Selecting an optimum threshold with the Kullback-Leibler deviance measure

A. Verster, D.J. De Waal and S. van der Merwe

Department of Mathematical Statistics and Actuarial Science, University of the Free State

Abstract

It is common practice in Extreme Value Theory to model observations, above a sufficiently large threshold, with a Peaks-Over-Threshold distribution, such as the Generalized Pareto distribution. The question remains: where should one choose the optimum threshold? Various literature discusses the problem of threshold selection, and the best method is still to be found. In this paper we consider choosing an optimum threshold using the Kullback-Leibler deviance measure. The deviance between a Peaks-Over-Threshold distribution and its posterior predictive distribution is obtained at various threshold levels. We consider the optimum threshold as that threshold value where the Kullback-Leibler deviance measure is the closest to zero. Thus, the optimum threshold is chosen where the POT distribution, fitted to the observations above the threshold, and its posterior predictive distribution is the closest to one another.

Keywords: Threshold, Kullback-Liebler, Extreme Value Theory, Peaks-Over-Threshold, Generalized Pareto distribution.

Introduction

In Extreme Value Theory (EVT) the exceedences above a high threshold is modelled through a Peaks-Over-Threshold (POT) distribution such as the Generalized Pareto distribution (GPD). The choice of optimum threshold remains debatable and much research has been on how to choose the best possible threshold for a given dataset. A threshold is often chosen on a visual basis such as by means of the Pareto quantile plot (as discussed later in Sections 1 and 2) and the Mean Residual Life plot. See for example, Beirlant *et al*. (2004), Coles (2001) and Smith (1985). Other literature on threshold selection include work done by Dupluis (1998), Guillou and Hall (2001), Thompson *et al.* (2009), Trancredi *et al.* (2006) and Verster and De Waal (2011) to name a few.

In the present paper we introduce a new method for selecting the optimum threshold by considering the Kullback-Leibler (KL) deviance measure. The KL deviance measure (Kullback and Leibler, 1951) is a well-known measure from considering the deviance between models. In this paper we consider fitting a POT distribution to the observations above a selection of thresholds and the KL is used to select the best threshold. A Bayesian approach is considered and the KL deviance measure between the POT distribution and its posterior

predictive distribution is calculated for different thresholds. The threshold that results in the smallest (closest to zero) KL value is considered as the optimum threshold.

Assume that the correct threshold is given by t and that the observations above t follow a known POT distribution with unknown parameter values. The KL deviance measure between the POT distribution and its posterior predictive distribution should be very close (in fact equal) to zero. If the threshold is incorrectly chosen at some value less than t it means that the POT distribution will be fitted to some observations that should not follow a POT distribution, thus the estimation of the parameters of the POT distribution will be incorrect and the KL between the POT distribution and its posterior predictive distribution will deviate more resulting in some value other than zero. If the threshold is incorrectly chosen at some value larger than t, the number of observations on which the POT distribution is fitted becomes smaller which leads to vague posteriors of the parameters and unstable estimates of the parameters. The KL between the POT distribution and its posterior predictive distribution will again deviate more from zero. Thus, we are searching for a threshold value with a KL close to zero.

In Section 1 the Bounded Pareto is chosen as the POT distribution. The KL deviance measure between the bounded Pareto distribution and its posterior predictive distribution can be derived explicitly as shown in Section 1. In Section 2 we consider the more general case where the GPD is chosen as the POT distribution. Again the KL deviance measure between the GPD and its posterior predictive distribution is calculated. No explicit formula for the KL measure is available in the GPD case and it is therefore investigated through various simulation examples of well-known Pareto type distributions such as the t and the F distributions (where the extreme value index is known).

In Section 3 a truncated Normal dataset where the tail is replaced with GPD values is investigated. The KL deviance measure method is applied to see whether the method correctly identifies the threshold. In Section 4 we continue to investigate the appropriateness of choosing the optimum threshold with the KL deviance measure in an extended simulation study where the threshold is known in advance (similar to the article by Thompson *et al.*, 2009). We compare results with those of Thompson *et al.* (2009) and show that our method gives more reliable results than those obtained in previous papers.

1 KL measure in the bounded Pareto case

The density function of the bounded Pareto is given as follows:

$$f(y) = \frac{1}{\gamma} y^{-\frac{1}{\gamma}-1} \tag{1}$$

where $Y = \frac{X}{t}$ $(X > \tau)$ describes the relative excesses over the threshold (τ). The posterior predictive density of a future $\tilde{Y} = \log(Y)$ given a dataset $\boldsymbol{x} = (x_{1,n}, x_{2,n}, \dots, x_{n,n})$ ordered under a Jeffreys prior, $\pi(\gamma) \propto \frac{1}{\gamma}$, is given as

$$g(\tilde{y}|\boldsymbol{x}) = \frac{1}{H_{k,n}}\left(1 + \frac{\tilde{y}}{kH_{k,n}}\right)^{-k-1} \quad (2)$$

(Beirlant *et al.*, 2004). Equation 2 is the density function of the GPD with shape parameter $\frac{1}{k}$ and scale parameter $H_{(k,n)} = 1/k \sum_{j=1}^{k} \log\left(x_{n-j+1,n}\right) - \log(x_{n-j+1,n})$. The scale parameter is called the Hill estimate of γ. The number of exceedences above the threshold τ is given by k (Beirlant *et al.*, 2004).

At the optimum threshold level one would expect that the posterior predictive density of the bounded Pareto and the true bounded Pareto density, fitted above τ, to be close to each other. Thus, the KL measure of deviance between the two densities should be close to zero. The KL deviance measure is defined as follows:

$$KL = E_Y \log \frac{f(Y)}{\mathrm{g}(Y|\boldsymbol{x})} \quad (3)$$

where $f(Y)$ is the density function and $g(Y|\boldsymbol{x})$ is the posterior predictive density given the data. It is known that the posterior predictive density of the log of the variable is GPD. One would expect that the KL measure between the density of the log of the variable and the posterior predictive density of the log of the variable should also be close to zero. Thus, we need to derive the KL measure of deviance between $f(\tilde{y})$ and $g(\tilde{y}|\boldsymbol{x})$, defined as

$$KL = E_{\tilde{Y}} \log \frac{f(\tilde{Y})}{\mathrm{g}(\tilde{Y}|\boldsymbol{x})} \quad (4)$$

where $E_{\tilde{Y}} \log\{f(\tilde{y})\} = -\log(\gamma) - (1+\gamma)E\{\log(1 - F(\tilde{Y})\} = -\log(\gamma) - 1$, since $\tilde{Y} \sim Exp\left(\frac{1}{\gamma}\right)$ and $E_{\tilde{Y}} \log\left\{g\left(\tilde{Y}|\boldsymbol{x}\right)\right\} = -\log H_{(k,n)} - 1 - \frac{1}{k}$. Therefore

$$KL = \log\left(\frac{H_{(k,n)}}{\gamma}\right) + \frac{1}{k}. \quad (5)$$

For different values of k (the number of exceedences above the threshold) the KL measure closest to zero can be obtained. This is done numerically, as shown in the next example.

1.1 Example 1

It is known that the absolute t distribution is a heavy tailed distribution that belongs to the Fréchet-Pareto class of Extreme Value distributions with an extreme value index (EVI) of $\frac{1}{\nu}$, where ν is the degrees of freedom (Beirlant *et al.*, 2004). We simulate $n = 500$ $|t_{\nu=4}|$ values and we assume that the observations above the threshold, τ, follow a bounded Pareto distribution. Our aim is to estimate the optimum threshold by calculating the KL-

measure (according to Equation 5) between the bounded Pareto densities and their predictive densities for different γ ($\gamma = 0.25: 0.1: 0.33$) and different threshold values ($k = 2: 100$). Each simulation is repeated 1000 times and the mean of the KL values for the different $\gamma's$ against the different thresholds are shown in Figure 1. The KL measure that is the closest to zero (KL = 0) is obtained for $\gamma = 0.32$ and $k = 24$.

A summary of the results are given in Table 1. Column 2 gives the KL values that are the closest to zero for the given values of γ and Column 3 gives the number of observations above the threshold that corresponds to the KL measure in Column 2. Our conclusion is therefore that the optimum threshold should be chosen at the 24th largest observation. This result can now be compared with the threshold that would have been chosen if a visual approach was taken such as the method of the Pareto quantile plot.

The Pareto quantile plot is defined as the scatter plot of the following points: $\left(-\log\frac{j}{n+1}, \log X_{n-j+1}\right)$, $j = 1,2, \dots, k$. The Pareto quantile plot for a simulation of 500 $|t_{\nu=4}|$ values are shown in Figure 2. An optimum threshold is considered as the observation on the y-axis where the plot starts to follow a linear pattern. We can estimate the threshold by looking at the leftmost point of the linear pattern. A straight line is drawn through the top 24 observations, as shown in Figure 2, and it seems visually appropriate to choose the threshold at the 24th observation.

Since we know that the true EVI (γ) is 0.25 ($1/\nu$) it seems as if the fit of the bounded Pareto overestimates the extreme value index in the case of the $|t_{\nu=4}|$ data. The GPD, which is a more general POT distribution, might be a better fit to consider. This is discussed in the next section.

Table 1 **Summary of simulation results**

γ	KL measure	k
0.2500	0.2024	6
0.2600	0.1632	6
0.2700	0.1254	6
0.2800	0.0891	6
0.2900	0.0540	6
0.3000	0.0201	6
0.3100	0.0004	16
0.3200	0.0000	24

0.3300	0.0014	32
0.3400	0.0008	39
0.3500	0.0005	48

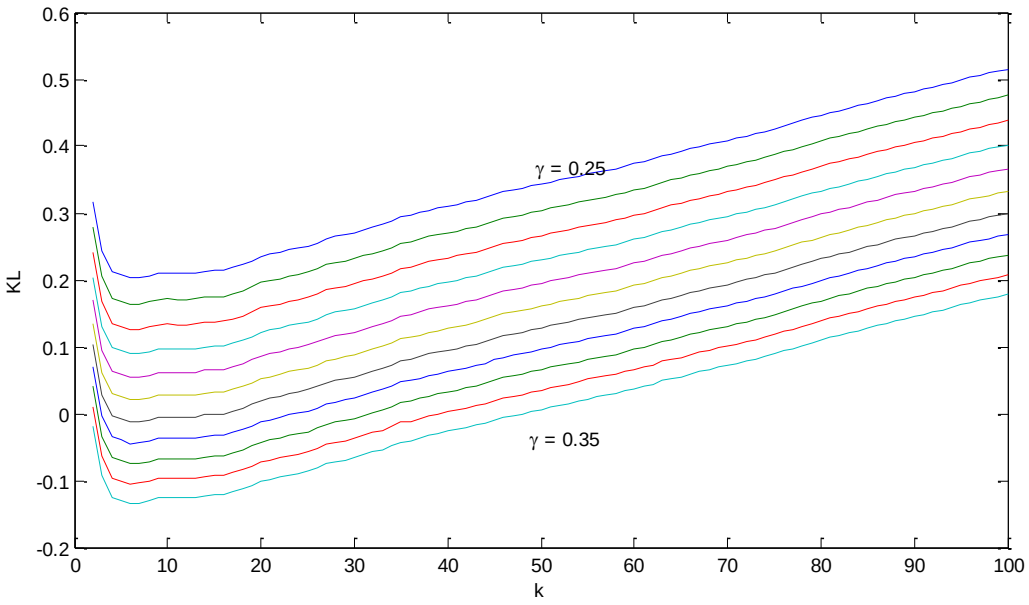

Figure 1 **Mean KL values for different values of γ and k.**

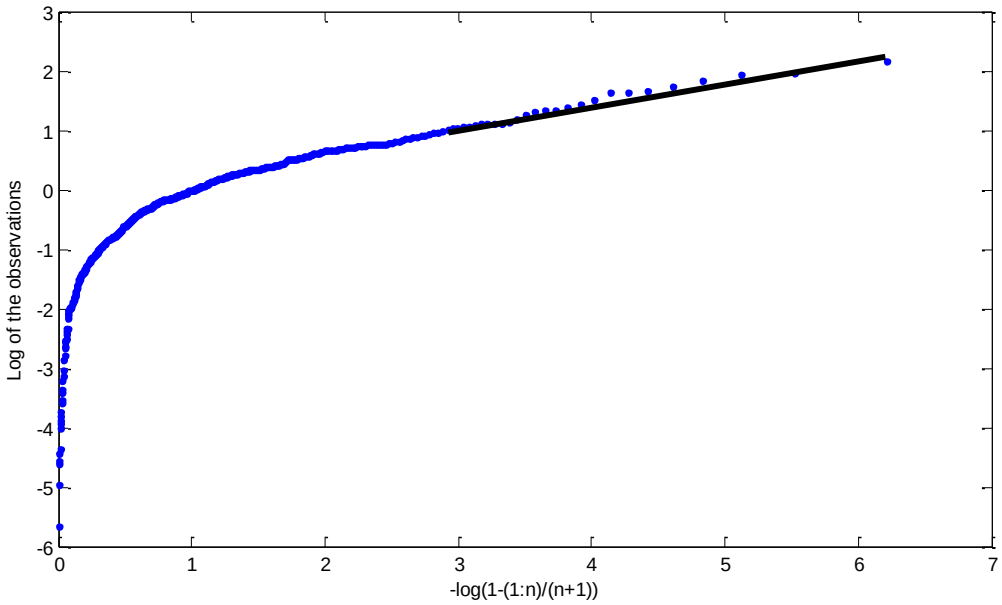

Figure 2 **The Pareto quantile plot to choose the optimum threshold.**

2 KL measure in the GPD case

In this section we assume that the observations above a threshold are modelled through a GPD. A GPD is one of the most popular and well-known POT distributions in Extreme Value Theory. See for example Coles (2001). The distribution function of the GPD is given as follows:

$$F(y) = 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{-1/\gamma} \tag{6}$$

where $Y = X - \tau$ $(X > \tau)$ describes the excesses over the threshold, τ, and $1 + \frac{\gamma y}{\sigma} > 0$. The distribution function (6) can also be written in terms of two parameters as follows:

$$F(y) = 1 - \left(1 + \frac{\gamma y}{\gamma \tau}\right)^{-1/\gamma} \tag{7}$$

where $\sigma = \gamma\tau$ (Ledford and Tawn, 1996). If the threshold is considered as fixed we only have one unknown parameter, γ. This distribution (Equation 7) is more convenient to work with (since it has less parameters) and we consider this distribution throughout the paper.

The posterior predictive density of a future observation Y, given the data $\boldsymbol{x} = (x_{1,n}, x_{2,n}, \ldots, x_{n,n})$, is as follows:

$$g(y|\boldsymbol{x}) = E_{\gamma|\boldsymbol{x}} f(y|\gamma). \tag{8}$$

Equation 8 cannot be solved explicitly, as in the bounded Pareto case, but the posterior predictive density can be simulated by taking the mean of the densities at y_{n+1} for a large number of γ values that are simulated from the posterior distribution of the GPD (Equation 10) and plugged into the density function. Equation 8 thus simplifies to

$$\hat{g}(y|\boldsymbol{x}) = \frac{1}{m}\sum_{j=1}^{m} f(y|\gamma_j) \tag{9}$$

where *m* is a large number of simulated γ values.

The posterior distribution of the GPD is given as follows

$$\pi(\gamma|\boldsymbol{x}) = like(\gamma|\boldsymbol{x})\pi(\gamma) \propto \left(\frac{1}{\gamma t}\right)\prod_{i=1}^{n}\left(1 + \frac{y}{t}\right)^{-\frac{1}{\gamma}-1} . \frac{1}{\gamma t} e^{-\gamma} \tag{10}$$

where $\pi(\gamma) \propto \frac{1}{\gamma t} e^{-\gamma}$ is the maximal data information (MDI) prior of the GPD (Beirlant *et al.* 2004).

The KL measure of deviance between the GPD density and the posterior predictive density of the GPD is given as

$$KL = E_{\hat{\gamma}} \log \frac{f(Y)}{\hat{g}(Y|\boldsymbol{x})} = -\log(\hat{\gamma} t) - \hat{\gamma} - 1 - E[\log(\hat{g}(Y|\boldsymbol{x}))]. \tag{11}$$

2.1 Example 2

A similar dataset as in the bounded Pareto case is simulated from the absolute t distribution with $\nu = 4$ and $n = 500$. The Pareto quantile plot is shown in Figure 3. A straight line is drawn through the last observations on the plot and indicates that a threshold will more or less be chosen at $\exp(0.8) = 2.2255$ with $k = 37$.

We now consider choosing the threshold with the KL measure (Equation 11) for different values of k ($k = 5{:}\,100$) where k indicates the number of observations above τ. For each k value the observations above the threshold are assumed to be GPD distributed and 1 000 γ values are simulated from the GPD posterior distribution (Equation 10). These 1 000 simulated parameter values are plugged into Equation 9 to estimate the posterior predictive density. The KL measure (Equation 11) is then calculated where $\hat{\gamma}$ is taken as the mean of the 1 000 simulated $\gamma's$.

Figure 4 shown the KL measures at different k values for one simulated $|t_{\nu=4}|$ dataset. For this simulation the KL measure closest to zero (KL = -0.0882) is obtained at $k = 36$ (which is in line with the threshold chosen with the Pareto quantile plot). At $k = 36$, the parameter is estimated as: $\hat{\gamma} = 0.2835$. The estimate of γ is very close to the true EVI of 0.25.

Figure 5 shows the histogram of the 1 000 $\gamma's$ at the threshold, $k = 36$. Figure 6 shows the estimated $\hat{\gamma}'s$ (at the different values of k (1:200)) plotted against the different k values. From the figure one can see that an estimate of γ in the area of 0.25 will be obtained more or less at a k value between 5 and 100. For a k value outside this interval γ will be estimated incorrectly as some other value significantly larger than 0.25. Thus, if the threshold is selected incorrectly it negatively influences the estimation of the gamma parameter.

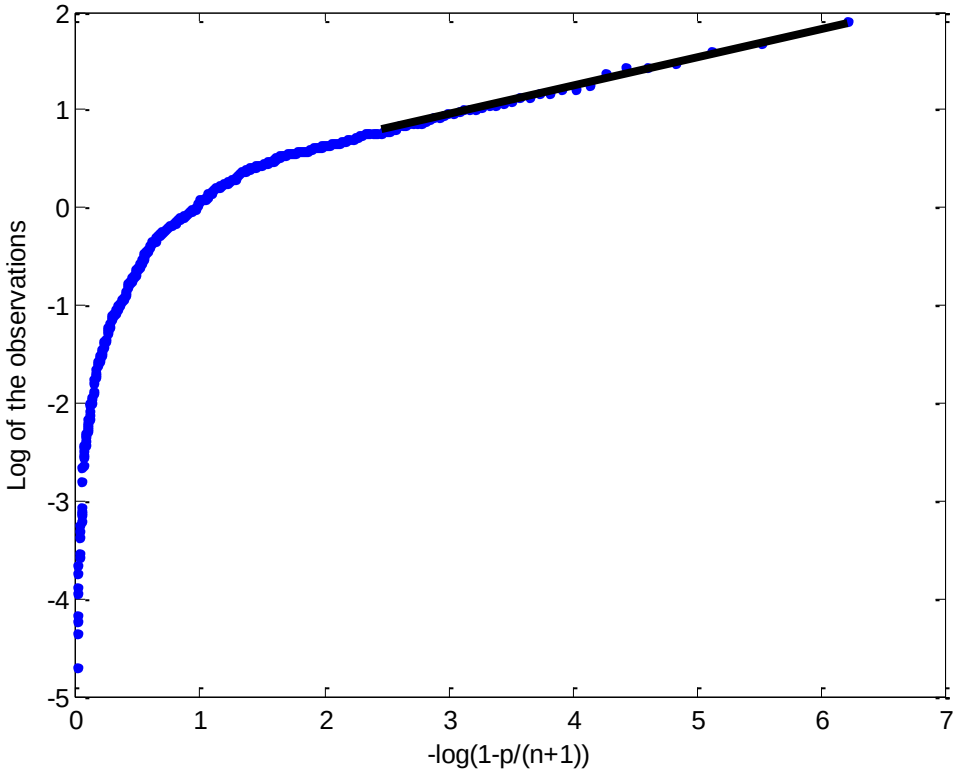

Figure 3 **The Pareto quantile plot to choose the optimum threshold visually.**

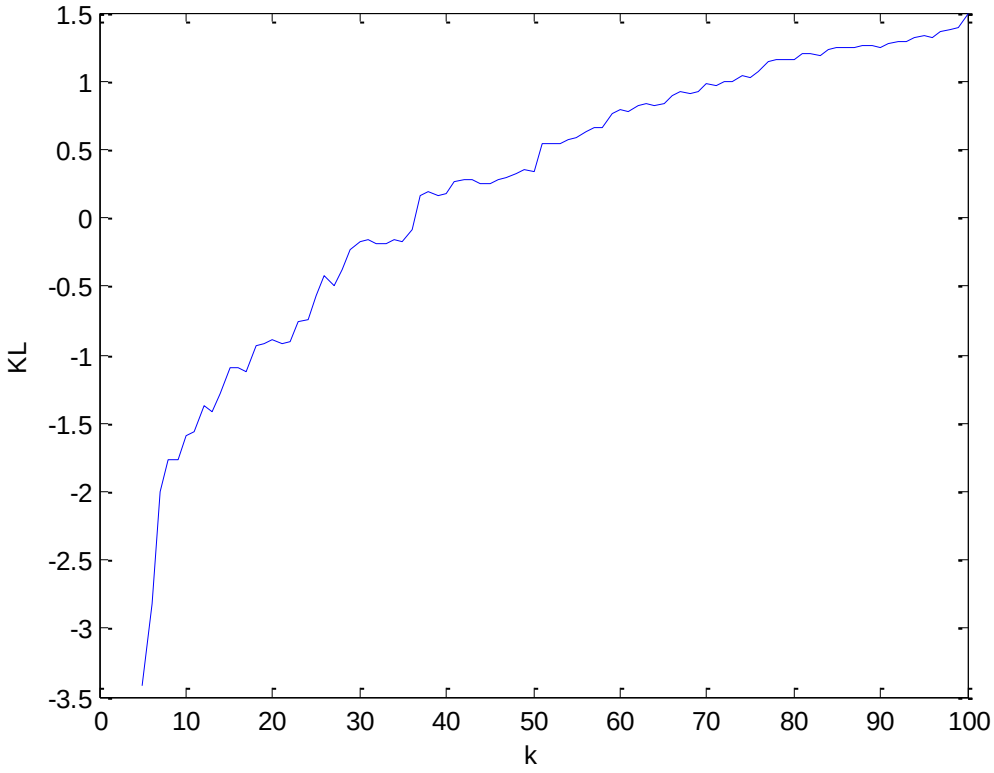

Figure 4 **KL values plotted against different k values.**

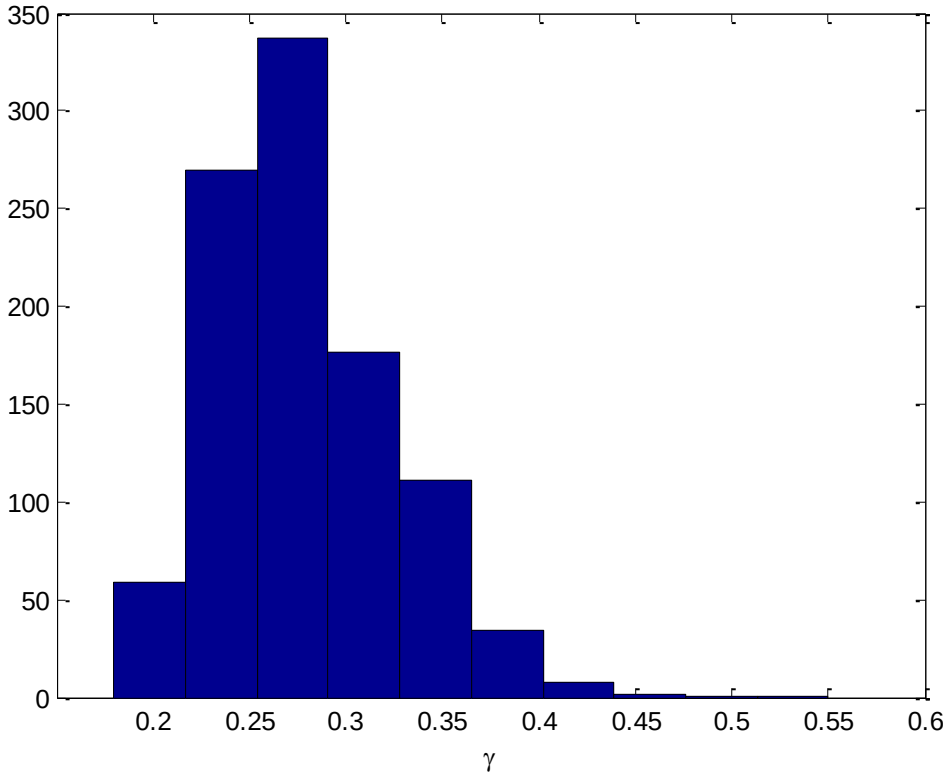

Figure 5 **The 1 000 simulated γ values.**

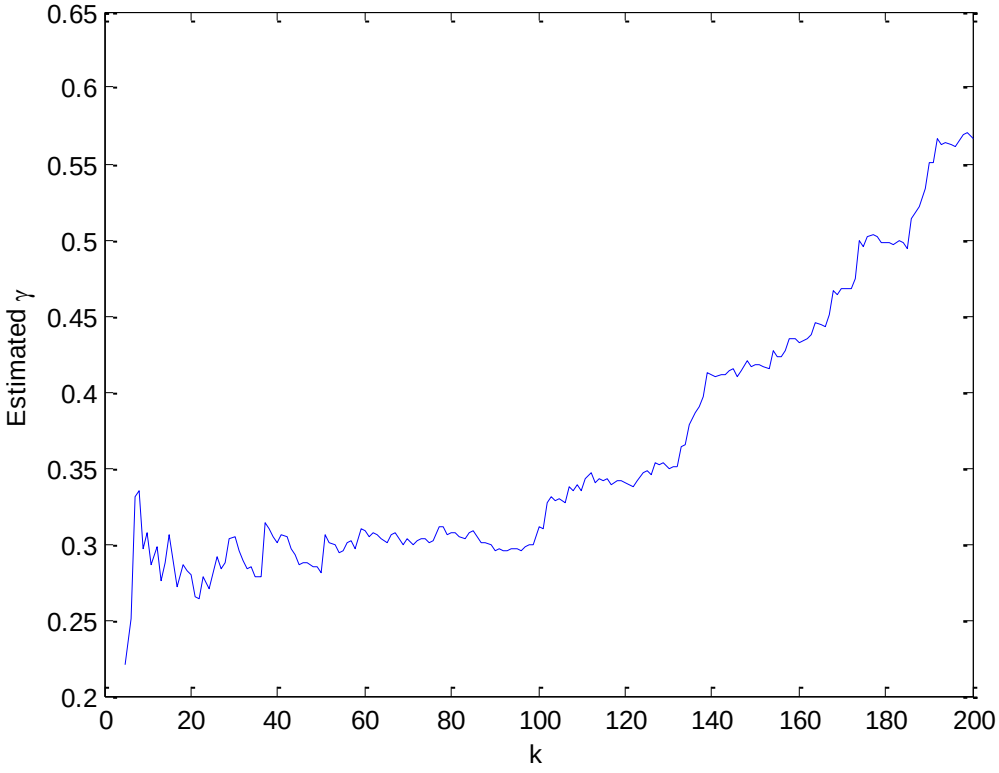

Figure 6 **k values plotted against different $\hat{\gamma}$ values.**

The simulation process is now repeated 1 000 times for the same dataset that was simulated at the beginning of Example 2. Each time the γ estimates are re-simulated and the optimum threshold is obtained through the KL measure. The outcomes are given in Table 2. The first column gives the mean of the 1 000 optimum $k's$ that were selected together with the 95% confidence interval in Column 2. Columns 3 and 4 give the mean estimate of the 1 000 $\hat{\gamma}'s$ together with the 95% confidence interval. The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$ are shown in Figure 7.

Table 2 **Summary of simulation results**

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
37.7340	[36; 39]	0.2869	[0.2793; 0.2908]

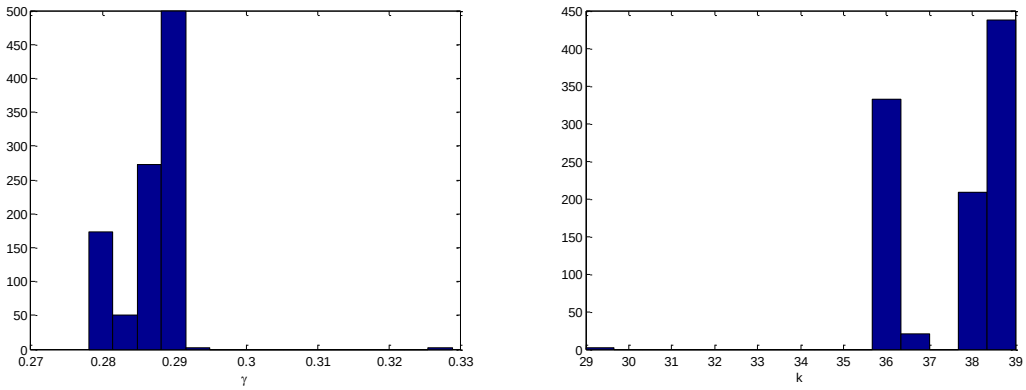

Figure 7 The histograms of the 1 000 simulated $\hat{\gamma}'s$ and $k's$

Next we simulate 1 000 different $|t_{\nu=4}|$ datasets each with $n = 500$ and we estimate the optimum threshold level through the KL measure. The results are shown in Table 3 and the histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$ are shown in Figure 8.

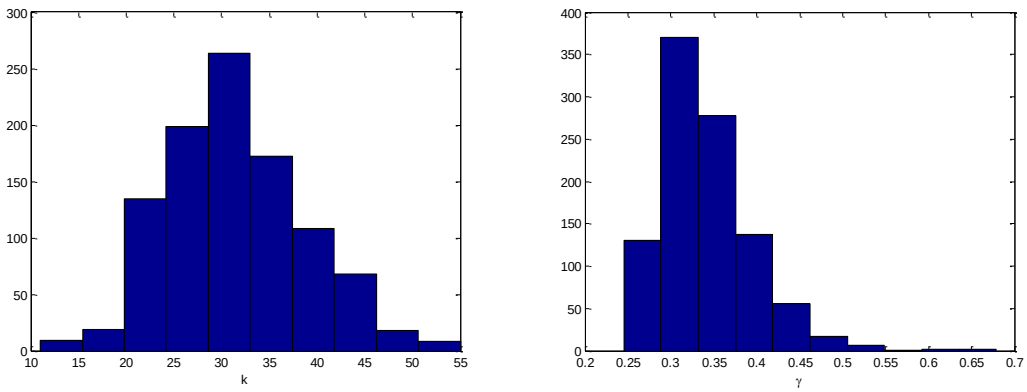

Figure 8 The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$

Table 3 **Summary of simulation results**

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
31.5750	$[19; 47]$	0.3422	$[0.2666; 0.4679]$

The degrees of freedom is now changed to $\nu = 2,$ 1 000 different $|t_{\nu=2}|$ datasets each with $n = 500$ were simulated and we estimated the optimum threshold level through the KL measure. The true EVI is 0.5. The results are shown in Table 4 and the histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$ are shown in Figure 9.

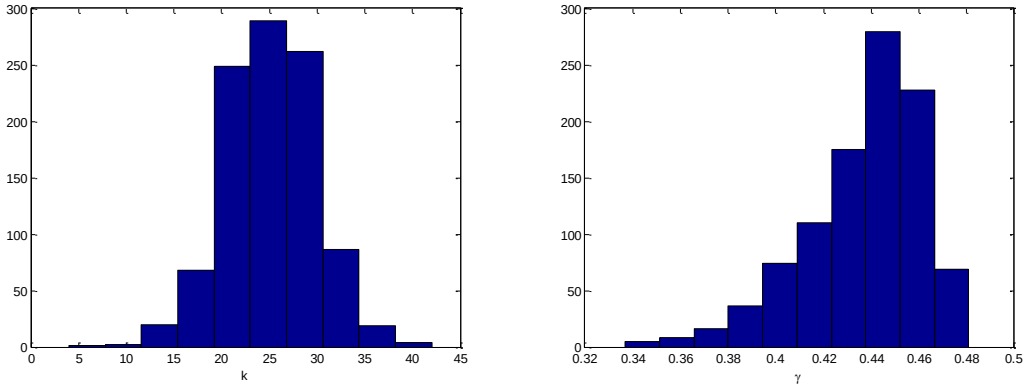

Figure 9 The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$

Table 4 **Summary of simulation results**

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
25.1810	$[16; 34]$	0.4369	$[0.3777; 0.4716]$

From these simulation examples the KL deviance measure between the GPD and the posterior predicted distribution (fitted to observations above the threshold of the absolute t distribution) seems to be working well in choosing the optimum threshold. The $\hat{\gamma}$ estimate at the optimum chosen threshold does not seem to be too far off from the true EVI.

2.2 Example 3

A dataset of size $n = 500$ is now simulated from the F distribution with $\nu_1 = 3$ and $\nu_2 = 4$. The F distribution is a heavy tailed distribution that belongs to the Fréchet-Pareto class of

Extreme Value distributions with an extreme value index (EVI) of $\frac{2}{\nu_2}$ (Beirlant *et al.*, 2004). We again consider choosing the threshold with the KL measure for different values of k ($k = 5{:}\,100$). A similar simulation study is conducted as in the case of the absolute t distribution.

Figure 11 shows the KL measures at different k values for one simulated $F_{3,4}$ dataset. For this simulation the KL measure closest to zero (KL = 0) is obtained at $k = 29$ (which is in line with the threshold chosen with the Pareto quantile plot in Figure 10, where the threshold is chosen roughly at $\exp(1.7)$). At $k = 29$, the parameter is estimated as $\hat{\gamma} = 0.53936$. The estimate of γ is very close to the true EVI of 0.5 which is an indication that our threshold is chosen accurately. Figure 12 shows the histograms of the 1 000 $\gamma's$ simulated at the threshold $k = 29$.

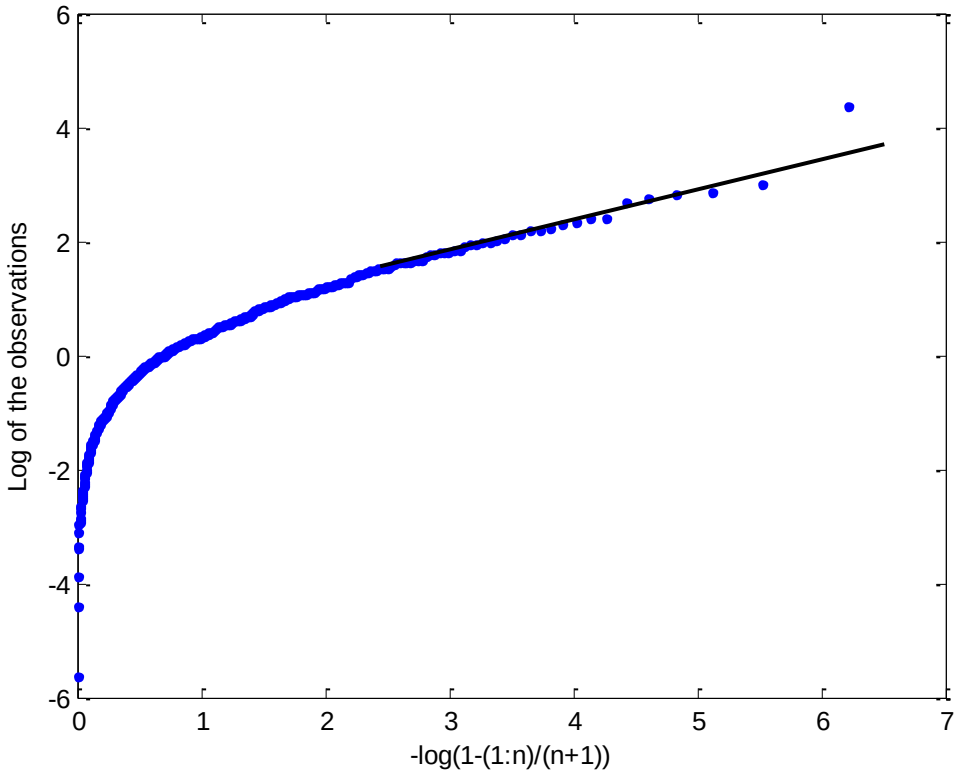

Figure 10 The Pareto quantile plot to choose the optimum threshold visually.

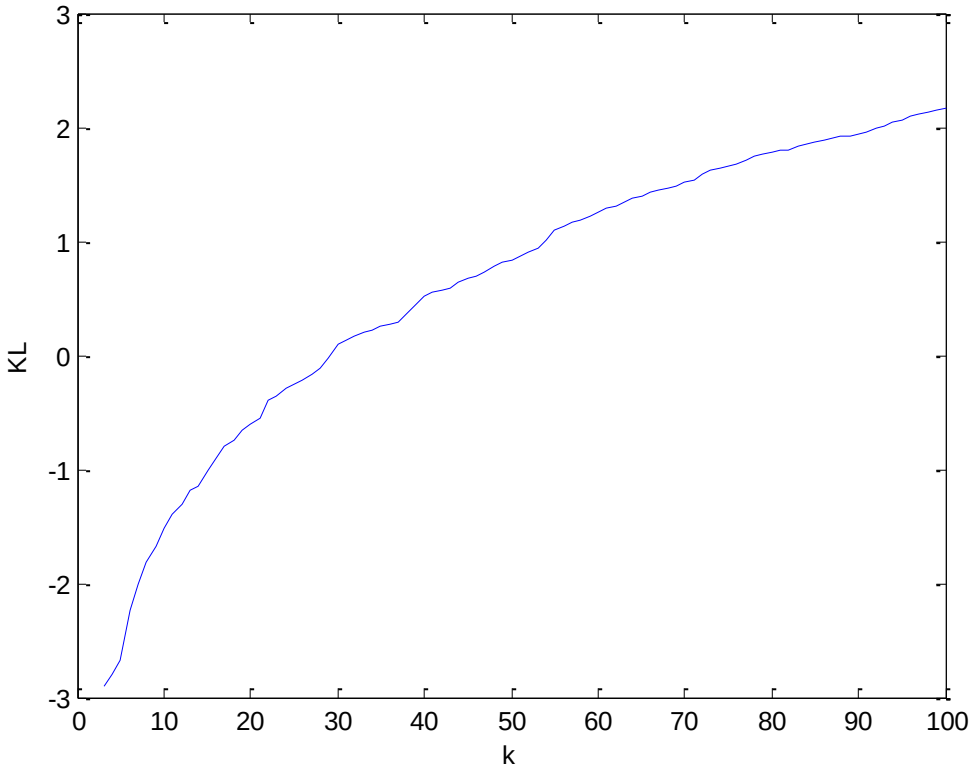

Figure 11 **KL values plotted against different k values.**

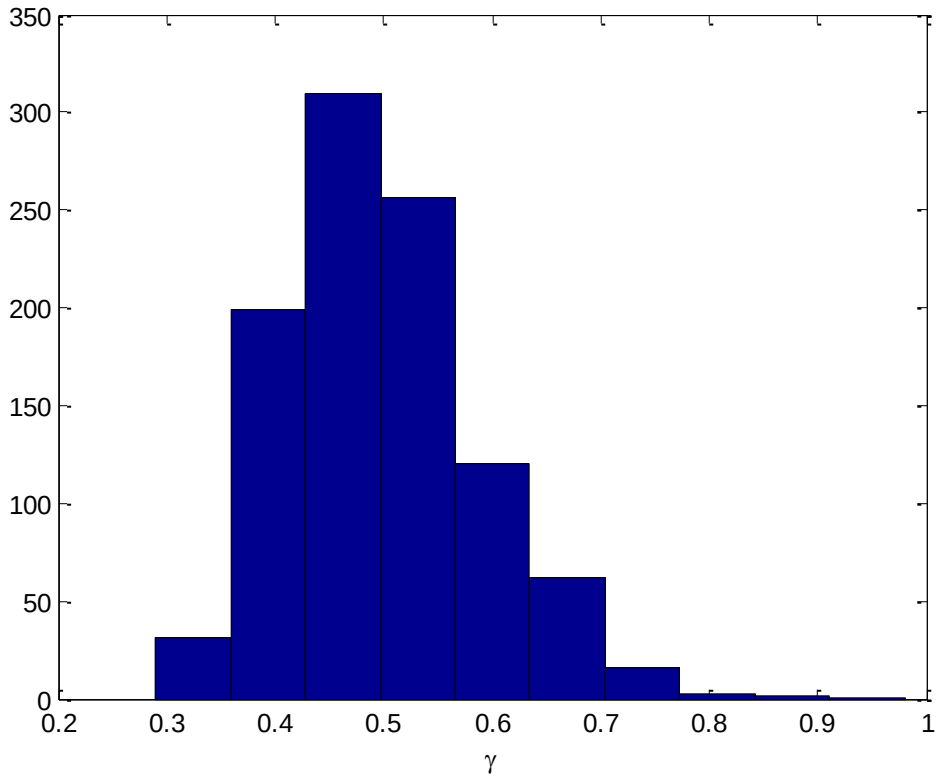

Figure 12 **The 1 000 simulated γ values.**

Next we simulate 1 000 different $F_{3;4}$ datasets each with $n = 500$ and we estimate the optimum threshold level through the KL measure. The results are shown in Table 5 and the histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$ are shown in Figure 13.

Table 5 **Summary of simulation results**

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
29.3060	$[22; 37]$	0.5952	$[0.4461; 0.7721]$

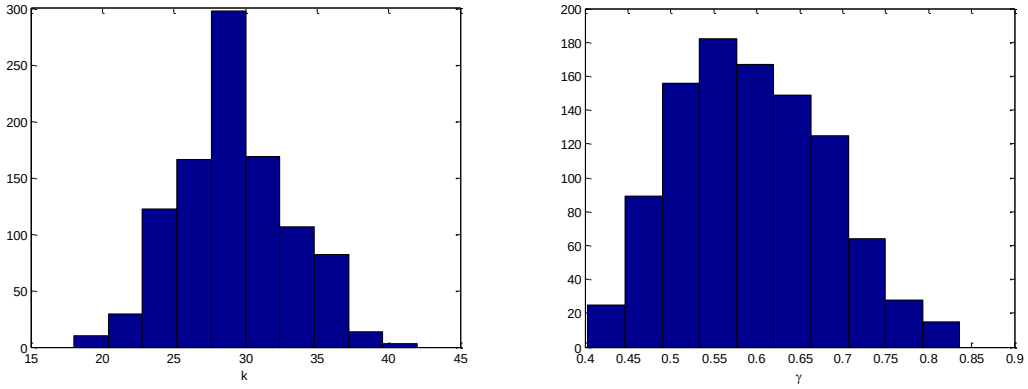

Figure 13 The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$

The degrees of freedom is now changed to $\nu_1 = 3$ and $\nu_2 = 5$, 1 000 different $F_{3;5}$ datasets each with $n = 500$ were simulated and we estimate the optimum threshold level through the KL measure. The true EVI is 0.4. The results are shown in Table 6 and the histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$ are shown in Figure 14.

Table 5 **Summary of simulation results**

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
28.9670	$[21; 38]$	0.5240	$[0.3861; 0.6974]$

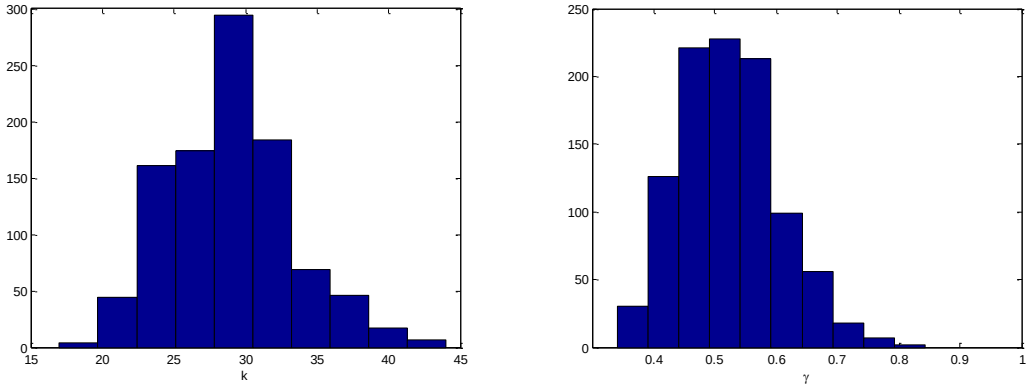

Figure 14 The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$

Again the KL deviance measure between the GPD and the posterior predicted distribution (fitted to observations above the threshold of the F distribution) seems to be working well in choosing the optimum threshold. The $\hat{\gamma}$ estimate at the optimum chosen threshold does not seem to be too far off from the true EVI.

3 KL measure in a truncated case (the threshold is known)

In the previous simulation examples an optimum threshold was chosen and although we are confident that the threshold was chosen appropriately with the KL deviance measure we have no proof that the optimum threshold is indeed the true threshold. In the next example we create the scenario where a data set from a Normal distribution is simulated. The tail of the distribution is then replaced with observations simulated from a GPD distribution. Since the Normal distribution is not heavy tailed but the GPD is, the threshold is the point above which the Normal data is replaced with the GPD. In the next simulation example we investigate whether the KL deviance measure chooses the threshold correctly.

3.1 Example 4

A standard Normal dataset of 500 observations is simulated and the largest 40 observations are replaced with simulated GPD observations with $\gamma = 0.25$ and τ the largest standard Normal observation (from the 460 that are left after truncation) + 0.1.

Figure 15 shows the KL measures at different values of k (k = 3:100) for one simulation. The KL measure closest to zero is obtained at $= 41$, which is close to the true threshold of $k = 40$. At $k = 41$ the parameter was estimated as $\hat{\gamma} = 0.2584$, which is also close to the true EVI of 0.25.

The simulation is repeated 1000 times, each time a new standard Normal dataset was simulated and the optimum threshold was chosen through the KL measure. The results are shown in Table 6 and the histograms of the 1000 simulated $k's$ and $\gamma's$ are shown in Figure 16.

Table 6 Summary of simulation results

$\overline{k}$	$CI_{0.95}$ for k	$\overline{\hat{\gamma}}$	$CI_{0.95}$ for $\hat{\gamma}$
34.2190	$[17; 43]$	0.2692	$[0.2245; 0.3612]$

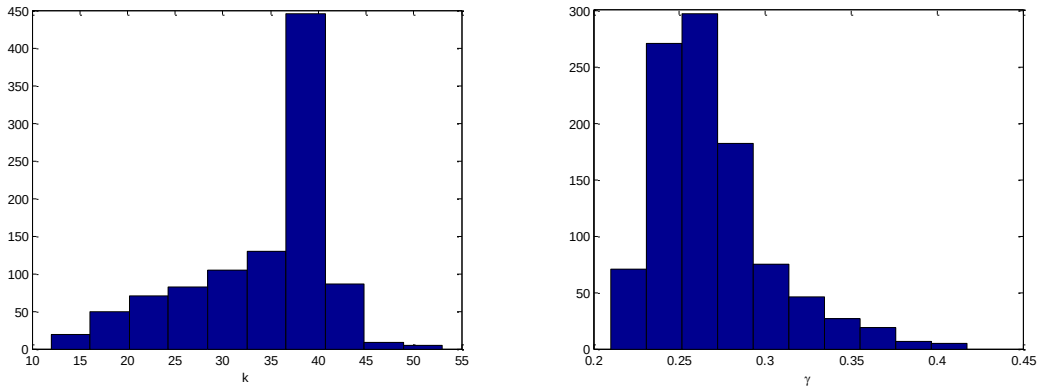

Figure 16 The histograms of the 1 000 simulated $k's$ and $\hat{\gamma}'s$

From the simulation example above the threshold was chosen fairly accurately with the KL measure the true threshold and EVI are both included in the 95% confidence interval.

4 A comparative simulation study between the KL measure and a method used by Thompson *et al.* 2009

In this section we repeat the automated threshold selection experiment described in Thompson *et al.* (2009) and perform our own automated selection procedure in tandem on every simulated data set. Thus, we obtain both a formal evaluation of the effectiveness of our approach and a direct comparison to the existing method.

In their experiment the samples are generated as follows: Simulate a large number of values from a $N(2,0.7^2)$ distribution. Discard the values less than zero and replace all the values above $\tau = 2.9$ with values from a GPD with parameters $\tau = 2.9$, $\gamma = 0.2$ and $\sigma = 0.4$. In their experiment they used a final sample size of 10 000; we reduced this to $n = 2\,000$.

Their algorithm begins by selecting 100 candidate thresholds evenly spaced between the median and the 98th percentile. For each threshold they estimate the parameters using the method of maximum likelihood and apply transformations to the sets of parameters to arrive at a set of numbers that should be Normally distributed. They apply tests for Normality over the range of thresholds and base their threshold choice on the p-values generated by the test.

For our algorithm we consider only the last 50 of these candidate thresholds. For each threshold we simulate 2 000 sets of parameters based on the observations exceeding the threshold only. For each set of parameters we obtain a value from the predictive posterior. We then combine the parameter and predictive posterior simulations to obtain a value for the KL deviance measure. Finally, we select the threshold that produced the minimum absolute KL value.

In total we considered 10 000 samples (of size 2 000 each) and applied both algorithms to each sample. For our algorithm we considered both the standard GPD where σ is free to vary, and the simplified GPD ($\sigma = \gamma\tau$).

We calculated the root mean square error (RMSE) and mean error (ME) as measures of accuracy and bias respectively. The error is calculated as the difference between the chosen threshold and the true threshold (known from construction) in terms of number of exceedences above the threshold. The results are given in Table 7 below.

Table 7 Summary of experiment results (10 000 samples)

Measure	Method of Thompson *et al.*	KL of 3 parameter GPD	KL of 2 parameter GPD
RMSE	269.6	141.6	40.8
ME	173.5	-139.5	30.0

It is clear from the observed results, viewed either per sample or in summary (as above) that the KL measure performs very well in terms of its ability to isolate the correct threshold value.

5 Conclusion

Often in Extreme Value Theory observations above a sufficiently large threshold are modelled with a Peaks-Over-Threshold distribution, such as the Generalized Pareto distribution. The choice of optimum threshold and the method used to obtain the optimum threshold has been the topic of discussion, with a lot of debate around it.

In this paper the emphasis falls on choosing an optimum threshold by using the Kullback-Leibler deviance measure between the GPD and its posterior predictive distribution, fitted to observations above a threshold. The threshold that results in a KL measure closest to zero is considered the optimum threshold. We have shown through various simulation examples and simulation studies that the KL deviance measure can successfully be used in a Bayesian context to select an optimum threshold. When compared with the method of Thompson *et al.*, 2009 our method gave more reliable results.

6 References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L. (2004). *Statistics of extremes: Theory and applications*. Chichester: Wiley.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Dipuis, D.J. (1998). Exceedences over high thresholds: a guide to threshold selection,

Extremes, Vol. 1, 251- 261.

Guillou, A., Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis, *Journal of the Royal Statistical Society B*, Vol. 63, 293-305.

Kullback, S., Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, Vol. 22, 79 – 86.

Ledford, A.W., Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83:169–187.

Thompson, P., Cai, Y., Reeve, D., Stander, J. (2009). Automated threshold selection methods for extreme wave analysis, *Coastal Engineering*, Vol. 56, 1013 – 1021.

Tancredi, A., Anderson, C.W., O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation, *Extremes*, Vol. 9, 87 – 106.

Verster, A., De Waal, D.J. (2011). A method for choosing an optimum threshold if the underlying distribution is generalized Burr-Gamma, *South African Statistical Journal*, 273 – 292.