Probability weighted moments bootstrap estimation: a case study in the field of insurance*

Frederico Caeiro
Universidade Nova de Lisboa, FCT, DM, and CMA
E-mail: ivette.gomes@fc.ul.pt

M. Ivette Gomes
Universidade de Lisboa, FCUL, DEIO, and CEAUL
E-mail: fac@fct.unl.pt

Abstract

We make use of *probability weighted moments* of largest observations, in order to build classes of estimators of the *extreme value index*. Due to the specificity of the estimators, we propose the use of bootstrap computer intensive methods for an adaptive choice of the optimal number of order statistics to be used in the estimation. The methodology is applied to data in the field of insurance.

1 Introduction and preliminaries

The *extreme value index* (EVI) is the real parameter γ in the general *extreme value* (EV) distribution function (d.f.), $G_\gamma(x) := \exp(-(1+\gamma x)^{-1/\gamma})$, $1 + \gamma x > 0$. Let $\underline{X}_n = (X_1, \ldots, X_n)$ denote a random sample of size n, and consider the associated sample of ascending order statistics (o.s.'s) $(X_{1:n} \le \cdots \le X_{n:n})$. One of the first classes of semi-parametric estimators of a positive EVI was the Hill (H) estimator ([4]), given by

*Research partially supported by FCT/OE, POCI 2010 and PTDC/FEDER.

$\hat{\gamma}^H_{k,n} := \sum_{i=1}^k \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}/k,\ k = 1, 2, \ldots, n-1$. We shall also deal with the *Pareto probability weighted moments* (PPWM) EVI-estimators, recently introduced in [1]. They are valid for $0 < \gamma < 1$, compare favourably with the Hill estimator, and are given by

$$\widehat{\gamma}^{PPWM}_{k,n} := 1 - \hat{a}_1(k)/\big(\hat{a}_0(k) - \hat{a}_1(k)\big), \tag{1}$$

with $\hat{a}_0(k) := \sum_{i=1}^k X_{n-i+1:n}/k$ and $\hat{a}_1(k) := \sum_{i=1}^k (i/k)\ X_{n-i+1:n}/k$. Consistency of these EVI-estimators is achieved if $X_{n-k:n}$ is an *intermediate* o.s., i.e., if $k = k_n \to \infty$ and $k/n \to 0$, as $n \to \infty$. In order to derive the asymptotic normality of these EVI-estimators, and with the notation $U(t) := \inf\{x : F(x) \geq 1 - 1/t\}$, $t \geq 1$, it is often assumed the validity of a second-order condition, like $\lim_{t\to\infty} \big(\ln U(tx) - \ln U(t) - \gamma \ln x\big)/A(t) = (x^\rho - 1)/\rho$, where $|A| \in RV_\rho$, $\rho \leq 0$. Under such a second-order framework, if $\sqrt{k}A(n/k) \to \lambda_{_A}$, finite, as $n \to \infty$, these EVI-estimators are asymptotically normal. Denoting $\hat{\gamma}^\bullet_{k,n}$, any of the estimators above, we have, with $Z^\bullet_k$ asymptotically standard normal and for adequate $(b_\bullet,\ \sigma_\bullet) \in (\mathbb{R},\ \mathbb{R}^+)$,

$$\hat{\gamma}^\bullet_{k,n} \overset{d}{=} \gamma + \sigma_\bullet Z^\bullet_k/\sqrt{k} + b_\bullet\ A(n/k)(1 + o_p(1)), \quad \text{as} \quad n \to \infty. \tag{2}$$

After a review, in Section 2, of the role of the bootstrap methodology in the estimation of optimal sample fractions, we provide a reference to an algorithm for the adaptive estimation through the Hill estimators, also valid for the PPWM EVI-estimators. In Section 3, as an illustration, we apply such a data-driven estimation to a data set in the field of insurance.

2 The bootstrap methodology and optimal levels

Under the above mentioned second-order framework, with $\rho < 0$, let us use the parameterization $A(t) = \gamma\beta t^\rho$, where β and ρ are generalized scale and shape second-order parameters. Given the EVI-estimator, $\hat{\gamma}^\bullet_{k,n}$, let us denote $k_0^{\hat{\gamma}^\bullet}(n) := \arg\min_k \mathrm{MSE}(\hat{\gamma}^\bullet_{k,n})$, with MSE standing for *mean squared error*. With $\mathbb{E}$ denoting the mean value operator and AMSE standing for *asymptotic mean squared error*, a possible substitute for $\mathrm{MSE}(\hat{\gamma}^\bullet_{k,n})$ is $\mathrm{AMSE}(\hat{\gamma}^\bullet_{k,n}) := \mathbb{E}\big(\sigma_\bullet\ \overline{Z}_k/\sqrt{k} + b_\bullet A(n/k)\big)^2 = \sigma^2_\bullet/k + b^2_\bullet\ \gamma^2\ \beta^2\ (n/k)^{2\rho}$, cf.

equation (2). Then, with the notation $k_{0|\hat{\gamma}^\bullet}(n) := \arg\min_k \mathrm{AMSE}\big(\hat{\gamma}^\bullet_{k,n}\big)$, we get $k_{0|\hat{\gamma}^\bullet}(n) = k_0^{\hat{\gamma}^\bullet}(n)(1+o(1))$. For the Hill estimator, we have, in (2), $(b_{_H}, \sigma_{_H}) = (1/(1-\rho), \gamma)$. Consequently, with $(\hat{\beta}, \hat{\rho})$ a consistent estimator of (β, ρ) and $[x]$ denoting the integer part of x, we have an asymptotic justification for the estimator $\hat{k}_0^H := [((1-\hat{\rho})^2 n^{-2\hat{\rho}}/(-2\hat{\rho}\hat{\beta}^2))^{1/(1-2\hat{\rho})}] + 1$. The same does not happen with the PPWM EVI-estimators, due to the fact that $\sigma_{_{PPWM}}$ and $b_{_{PPWM}}$ depend both on γ. It is sensible to use the bootstrap methodology for the adaptive PPWM EVI-estimation. Similarly to what has been done in [3], for the H estimator, we can use the algorithm in [2], considering the auxiliary statistic, $T^\bullet_{k,n} := \hat{\gamma}^\bullet_{[k/2],n} - \hat{\gamma}^\bullet_{k,n}$, $k = 2, \ldots, n-1$, which converges to the known value zero, and double-bootstrap it adequately on the basis of samples of sizes $n_1 = o(n)$ and $n_2 = [n_1^2/n]$, in order to estimate $k_{0|\hat{\gamma}^\bullet}(n)$, through a bootstrap estimate $\hat{k}^\bullet_{0*}$. Note also that bootstrap confidence intervals (CIs) are easily associated with the bootstrap EVI-estimates, through the replication of the above-mentioned algorithm r times.

3 A case study

We shall next consider an illustration of the performance of the adaptive PPWM EVI-estimates under study, comparatively with the same methodology applied to the Hill EVI-estimates, again through the analysis of $n = 371$ automobile claim amounts exceeding 1,200,000 Euro over the period 1988-2001, gathered from several European insurance companies co-operating with the same re-insurer, Secura Belgian Re. The above-mentioned algorithm led us to $\hat{\rho}_0 = -0.74$ and $\hat{\beta}_0 = 0.80$. For a sub-sample size $n_1 = [n^{0.955}] = 284$, and $B = 250$ bootstrap generations, we were led to $\hat{k}_{0*}^{PPWM} = 58$ and to $PPWM^* = 0.272$. This same algorithm applied to the Hill estimates leads us to $\hat{k}_{0*}^H = 52$ and to $H^* = 0.299$.

In Figure 1, as a function of the sub-sample size n_1, ranging from $n_1 = [n^{0.95}] = 275$ until $n_1 = [n^{0.9999}] = 370$, we picture, at the left, the estimates $\hat{k}_{0*}(n_1)/n$ of the *optimal sample fraction* (OSF), $k_0^\bullet/n$, for the adaptive bootstrap estimation of γ through the Hill and the PPWM estimators.

Associated bootstrap EVI-estimates are pictured at the right. Contrarily to the bootstrap Hill, the bootstrap PPWM EVI-estimates are quite stable as a function of the sub-sample size n_1 (see Figure 1, *right*).

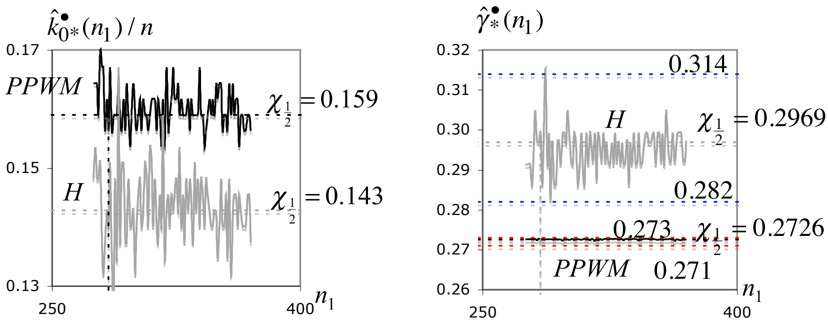

Figure 1: Estimates of the OSF's $\hat{k}^\bullet_0/n$ (*left*) and the bootstrap adaptive extreme value index estimates $\hat{\gamma}^\bullet_*$ (*right*), as functions of the sub-sample size n_1, for the SECURA data.

The running of the above mentioned algorithm $r = 100$ times, for $n_1 = [n^{0.955}]$, provided, for the PPWM-estimates, a median 0.2726, an average 0.2725, and a 95% bootstrap CI for γ given by $(0.271, 0.273)$, as shown in Figire 1. The equivalent indicators for the bootstrap Hill estimates were 0.2969, 0.2949 and $(0.282, 0.314)$, also shown in Figure 1. The size of the CIs are in favour of the PPWM estimation. Indeed, the H-estimates are clearly over-estimating the true value of the EVI, and should be used with care.

References

[1] Caeiro F., Gomes M.I. (2011). Semi-parametric tail inference through probability-weighted moments. *J. Statistical Planning and Inference* **141**, 937–950.

[2] Caeiro, F., Gomes, M.I. and Vandewalle, B. (2011). *Semi-Parametric Probability-Weighted Moments Estimation Revisited.* Notas e Comunicações 03/2011. Submitted.

[3] Gomes M.I., Oliveira O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction. *Extremes* **4**:4, 331–358.

[4] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**:5, 1163–1174.