ENVIRONMETRICS
Environmetrics 2010; **21**: 661–686
Published online 12 February 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/env.1041

Change point analysis of extreme values‡

G. Dierckx[1,2*,†] and J. L. Teugels[2,3]

[1]*Department of Mathematics, Hogeschool Universiteit Brussel, Stormstraat, B-1000 Brussel, Belgium*
[2]*Departement of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium*
[3]*EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

SUMMARY

In a sample from the distribution of a random variable, it is possible that the tail behavior of the distribution changes at some point in the sample. This tail behavior can be described by absolute or relative excesses of the data over a high threshold, given that the random variable exceeds the threshold. The limit distribution of the absolute excesses is given by a generalized Pareto distribution with an extremal parameter γ and a scale parameter σ. When the extreme value index γ is positive, then the relative excesses can be described in the limit by a Pareto distribution with this index as parameter.

In this paper we concentrate on testing whether changes occur in the value of the extreme value index γ and/or the scale parameter. To this end, appropriate test statistics are introduced based on the likelihood approach for independent data. Asymptotic properties of these test statistics are studied leading to adequate critical values. After giving a practical test procedure, we apply our results to a series of simulations and real life examples.

2000 MATHEMATICS SUBJECT CLASSIFICATION: primary 62G32; secondary 60G70; 62P05; 62P12; 92F05

KEY WORDS: change point analysis; earthquake analysis; extreme value index; extreme value theory; generalized Pareto distribution; maximum likelihood; precipitation; stock market index

1. INTRODUCTION

When a phenomenon X is examined through the sample $X_1, \ldots X_n$, the distribution of X might change at some point. To investigate whether or not such a change point occurs, a change point analysis can be performed. In the literature, some authors follow the sequential approach of Gut and Steinebach (2005). However, in our approach, we will assume that the sample size n is fixed.

*Correspondence to: G. Dierckx, Department of Mathematics, Hogeschool Universiteit Brussel, Stormstraat, B-1000 Brussel, Belgium and Departement of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium.
†E-mail: goedele.dierckx@hubrussel.be
‡This article is published in *Environmetrics* as a special issue on TIES 2008: *Quantitative Methods for Environmental Sustainability*, edited by Sylvia R. Esterby, University of British Columbia Okanagan, Canada.

Received 21 October 2008
Accepted 24 December 2009

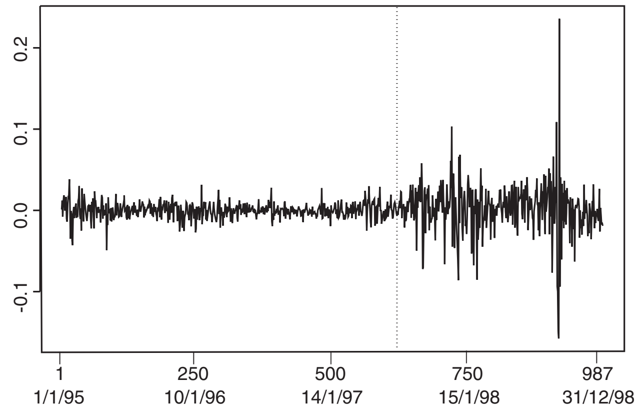

Figure 1. Daily stock market returns of the Malaysian index from January 1995 until December 2001. The dotted line is indicating July 1, 1997 (the start of the Asian crisis)

Example 1. An example where it is generally accepted that a change point occurred somewhere in July 1997, is provided by the daily stock market returns of the Malaysian stock index. In Figure 1, these returns are plotted from January 1995 until December 1998, covering the Asian financial crisis in July 1997. Malaysia was one of the countries that was most affected by the crisis. The Malaysian currency rate was linked to the US dollar and switched to a floating rate system at that time. Clearly from that point onward, the distribution of the returns changed and more extreme values occurred from July 1997 onward.

Although the change in the distribution can be exhibited through a variety of means, one mostly looks for changes in location or scale when performing a change point analysis. However, this might not always be the only possibility. In the Malaysian stock index example, the extreme values of the distribution might be of greater importance than their mean values. This is even more transparent in the discussion on climate change as performed by Crisci *et al*. (2002) and Katz and Brown (1992). In the latter case, the authors refer to the literature to motivate the study of a change in the extreme events, since any climate change will first lead to changes in the frequency and intensity of extreme events. This is also illustrated through simulation by Mason and Turova (1994).

Besides the fact that location or scale are not necessarily the main interests in an extreme value context, it might be problematic to study changes in mean and variance for other reasons. First of all, concentration on location and scale, often does not adequately describe the tail of the distribution. Secondly, one cannot always use the mean and the variance as location-scale parameters, as they might not even exist for some very heavy tailed distributions.

A solution might be to study changes in the distribution through quantiles as has been investigated by Ferro *et al*. (2005) by comparison of two time series. Apart from diagnostic plots, the authors provide a test based on a bootstrap method, to check whether a number of quantiles for the two time series coincide. Also here, the authors do not always find a good description of the tails.

In this paper, we concentrate on tests that deal with changes of the parameters γ and/or σ that better describe the tail of a distribution. These parameters occur in the limit distributions of relative and absolute excesses over a high threshold. More specifically, we will deal with two specific examples of such distributions.

DOI: 10.1002/env

- (i) X has a *Pareto-type distribution* with parameter $\gamma > 0$, when the relative excesses of X over a high threshold u, given that X exceeds u satisfy the condition

$$P\left(\frac{X}{u} > x | X > u\right) \to x^{-\frac{1}{\gamma}}, \quad u \to \infty \tag{1}$$

- (ii) More generally X follows a *generalized Pareto distribution* (GPD) with parameters γ and σ if the behavior of the absolute excesses over a high threshold u satisfies the condition

$$P(X - u > x | X > u) \to \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}, \quad u \to \infty \tag{2}$$

  This limiting result appears naturally in applications of the *peaks over threshold method* (POT).

Extensive treatments of both distributions can be found in Beirlant *et al*. (1999, 2004).

In Section 2, we construct test statistics that will test whether γ and/or σ have changed. The construction is based on a likelihood approach introduced by Csörgő and Horváth (1997) for independent data but adapted to an extreme value context. A similar approach was applied in Dias and Embrechts (2004) where dependence structures for extremal data have been studied using parametric copulas. In Jarušková and Rencová (2008), the likelihood approach is used as well to find changes in the location, scale and shape of the generalized extreme value distributions, describing minima or maxima of data. In the current text we base our analysis on all extreme data points, not just minima and maxima over a certain period in time. Apart from applying the likelihood approach of Csörgő and Horváth (1997), we also provide a clear link between our results and some well known statistics with their properties from extreme value theory. The arguments of the paper and theorems are heuristic and not claimed to be complete. However they are backed up by simulations.

The required asymptotic theory is provided in Section 3, where we derive the asymptotic distribution of the test statistics under the null hypothesis. Based on these results, critical values for the tests can be derived.

Finally, a practical test procedure is given in Section 4. The procedure is applied to a simulation study in Section 5 and to a number of real life examples in Section 6, some having an environmental context. In the Malaysian stock index example, we also illustrate how the condition of independence of the data often can be circumvented by first performing a declustering procedure.

2. CONSTRUCTION OF THE TEST STATISTICS

We first introduce the test statistic proposed by Csörgő and Horváth (1997). In Sections 2.2 and 2.3, we adapt the statistic to a more specific extreme value context.

2.1. General distribution

Csörgő and Horváth (1997) study the change of parameters in a distribution. Based on a sample $X_1, \ldots, X_{m^*}, X_{m^*+1}, \ldots X_n$, with density function $f(x; \theta_i, \eta)$ for X_i, their goal is to test whether θ_i

DOI: 10.1002/env

changes at some point m^*

$$
\begin{aligned}
&H_0 : \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \\
&H_1 : \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \text{ for some } m^*
\end{aligned} \tag{3}
$$

It should be noted that the parameters θ and η can be vectors.

In order to test hypothesis (3), Csörgő and Horváth (1997) introduce the test statistic

$$
Z_n = \sqrt{\max_{1 \leqslant m < n} (-2 \log \Lambda_m)} \tag{4}
$$

where

$$
\Lambda_m = \frac{\sup_{\theta,\eta} \prod_{i=1}^{n} f(X_i;\theta,\eta)}{\sup_{\theta,\tau,\eta} \prod_{i=1}^{m} f(X_i;\theta,\eta) \prod_{i=m+1}^{n} f(X_i;\tau,\eta)}
$$

The sample is split repeatedly in two groups $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots X_n$, for $m = 1 \ldots n-1$. Clearly, a change in parameters at $m = m^*$ results in a small Λ_{m^*} and thus in a large Z_n. As such, we can reject H_0 for a test statistic value Z_n that is larger than some critical value. In Section 3, this critical value will be determined, based on the asymptotic theory. Thus, Z_n looks a fair candidate for a test statistic to test hypothesis (3).

Naturally, the value m for which the maximum Z_n is attained, can be used to estimate when the change, if any, occurred and the corresponding estimate is denoted by $\hat{m}$.

From a practical point of view, it is more convenient to rewrite Λ_m using the link to the likelihood functions L_m for $X_1, \ldots, X_m$, L_m^+ for $X_{m+1}, \ldots, X_n$ and L_n for$X_1, \ldots, X_n$. Csörgő and Horváth (1997) state that

$$
-2 \log \Lambda_m = 2\left[L_m(\hat{\theta}_m, \hat{\eta}_m) + L_m^+(\hat{\theta}_m^+, \hat{\eta}_m^+) - L_n(\hat{\theta}_n, \hat{\eta}_n)\right] \tag{5}
$$

where the likelihood estimator $\hat{\theta}_m$ is based on $(X_1, \ldots, X_m)$, $\hat{\theta}_m^+$ on $(X_{m+1}, \ldots, X_n)$ and $\hat{\theta}_n$ on $(X_1, \ldots, X_n)$.

We first calculate $-2 \log \Lambda_m$ in Equation (5) for the two specific cases that will be used in the next sections.

2.1.1. Exponential distribution. Suppose $X_1, \ldots, X_{m^*}, X_{m^*+1}, \ldots X_n$ are exponential random variables, with mean θ_i, $i = 1, \ldots, n$.

The likelihood estimators of the mean θ based on the samples $(X_1, \ldots, X_m)$, $(X_{m+1}, \ldots, X_n)$, and $(X_1, \ldots, X_n)$, respectively, are given by

$$
\hat{\theta}_m = \frac{1}{m} \sum_{i=1}^{m} X_i, \quad \hat{\theta}_m^+ = \frac{1}{n-m} \sum_{i=m+1}^{n} X_i, \textit{ and} \quad \hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i
$$

DOI: 10.1002/env

Evaluation of the likelihood functions at these likelihood estimators results in

$$L_m(\hat{\theta}_m) = -m\log\hat{\theta}_m - \frac{1}{\hat{\theta}_m}\sum_{i=1}^{m} X_i$$

$$L_m^+(\hat{\theta}_m^+) = -(n-m)\log\hat{\theta}_m^+ - \frac{1}{\hat{\theta}_m^+}\sum_{i=m+1}^{n} X_i$$

$$L_n(\hat{\theta}_n) = -n\log\hat{\theta}_n - \frac{1}{\hat{\theta}_n}\sum_{i=1}^{n} X_i$$

Introduction of these expressions in formula (5) shows that $-2\log\Lambda_m$ equals

$$2\left[-m\log\frac{1}{m}\sum_{i=1}^{m} X_i - (n-m)\log\frac{1}{n-m}\sum_{i=m+1}^{n} X_i + n\log\frac{1}{n}\sum_{i=1}^{n} X_i\right] \tag{6}$$

2.1.2. GPD. Suppose $X_1, \ldots, X_{m^*}, X_{m^*+1}, \ldots X_n$ follow a GPD with parameters $\theta_i = (\gamma_i, \sigma_i)$, $i = 1, \ldots, n$. The likelihood functions equal

$$L_m(\hat{\theta}_m) = -m\log\hat{\sigma}_m - \left(\frac{1}{\hat{\gamma}_m}+1\right)\sum_{i=1}^{m}\log\left(1+\hat{\gamma}_m\frac{x}{\hat{\sigma}_m}\right),$$

$$L_m^+(\hat{\theta}_m^+) = -(n-m)\log\hat{\sigma}_m^+ - \left(\frac{1}{\hat{\gamma}_m^+}+1\right)\sum_{i=m+1}^{n}\log\left(1+\hat{\gamma}_m^+\frac{x}{\hat{\sigma}_m^+}\right) \tag{7}$$

for the likelihood estimators $(\hat{\gamma}_m, \hat{\sigma}_m)$ and $(\hat{\gamma}_m^+, \hat{\sigma}_m^+)$, respectively, based on the samples $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots X_n$. These maximum likelihood estimators can be calculated numerically.

2.2. Pareto-type distributions

We now adapt the test statistic of Csörgő and Horváth (1997) to an extreme value setting where we want to check possible changes in the parameters in Equations (1) and (2).

Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and Pareto-type distributed. We denote the extreme value index for X_i with γ_i. In order to determine whether the index γ changes or not, we test the following hypothesis

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_n = \gamma \quad \text{versus}$$
$$H_1 : \gamma_1 = \ldots = \gamma_{m^*} \neq \gamma_{m^*+1} = \ldots = \gamma_n \text{ for some } m^* \tag{8}$$

We first show that this hypothesis can be reformulated in terms of the mean parameter of an exponential distribution. To do that, we look at the logarithm of the relative excess of X over a threshold u, given that X exceeds u. From Equation (1) it follows that this logarithm is asymptotically exponentially distributed,

DOI: 10.1002/env

as $u \to \infty$. Moreover, the mean of the exponential distribution is given by the extreme value index γ. As is well known in extreme value analysis, one needs to restrict the number of extremes. In practice, one uses the order statistic $X_{n-k,n}$ as threshold u. As such, only the k excesses above this threshold appear in the subsequent analysis. As traditional in extreme value theory, *asymptotically ($u \to \infty$)* has to be interpreted as $k = k_{(n)} \to \infty$ such that $\frac{k_{(n)}}{n} \to 0$, as $n \to \infty$. Then the sequence $k_{(n)}$ is called an *intermediate* sequence.

We denote the logarithms of the relative excesses of the first group $X_1, X_2, \ldots, X_m$ over X_{n-k} by $E_1, \ldots, E_{k_1}$ where we assume that there are k_1 excesses in this group. In the same way $E_{k_1+1}, \ldots, E_k$ denote the logarithms of the relative excesses in the second group $X_{m+1}, X_{m+2}, \ldots, X_n$ and k_2 their number. Further put $k = k_1 + k_2$. It therefore follows that the a test for the hypothesis in Equation (8), based on the sample $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$, is asymptotically equivalent to a test for a change in mean of an exponential distribution based on the sample $E_1, \ldots, E_{k_1}, E_{k_1+1}, \ldots E_k$.

To test Equation (8) we can use the test statistic in Equation (4) based on the exponential sample $E_1, \ldots, E_{k_1}, \; E_{k_1+1}, \ldots, E_k$. Applying formula (6), it immediately follows that $-2 \log \Lambda_m$ can be calculated as

$$2\left[-k_1 \log \frac{1}{k_1}\sum_{i=1}^{k_1} E_i - (k-k_1)\log \frac{1}{k-k_1}\sum_{i=k_1+1}^{k} E_i + k \log \frac{1}{k}\sum_{i=1}^{k} E_i\right] \tag{9}$$

Note that $\frac{1}{k_1}\sum_{i=1}^{k_1} E_i$ is asymptotically equivalent to the Hill (1975) estimator $H_{k_1,m}$ based on the sample $X_1, X_2, \ldots, X_m$. Indeed, this estimator is defined as

$$H_{k_1,m} = \frac{1}{k_1}\sum_{i=1}^{k_1} \log X_{m-i+1,m} - \log X_{m-k_1,m}$$

In the same way, $\frac{1}{k-k_1}\sum_{i=k_1+1}^{k} E_i$ is asymptotically equal to the Hill estimator based on the sample $X_{m+1}, \ldots X_n$.

Therefore, we propose the following test statistic to test hypothesis (8)

$$Z_n = \sqrt{\max_{1 \leqslant m < n} (-2 \log \Lambda_m)}$$

where in turn

$$\begin{aligned} -2 \log \Lambda_m = & -2\left[k_1 \log H_{k_1,m} + (k-k_1)\log H_{k-k_1,n-m} - k \log H_{k,n}\right] \\ & - \left[\frac{1}{H_{k,n}}(k_1 H_{k_1,m} + (k-k_1)H_{k-k_1,n-m} - k H_{k,n})\right] \end{aligned} \tag{10}$$

The equivalence, under H_0, of $-2 \log \Lambda_m$ in Equations (9) and (10) follows from the Appendix.

Example 2. We illustrate the use of our test statistic for simulated data from the heavy-tailed Burr (1942) distribution. Recall that a Burr(β, τ, λ) distribution F has the form

$$1 - F(x) = \beta^{\lambda}/(\beta + x^{\tau})^{\lambda}$$

DOI: 10.1002/env

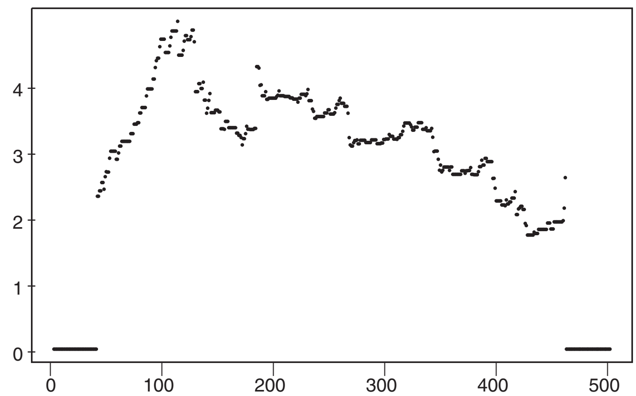

Figure 2. $(m, \sqrt{-2 \log \Lambda_m})$ for the simulated data set in Example 2, where m indicates the group which is split into two groups. The likelihood expression is calculated as in Equation (10)

We consider a simulation from a Burr(1,1,1) and Burr(1,0.5,1) distribution, respectively, with $n = 500$ and $m^* = 100$. This means that the first 100 data come from a Burr(1,1,1) distribution with $\gamma = 1$ while the next 400 data come from a Burr(1,0.5,1) distribution with γ equals 2. Using for example the threshold $u = 1.64$, the square root of $-2 \log \Lambda_m$ in Equation (10) is plotted in Figure 2, for $m = 1, \ldots n - 1$.

The test statistic Z_n for the test in Equation (8) is given by the maximum value of this square root, which equals $z_n = 4.67$. Whether this value will lead to the rejection of H_0 in Equation (8) depends on the critical value of the test statistic which will be derived in Section 3. But if H_0 has to be rejected, then the time of change can be estimated by that m-value for which the maximum of $-2 \log \Lambda_m$ is attained. In this example, this maximum is found for $m = 112$. In any case $m = 112$, is an acceptable estimate for the true value $m^* = 100$.

2.3. GPD

We turn to the more general GPD. Suppose that $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent but not necessarily heavy-tailed. Instead of investigating the relative excesses over a threshold u, we now turn to absolute excesses. Again we will use as threshold $u = X_{n-k,n}$ so that there are k excesses in the sample. From Equation (2), we observe that this absolute excess asymptotically ($u \to \infty$) follows a generalized Pareto distribution with parameters $\theta = (\gamma, \sigma)$.

In the same way as in Section 2.2, we introduce the notations $E_1, \ldots, E_{k_1}$ for the absolute excesses over X_{n-k} of the group $X_1, \ldots, X_m$ and we assume that there are k_1 excesses in number. Also, $E_{k_1+1}, \ldots, E_k$ denote the absolute excesses in the group $X_{m+1}, X_{m+2}, \ldots, X_n$ and $k_2 := k - k_1$ their number.

To test the hypothesis

$$
\begin{aligned}
&H_0 : \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \\
&H_1 : \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \text{ for some } m^* \qquad (11)
\end{aligned}
$$

DOI: 10.1002/env

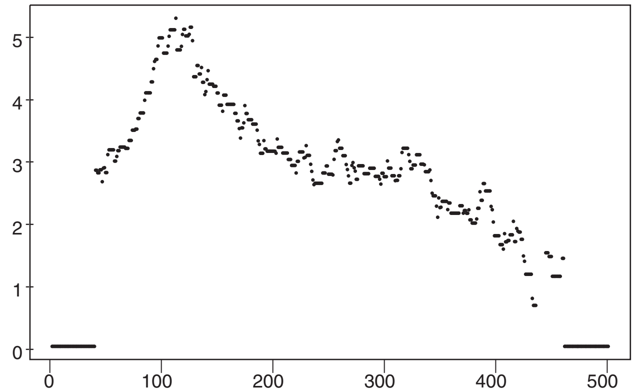

Figure 3. $(m, \sqrt{-2\log\Lambda_m})$ for the simulated data set in Example 2, where m is indicating the group which is split up into two groups. The likelihood expression is calculated as in Equation (12)

we can immediately apply the method in Section 2.1 for the GPD distribution based on the sample $E_1, \ldots, E_{k_1}, E_{k_1+1}, \ldots, E_k$. This 'Peaks over Threshold' technique has been studied well in the literature. For more information we refer for example to Davison and Smith (1990).

Henceforth, we can use as test statistic to test hypothesis (11)

$$Z_n = \sqrt{\max_{1 \leqslant m < n} (-2\log\Lambda_m)}$$

where

$$-2\log\Lambda_m = 2\left[L_{k_1}(\hat{\theta}_{k_1}) + L_{k_1}^{+}(\hat{\theta}_{k_1}^{+}) - L_k(\hat{\theta}_k)\right] \tag{12}$$

and with the likelihood expressions as in Equation (7).

Example 2 (continued). As an illustration, we use the same simulated data as in Figure 2. Again the threshold is set at $u = 1.64$ and the square root of $-2\log\Lambda_m$ is pictured in Figure 3, for $m = 1, \ldots n-1$.

The test statistic to test hypothesis (11), the maximum of this square root, equals 5.26. As before, this maximum is attained at $\hat{m} = 112$, the estimate of the time of change found in Section 2.2.

Note that hypothesis (11) involves a vector of parameters. Of course the method could be adapted to test for a change in only one of the parameters.

3. ASYMPTOTIC THEORY OF THE TEST STATISTIC

In this section, we investigate the asymptotic null distributions of the test statistics of Section 2.2 and Section 2.3. Based on the limiting distributions, an approximate critical value can be obtained to test changes in the parameters γ and/or σ.

DOI: 10.1002/env

3.1. Pareto-type distributions

We closely follow the approach in Theorem 1.3.2 in Csörgő and Horváth (1997).

Theorem 3.1. *Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and Pareto-type distributed. We assume the von Mises condition $xf(x)/(1-F(x)) \to \gamma$. We set the threshold at $u = X_{n-k,n}$. Define*

$$Z_n = \sqrt{\max_{c_n \leqslant m < n-d_n} (-2\log\Lambda_m)}$$

with $-2\log\Lambda_m$ as in Equation (10). Let $n, k \to \infty$ such that $k/n \to 0$. Further, let c_n and d_n be intermediate sequences, but bounded away from 0 by a small amount $\epsilon > 0$. Then, under H_0 in Equation (8), the following weak convergence holds

$$Z_n \to_D \sqrt{\sup_{\epsilon \leqslant t < 1-\epsilon} \frac{B^2(t)}{t(1-t)}}$$

with $B(t)$ a Brownian bridge.

Proof. From Equation (10), it follows that $-2\log\Lambda_m$

$$\begin{aligned}
&= -2\left\{k_1 \log\left(1 + \left(\frac{H_{k_1,m}}{\gamma} - 1\right)\right) + (k-k_1)\log\left(1 + \left(\frac{H_{k-k_1,n-m}}{\gamma} - 1\right)\right)\right. \\
&\quad -k\log\left(1 + \left(\frac{H_{k,n}}{\gamma} - 1\right)\right) \\
&\quad \left. -\frac{1}{H_{k,n}}\left[k_1 H_{k_1,m} - (k-k_1)H_{k-k_1,n-m} + kH_{k,n}\right]\right\} + o_P(1)
\end{aligned} \tag{13}$$

The first order terms in Equation (13) weakly converge to 0, since the Hill estimator is known to be a consistent estimator of γ.

The second order terms can be approximated in terms of a Wiener process. Indeed, the Hill estimator process has been studied in the literature. For more information on the limiting results and the error terms we refer for example to Mason and Turova (1994) and Kaufmann and Reiss (1998). Under the von Mises condition, weak convergence follows

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor kt \rfloor} \frac{E_i - \gamma}{\gamma} \to_D W(t)$$

with a limiting Wiener process $W(t)$, for $0 < t < 1$. For $t = \frac{k_1}{k}$ this result can be translated as

$$\sqrt{k_1}\left(\frac{H_{k_1,m}}{\gamma} - 1\right) \to_D \frac{W(t)}{\sqrt{t}}$$

DOI: 10.1002/env

which for $k_1 = k$ becomes

$$\sqrt{k}\left(\frac{H_{k,n}}{\gamma} - 1\right) \to_D W(1)$$

In a similar fashion, we can find a Wiener process $\tilde{W}(t)$ such that

$$\frac{1}{\sqrt{k}} \sum_{i=\lfloor kt \rfloor+1}^{k} \frac{E_i - \gamma}{\gamma} \to_D \tilde{W}(t)$$

and hence again for $t = \frac{k_1}{k}$,

$$\sqrt{k - k_1}\left(\frac{H_{k-k_1,n-m}}{\gamma} - 1\right) \to \frac{\tilde{W}(t)}{\sqrt{1-t}}$$

Clearly $\tilde{W}(t)$ and $W(t)$ are linked as $\tilde{W}(t) = W(1) - W(t)$, because $\sum_{i=k_1+1}^{n} E_i = \sum_{i=1}^{k} E_i - \sum_{i=1}^{k_1} E_i$.

When we use these limiting Wiener processes for $t = \frac{k_1}{k}$, the second order terms in Equation (13) weakly converge to

$$\begin{aligned}
&\frac{W^2(t)}{t} + \frac{[W(1) - W(t)]^2}{1-t} - W^2(1) \\
&= \frac{1}{t(1-t)}\left[W^2(t)(1-t) + t[W(1) - W(t)]^2 - t(1-t)W^2(1)\right] \\
&= \frac{1}{t(1-t)}[W(t) - tW(1)]^2 = \frac{1}{t(1-t)} B^2(t)
\end{aligned}$$

where $W(t) - tW(1) = B(t)$ is called a Brownian bridge. Theorem 3.1 now follows immediately.

■

Theorem 3.1 can now be used to approximate critical values to test hypothesis(8). To this end, we calculate the corresponding critical values for sup $B(t)/\sqrt{t(1-t)}$. These critical values for sup $\frac{B(t)}{\sqrt{t(1-t)}}$ can be calculated from a result in Csörgő and Horváth (1997) where it is stated that the distribution of the limit $P\left(\sup_{c \leqslant t < 1-d} \frac{B(t)}{\sqrt{t(1-t)}} \geqslant x\right)$ can be approximated by

$$\frac{xe^{-x^2/2}}{2^{1/2}\Gamma(\frac{1}{2})}\left[\log\frac{(1-c)(1-d)}{cd} - \frac{1}{x^2}\log\frac{(1-c)(1-d)}{cd} + \frac{4}{x^2} + O\left(\frac{1}{x^4}\right)\right] \tag{14}$$

These authors also found that $c = d = (\log k)^{3/2}/k$ is a good choice.

For example, when $\alpha = 0.05$ these critical values lead to the critical values to test hypothesis (8) as shown in Figure 4.

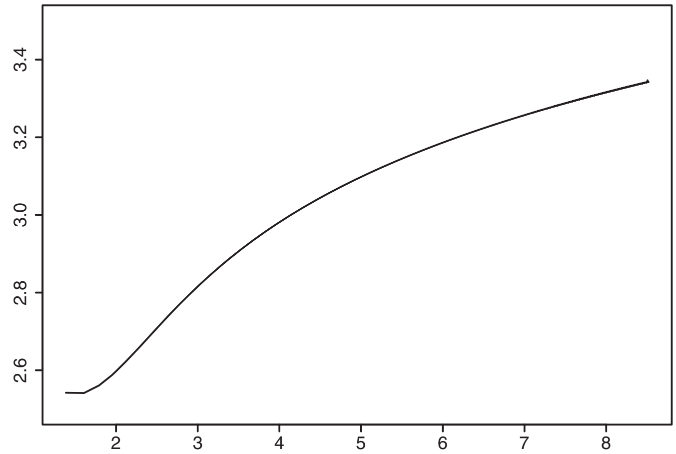

Figure 4. Critical values for the test statistic Z_n to test Equation (8) as a function of $\ln(k)$

Remark: In Theorem 1.3.1 in Csörgő and Horváth (1997), the authors also provide a limiting result in terms of an extreme value distribution. The rate of convergence to extreme value distributions, however, is believed to be usually very slow. The authors argue that rate of the Brownian bridge approximation in their Theorem 1.3.2 is fairly good.

3.2. GPD

We succinctly indicate how to deal with this more general case. Following again the approach as in the previous theorem but with the set up of Section 2.3, we arrive at our next result.

Theorem 3.2. *Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are identically distributed and independent following the same von Mises condition as in Theorem 3.1. Let the threshold be set at $u = X_{n-k,n}$. Define*

$$Z_n = \sqrt{\max_{c_n \leqslant m < n-d_n} (-2 \log \Lambda_m)}$$

with $-2 \log \Lambda_m$ as in Equation (12). Let $n, k \to \infty$ such that $k/n \to 0$. Further, let c_n and d_n be intermediate sequences, but bounded away from 0 by a small amount $\epsilon > 0$. Then, under H_0 in Equation (11), the following weak convergence holds

$$Z_n \to_D \sqrt{\sup_{\epsilon \leqslant t < 1-\epsilon} \frac{B_2^2(t)}{t(1-t)}}$$

with $B_2(t)$ a sum of two Brownian bridges.

DOI: 10.1002/env

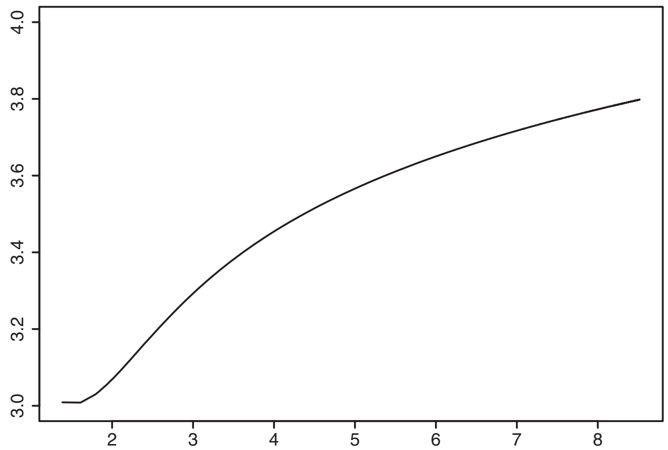

Figure 5. Critical values for the test statistic Z_n to test Equation (11) as a function of ln(k)

To determine the critical values we appeal to another result in Csörgő and Horváth (1997) where it is stated that the distribution of the limit $P\big(\sup_{c\leqslant t<1-d}\frac{B_2(t)}{\sqrt{t(1-t)}}\geqslant x\big)$ *can be approximated by*

$$\frac{x^2\exp\left(-\frac{x^2}{2}\right)}{2}\left[\log\frac{(1-c)(1-d)}{cd}-\frac{2}{x^2}\log\frac{(1-c)(1-d)}{cd}+\frac{4}{x^2}+O\left(\frac{1}{x^4}\right)\right] \tag{15}$$

The critical values to test hypothesis (11) are illustrated in Figure 5.

4. PRACTICAL TEST PROCEDURE

In practice, one can use the following procedure to test hypothesis (8) for a change in γ of a Pareto-type distribution. The procedure can be used as well to test hypothesis (11) for a change in γ and σ for a GPD.

- Select a threshold u. Under H_0, one can use the data as a whole to do this. One way to proceed is to choose that threshold — or that value of k — that minimizes the asymptotic mean square error of the Hill estimator as suggested in Beirlant *et al*. (2004). This approach has the additional advantage that it is closely related to the likelihood idea. When we use this *optimal* threshold, then we will denote it by $u = X_{n-k_{\text{opt}},n}$.
- As can be seen from Theorem 3.1 (respectively, from Theorem 3.2), c_n and d_n have to be chosen in such a way that the sample sizes of the two groups are not too small. The sample sizes k_1 and $k_2 = k - k_1$ should not be smaller than $k_{\min} = (\log k_{\text{opt}})^{3/2}/k$, as indicated in Equation (14). Therefore, one can define c_n as the smallest number such that at least $k_{\min}$ of the data points $X_1, \ldots X_{c_n}$ are larger than u. In the same way, d_n is the smallest integer such that at least $k_{\min}$ of the data points $X_{n-d_n+1}, \ldots X_n$ are larger than u.

DOI: 10.1002/env

- Repeat the next steps for $m = c_n, \dots, n - d_n$.
  - Split the data up in two groups $X_1, X_2, \dots, X_m$ and $X_{m+1}, \dots, X_n$.
  - Calculate $-2\log\Lambda_m$ as in Equation (10) (respectively, as in Equation (12)).
- Calculate $Z_n = \sqrt{\max_{c_n \leqslant m < n-d_n}(-2\log\Lambda_m)}$ and compare Z_n with the critical values for sample size k in Figure 4 (respectively, in Figure 5) to draw a conclusion for hypothesis in Equation (8) (respectively, as in Equation (11)).

5. SIMULATION STUDY

We simulate 1000 data sets of size n (with $n = 100$, $n = 500$, $n = 5000$) from different Burr and stable distributions. To this end the built-in random generator of the software package *R* is used. For the Burr distribution we refer to Section 2.2. For the stable distribution with $\gamma = \frac{1}{\alpha}$, we take

$$X = \frac{\sin(\alpha U)}{(\cos U)^{1/\alpha}}\left(\frac{\cos((1-\alpha)U)}{E}\right)^{(1-\alpha)/\alpha}$$

with $U \sim Uniform(-\pi/2, \pi/2)$ and $E \sim$ *exponential* with mean 1. The change point is situated at different places m^* as in Table 1.

5.1. Pareto-type distributions

We first apply the practical procedure to test Equation (8) on significance level $\alpha = 0.05$, using the test statistic of Equation (10).

Some simulation results are given in Table 1. The first column [1] in Table 1 corresponds to results for cases where H_0 holds. The probability of rejecting H_0 in those cases agrees well with the significance level of the test. For the smallest data sets the level is slightly too high. For larger data sets the test is conservative. To investigate the power of the test, columns [2]–[7] show the probability of rejection for

Table 1. Estimation by simulation of the probability of rejecting H_0 in Equation (8) for $\alpha = 0.05$, using test statistic of Equation (10)

	Burr						Stable	
	H_0 true		H_0 false					
	[1] $\gamma = 1$		[2] $\gamma_1 = 1$, $\gamma_2 = 2$	[3] $\gamma_1 = 2$, $\gamma_2 = 1$	[4] $\gamma_1 = 1$, $\gamma_2 = 0.5$	[5] $\gamma_1 = 0.5$, $\gamma_2 = 1$	[6] $\gamma_1 = 1.25$, $\gamma_2 = 2$	[7] $\gamma_1 = 2$, $\gamma_2 = 1.25$
n		m^*						
100	0.096	20	0.191	0.460	0.486	0.182	0.103	0.172
	0.075	50	0.517	0.512	0.519	0.559	0.201	0.189
500	0.029	50	0.181	0.782	0.799	0.144	0.058	0.199
	0.044	100	0.378	0.955	0.951	0.645	0.183	0.430
	0.019	250	0.894	0.951	0.966	0.909	0.475	0.426

DOI: 10.1002/env

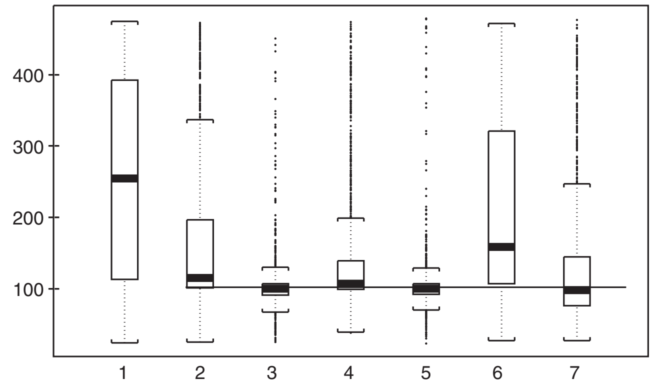

Figure 6. Box-plot of $\hat{m}$ for the cases in Table 1 for $n = 500$ and $m^* = 100$

cases where H_0 does not hold. These probabilities indicate that the test procedure works quite well for larger sample sizes n and when m^* is not too close to 1 or n. Especially for $m^* = n/2$, high powers are attained. Also note that the test works better if the group with the smallest extreme value index has the largest sample size. Clearly the stable distributions are more of a challenge than the Burr distribution.

When we estimate the location m^* of the change point from where $-2 \log \Lambda_m$ attains its maximum, the results are rather unstable. This can be noticed in Figure 6 for $n = 500$ and $m^* = 100$. From the box-plots of $\hat{m}$ for the same cases as in Table 1, we observe, however, that the median of $\hat{m}$ is close to the true m^*; still, the variability is large.

5.2. GPD

We turn to the practical procedure to test Equation (11) on significance level $\alpha = 0.05$, using test statistic of Equation (12).

Some simulation results are given in Table 2. Also here we notice in the first column [1] that the procedure falsely rejects H_0 in 5% of the cases. In this case more cases are falsely rejected than in

Table 2. Estimation by simulation of the probability of rejecting H_0 in Equation (11) for $\alpha = 0.05$, using test statistic of Equation (12)

	Burr						Stable	
	H_0 true		H_0 false					
	[1]		[2]	[3]	[4]	[5]	[6]	[7]
	$\gamma = 1$		$\gamma_1 = 1$	$\gamma_1 = 2$	$\gamma_1 = 1$	$\gamma_1 = 0.5$	$\gamma_1 = 1.25$	$\gamma_1 = 2$
			$\gamma_2 = 2$	$\gamma_2 = 1$	$\gamma_2 = 0.5$	$\gamma_2 = 1$	$\gamma_2 = 2$	$\gamma_2 = 1.25$
n		m^*						
100	0.143	20	0.354	0.430	0.484	0.369	0.204	0.193
	0.158	50	0.598	0.621	0.627	0.679	0.265	0.252
500	0.149	50	0.361	0.823	0.835	0.418	0.249	0.317
	0.188	100	0.543	0.958	0.968	0.777	0.409	0.495
	0.149	250	0.933	0.971	0.978	0.939	0.584	0.562

DOI: 10.1002/env

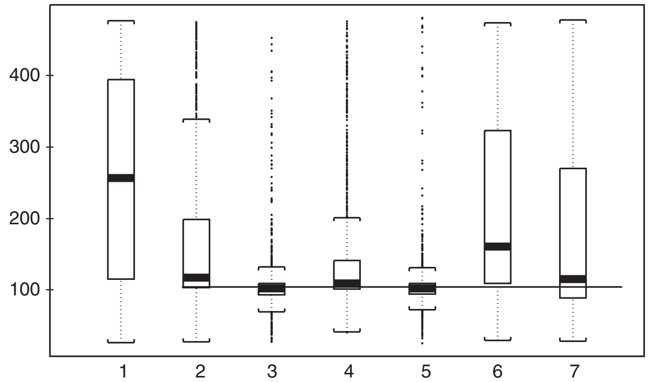

Figure 7. Box-plot of $\hat{m}$ for the cases in Table 2 for $n = 500$ and $m^* = 100$

Table 1. In columns [2]–[7], the probability of rejection for cases where H_0 does not hold, indicates a reasonable power of the test, especially for the larger data sets. The powers are usually higher than in Table 1. As before, the test works better for $m^* = n/2$ and also when the group with the smallest extreme value index has the largest sample size.

The estimation of m^* is shown in Figure 7 and leads to similar conclusions as for the Pareto-type case.

6. EXAMPLES

We now pass to a number of real life examples.

6.1. Earthquake data

We refer here to Beirlant *et al*. (2004), pp. 200–208. The data contain seismic moment measurements of shallow earthquakes over the period from 1977 to 2000 for subduction zones and midocean ridge zones. The first 6458 data points correspond to measurements for subduction zones. The data for the midocean ridge zone are given by the subsequent 1664 figures in the data set.

In Figure 8(a), the separate Pareto QQ plots for the measurements of the two different zones ultimately show a linear behavior, suggesting Pareto type behavior. Using the Hill estimator, as illustrated in Figure 8(b), for the separate groups, the extreme value index γ_1 for subduction zones can be estimated by 1.4, while γ_2 for the midocean ridge zones is estimated to be smaller, namely 0.9. To obtain these estimators, optimal thresholds leading to a minimal asymptotical mean squared error are choosen in both groups. The mean squared error of the Hill estimator, based on the whole data set, is minimized for the threshold u given by $X_{8123-1333,8123} = 1.28\ 10^{25}$ so that $k = k_{\text{opt}} = 1333$.

(1) *Pareto-type distribution*
We test for a change in γ as in Section 2.2. Therefore, based on the logarithm of the relative excesses over the threshold u, we calculate $-2 \log \Lambda_m$, $1 \leqslant m \leqslant n - 1$ as in Equation (10). The square root of this expression is plotted in Figure 9.

DOI: 10.1002/env

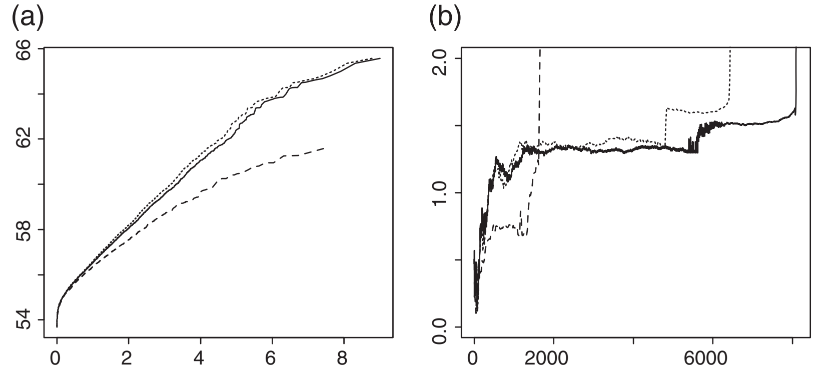

Figure 8. (a) Pareto quantile plots and (b) biased reduced estimators for γ (see Beirlant *et al*. (1995)) for the earthquake data. In these plots the results for the whole group are plotted with a full line. The dotted lines correspond to the first group and the dashed lines correspond to the second group

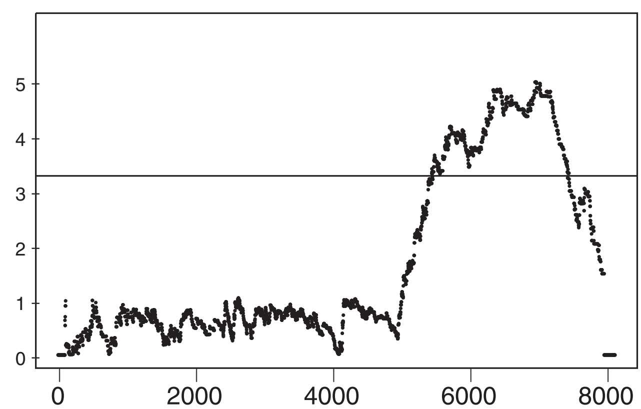

Figure 9. $(m, \sqrt{-2\log\Lambda_m})$ for the eartquake data, where the likelihood expression is calculated as in Equation (10). The critical value to test hypothesis (8) is indicated with a horizontal line

To draw a conclusion for hypothesis (8), we compare $Z_n = \sqrt{\max(-2\log\Lambda_m)} = 4.98$ to the appropriate critical value. This critical value 3.27 is added in Figure 9. Clearly the maximum of $\sqrt{-2\log\Lambda_m}$ is larger than the critical value. As a result we reject H_0. The time of change can be estimated by $\hat{m} = 6957$.

(2) *GPD*

Next, we test for a change point in $\theta = (\gamma, \sigma)$ as in Section 2.3. Therefore, based on the absolute excesses over the threshold u, we calculate $-2\log\Lambda_m$, $1 \leqslant m \leqslant n-1$ as in Equation (12). The square root of the expression is plotted in Figure 10.

To test hypothesis (11), we compare $Z_n = \sqrt{\max(-2\log\Lambda_m)} = 6.05$ to the appropriate critical value 3.31 which has been added to the Figure 10. Clearly the maximum of $\sqrt{-2\log\Lambda_m}$ is larger than the critical value, leading to the rejection of H_0. Note that Z_n is obtained for $\hat{m} = 7025$.

DOI: 10.1002/env

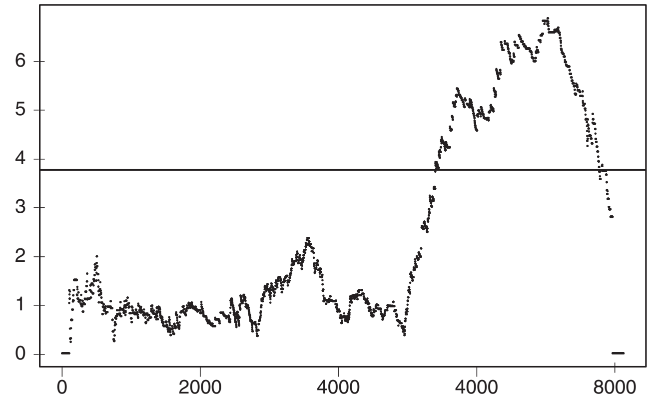

Figure 10. $(m, \sqrt{-2 \log \Lambda_m})$ for the earthquake data where the likelihood expression is calculated as in Equation (12). The critical value to test hypothesis (11) is indicated with a horizontal line

6.2. Malaysian stock index

We return to the example of the Malaysian stock index, initiated in the introduction.

6.2.1. Classical approach. In Figure 11(a) a separate Pareto QQ plot based on the data before July 1997 and one based on the data after July 1997 is provided. In both plots, a linear behavior might be observed for the extreme returns. As before the extreme value index for the two separate groups can be estimated using the Hill estimator as in Figure 11(b). Clearly, the extreme value index γ_1 before July 1997 can be estimated between 0.1 and 0.2, while γ_2 for after July 1997 is estimated around 0.5. These values can be obtained when we only use a small number of the largest data in each group. When more data are used, the estimates are similar for both groups. This example illustrates that one should be careful not to incorporate too many large data points into the analysis. The mean squared error of the Hill estimator based on the whole data set attains a local minimum for the threshold u given by $X_{987-224,987} = 0.0099$ so that $k = k_{\text{opt}} = 224$.

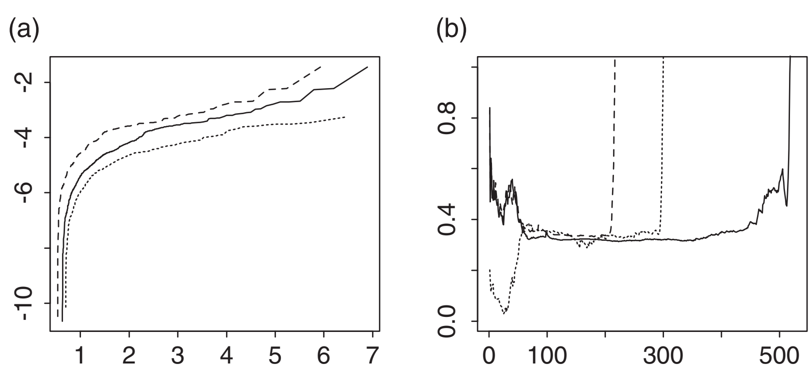

Figure 11. (a) Pareto quantile plot of the two separate groups of the Malaysian stock index data and (b) Biased reduced estimators of γ. In these plots, the results for the whole group are plotted with a full line. The dotted lines correspond to the first group and the dashed lines correspond to the second group

DOI: 10.1002/env

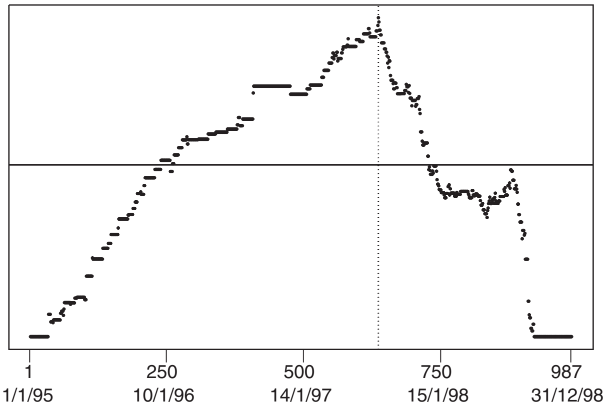

Figure 12. $(m, \sqrt{-2\log\Lambda_m})$ for the Malaysian stock index data, where m is indicating the group which is split up into two groups. The likelihood expression is calculated as in Equation (10). The critical value for the test is indicated with a horizontal line

(1) *Pareto-type distribution*
First we investigate the logarithm of the relative excesses and calculate $-2\log\Lambda_m$, $1 \leqslant m \leqslant n-1$ as in Equation (10). The square root of this expression is plotted in Figure 12.

To draw a conclusion for hypothesis (8), we compare $Z_n = \sqrt{\max(-2\log\Lambda_m)} = 5.8$ to the appropriate critical value 3.14. This critical value is added in Figure 12 with a horizontal line. The maximum of $\sqrt{-2\log\Lambda_m}$ is larger than the critical value 3.31 so that we reject H_0. Also note that a maximum is attained at $m = 635$, which corresponds to 1/08/1997, only shortly after the start of the Asian crisis.

(2) *GPD*
Next, we test for a change point in $\theta = (\gamma, \sigma)$ as in Section 2.3. We use the absolute excesses over the threshold u and calculate $-2\log\Lambda_m$, $1 \leqslant m \leqslant n-1$ as in Equation (12). The square root of the expression is plotted in Figure 13.

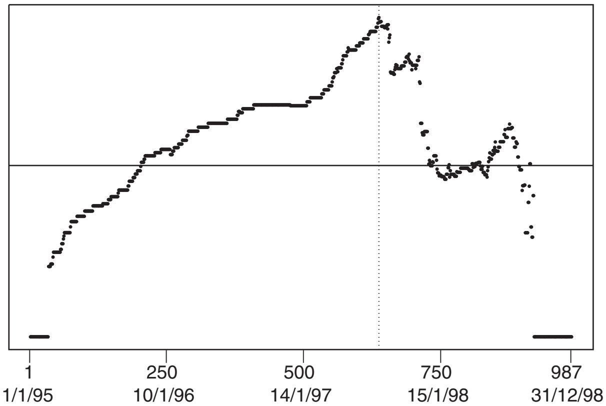

Figure 13. $(m, \sqrt{-2\log\Lambda_m})$ for the Malaysian stock index data, where the likelihood expression is calculated as in Equation (12). The critical value to test Equation (11) is indicated with a horizontal line

DOI: 10.1002/env

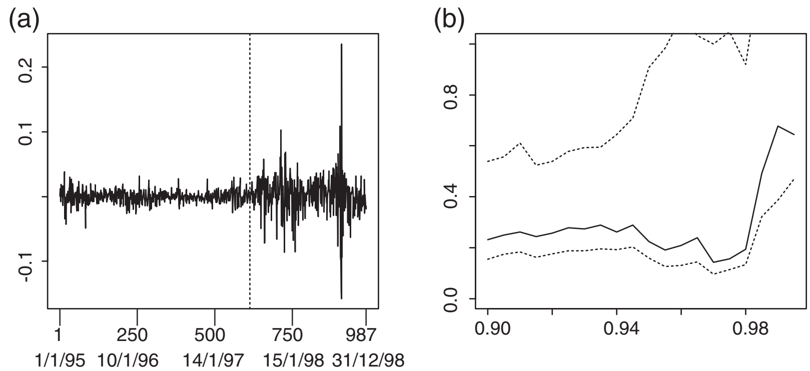

Figure 14. (a) Time series plot of the Malaysian stock return from January 1995 to December 2001 and (b) Extremal index as a function of the threshold with 95% bootstrap confidence intervals

The conclusion for hypothesis (11) is based on the comparison of Z_n which equals $\sqrt{\max(-2\log\Lambda_m)} = 5.93$ to the appropriate critical value 3.18 and added in Figure 13. Since the maximum of $\sqrt{-2\log\Lambda_m}$ is larger than the critical value, we again are forced to reject H_0. Also the instant of change is very close to the value before as $\hat{m} = 636$.

6.2.2. Improved approach. The above analysis has been performed under the assumption that the data were independent. However, it is a known fact that market data are not independent as can be seen from the time series plot given in Figure 14. The extremal index (see Beirlant *et al*. (1995)) seems to differ from the value 1.

Since our methods are closely related to that for the Hill estimator, we are confident that our methods are robust for some types of dependence. Indeed, it is known that the Hill estimator can withstand some dependence.

If one does not want to rely on this fact, then one can proceed as follows. Using the declustering scheme from Beirlant *et al*. (2004) that is justified by an asymptotic result for the times between threshold exceedances, 76 clusters can be identified. We now assume that these 76 cluster maxima are approximately independent. From there on, one can apply the previous procedures to test for a change in parameter, although it will be on a much smaller set of data.

(1) *Pareto-type distribution*
We test for a change in γ using the likelihood ratio $\sqrt{-2\log\Lambda_m}$ in Equation (10) for the declustered data as given in Figure 15. In this case, a local maximum is attained for cluster maximum 48 which corresponds to $m = 631(28/07/1997)$. However, this local maximum is not larger than the critical value. The actual maximum Z_n is attained for cluster maximum 66 which corresponds to $m = 854$ (22/6/98).
(2) *GPD*
We repeat the procedure for the case of the GPD where we test for a change in $\theta = (\gamma, \sigma)$ as in Section 2.3, but based on the declustered data. Again using the the logarithm of the absolute excesses, we calculate $-2\log\Lambda_m$, $1 \leqslant m \leqslant n-1$ as in Equation (12). The square root of this expression is plotted in Figure 16. The maximum Z_n is attained for cluster maximum 48 which corresponds to $m = 631(28/07/1997)$.

DOI: 10.1002/env

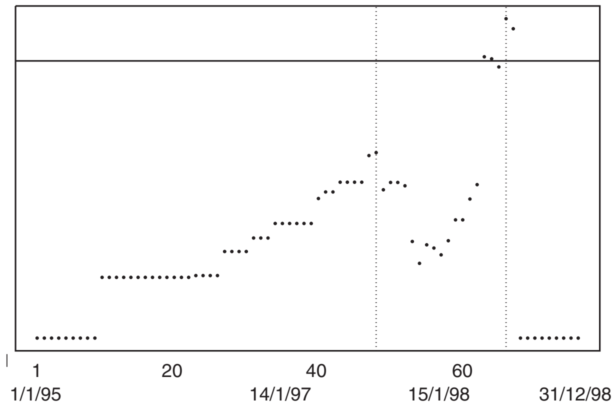

Figure 15. $(m, \sqrt{-2 \log \Lambda_m})$ for the declustered data of the Malaysian stock index, where m is indicating the group which is split into two groups. The likelihood ratio is calculated for the "exponential" log spacings of the cluster maxima. The critical value for the test is indicated with a horizontal line

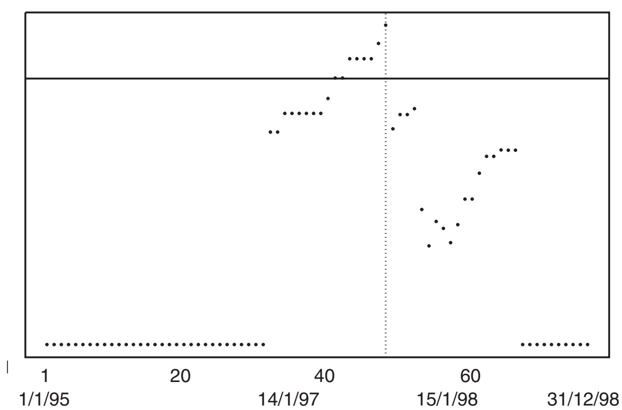

Figure 16. $(m, \sqrt{-2 \log \Lambda_m})$ for the declustered data of the Malaysian stock index, where m is indicating the group which is split into two groups. The likelihood expression is calculated as in Equation (12) for the cluster maxima. The critical value 2.95 for the test is indicated with a horizontal line

6.3. Climatological data

In our next example, the data consist of daily total precipitation and daily minimum temperatures of the climatological station in Quebec City. The daily observations from 1915 to 2007, covering 34596 data points are illustrated in Figure 17. During this period the instruments might have been relocated several times and each relocation might have created changes in the time series.

For the detection of the change points, we take into account the observations of four neighboring stations. For every day, the measurement in Quebec is compared to the measurements in four neighboring stations. More specifically, we investigate the average of absolute deviations of four neighboring stations with the observations in Quebec itself. A large deviation means that the measurement in Quebec differs significantly from the measurements in the neighboring stations.

DOI: 10.1002/env

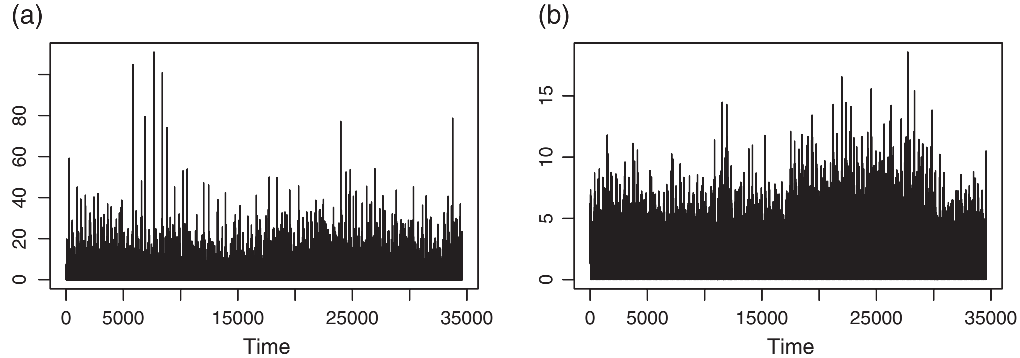

Figure 17. Time series plot, from January 1915 to December 2007, of the absolute deviation from measurements in Quebec with four neighboring stations for (a) daily total precipitation and (b) daily minimum temperature

6.3.1. Precipitation. We will base our analysis on the $k = 907$ largest deviations, since the AMSE of the Hill estimator is minimal for this k-value.

(1) *Pareto-type distribution*
When applying the change point detection method based on the exponential model, the test statistic is given in Figure 18(a). The test statistic attains a maximum near the end of the time series. Also at the beginning of the time series a local maximum is attained. If we discard these maxima at the borders, another local maximum can be observed for m=10595 (corresponding to the year 1943). However, the change is not significant since the critical value of 2.02 is not reached.

(2) *GPD*
The results for the change point detection method based on the GPD can be found in Figure 18(b). The global maximum now occurs at m=14 685 (year 1954), but at m=10 595 a local maximum occurs which was found with the exponential model but now the test statistic overshoots the critical value.

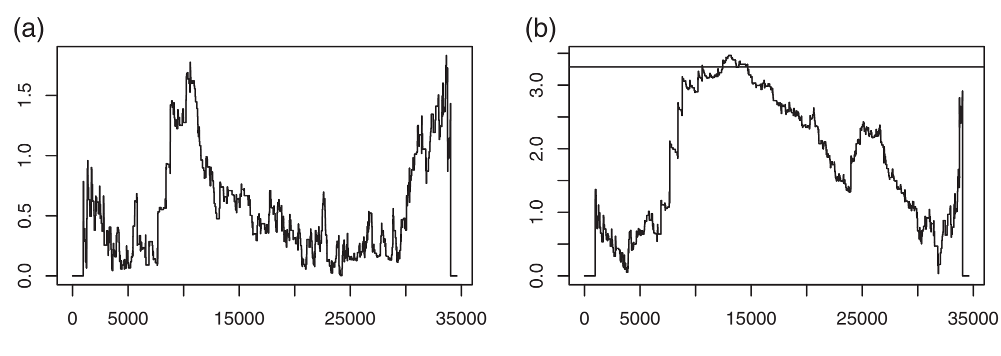

Figure 18. (a) $(m, \sqrt{-2 \log \Lambda_m})$ for the precipitation deviations, where m indicates the group which is split into two groups. The likelihood expression is calculated as in (a) exponential model and (b) GPD model. The critical value for the test is added with a horizontal line

DOI: 10.1002/env

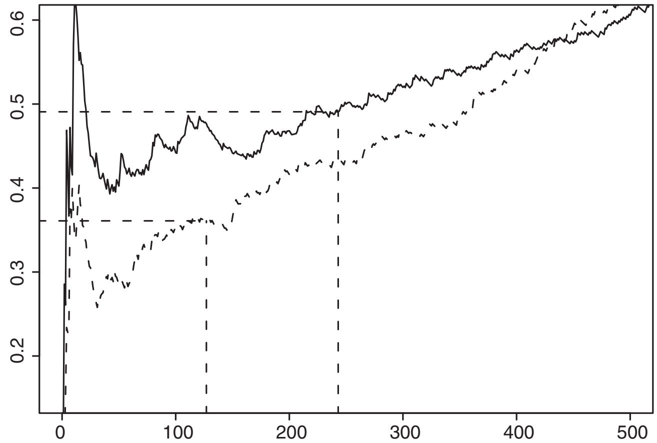

Figure 19. Hill estimator as function of k for two groups of the precipitation deviations (a) from 1933 until 1943 (full line) and from 1943 until 1953 (dashed line). Based on the minimal AMSE of the Hill estimator the optimal values are indicated as well

We have a closer look at the change in 1943, that emerged from both methods. We consider the data from 1933 until 1953 and split these data at the change point in 1943. The corresponding Hill estimators of both groups are shown in Figure 19. They seem quite different for both groups. Based on the AMSE of the Hill estimator, the optimal estimators for γ are 0.49 and 0.36, respectively. The deviations of Quebec compared to the neighboring stations could be more extreme before 1944, because of the larger extreme value index in that period.

6.3.2. Temperatures. A similar treatment can be given to the temperatures in Quebec. Here the analysis will be based on the $k = 893$ largest deviations.

(1) *Pareto-type distribution*
When applying the change point detection method based on the exponential model, the test statistic is given in Figure 20(a). The test statistic shows a significant change as the test statistic overshoots

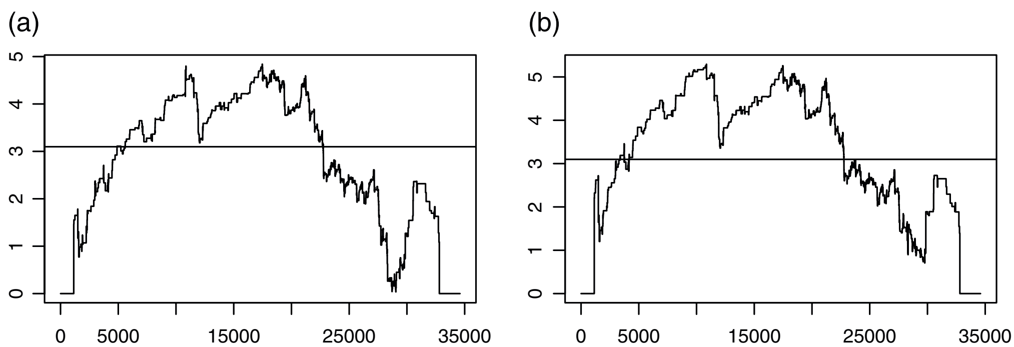

Figure 20. (a) $(m, \sqrt{-2 \log \Lambda_m})$ for the temperature deviations, where m indicates the group which is split into two groups. The likelihood expression is calculated as in (a) exponential model and (b) GPD model. The critical value for the test is added with a horizontal line

DOI: 10.1002/env

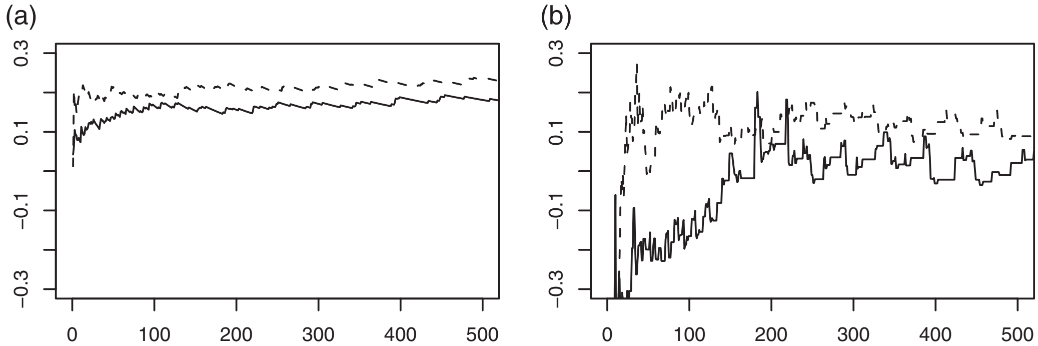

Figure 21. Estimators of γ for the temperature deviations as function of k for two groups from 1933 until 1943 (full line) and from 1943 until 1953 (dashed line): (a) Hill estimator and (b) POT estimator

the critical value for the change point test. The test statistic attains a maximum at m=17 474. A second local maximum (which is nearly as high) is attained at m=10 861. Both time points are good candidates for change points and correspond to the years 1961 and 1944, respectively.

(2) *GPD*

When applying the change point detection method based on the GPD, the same conclusions can be drawn, as is illustrated in Figure 20(b). Only now m=10 861 corresponds to the global maximum, while m=17 470 leads to a slightly smaller local maximum.

As happened for precipitation, there might be a change point for temperature at the year 1943. We again investigate the change in 1943 by considering the data from 1933 until 1953 and splitting then at the change point 1943. The corresponding Hill estimators of both groups are shown in Figure 21(a) and seem quite different for both groups. However, note that for the data after 1943, the extreme value index gradually keeps getting smaller approaching 0 as k decreases. This might indicate that the data are not Pareto-type distributed, but follow a distribution with zero, or negative extreme value index. Indeed when calculating the extreme value indices using the POT method, we find the results in Figure 21(b). Now the extreme value index of the deviations is larger after 1944. From 1944 onward, the extreme value index is close to 0.1 using the largest data points, whereas γ is fluctuating around -0.2 before 1944.

In all the above analysis of precipitation and temperature, there is a distinctive time point that consistently shows up as a possible change point and that is near the year 1943. We were informed that there has been a relocation of the instruments in 1943. Before 1943, the observations were taken in the city and they were representative of an urban environment. The observation site closed in 1943 and the instruments were relocated to the airport a few kilometers away from the original site and within a less urban environment. This relocation implied a change in elevation from 90 to 74 m and a change in observers and observing agencies.

7. CONCLUSIONS

The likelihood approach of Csörgő and Horváth (1997) can be adapted to the case where the parameters in the change point analysis refer to the tail behavior of the underlying distribution. In this way it can

DOI: 10.1002/env

be tested whether or not a change of the extreme value index γ occurs. Also changes of the parameters in the GPD distribution can be examined. The method was tested in simulation settings and works quite well. Also in real life examples, change points could be detected.

One observation that needs further investigation is that the choice between the exponential model and the GPD leads to different conclusions. We observed this phenomenon in the example of the Malaysian stock exchange data and in the Quebec figures. Since the exponential model only uses a single parameter (the extreme value index), the fit to the real data might be too weak to draw reliable conclusions. In both examples the reader can visually check that there is quite some volatility in the data. Hence, fitting a one-parameter model might not be appropriate. Introducing the shape parameter σ in the GPD model gives us a chance to also keep track of at least part of the volatility. It is then a fortunate circumstance that the test leads to decisive conclusions. It needs further investigation to find out when one needs a more parameterized model.

But a number of other important questions remain. A further step in the analysis would be to construct confidence intervals around the instant of change, given that it happened. Also, our analysis is essentially based on the underlying assumption of independence, if not given, then by declustering. What happens under precise forms of dependence needs further research.

ACKNOWLEDGEMENTS

Part of this work was done while the second author was on leave at the University of Stellenbosch (South Africa). He thanks the Department of Statistics and Actuarial Science for the hospitality. The authors thank Jan Beirlant for some useful suggestions. The authors are also grateful to Jeffrey Dambacher for providing interesting data. The climatology data from Quebec have been extracted from the National Climate Data Archive of Environment, Canada and were kindly provided by Lucie Vincent from the Climate Research Division. We express our special thanks to her. Finally, the authors thank the referee for his constructive comments.

REFERENCES

Gut A, Steinebach J. 2005. A two-step sequential procedure for detecting an epidemic change. *Extremes* **8**: 311–326.

Crisci A, Gozinni B, Meneguzzo F, Pagliara S, Maracchi G. 2002. Extreme reinfall in a changing climate: regional analysis and hydrological implications in Tuscany.*Hydological Processes* **16**: 1261–1274.

Katz RW, Brown BG. 1992. Extreme events in a changing climate: variability is more important than averages. *Climatic Change* **21**: 289–302.

Mason DM, Turova TS. 1994. Weak convergence of the hill estimator process. In *Extreme value theory and Applications*. Galambos J, Lechner J, Simiu E (eds). KLuwer: Dordrecht; 419–431.

Ferro C, Hannachi A, Stephenson D. 2005. Simple nonparametric techniques for exploring changing probability distributions of weather. *Journal of Climate* **18**: 4344–4354.

Beirlant J, Dierckx G, Goegebeur Y, Matthys G. 1999. Tail index estimation and an exponential regression model. *Extremes* **2**: 177–200.

Beirlant J, Goegebeur Y, Segers J, Teugels JL. 2004. Statistics of Extremes. *Theory and Applications*. Wiley: Chichester; 490.

Dias A, Embrechts P. 2004. Change-point analysis for dependence structures in finance and insurance. In *Risk Measures for the 21st Century*, Giorgio Szegoe (ed.). Wiley Finance Series; 321–335.

Jarušková D, Rencová M. 2008. Analysis of annual maximal and minimal temperatures for some European cities by change point methods. *Environmetrics* **19**: 221–233.

Csörgő M, Horváth L. 1997. *Limit Theorems in Change Point Analysis*. Wiley: Chichester.

Burr IW. 1942. Cumulative frequency functions. *Annals of Mathematical Statistics* **13**: 215–232.

Davison AC, Smith RL. 1990. Models of exceedances over high thresholds. *Journal of the Royal Statistical Society B* **52**: 393–442.

Kaufmann E, Reiss RD. 1998. Approximation of the hill estimator process. *Statistics and Probability Letters* **39**: 347–354.

Hill BM. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**: 1163–1174.

DOI: 10.1002/env

APPENDIX

We rewrite expression (9) in terms of the Hill estimator, under the assumptions of Theorem 3.1.

The sample contains two subsamples. Group 1 consists of $X_1, \dots X_m$, which can be ordered as $X_{1,m} \leqslant \dots \leqslant X_{m,m}$; whereas $X_{m+1}, \dots X_n$ form a second group, with order statistics $X_{1,n-m} \leqslant \dots \leqslant X_{n-m,n-m}$. Remember that E was defined as the logarithm of relative excesses over a threshold $u = X_{n-k,n}$, such that

$$\frac{1}{k_1}\sum_{i=1}^{k_1} E_i = \frac{1}{k_1}\sum_{i=1}^{k_1} \log X_{m-i+1,m} - \log X_{n-k,n}$$

$$\frac{1}{k-k_1}\sum_{i=k_1+1}^{k} E_i = \frac{1}{k-k_1}\sum_{i=1}^{k_1} \log X_{n-m-i+1,n-m} - \log X_{n-k,n}$$

$$\frac{1}{n}\sum_{i=1}^{n} E_i = \frac{1}{k}\sum_{i=1}^{k_1} \log X_{n-i+1,n} - \log X_{n-k,n} \tag{16}$$

We suppose u belongs to group 1 (an analogous argument can be made in case u belongs to group 2), yielding $u = X_{n-k,n} = X_{m-k_1,m}$. It follows immediately that $\frac{1}{k_1}\sum_{i=1}^{k_1} E_i$ is exactly equal to $H_{k_1,m}$. In the same way $\frac{1}{n}\sum_{i=1}^{n} E_i$ equals $H_{k,n}$. However, $\frac{1}{k-k_1}\sum_{i=k_1+1}^{k} E_i$ is only asymptotically equal to $H_{k-k_1,n-m}$. Indeed, $X_{n-m-(k-k_1)+1,n-m} < u = X_{n-k,n} < X_{n-m-(k-k_1)+1,n-m}$, such that $\frac{1}{k-k_1}\sum_{i=k_1+1}^{k} E_i$

$$= \frac{1}{k-k_1}\sum_{i=k_1+1}^{k} \log \frac{X_{n-m-i+1,n-m}}{X_{n-m-(k-k_1),n-m}} + \log \frac{X_{n-m-(k-k_1),n-m}}{X_{n-k,n}}$$

$$= H_{k-k_1,n-m} + \log \frac{X_{n-m-(k-k_1),n-m}}{X_{n-k,n}}$$

Denote $U_m = |\log X_{n-m-(k-k_1),n-m} - \log X_{n-k,n}|$. Note that $(k-k_1)U_m$ is bounded by $(k-k_1)|\log X_{n-m-(k-k_1),n-m} - \log X_{n-m-(k-k_1)+1,n-m}|$ which is known to be asymptotically exponentially distributed with mean γ. Therefore, $U_m = O_P(1/(k-k_1)) = o_P(1)$.

Thus Equation (9) becomes,

$$2\big[-k_1 \log H_{k_1,m} - (k-k_1)\log(H_{k-k_1,n-m} + U_m) + k\log H_{k,n}\big]. \tag{17}$$

It is easy to calculate that $kH_{k,n} - k_1 H_{k_1,m} - (k-k_1)H_{k-k_1,n-m}$ equals $(k-k_1)U_m$. Therefore, Equation (9) asymptotically results in

$$2\left[-k_1 \log H_{k_1,m} - (k-k_1)\log H_{k-k_1,n-m} + k\log H_{k,n} - (k-k_1)\log\left(1 + \frac{U_m}{H_{k-k_1,n-m}}\right)\right]$$
$$= 2\left[-k_1 \log H_{k_1,m} - (k-k_1)\log H_{k-k_1,n-m} + k\log H_{k,n} - (k-k_1)\frac{U_m}{H_{k-k_1,n-m}}\right] + O_P\left(\frac{1}{k-k_1}\right)$$

$$= 2\Big[-k_1 \log H_{k_1,m} - (k-k_1) \log H_{k-k_1,n-m} + k \log H_{k,n}$$
$$- \frac{kH_{k,n} - k_1 H_{k_1,m} - (k-k_1) H_{k-k_1,n-m}}{H_{k-k_1,n-m}}\Big] + O_P\left(\frac{1}{k-k_1}\right)$$
$$= 2\Big[-k_1 \log H_{k_1,m} - (k-k_1) \log H_{k-k_1,n-m} + k \log H_{k,n}$$
$$- \frac{kH_{k,n} - k_1 H_{k_1,m} - (k-k_1) H_{k-k_1,n-m}}{H_{k,n}}\Big] + o_P(1)$$

The last step follows from the fact that both $H_{k-k_1,n-m}$ and $H_{k,n}$ are asymptotically consistent estimators of γ.

DOI: 10.1002/env