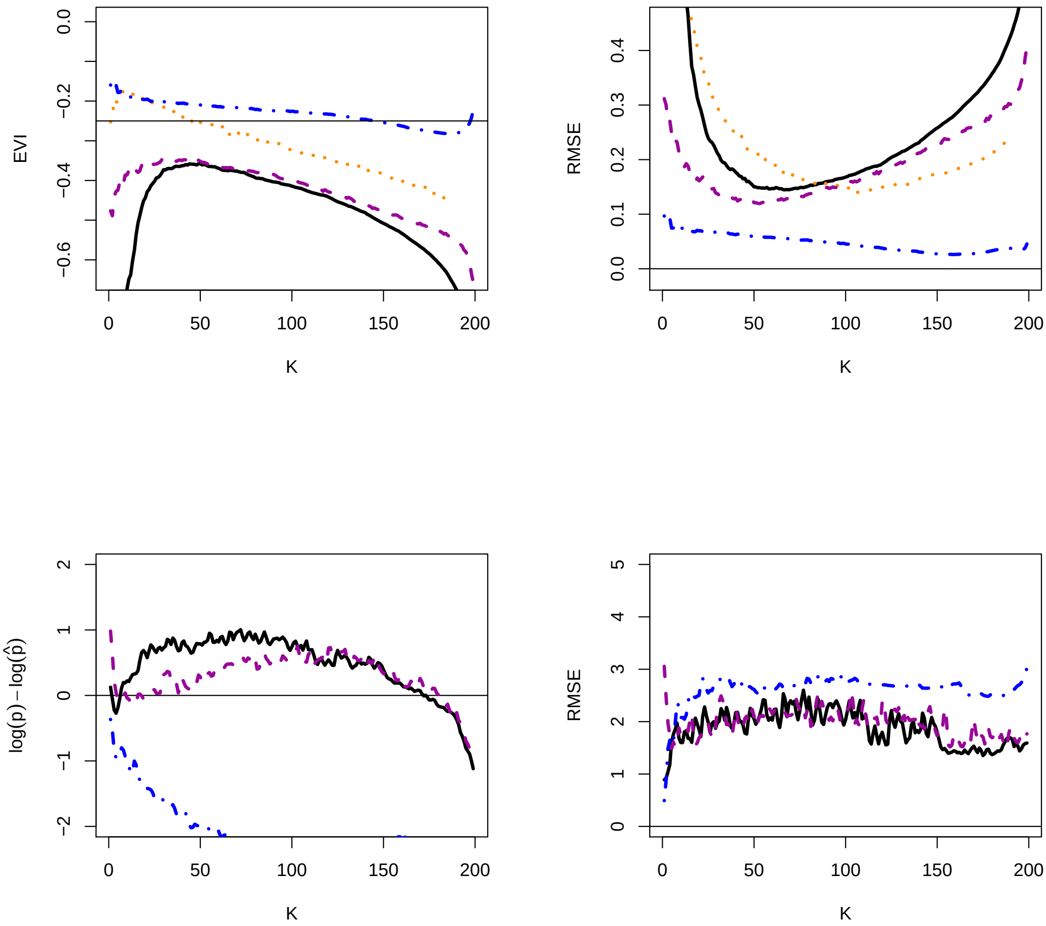

Figure 8: Extreme value Weibull distribution ($\xi = -0.25$ and $\tilde{\rho} = -1$). Estimation of ξ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), Ep (dash-dotted), $E\bar{p}$ (dashed) and ridge regression estimator (dotted).

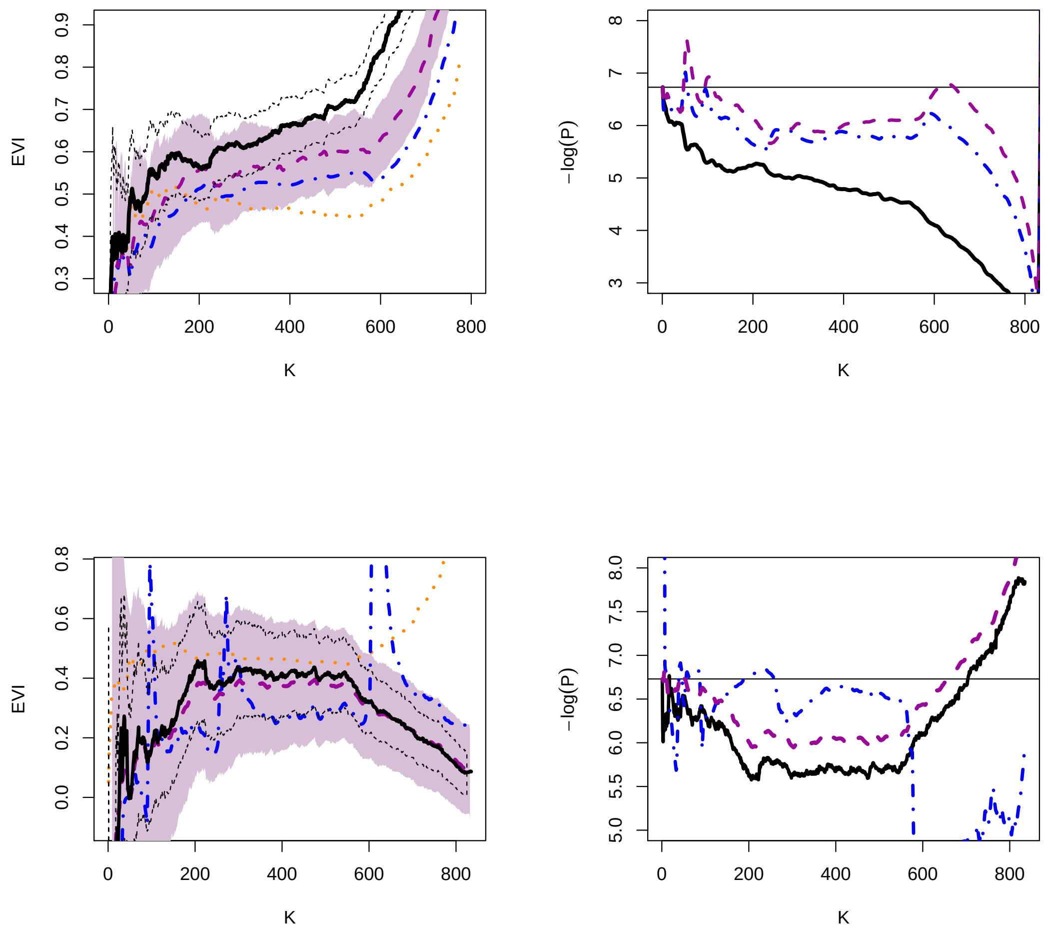

Figure 9: Ultimates of Belgian car insurance claims: estimation of ξ with asymptotic confidence intervals (left), tail probability estimation at maximum observation (right), Pareto-based analysis (top) and GPD-based analysis (bottom): classical ML estimation (full line with dotted confidence intervals), Ep (dashed with shaded confidence intervals) and $E\bar{p}$ (dash-dotted). CH (top left) and ridge regression (bottom left) estimators are indicated by dotted lines.

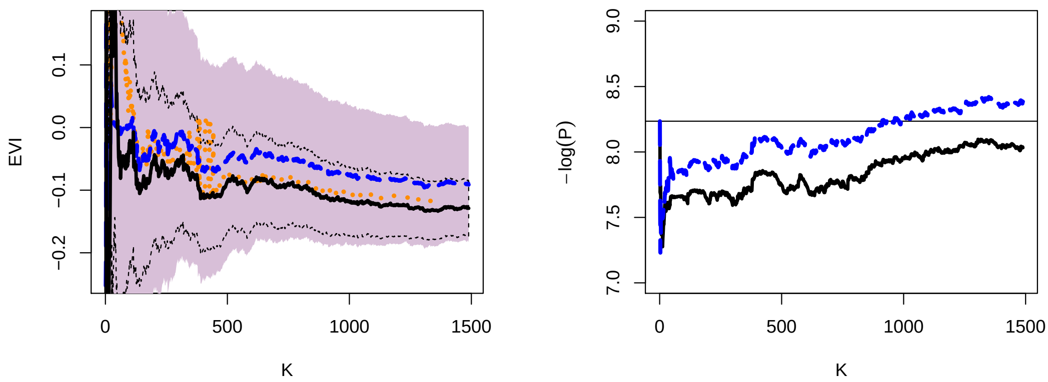

Figure 10: Lifetime data from the Netherlands, female persons who died in 1986. Left: estimation of ξ with asymptotic confidence intervals for classical ML estimation (full line with dotted confidence intervals), *Ep* (dashed with shaded confidence intervals, $\tilde{\rho} = -0.5$) and ridge regression (dotted). Right: tail probability estimation at maximum observation for classical ML estimation (full line) and *Ep* (dashed).