ProbStat Forum, Volume 08, April 2015, Pages 45–53
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

Estimating the tail index: Another algorithmic method

Marta Ferreira

Center of Mathematics of Minho University and Department of Mathematics and Applications of Minho University, Campus de Gualtar 4710 - 057 *Braga, Portugal*

Abstract. The tail index is a determinant parameter within extreme value theory. Under a semi-parametric approach, one has often to choose the number of the largest order statistics to include in estimates. This is a hard task since it is not possible to know for sure where the tail of data really begins. This crucial topic has been largely addressed in literature and several methods were developed. In this paper we analyze, through simulation, a heuristic method and compare it with two very popular methodologies. It will be seen that the new method can be a good alternative.

1. Introduction

Extreme value theory (EVT) is focused on the occurrence of extreme events. This issue is on the agenda of areas like climatology and engineering due to climate changes, finances given the instability in the markets due to globalization, among others.

A crucial parameter in the estimation of the probability of rare events is the tail index, which defines the type of tail of a model. Consider a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s), $\{X_n\}_{n\geq 1}$, with marginal distribution function (d.f.) F. If for some real constants $a_n > 0$ and b_n the limit of the linearly normalized maxima exists and satisfies

$$P(\max(X_1, \ldots, X_n) \leq a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G_\xi(x)$$

for some non-degenerate function G_ξ, then we say that F belongs to the max-domain of attraction of G_ξ, in short, $F \in \mathcal{D}(G_\xi)$. This function is called generalized extreme value (GEV) and is given by

$$G_\xi(x) = \exp(-(1 + \xi x)^{-1/\xi}), \quad 1 + \xi x > 0, \xi \in \mathbb{R},$$

($G_0(x) = \exp(-e^{-x})$). The shape parameter ξ is the tail index and, whenever positive means that F has a heavy tail (Fréchet max-domain of attraction) with infinite right-end-point, if null means an exponential tail (Gumbel max-domain of attraction) and whenever negative corresponds to a light tail (Weibull max-domain of attraction) with finite right-end-point. Several estimators of ξ require that $F \in \mathcal{D}(G_\xi)$ (see, e.g., Hill (1975), Beirlant et al. (1996), among others) and depend on the k larger order statistics. We shall denote them $\widehat{\xi}_{k,n}$. Let $X_{n:n} \geq X_{n-1:n} \geq \cdots \geq X_{1:n}$ be the order statistics of $\{X_n\}_{n\geq 1}$. The Hill estimator (Hill, 1975), only valid for $\xi > 0$, is given by

$$\widehat{\xi}^H_{k,n} := \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} .$$

2010 *Mathematics Subject Classification.* Primary: 60G70; Secondary: 62G32
Keywords. Extreme value theory, Hill estimator, Generalized Hill estimator
Received: 02 February 2015; Revised: 17 February 2015; Accepted: 09 March 2015
Email address: msferreira@math.uminho.pt (Marta Ferreira)

The generalized Hill estimator (Beirlant et al., 1996) holds for all real values of ξ and is defined by

$$\widehat{\xi}^{GH}_{k,n} := \frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{X_{n-i:n}\,\widehat{\xi}^{H}_{i,n}}{X_{n-(k+1):n}\,\widehat{\xi}^{H}_{k+1,n}}\right).$$

Both estimators are consistent and asymptotically normal under quite general conditions, in particular, $k \equiv k_n \to 0$, $k_n/n \to 0$, $n \to \infty$, with asymptotical variances given by, respectively,

$$Avar(\widehat{\xi}^{H}_{k,n}) = \frac{\xi^2}{k} \quad (1)$$

and

$$Avar(\widehat{\xi}^{GH}_{k,n}) = \begin{cases} \frac{1+\xi^2}{k} & ,\ \xi \geq 0 \\ \frac{(1-\xi)(1+\xi+2\xi^2)}{(1-2\xi)k} & ,\ \xi < 0. \end{cases} \quad (2)$$

The choice of the k upper values (defining the tail of F) is the crucial issue and is not easy. On one hand, small values of k lead to larger variances of the estimators and, on the other hand, large values increase the bias. By plotting $(k, \widehat{\xi}_{k,n})$, $1 \leq k < n$, we can find a range of k where $\widehat{\xi}_{k,n}$ is approximately constant around the true value (see Figure 1). Many estimation procedures are based on this graphical tool (see e.g., Drees and Kaufmann (1998), Neves and Fraga Alves (2004) and references therein).

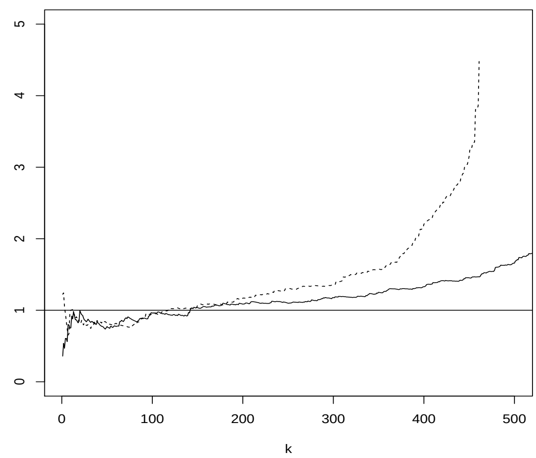

Figure 1: Hill plots of 1000 realizations of a GEV with $\xi = 1$ (full line) and a Cauchy (dashed line). The horizontal line corresponds to the true value of ξ.

A heuristic procedure to infer a stable region from a similar graphical tool within the context of bivariate tail dependence was introduced in Frahm et al. (2005) and further applied in Ferreira and Silva (2014). In Ferreira (2014) the method was analyzed for the Hill estimator in order to evaluate if the smooth bandwidth affected the results. Here we use this procedure and present a pontual and intervalar estimation of the tail index, based on the Hill and Generalized Hill estimators. The variance estimation is conducted as in Ferreira and Silva (2014). Our simulation results are also compared with two other graphical-based estimation procedures well-known and applied in literature.

2. The Methods

We start by presenting our method which is based in the one introduced in Frahm et al. (2005) within the context of bivariate tail dependence. More precisely, the following algorithm looks for the stable region of the estimator's path:

Algorithm 1:
Smooth the plot $(k, \widehat{\xi}_{k,n})$ by considering the means of $2b+1$ successive points of $\widehat{\xi}_{k,n}$, $1 \le k < n$, with bandwidth $b = \lfloor wn \rfloor \in \mathbb{N}$ (here w is a parameter for the degree of smoothing to be used, e.g., if $w = 0.005$ then each moving average is about 1% of the data). Define the plane regions $p_k = (\overline{\widehat{\xi}}_{k,n}, \ldots, \overline{\widehat{\xi}}_{k+m-1,n})$, $k = 1, \ldots, n-2b-m+1$, with length $m = \lfloor \sqrt{n-2b} \rfloor$ in the smoothed moving average values, $\overline{\widehat{\xi}}_{1,n}, \ldots, \overline{\widehat{\xi}}_{n-2b,n}$. The algorithm stops at the first plane region such that

$$\sum_{i=k+1}^{k+m-1} \left| \overline{\widehat{\xi}}_{i,n} - \overline{\widehat{\xi}}_{k,n} \right| \le 2s,$$

where s is the empirical standard deviation of $\overline{\widehat{\xi}}_{1,n}, \ldots, \overline{\widehat{\xi}}_{n-2b,n}$. Estimate ξ by the mean of the values of the plane region chosen in the previous step (consider the estimate zero if no stable region fulfills the stopping condition).

In Ferreira (2014) it was shown that, in general, the Hill estimates do not change much for a wide range of values of w. Our simulation results also support this to the generalized Hill (we have considered $w = 0.005$, 0.01, 0.015, 0.02, 0.025 and 0.03). Thus we only present the results of $w = 0.005$ (this value was also suggested in Frahm et al. (2005)). Besides point estimation, we analyze interval estimation. In order to estimate the variance, we are going to apply Algorithm 1 to the asymptotical variance's path $(k, Avar(\widehat{\xi}^H_{k,n}))$, but we choose the plane region in step 3 at the same position of the one found for the ξ estimation. This strategy was also followed in Ferreira and Silva (2014).

Now we describe the other two graphical procedures that look for a possibly k optimal value, which are based on the following model derived by Beirlant et al. (2002), under quite mild assumptions, for $\xi \ne 0$:

$$Y_i := (i+1) \log \frac{X_{n-i:n}\, \widehat{\xi}^H_{i,n}}{X_{n-(i+1):n}\, \widehat{\xi}^H_{i+1,n}} = \xi + b(n/k) \left(\frac{i}{k}\right)^{-\rho} + \epsilon_i,\ i = 1, \ldots, k, \tag{3}$$

where ϵ_i are considered zero-centered error terms and b is a positive function such that $b(x) \to 0$, as $x \to \infty$. Parameter ρ is also a tail parameter that can be estimated (see Beirlant et al. (2002)). However, in practice, it can be replaced by the value $\rho = -1$ (more details can be seen in Beirlant et al. (2004) and references therein). Therefore, under fixed ρ, estimates of ξ and $b(n/k)$ can be obtained from (3) by applying least squares, leading to

$$\begin{aligned} \widetilde{\xi}^{LS}_{k,n} &= \overline{Y}_k - \widetilde{b}^{LS}_{k,n}/(1-\rho) \\ \widetilde{b}^{LS}_{k,n} &= \tfrac{(1-\rho)^2(1-2\rho)}{\rho^2} \tfrac{1}{k} \textstyle\sum_{i=1}^k \left(\left(\tfrac{i}{k}\right)^{-\rho} - \tfrac{1}{1-\rho}\right) Y_i. \end{aligned} \tag{4}$$

The asymptotic mean squared error (AMSE) of Hill and generalized Hill are, respectively, given by

$$\text{AMSE}(\widehat{\xi}^H_{k,n}) = \tfrac{\xi^2}{k} + \left(\tfrac{b(n/k)}{1-\rho}\right)^2 \tag{5}$$

and

$$\text{AMSE}(\widehat{\xi}^{GH}_{k,n}) = \begin{cases} \tfrac{1+\xi^2}{k} + \left(\tfrac{b(n/k)}{1-\rho}\right)^2 & ,\ \xi \ge 0 \\ \tfrac{(1-\xi)(1+\xi+2\xi^2)}{(1-2\xi)k} + \left(\tfrac{b(n/k)}{1-\rho}\right)^2 & ,\ \xi < 0. \end{cases} \tag{6}$$

The optimal values of k for which the respective AMSE is minimal have the expressions

$$k^H_{opt} \sim b(n/k)^{-2/(1-2\rho)} k^{-2\rho/(1-2\rho)} \left(\tfrac{\xi^2(1-\rho)^2}{-2\rho}\right)^{1/(1-2\rho)} \tag{7}$$

and

$$k_{opt}^{GH} \sim \begin{cases} b(n/k)^{-2/(1-2\rho)} k^{-2\rho/(1-2\rho)} \left(\frac{(1+\xi^2)(1-\rho)^2}{-2\rho}\right)^{1/(1-2\rho)} & , \xi > 0 \\ \frac{1}{4} b(n/k)^{-5/2} k^{-5\rho/2} & , \xi = 0 \\ \left(\frac{(1-\xi)(1-\rho)^2(1+\xi+2\xi^2)}{(1-2\xi)(-2\rho)}\right)^{1/(1-2\rho)} b(n/k)^{-2/(1-2\rho)} k^{-2\rho/(1-2\rho)} & , \xi < 0. \end{cases} \quad (8)$$

More details can be found in Beirlant et al. (2004).

For comparison, the following two algorithms (see Beirlant et al. (2002) and Matthys and Beirlant (2000)) were also applied:

Algorithm 2:
Consider $\rho = -1$ and obtain estimates for ξ and $b(n/k)$ by least squares using (4), for $k = 3, \ldots, n$. Compute $\widehat{k}_{opt,k}$ according to the expression (7) or (8), whether you use Hill or generalized Hill, by replacing ξ and $b(n/k)$ with the values calculated in the previous step, for $k = 3, \ldots, n$. Take $\widehat{k}_{opt}^2 = median\{\widehat{k}_{opt,k},\ k = 3, \ldots, \lfloor \frac{n}{2} \rfloor\}$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. Estimate ξ by $\widehat{\xi}^H_{\widehat{k}^2_{opt},n}$.

Algorithm 3:
Consider $\rho = -1$ and obtain an estimate of ξ and $b(n/k)$ by least squares using (4). Compute the AMSE according to the expression (5) or (6), whether you use Hill or generalized Hill, by replacing ξ and $b(n/k)$ with the values calculated in the previous step. Take $\widehat{k}_{opt}^1$ as the value of k that minimizes the estimates of the AMSE in the previous step. Estimate ξ by $\widehat{\xi}^H_{\widehat{k}^1_{opt},n}$.

The asymptotical variances of Hill and generalized Hill can be obtained within these two algorithms by, respectively, $Avar(\widehat{\xi}^H_{\widehat{k}^1_{opt},n})$ and $Avar(\widehat{\xi}^H_{\widehat{k}^2_{opt},n})$ using (1), and $Avar(\widehat{\xi}^{GH}_{\widehat{k}^1_{opt},n})$ and $Avar(\widehat{\xi}^{GH}_{\widehat{k}^2_{opt},n})$ using (2).

In the next section, we analyze the performance of the heuristic method presented in Algorithm 1, based on the Hill and generalized Hill estimators, through simulation and compare the results with the ones obtained from the methods described in Algorithm 2 and Algorithm 3, both widely known and applied in literature. The asymptotical variances of Hill and generalized Hill are also computed in order to obtain interval estimates.

3. Simulations and results

The models used in the simulation study are:

- GEV(ξ) with d.f. $G_\xi(x)$: GEV(1), GEV(0) and GEV(-1);
- Generalized Pareto with d.f. $1 + \log G_\xi(x)$ (GP(ξ)): GP(1), GP(0) and GP(-1);
- Cauchy ($\xi = 1$),
- Burr(β,τ,λ), with d.f. $F(x) = 1 - (\beta/(\beta + x^\tau))^\lambda$): Burr(1,2,2) ($\xi = 1/4$) and Burr(1,1/2,2) ($\xi = 1$)
- Standard Normal ($\xi = 0$);
- Weib(λ,τ), with d.f. $F(x) = 1 - \exp(-\lambda x^{-\tau})$: Weibull(1,1/2) ($\xi = 0$);
- Reversed Burr, RBurr(β,τ,λ) with d.f. $F(x) = 1 - (\beta/(\beta + (x_+ - x)^{-\tau}))^\lambda$): RBurr(1,4,1), with $x_+ = 1$ ($\xi = -1/4$).

We have considered 1000 independent samples of size $n = 1000$. We denote the methods based on Algorithms 1, 2 and 3, respectively, A1, A2 and A3. The bias and the rmse of the Hill estimator are in Tables 1 and 3,

Table 1: Bias of $\widehat{\xi}^{H}_{k,n}$.

bias	A1	A2	A3
Cauchy	-0.0298	-0.0332	-0.0700
GP(1)	0.0031	0.0115	-0.0406
GEV(1)	-0.0023	0.0132	-0.0618
Burr(1,2,2)	0.0295	0.0253	0.1902
Burr(1,1/2,2)	0.1306	0.1555	-0.0295

Table 2: Bias of $\widehat{\xi}^{GH}_{k,n}$.

bias	A1	A2	A3
Cauchy	-0.0737	-0.0858	-0.0399
GP(1)	-0.0306	-0.0253	0.0012
GEV(1)	-0.0413	-0.0335	-0.0056
Burr(1,2,2)	-0.0569	-0.0708	-0.0603
Burr(1,1/2,2)	0.0480	0.0788	0.0685
GP(0)	-0.0042	0.0098	0.0170
GEV(0)	-0.0095	-0.0083	0.0077
Normal St.	-0.1551	-0.1653	-0.1521
Weibull(1,1/2)	0.2320	0.2615	0.2141
GP(-1)	-0.0122	-0.0077	0.2381
GEV(-1)	-0.0390	-0.0471	0.1597
R. Burr(1,4,1)	-0.0369	-0.0537	0.1748

Table 3: Root mean squared error (rmse) of $\widehat{\xi}^{H}_{k,n}$.

rmse	A1	A2	A3
Cauchy	0.1848	0.2048	0.2720
GP(1)	0.1444	0.1279	0.2675
GEV(1)	0.1440	0.1274	0.2780
Burr(1,2,2)	0.0551	0.0636	0.2159
Burr(1,1/2,2)	0.2277	0.2385	0.3860

respectively, and plotted in Figure 2. The results concerning the bias and the rmse of the generalized Hill estimator can be seen in Tables 2 and 4 and also in Figures 3 (case $\xi > 0$) and 4 (case $\xi \leq 0$).

For estimator $\widehat{\xi}^{H}_{k,n}$, the method A3 has the worst performance. Both methods A1 and A2 behave similar and reasonably. The results within these latter (except for Cauchy and GP(1) models) also compare favorably with the ones based on Drees and Kaufmann (1998) sequential method derived in Neves and Fraga Alves (2004). In the case of estimator $\widehat{\xi}^{GH}_{k,n}$ for $\xi > 0$, the three methods have a similar performance. The method A3 behaves better than in the case of the Hill estimator, while methods A1 and A2 are slightly worse. However, for $\xi < 0$, method A3 does not perform well. It is preferable in this case to use the methods A1 and A2, since both have a good performance, although not so good for Normal and Weibull models. Regarding the interval estimation (Tables 5 and 6 and Figure 5), the best method is the A1 for both estimators, with a quite reasonable performance except again in the Normal and Weibull models. In a future work, we will continue to exploit the heuristic method A1 presented here, in order to improve its performance in a larger range of models, and compare it with other procedures like, for instance, bootstrap methods (see, e.g., Gomes and Oliveira (2001) and references therein).

Table 4: Root mean squared error (rmse) of $\widehat{\xi}^{GH}_{k,n}$.

rmse	A1	A2	A3
Cauchy	0.2339	0.2415	0.2337
GP(1)	0.1775	0.1608	0.2036
GEV(1)	0.1746	0.1601	0.2032
Burr(1,2,2)	0.1163	0.1245	0.0848
Burr(1,1/2,2)	0.2299	0.2246	0.3760
GP(0)	0.1198	0.1079	0.1229
GEV(0)	0.1380	0.1232	0.1141
Normal St.	0.2189	0.2139	0.1879
Weibull(1,1/2)	0.2732	0.3001	0.3394
GP(-1)	0.1282	0.1117	0.2381
GEV(-1)	0.1370	0.1467	0.5847
R. Burr(1,4,1)	0.1422	0.1457	0.4282

Table 5: Empirical coverage probabilities of $\widehat{\xi}^{H}_{k,n}$.

coverage IC 95%	A1	A2	A3
Cauchy	0.937	0.836	0.660
GP(1)	0.967	0.869	0.591
GEV(1)	0.969	0.883	0.584
Burr(1,2,2)	0.960	0.696	0.580
Burr(1,1/2,2)	0.963	0.803	0.406

Table 6: Empirical coverage probabilities of $\widehat{\xi}^{GH}_{k,n}$.

coverage IC 95%	A1	A2	A3
Cauchy	0.905	0.832	0.865
GP(1)	0.951	0.892	0.834
GEV(1)	0.958	0.883	0.827
Burr(1,2,2)	0.908	0.840	0.707
Burr(1,1/2,2)	0.975	0.863	0.607
GP(0)	0.982	0.928	0.799
GEV(0)	0.982	0.935	0.841
Normal St.	0.865	0.755	0.564
Weibull(1,1/2)	0.837	0.463	0.406
GP(-1)	0.976	0.946	0.838
GEV(-1)	0.965	0.942	0.909
R. Burr(1,4,1)	0.953	0.909	0.874

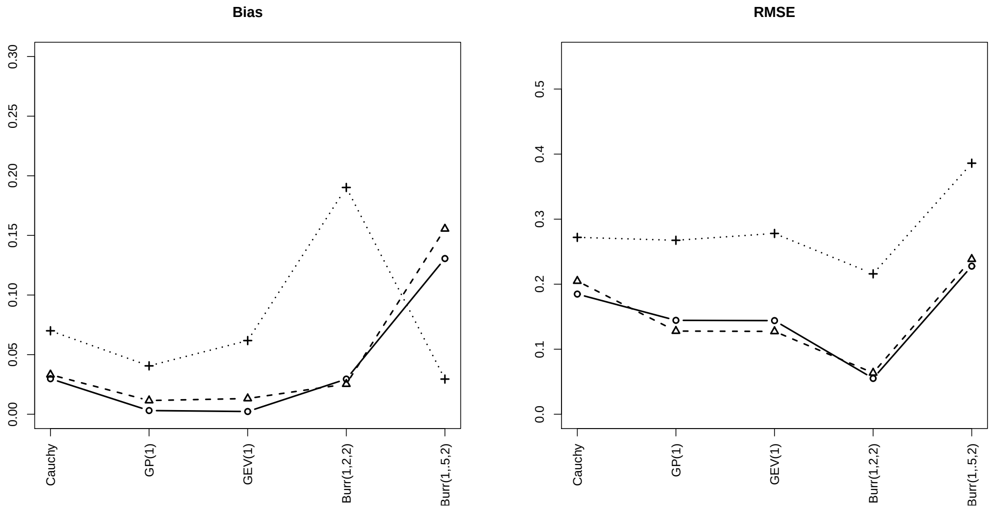

Figure 2: Absolute bias (left) and rmse (right) of estimator $\widehat{\xi}^{\text{H}}$, where the full line, the dashed line and the dotted line correspond to the methods, respectively, A1, A2 and A3.

Acknowledgments

The paper was partially supported by the Research Centre of Mathematics of the University of Minho with the Portuguese Funds from the "Fundao para a Cincia e a Tecnologia", through the Project PEstOE/ MAT/UI0013/2014.

References

Beirlant, J., Vynckier, P., Teugels, J.L. (1996) *Excess functions and estimation of the extreme value index*, Bernoulli 2, 293–318.

Beirlant, J., Dierckx, G., Guillou, A., Stărică, C. (2002) *On Exponential Representation of Log-Spacings of Extreme Order Statistics*, Extremes 5, 157–180.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L. (2004) *Statistics of Extremes: Theory and Applications*, J. Wiley & Sons.

Drees, H., Kaufmann, E. (1998) *Selecting the optimal sample fraction in univariate extreme value estimation*, Stochastic Process Appl. 75, 149–172.

Ferreira, M., Silva, S. (2014) *An Analysis of a Heuristic Procedure to Evaluate Tail (in)dependence*, Journal of Probability and Statistics, Vol. 2014, Article ID 913621, 15 pages.

Ferreira, M. (2014) *A Heuristic Procedure to Estimate the Tail Index*, Proceedings of the 14th International Conference in Computational Science and Its Applications - ICCSA 2014, June 30 - July 3 (2014), Guimares, Portugal, IEEE-Computer Society 4264a241, 241–245.

Frahm, G., Junker, M., Schmidt R. (2005) *Estimating the tail-dependence coefficient: properties and pitfalls*, Insurance: Mathematics & Economics 37(1), 80–100.

Gomes, M.I., Oliveira, O. (2001) *The bootstrap methodology in statistics of extremes-choice of the optimal sample fraction*, Extremes 4(4), 331–358.

Hill, B.M. (1975) *A Simple General Approach to Inference About the Tail of a Distribution* Ann. Stat. 3, 1163–1174.

Matthys, G., Beirlant, J. (2000) *Adaptive Threshold Selection in Tail Index Estimation*, in Extremes and Integrated Risk Management, ed. P. Embrechts, London: Risk Books.

Neves, C., Fraga Alves, M.I. (2004) *Reiss and Thomas' automatic selection of the numberof extremes*, Computational Statistics & Data Analysis 47, 689–704.

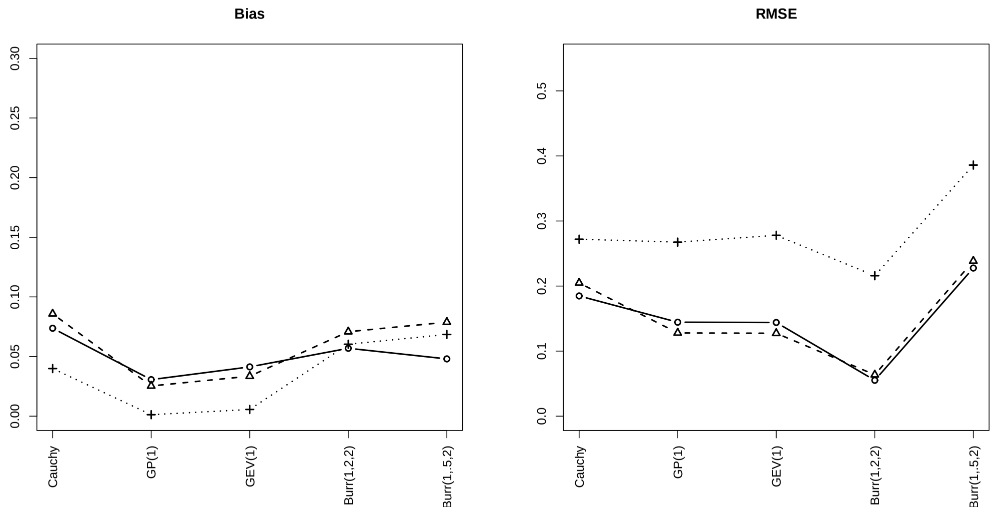

Figure 3: Absolute bias (left) and rmse (right) of estimator $\widehat{\xi}^{\mathtt{GH}}$ for $\xi > 0$, where the full line, the dashed line and the dotted line correspond to the methods, respectively, A1, A2 and A3.

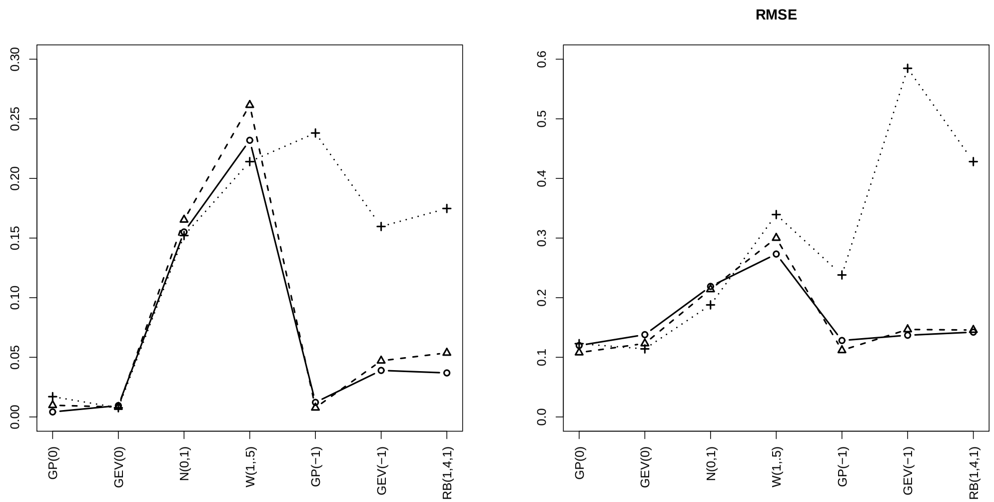

Figure 4: Absolute bias (left) and rmse (right) of estimator $\widehat{\xi}^{\mathtt{GH}}$ for $\xi \leq 0$, where the full line, the dashed line and the dotted line correspond to the methods, respectively, A1, A2 and A3.

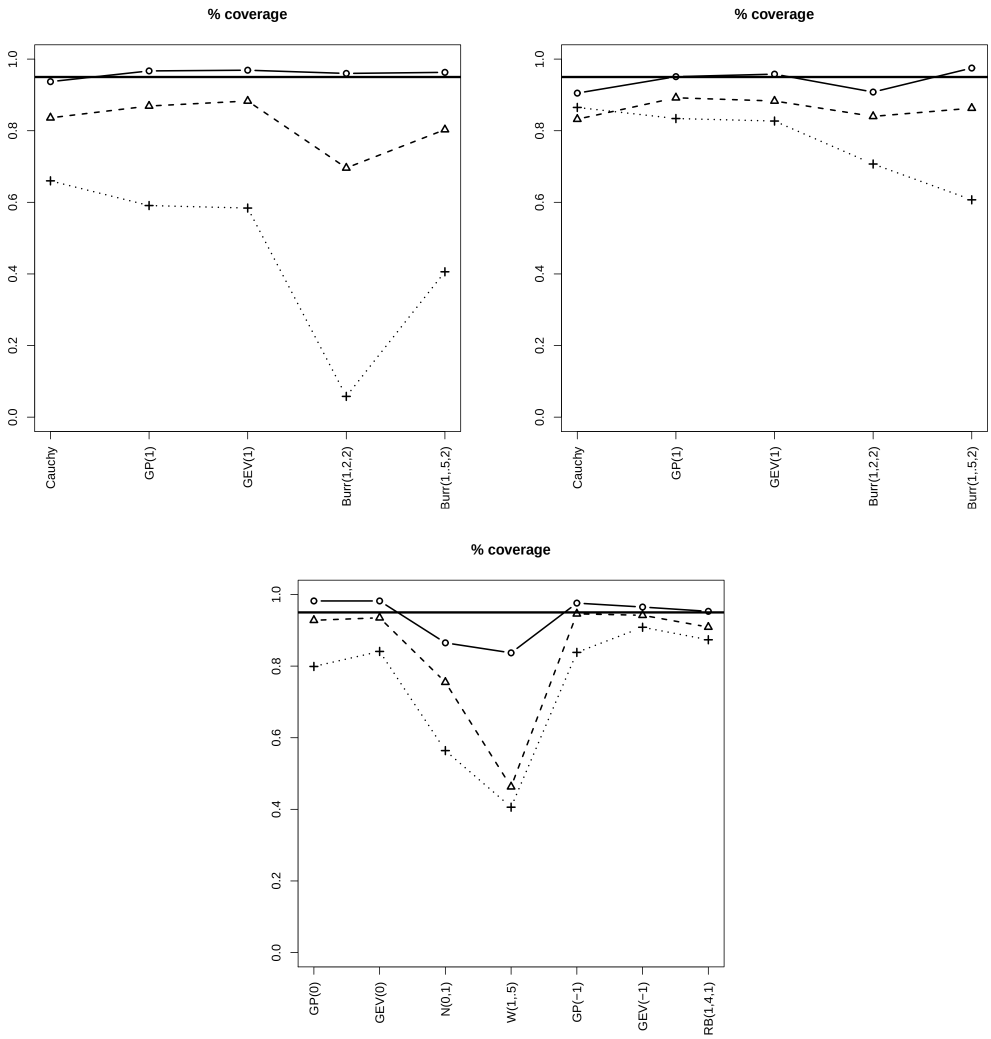

Figure 5: Empirical coverage probabilities of $\widehat{\xi}^{H}_{k,n}$ (top-left), $\widehat{\xi}^{GH}_{k,n}$ for $\xi > 0$ (top-right) and $\widehat{\xi}^{GH}_{k,n}$ for $\xi \leq 0$ (bottom), where the full line, the dashed line and the dotted line correspond to the methods, respectively, A1, A2 and A3.