

Chapter 4

Extreme Value Statistics



When studying peaks in electricity demand, we may be interested in understanding the risk of a certain large level for demand being exceeded. For example, there is potential interest in finding the probability that the electricity demand of a business or household exceeds the contractual limit. An alternative, yet in principle equivalent way, involves assessment of maximal needs for electricity over a certain period of time, like a day, a week or a season within a year. This would stem from the potential interested in quantifying the largest electricity consumption for a substation, household or business. In either case, we are trying to infer about extreme traits in electricity loads for a certain region assumed fairly homogeneous with the ultimate aim of predicting the likelihood of an extreme event which might have never been observed before. While the exact truth may be not be possible to determine, it may be possible come up with an educated guess (an estimate) and ascertain confidence margins around it.

In this chapter, not only we will list and describe mainstream statistical methodology for drawing inference on extreme and rare events, but we will also endeavour to elucidate what sets the semiparametric approach apart from the classical parametric approach and how these two eventually align with one another. For details on possible approaches and related statistical methodologies we refer to [1, 2].

We hope that, through this chapter, practitioners and users of extreme value theory will be able to see how the theoretical results in Chap. 3 translate in practice and how conditions can be checked. We will be mainly concerned with semi-parametric inference for univariate extremes. This statistical methodology builds strongly on the fundamental results expounded in the previous section, most notably the theory of extended regular variation (see e.g. [3], Appendix B).

Despite the numerous approaches whereby extreme values can be statistically analysed, these are generally classed into two main frameworks: methods for maxima over fixed intervals (blocks) and methods for exceedances (peaks) over high

thresholds. The former relates the oldest group of models, arising from the seminal work of [4] as block maxima models to be fitted to the largest observations collected from large samples (blocks) of identically distributed observations. The latter has often been considered the most useful framework in practical applications due to the widely proclaimed advantage over the former of a more efficient use of the often scarce extreme data.

The two main settings in univariate extreme value theory we are going to address are:

- The block maxima (BM) method, whereby observations are selected in blocks of equal length (assumed large blocks) and perform inference under the assumption that the maximum in each block (usually a year) is well approximated by the Generalised Extreme Value (GEV) distribution.
- The peaks over threshold method (POT) enables to restrict attention to those observations from the sample that exceed a certain level or threshold, supposedly high, and use mainstream statistical techniques such as maximum likelihood or method of moments to estimation and hypothesis testing, under the assumption that these excesses (values by each the threshold is exceeded) follow exactly a Generalised Pareto distribution (GPD).

Application to energy smart meter data is an important part of the challenge, in the sense of the impending application to extreme quantile estimation, i.e. to levels which are exceeded with only a small probability. A point to be wary of, when applying EVT is that, due to the wiggleness of the real world and the means by which empirical measurements are collected, observations hardly follow exactly an extreme value distribution. A recent application of extreme value theory can be found in [5]. Despite the underpinning theory to the physical models considered in this paper determines that the true probability function generating the data belongs to the Gumbel domain of attraction, the authors attempt statistical verification of this assumption with concomitant estimation of the GEV shape parameter γ via maximum likelihood, in a purely parametric approach. They find an estimate of -0.002 which indicates that their data is bounded from above, i.e., there is a finite upper endpoint. However this is merely a point estimate whose significance must be evaluated through a test of hypothesis. In the parametric setting, the natural choice would be the likelihood ratio test for the simple null hypothesis that $\gamma = 0$.

The semiparametric framework, where inference takes places in the domains of attraction rather than through of the prescribed limiting distribution to the data—either GEV or GPD depending on we set about to look at extremes in the data at our disposal—has proven a fruitful and flexible approach.

In this chapter, we discuss the choice of max-domains of attraction within the semiparametric framework where an EVI $\gamma = 0$ and finite upper endpoint are allowed to coexist. To this effect, we choose to focus on the POT approach as the methodology expounded here will be greatly driven and moderated by the statistical analysis of extreme features (peaks) exhibited by the Irish smart meter data. The description of the data has been given in Chap. 1. We recall that the data comprises 7 weeks of half-hourly loads (in kWh) for 503 households. Figure 4.1 is a rendering of the box-plots

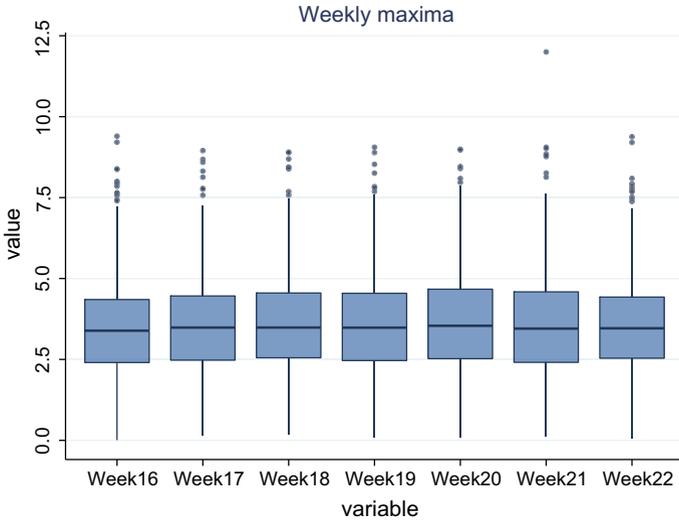


Fig. 4.1 Parallel box-plot representation of weekly maxima

for each week being analysed. All seven box-plots look very similar, both in terms of median energy demand and dispersion of values of energy around the median. There is however one absolute extreme which we will be taking notice in the next sections. This value was recorded in Week 21. An important consideration at this point is that there is a physical limit to the individual electricity loads, which not any less imposed by limitations of the electrical supply infrastructure than by contractual constraint on the upper bound for an individual load. In statistical terms, this means that the assumption of a finite upper endpoint to the d.f. underlying seems fairly reasonable. Nonetheless, the data might indicated otherwise, suggesting that households are actually operating far below the stipulated upper bound, a circumstance that can be potentially exploited by the energy supplier so as to improve management of energy supply.

Figure 4.2 is a scatter-plot representation of the actual observations in terms of the household number. With these two plots we intend to illustrate the two stated methods for the statistical analysis of extreme values, both BM and POT. The top panel shows all the data points. While proceeding with the BM method, one would take the largest observation at each $h = 1, 2, \dots, 503$, whereby one would have available a sample of 503 independent maxima. On the other hand, applying the POT method with the selected high threshold $t = 7$ kWh, we are left with fewer data points and more importantly with several observations originating from the same household. In this case, we find the observations fairly independent only because they are one week apart. We highlight that the POT method implies that some households are naturally discarded, which we find an important caveat to the POT-method, a method that has heralding the efficient use the available extreme data.

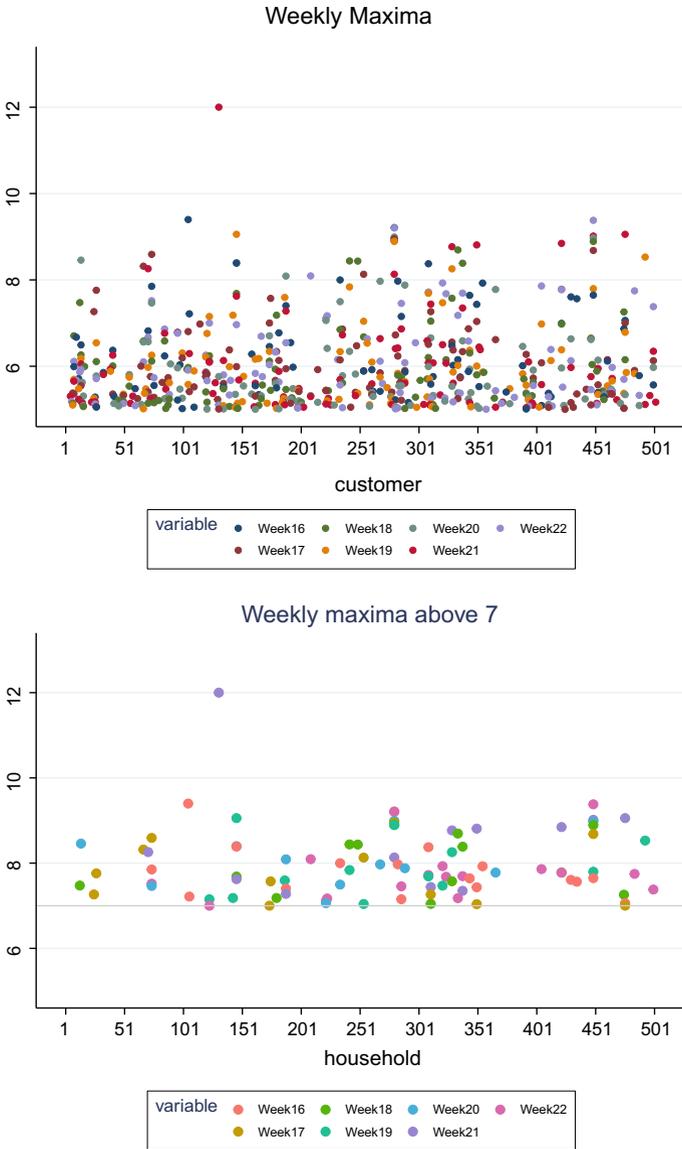


Fig. 4.2 Weekly maxima and exceedances above the threshold 7 kWh for the Iris smart meter data

4.1 Block Maxima and Peaks over Threshold Methods

Due to their nature, semi-parametric models, are never specified in detail by hand. Instead, the only assumption made is that F is in the domain of attraction of an extreme value distribution, i.e. $F \in \mathcal{D}(G_\gamma)$. In order to better understand what we mean by inference in extreme domains of attraction, let us remind ourselves of the well-known condition of extended regular variation [3, 6, 7], introduced in Chap. 3, as tantamount to the domain of attraction condition. Notably, $F \in \mathcal{D}(G_\gamma)$ if and only if there exists a positive measurable function a such that the pertaining tail quantile function $U \in ERV_\gamma$. The limit in (3.10) coincides with the U -function of the GPD, with distribution function $1 + \log G_\gamma$, which justifies the usual inference on the excesses above a high threshold ascribed to the POT method. In fact, the extreme value condition (3.10) on the tail quantile function U is the usual assumption in semi-parametric inference for extreme outcomes. We shall develop this aspect further in Sect. 4.3. In the next Sect. 4.2 we will start off with yet another equivalent extreme value condition to the extended regular variation of U that is provided in [8] for dealing with block length and/or block number as opposed to looking at a certain of upper order statistics above a sufficiently high (random) threshold. Let V be the left generalised inverse of $-1/\log F$, i.e. $V(-1/\log(1-t)) = F^\leftarrow(1-t)$. In other words, $V(t) = F^\leftarrow(e^{-1/t})$, for $0 \leq t < 1$, which conveys standardisation to the standard Fréchet. It is straightforward to see that the df F underlying the random sample (X_1, \dots, X_n) belongs to some max-domain of attraction if and only if there exist functions a and b , as defined in Theorem 3.2, such that

$$\lim_{t \rightarrow \infty} \frac{V(tx) - b(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{4.1}$$

for all $x > 0$. In contrast to the previous case of associating relation (3.1) with (3.10), there is now an asymptotically negligible factor creeping in when substituting $b(t)$ with $V(t)$. This is where we focus next, as this factor reflects the bias stemming from absorbing b (or V) into the location parameter of the GEV limit distribution (see 3.2). The common understanding is that such bias is somewhat difficult to control, but we will have a closer look at this in terms of the second order refinements. The theoretical development for working out the order of convergence in Eq. (4.1) and (3.10) in tandem is given in Proposition 4.1. For the proof, we refer the reader to [9].

Proposition 4.1 *Assume condition (3.10) (i.e. $F \in \mathcal{D}(G_\gamma)$) and that U is of extended regular variation of second order, that is, there exists a positive or negative function A with $\lim_{t \rightarrow \infty} A(t) = 0$ and a non-positive parameter ρ , such that for $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{A(t)} = \frac{1}{\rho} \left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right) =: H_{\gamma,\rho}(x). \tag{4.2}$$

Define

$$\tilde{A}(t) := \begin{cases} \frac{1-\gamma}{2}t^{-1}, & \gamma \neq 1, \rho < -1, \\ A(t) + \frac{1-\gamma}{2}t^{-1}, & \gamma \neq 1, \rho = -1, \\ A(t), & \rho > -1 \text{ or } (\gamma = 1, \rho > -2), \\ A(t) + \frac{1}{12}t^{-2}, & \gamma = 1, \rho = -2, \\ \frac{1}{12}t^{-2}, & \gamma = 1, \rho < -2. \end{cases}$$

If \tilde{A} is either a positive or negative function near infinity, then with

$$\tilde{a}(t) := \begin{cases} a(t)(1 + \frac{\gamma-1}{2}t^{-1}), & \gamma \neq 1, \rho \leq -1, \\ a(t), & \rho > -1 \text{ or } (\gamma = 1, \rho > -2), \\ a(t)(1 - \frac{1}{12}t^{-2}), & \gamma = 1, \rho \leq -2, \end{cases}$$

the following second order condition holds

$$\lim_{t \rightarrow \infty} \frac{\frac{V(tx) - V(t)}{\tilde{a}(t)} - \frac{x^\gamma - 1}{\gamma}}{\tilde{A}(t)} = H_{\gamma, \tilde{\rho}}(x), \quad (4.3)$$

for all $x > 0$, where $\tilde{\rho} = \max(\rho, -1)$ if $\gamma \neq 1$, and $\tilde{\rho} = \max(\rho, -2)$ if $\gamma = 1$.

We now provide four examples of application of Proposition 4.1 alongside further details as to how the prominent GPD can, at a first glance, escape the grasp of this proposition.

Example 4.1

Burr(1, τ , λ). This example develops along similar lines to the proof of Proposition 4.1. The Burr distribution, with d.f. $1 - (1 + x^\tau)^{-\lambda}$, $x \geq 0$, $\lambda, \tau > 0$, provides a very flexible model which mirrors well the GEV behaviour in the limit of linearly normalised maxima, also allowing a wide scope for tweaking the order of convergence through changes in the parameter λ . The associated tail quantile function is $U(t) = (t^{1/\lambda} - 1)^{1/\tau}$, $t \geq 1$. Upon Taylor's expansion of U , the extreme tail condition up to second order (Eq. 4.2) arises:

$$U(tx) - U(t) = \frac{t^{1/\lambda}}{\lambda\tau} \left[\frac{x^{1/\lambda} - 1}{1/\lambda} - \lambda t^{-1/\lambda} (x^{1/\lambda} (x^{1/\lambda} - 1)^{-1} - 1) + o(t^{-1/\lambda}) \right],$$

as $t \rightarrow \infty$. Whence, the second order condition on the tail given in Eq. (4.2) holds for $\gamma = 1/(\lambda\tau)$ and $\rho = -1/\lambda$, $\gamma + \rho \neq 0$, with

$$a(t) = \frac{t^{1/\lambda}}{\lambda\tau} \left(1 - \left(\frac{1}{\tau} - 1 \right) t^{-1/\lambda} \right) \text{ and } A(t) = \frac{1}{\lambda} \left(\frac{1}{\tau} - 1 \right) t^{-1/\lambda} = (\gamma + \rho)t^\rho.$$

Proposition 4.1 is clearly applicable and therefore the Burr distribution satisfies the extreme value condition of second order (Eq.4.3) with $\gamma = 1/(\gamma\tau)$ and $\tilde{\rho} = \max(-1/\lambda, -1)$ if $\tau \neq 1$.

Example 4.2

Cauchy. The relevant d.f. is $F(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$, $x \in \mathbb{R}$. The corresponding tail quantile function is $U(t) = \tan(\pi/2 - \pi/t) = t/\pi - \pi/3 t^{-1} + O(t^{-3})$, as $t \rightarrow \infty$, and admits the representation $U(tx) - U(t) = \frac{t}{\pi} [x - 1 - \frac{\pi^2}{3} t^{-2} (x^{-1} - 1) + O(t^{-4})]$, $x > 0$. Hence, we have that $\gamma = 1$, $\rho = -2$ in Eq.(4.2) with auxiliary function $a(t) = t/\pi$. Proposition (4.1) thus ascertains that (Eq.4.3) also holds true for the Cauchy distribution where $\gamma = 1$ and $\tilde{\rho} = -2$.

Example 4.3

GPD(γ). The relevant d.f. is $W_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, for all x such that $1 + \gamma x > 0$. The pertaining tail quantile function is $U(t) = (t^\gamma - 1)/\gamma$ which is also born out of the exact tail condition (3.10). Clearly, U does not satisfy the second order condition (Eq.4.2) in a straightforward fashion, however we are going to show that the corresponding $V(t) = U(1/(1 - e^{-1/t}))$ satisfies (Eq.4.3). To this end, we shall deal with the cases $\gamma = 1$ and $\gamma \neq 1$ separately.

Case $\gamma = 1$: Applying Laurent series expansion upon $(1 - e^{-1/t})^{-1}$, we get

$$V(tx) - V(t) = \left(1 - \frac{1}{12t}\right)(x - 1) + \frac{2}{12t} H_{1,-2}(x) + O(t^{-3}),$$

as $t \rightarrow \infty$. Whence, the second order condition (Eq.4.3) holds with $\gamma = 1$ and $\tilde{\rho} = -2$, where $\tilde{A}(t) = t^{-2}/6$ and $\tilde{a}(t) = t(1 + \tilde{A}(t)/\tilde{\rho})$.

Case $\gamma \neq 1$: Upon Taylor's expansion around zero, we obtain

$$V(tx) - V(t) = t^\gamma \left[\left(1 + \frac{\gamma - 1}{2t}\right) \frac{x^\gamma - 1}{\gamma} - \frac{\gamma - 1}{2t} H_{\gamma,-1}(x) \right] + O(t^{-3}),$$

as $t \rightarrow \infty$. Whence, the second order condition (Eq.4.3) holds with $\rho = -1$, where $t\tilde{A}(t) = (1 - \gamma)/2$ and $\tilde{a}(t) = t^\gamma(1 + \tilde{A}(t)/\tilde{\rho})$.

Therefore, the GPD verifies Proposition 4.1 if one tunnels through the consideration that the GDP satisfies (Eq.4.2) with $\rho = -\infty$.

Example 4.4

Pareto(α). This distribution is a particular case of the GPD d.f. in Example 4.1 with $\gamma = 1/\alpha > 0$ and $U(t) = t^{1/\alpha}$, that is U does not satisfy the second order condition (Eq. 4.2) and thus Proposition 4.1 stands applicable provided similar interpretation to Example 4.1.

Example 4.5

Contaminated Pareto(α). We now consider the Pareto distribution with a light contamination in the tail by a slowly varying function $L(t) = (1 + \log t)$, that is, $L(tx)/L(t) \rightarrow 1$, as $t \rightarrow \infty$, for all $x > 0$. This gives rises to the quantile function $U(t) = t^{1/\alpha}(1 + \log t)$, with $\alpha > 0$. For the sake of simplicity, we shall use the identification $\gamma = 1/\alpha$. With some rearrangement, we can write the spacing $U(tx) - U(t)$ in such a way that the first and second order parameters in condition (Eq. 4.2), both γ and $\rho \leq 0$, crops up: $U(tx) - U(t) = \gamma t^\gamma (\log t + 1) \left[\left(1 + \frac{1}{\gamma \log t + 1}\right) \frac{x^\gamma - 1}{\gamma} + \frac{1}{1 + \log t} H_{\gamma,0}(x) \right]$, where $H_{\gamma,0}(x) := \frac{1}{\gamma} (x^\gamma \log x - \frac{x^\gamma - 1}{\gamma})$. Note that we have provided an exact equality, i.e. there is no error term. We thus find that tampering with the Pareto distribution, by contaminating its tail-related values with a slowly varying factor, is just enough bring the convergence (Eq. 4.2) to a halt which is flagged-up by the lowest possible $\rho = 0$. This stalling of the Pareto distribution enables to fulfill the conditions in Proposition (4.1) thus ensuring that this contaminated Pareto distribution belongs to the max-domain of attraction of the GEV distribution with $\gamma = 1/\alpha > 0$ and $\tilde{\rho} = 0$.

4.2 Maximum Lq-Likelihood Estimation with the BM Method

Let us define the random sample consisting of k i.i.d. block maxima as

$$M_i = \max_{(i-1)m < j \leq im} X_j, \quad i = 1, 2, \dots, k, \quad m = 1, 2, \dots \quad (4.4)$$

The above essentially states that we are dividing the whole sample of size n into k blocks of equal length (time) m . For the Extreme Value theorem to hold within each block, the block length must be sufficiently large, i.e. one needs to impose m tending to infinity to able to proceed with inference. We are then led to the reasonable assumption that the sample of k -maxima behaves approximately as though it stems from the GEV.

Under a semi-parametric approach, maximum likelihood estimators for the vector-valued parameter $\theta = (\mu, \sigma, \gamma)$ are obtained by pretending (which is approximately

true) that the random variables M_1, M_2, \dots, M_k are independent and identically distributed as maxima of the GEV distribution with d.f. given by

$$G_\theta(x) = \exp\left\{-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right\},$$

for those x such that $\sigma + \gamma(x - \mu) > 0$. The density of the parametric fit to the BM framework is the GEV density, which we denote by g_θ , may differ slightly from the true unknown p.d.f. f underlying the sampled data. We typically estimate these constants $a(m)$ and $b(m)$ via maximum likelihood, despite these being absorbed into the scale $\sigma > 0$ and location $\mu \in \Theta$ parameters of the parametric limiting distribution thus assumed fixed, eventually. As a result, BM-type estimators are not so accurate for small block sizes since these estimators must rely on blocks of reasonable length to fulfill the extreme value theorem.

The criterion function ℓ_q that gives rise to a maximum Lq-likelihood (MLq) estimator,

$$\tilde{\theta} := \arg \max_{\theta \in \Theta} \sum_{i=1}^k \ell_q(g_\theta(x_i)),$$

for $q \geq 0$, makes use of the Tsallis deformed logarithm as follows:

$$\tilde{\theta} = \arg \max_{\theta \in \Theta} \sum_{i=1}^k \frac{(g_\theta(x_i))^{1-q} - 1}{1 - q}, \quad q \geq 0, \tag{4.5}$$

This MLq estimation method picks up the standard maximum likelihood estimator (MLE) if one sets the distortion parameter $q = 1$. This line of reasoning can be stretched on to a continuous path, that is, as q tends to 1, the MLq estimator approaches the usual MLE. The common understanding is that values of q closer to one are preferable when we have numerous maxima drawn from large blocks since this will give enough scope for the EVT to be accessible and applicable. In practice, we often encounter limited sample sizes $n = m \times k$ in the sense that either a small number of extremes (k sample maxima) or blocks of insufficient length m to contain even one extreme are available. MLq estimators have been recognised as particularly useful in dealing with small sample sizes, which is often the situation in the context of the analysis of extreme values due to the inherent scarcity of extreme events with catastrophic impact. Previous research by [10, 11] shows that the main contribution towards the relative decrease in the mean squared error stems from the variance reduction, which is the operative statement in small sample estimation. This is in contrast with the bias reduction often sought after in connection with large sample inference. Large enough samples tend to yield stable and smooth trajectories in the estimates-paths, allowing scope for bias to set in, and eventually reflecting the regularity conditions in the maximum likelihood sense. Once these asymptotic conditions are attained, the dominant component of the bias starts to emerge, and by then it can be efficiently removed. This is often implemented at the expense of

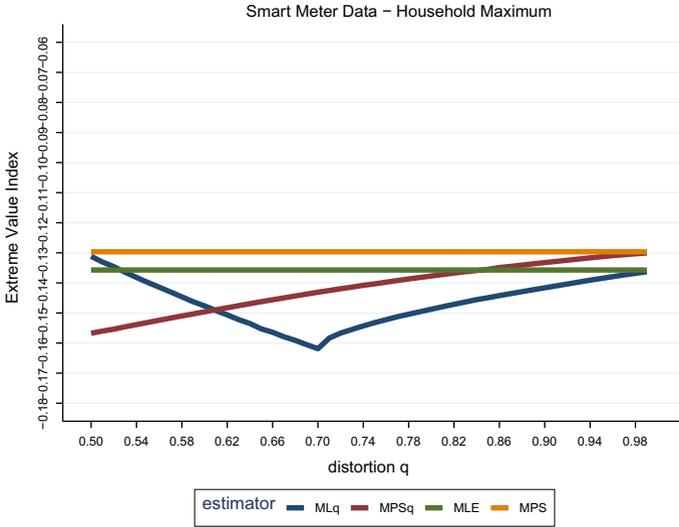


Fig. 4.3 Estimation of the EVI with the BM method

an increased variance, for the usual bias/variance trade off in statistics seems never to offer anything worthwhile on one side without also providing a detriment to the other.

Furthermore, we will address how the maximum likelihood compares with the maximum product of spacings (MPS) estimator in this case study. The MPS estimator of θ maximises the product of spacings

$$\prod_{i=1}^{k+1} D_i(\theta) = \prod_{i=1}^{k+1} \left\{ G_\theta(x_{i,k}) - G_\theta(x_{i-1,k}) \right\},$$

with $G_\theta(x_{0,k}) = 0$ and $G_\theta(x_{k+1,k}) = 1$, or equivalently the log-spacings

$$L^{MPS}(\theta; \mathbf{x}) = \sum_{i=1}^{k+1} \log D_i(\theta). \tag{4.6}$$

The MPS method was introduced by [12], and independently by [13]. A generalisation of this method is proposed and studied in great depth by [14]. The MPS method was further exploited by [15] in estimating and testing for the only possible three types of extreme value distributions (Fréchet, Gumbel and Weibull), all unified in the GEV distribution.

Figure 4.3 displays the sample paths of the adopted extreme value index estimators, plotted against several values of the distortion parameter $q \in [0.5, 1]$. As q increases to 1, the deformed version of both ML and MPS estimators approach their

classical counterpart which stem from stipulating the natural logarithm as criterion function. The estimates for the EVI seem to consolidate between -0.14 and -0.13 . The negative estimates obtained for all values of q provide evidence that the true d.f. F generating the data belongs to the Weibull domain of attraction and therefore we can reasonable conclude that we are in the presence of a short tail with finite right endpoint. The next section concerns the estimation of this *ultimate return-level*.

4.2.1 Upper Endpoint Estimation

A class of estimators for the upper endpoint x^F stems from the extreme value condition (Eq. 4.1) via the approximation $V(\infty) \approx V(m) - a_m/\gamma$, as $m \rightarrow \infty$, by noticing that $V(\infty) = \lim_{t \rightarrow \infty} V(t) = F^{\leftarrow}(1) = x^F$. The existing finite right endpoint x^F can be viewed as the ultimate return level. When estimating extreme characteristics of this sort, we are required to replace all the *unknowns* in the above by their empirical analogues, yielding the estimator for the right endpoint:

$$\hat{x}^F := \hat{V}(m) - \frac{\hat{a}(m)}{\hat{\gamma}}, \tag{4.7}$$

where quantities \hat{a} , \hat{V} and $\hat{\gamma}$ stand for the MLq estimators for the scale and location functions $a(m)$ and $V(m)$, and for the EVI, respectively.

Figure 4.4 displays the endpoint estimates for several values of $q \leq 1$ with respect to the tilted version of both ML and MPS estimators. The latter always finds larger estimates than the former, with a stark distance from the observed overall maximum.

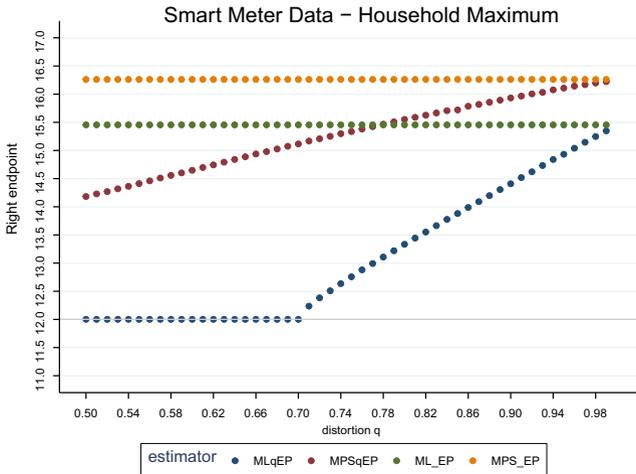


Fig. 4.4 Estimation of the upper endpoint through the BM method

Since the adopted estimators do not seem to herd towards one value, it is not easy to conciliate between them. Given the maximum likelihood estimator has been widely used and extensively studied in the literature, it is sensible to ascribe preference to this estimator. Furthermore, since we are dealing with small sample sizes (we are taking the maximum over 7 weeks), the distorted version, i.e., the MLq estimator must be taken into account. Therefore, we find reasonable to take as estimate for the upper bound the midpoint of the region where the MLq changes way and travels across towards the plain ML estimator with increasing q . Thus, we get an upper bound of 14.0 kWh, approximately.

4.3 Estimating and Testing with the POT Method

In Sect. 4.1, we noticed that the function appearing in the limit of the extended regular variation of U matches the tail quantile function of the Generalised Pareto distribution. This fact reflects indeed the exceptional role of the GPD in the extreme value theory for exceedances [16, 17] and prompts the need for classifying of the tails of all possible distributions in $\mathcal{D}(G_\gamma)$ into three classes in accordance to the sign of the extreme value index γ . For positive γ , the power-law behaviour of the underlying tail distribution function $1 - F$ has important implications one of which being the presence of infinite moments. Because when $\gamma > 0$ the first order condition (3.10) can be rephrased as $\lim_{t \rightarrow \infty} U(tx)/U(t) = x^\gamma$, for all $x > 0$, that is, U is γ -regularly varying at infinity (notation: $U \in RV_\gamma$), then Karamata's theorem ascertains that $E(X_1^+)^p$ is infinite for $p > 1/\gamma$, where $X_1^+ = \max(0, X_1)$. Hence, heavy-tailed distributions in a max-domain of attraction not only have an infinite right endpoint, but also the order of finite moments is determined by order of the magnitude of the EVI $\gamma > 0$. The Fréchet domain of attraction contains distributions with polynomially decay tails such as the Pareto, Cauchy, Student's and Fréchet itself. All d.f.'s belonging to $\mathcal{D}(G_\gamma)$ with $\gamma < 0$ —Weibull domain of attraction—are light tailed distributions with finite right endpoint. Such domain of attraction encloses Uniform and Beta distributions. The intermediate case $\gamma = 0$ is of particular interest in many applied sciences where extremes are relevant. At a first glance, the Gumbel domain of attraction seems quite appealing for the simplicity of inference in connection to G_0 that dispenses the estimation of γ . But a closer inspection allows to appreciate the great variety of distributions possessing such an exponential tail related behaviour, whether having finite upper endpoint or not. Normal, Gamma and Lognormal distributions can all be found in Gumbel domain. The negative Fréchet distribution also belongs to the Gumbel domain, albeit with finite endpoint (cf. [18]). Therefore, a statistical test for assessing significance of the EVI would be of great use and most consequential. Looking for the most propitious type of tail before estimating tail-related features of the distribution underlying the data can mark the difference between aiming at the estimation of extreme quantile or sprinting to the estimation of the upper endpoint estimation. In fact, adopting tailored statistical methodology to the suitable domain of attraction for the underlying d.f. F has become a regular practice.

4.3.1 Selection of the Max-Domain of Attraction

A test for Gumbel domain *versus* Fréchet or Weibul max-domains has received in the literature the general designation of statistical choice of extreme domains of attraction. References in this respect are [19–29].

The present section primarily deals with the two-sided problem of testing Gumbel domain against Fréchet or Weibull domains, i.e.,

$$F \in \mathcal{D}(G_0) \quad \text{vs} \quad F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0}. \quad (4.8)$$

Bearing on the sample maximum as the dominant and most informative statistic in any analysis of extreme value, we shall consider the ratio statistic

$$T_n^*(k) := T_n(k) - \log k = \frac{X_{n,n} - X_{n-k,n}}{\frac{1}{k} \sum_{i=1}^k (X_{n-i+1,n} - X_{n-k,n})} - \log k, \quad (4.9)$$

for the testing problem in Eq.(4.8). According to the asymptotic results stated in [27], the null hypothesis $F \in \mathcal{D}(G_0)$ is rejected, against the bilateral alternative $F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0}$, at an asymptotic level $\alpha \in (0, 1)$ if $T_n^*(k) < g_{\alpha/2}$ or $T_n^*(k) > g_{1-\alpha/2}$, where g_ε denotes the ε -quantile of the Gumbel distribution, i.e., $g_\varepsilon = -\log(-\log \varepsilon)$. One sided tests are also within reach of this test statistic. We reject the null hypothesis in favor of either unilateral alternatives $H_1' : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}$ or $H_1'' : F \in \mathcal{D}(G_\gamma)_{\gamma > 0}$ on either side $T_n^*(k) < g_\alpha$ or $T_n^*(k) > g_{1-\alpha}$, respectively. Neves et al. [27] show that this statistic provides a consistent test to discriminate between light tails and heavy tails, departing from the null hypothesis of an exponential tail.

Built on Shapiro–Wilk goodness-of-fit statistic, the well-known Hasofer and Wang test statistic, embodies the reciprocal squared empirical coefficient of variation associated with the sample of the excesses above the random threshold $X_{n-k,n}$. More concretely,

$$W_n(k) := \frac{1}{k} \frac{\left(k^{-1} \sum_{i=1}^k Z_i\right)^2}{k^{-1} \sum_{i=1}^k Z_i^2 - \left(k^{-1} \sum_{i=1}^k Z_i\right)^2}, \quad (4.10)$$

where $Z_i := X_{n-i+1,n} - X_{n-k,n}$, $i = 1, \dots, k$. Giving heed to the problem of the statistical choice of domains of attraction postulated in Eq.(4.8), we define for $j = 1, 2$

$$N_{n,k}^{(j)} := \frac{1}{k} \sum_{i=1}^k (X_{n-i+1,n} - X_{n-k,n})^j \quad (4.11)$$

and use it to write Hasofer and Wang, $W_n(k)$, and Greenwood $R_n(k)$ statistics in the following way:

$$R_n(k) = \frac{N_n^{(2)}(k)}{(N_n^{(1)}(k))^2}, \quad (4.12)$$

$$W_n(k) = \frac{1}{k} \left[1 - \frac{R_n(k) - 2}{1 + (R_n(k) - 2)} \right]. \quad (4.13)$$

Considering, as before, the k upper order statistics from a sample of size n such that $k = k_n$ is an intermediate sequence, i.e., $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$, define

$$R_n^*(k) := \sqrt{k/4} (R_n(k) - 2) \quad (4.14)$$

$$W_n^*(k) := \sqrt{k/4} (kW_n(k) - 1). \quad (4.15)$$

These normalized versions, $R_n^*(k)$ and $W_n^*(k)$, are eventually the main features to take part in the testing procedure. The critical region for the two-sided test of nominal size α is given by $|T_n^*(k)| > z_{1-\alpha/2}$, with z_ε denoting the ε -quantile of the standard normal distribution and where T has to be conveniently replaced by R or W .

In addition, one-sided testing problems

$$F \in \mathcal{D}(G_0) \quad \text{vs} \quad F \in \mathcal{D}(G_\gamma)_{\gamma < 0} \quad (\text{or } F \in \mathcal{D}(G_\gamma)_{\gamma > 0}),$$

can also be tackled with both these test statistics. Here, the null hypothesis is rejected in favor of either unilateral alternatives $H_1' : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}$ or $H_1'' : F \in \mathcal{D}(G_\gamma)_{\gamma > 0}$ on either side $T_n^*(k) < z_\alpha$ or $T_n^*(k) > z_{1-\alpha}$, respectively, with z_ε denoting again the ε -quantile of the standard normal distribution and whereas T has to be conveniently replaced by R or W .

We remark that the test based on the very simple Greenwood-type statistic R^* is shown to good advantage when testing the presence of heavy-tailed distributions in demand. While the R^* -based test barely detects small negative values of γ , the Hasofer and Wang's is the most powerful test under study concerning alternatives in the Weibull domain of attraction. The three testing procedures will be illustrated with the Iris smart meter data in the next section.

4.3.2 Testing for a Finite Upper Endpoint

The aim of this section is to assess finiteness in the right endpoint of the actual d.f. F underlying the Irish smart meter data. The basic assumption is that F belongs to some max-domain of attraction. We then consider the usual asymptotic setting, where assume an intermediate number k of order statistics to draw inference upon, that is, we take $k = k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$, as $n \rightarrow \infty$, and hence the corresponding random threshold $X_{n-k,n} \rightarrow x^F$ a.s..

The statistical judgment on whether a finite upper bound exists finite will be informed by the testing procedure introduced in [30]. Notably, the testing problem

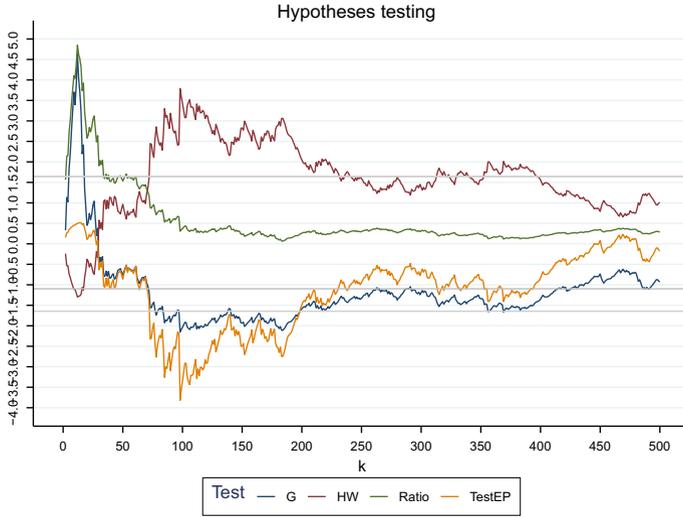


Fig. 4.5 Statistical choice of max-domain of attraction and detection of finite endpoint

$$H_0 : F \in \mathcal{D}(G_0), x^F = \infty \quad \text{vs} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \leq 0}, x^F < \infty$$

can be tackled using the log-moments

$$M_{n,k}^{(j)} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j, \quad j = 1, 2. \quad (4.16)$$

The test statistic T_1 being used is defined as

$$T_1 := \frac{1}{k} \sum_{i=1}^k \frac{X_{n-i,n} - X_{n-k,n} - T}{X_{n,n} - X_{n-k,n}}, \quad \text{with } T := X_{n-k,n} \frac{M_1}{2} \left(1 - \frac{[M_1]^2}{M_2} \right)^{-1}.$$

Under H_0 , the standardised version of the test, $T_1^* := \sqrt{k} \log k T_1$, is asymptotically normal. Moreover, T_1^* tends to inflect to the left for bounded tails in the Weibull domain and to the right if the underlying distribution belongs to the Gumbel domain. The rejection region of the test is given by $|T_1^*| \geq z_{1-\alpha/2}$, for an approximate α significance level. Figure 4.5 displays the sample path of T_1^* , labeled TestEP, alongside the observed values of the three test for selecting max domain of attraction presented in Sect. 4.3.1. The horizontal grey lines mark the critical barriers for the one-sided test at a $\alpha = 5\%$ significance level. When these critical bounds are crossed, the null hypothesis of the Gumbel domain is rejected in favour of the Weibull domain. The statement for the testing problem on the upper endpoint is slightly different as it entails a different breakdown of the Gumbel domain. The choice of the optimal number k of

intermediate order statistics is of paramount importance to any inference problem in extremes. Many methods have been proposed, but sadly there is not universal solution that can hold for the multitude of estimators and testing procedures available. Here, we loosely follow the guideline of [23] in that the most adequate choice of the intermediate number k (which carries over to the subsequent semi-parametric inference) should set on the lowest k at which the critical barriers are overtaken. The ratio test (Eq. 4.9), which is known to be the most conservative of all three tests for choice of domains, does not reject the null hypothesis of the Gumbel domain since the green trajectory remains above the second horizontal line from below, for all intermediate values of k considered. We have remarked that the Hasofer and Wang (HW) test defined in Eq. (4.13) is the most powerful test for detecting distributions in the Weibull domain of attraction. The application of the HW test seems to do justice to this Irish smart meter data set and finds sufficient evidence to reject the null hypothesis of the Gumbel domain in favour of entertaining estimation procedures suited to the Weibull domain. Therefore we will proceed on to the estimation of the finite upper bound in the POT framework. This will be tackled in the next section.

4.3.3 Upper Endpoint Estimation

Along similar lines to Sect. 4.2, a valid estimator for the upper endpoint $x^F = U(\infty)$ arises by making $t = n/k$ in the approximate equality corresponding to (3.10), and then replacing $U(n/k)$, $a(n/k)$ and γ by suitable consistent estimators, i.e.

$$\hat{x}^F = \hat{U}\left(\frac{n}{k}\right) - \frac{\hat{a}\left(\frac{n}{k}\right)}{\hat{\gamma}}$$

(cf. Sect. 4.5 of [3]). Typically we consider the semiparametric class of endpoint estimators as follows:

$$\hat{x}^F = X_{n-k,n} - \frac{\hat{a}(n/k)}{\hat{\gamma}}. \quad (4.17)$$

For the application of EVT to the class (Eq. 4.17) of upper endpoint estimators it is thus necessary to estimate the parameter γ . We mention the estimators: Pickands' estimator [17] for $\gamma \in \mathbb{R}$; the so-called ML estimator [31], valid for $\gamma > -1/2$; the moment estimator [32] and the mixed moment estimator, both valid for any $\gamma \in \mathbb{R}$. These moment estimators are purely semiparametric estimators for they are devised upon conditions of regular variation underpinning the max-domain of attraction characterisation. Since these are rather qualitative conditions, with functions U and a specific to the underlying d.f. F , then inference must be built on summary statistics that can capture the most distinctive traits in tail-related observations. The method of moments is ideally suited to this purpose.

In order to develop a novel estimator for the extreme value index $\gamma \in \mathbb{R}$, [33] considered a combination of Theorems 2.6.1 and 2.6.2 of [7] and went on with replacing

F with its empirical counterpart F_n and t by the order statistic $X_{n-k,n}$ with $k < n$. This led to the statistic

$$\hat{\varphi}_n(k) := \frac{M_n^{(1)}(k) - L_n^{(1)}(k)}{(L_n^{(1)}(k))^2}, \quad (4.18)$$

where we define, for $j = 1, 2$,

$$L_n^{(j)}(k) := \frac{1}{k} \sum_{i=0}^{k-1} \left(1 - \frac{X_{n-k,n}}{X_{n-i,n}}\right)^j \quad (4.19)$$

and with $M_n^{(j)}(k)$ given in Eq. (4.16). The statistic in (Eq. 4.18) is easily transformed into the so-called mixed moment estimator (MM) for the extreme value index $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_n^{MM}(k) := \frac{\hat{\varphi}_n(k) - 1}{1 + 2 \min(\hat{\varphi}_n(k) - 1, 0)}. \quad (4.20)$$

The most attractive features of this estimator are:

- like Pickands and Moment estimators, it is valid for any $\gamma \in \mathbb{R}$ and, contrary to the maximum likelihood estimator, it has a simple explicit functional form;
- it is very close to the maximum likelihood estimator for $\gamma \geq 0$;
- If $\gamma \leq 0$, the asymptotic variance is the same as of the Moment estimator; if $\gamma > 0$, its asymptotic variance is equal to that of the maximum likelihood estimator;
- A shift invariant version with similar properties is available with the same asymptotic variance without sacrificing the dominant component of the bias, which is never increased as long as we keep a suitable number k ;
- there are accompanying shift and scale estimators that make e.g. high quantile and upper endpoint estimation straightforward.

Figure 4.6 shows the sample paths of several estimators of the EVI as the upper number k of order statistics embedded in the estimators increases, concomitantly lowering the threshold until the value 5 khW is reached. The standard practice for drawing meaningful conclusions from this type of plots is by eyeballing the trajectories and seek for a plateau of stability at the confluence of the adopted estimators. In the top panel of Fig. 4.6, the MLq estimator of the extreme value index γ , which has no explicit closed form and thus delivers estimates numerically, experiences convergence issues. This is often the case with maximum likelihood estimation for the GPD when the true value is negative but close to zero.

In the semi-parametric setting, whilst working in the domain of attraction rather than dealing with the limiting distribution itself, the upper intermediate order statistic $X_{n-k,n}$ plays the role of the high deterministic threshold $u \uparrow x^F \leq \infty$ above which the parametric fit to the GPD is applicable. For the asymptotic properties of the POT maximum likelihood estimator of the EVI under a semi-parametric approach, see e.g. [34–36]. Although theoretically well determined, even when $\gamma \uparrow 0$, the non-convergence to a ML-solution can be an issue when γ is close to zero. There are also

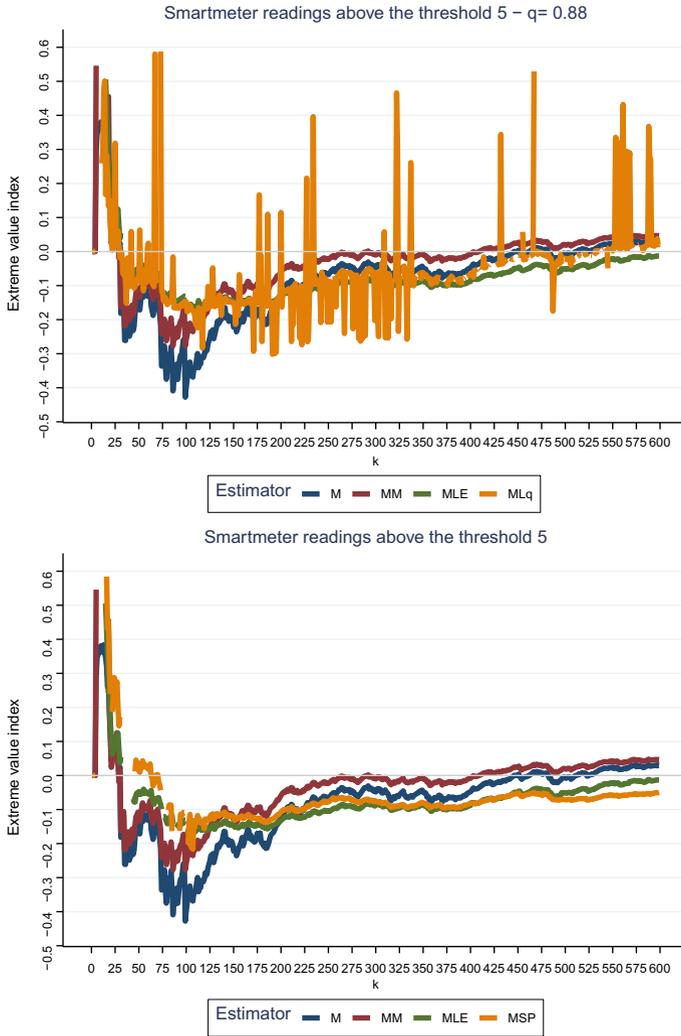


Fig. 4.6 Estimation of the EVI with the POT method

irregular cases which may compromise the practical applicability of ML. Theoretical and numerical accounts of these issues can be found in [37, 38] and references therein.

In the second panel in Fig. 4.6, we swap the MLq estimator with the MPS (or MSP) estimator for the GPD. Although there are issues in the numerical convergence for small values of k , where the variance is stronger, this estimator shows enhanced behavior returning estimates of the EVI in agreement with the remainder estimators. Therefore, it seems reasonable to settle with the point estimate $\hat{\gamma} = -0.01$. It is worth highlighting that the MLq shows its best performance within the corresponding region

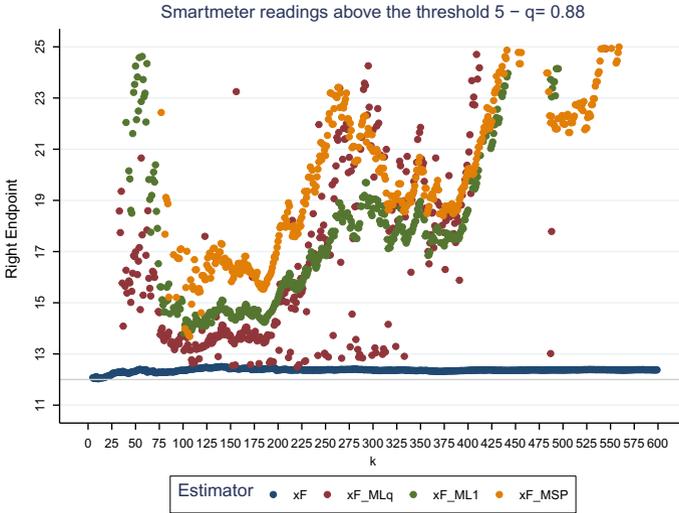


Fig. 4.7 Estimation of the upper endpoint with the POT method

of values of, that is, for k between 125 and 175, a region that also holds feasible for the tests expounded in Sect. 4.3.1.

There is however one estimator for the upper endpoint x^F that does not depend on the estimation of the EVI γ , worked out in [18], this being designed for distributions with finite upper endpoint enclosed in the Gumbel domain of attraction. The so-called general right endpoint estimator is defined as

$$\hat{x}^F := X_{n,n} + X_{n-k,n} - \frac{1}{\log 2} \sum_{i=0}^{k-1} \log\left(1 + \frac{1}{k+i}\right) X_{n-k-i,n}. \quad (4.21)$$

Figure 4.7 displays the estimate-yields for several endpoint estimators in the class (Eq. 4.17), with accompanying general endpoint estimator. Again, the corresponding maximum Lq-likelihood estimator is found with the distortion parameter q being set equal to 0.88, where the k -values for which the Lq-likelihood method experienced convergence issues in the estimation of the EVI are now omitted. The value $q = 1$ determines the mainstream ML estimator for the endpoint which the class of estimators defined in Eq. (4.17) also encompasses. The relative finite sample performance of these endpoint estimators is here compared with the naïve maximum estimator $X_{n,n}$. We recall that the observed maximum is 12.01. The general endpoint estimator consistently returns values around 12.4 for almost all values of k . All the other estimators, expect the MLq estimator for the upper endpoint seem to exacerbate the upper bound for the electricity load. Therefore, we find reasonable to advise that the estimate for the upper endpoint of $\hat{x}^F = 13.0$ kWh should be taken as benchmark for

future work concerning technologies to enable end use energy demand. We also issue the cautionary remark that this is a mere point estimate, disposed of any confidence bounds.

4.4 Non-identically Distributed Observations—Scedasis Function

Thus far we have assume that our data consists of observations of i.i.d random variables X_1, \dots, X_n are i.i.d random variables. In reality, this may not be the case. Thus far in this chapter, we have been treating the Irish smart meter data as if these comprise independent and identically distributed observations, but intuitively at least one of these assumptions may not hold. Despite the ways we are sampling extremes from the original raw data, through either BM or POT approaches, inevitably though a few households may end up as main contributors to the extreme values being taken up to the statistical analysis. Figure 4.8 is a representation of the excess load per household above 7 kWh. The top panel shows the exceedances in terms of their cumulative yields over 7 weeks, whilst the bottom panel is the same scatter plot already presented in Fig. 4.2. It is noticeable that the household yielding the absolute maximum of 12.01 kWh does not show the largest total in the cumulative excess plot but two other households are sparking up totals by consistently exceeding the threshold over the course of the 7 weeks. Despite this does not shift the upper endpoint itself (as this is kept fixed by structural settlements with the energy provider), it may push the observations closer to the true upper bound signifying a trend is present in the extreme layers of the process generating the data.

Social demographics and behavioural changes are not likely to occur within the time span considered to this illustrative example, but we can see that in other applications to electricity consumption even a decade is enough time for human behaviour to be vastly different. Regulation of the gas and electricity industry, introduction of software applications that can monitor and incentivise certain consumption over others, time-of-use tariffs, new low carbon technologies and the variety of electronic devices in homes will change the way consumers interact with the grid and consume their electricity and most probably has already changed.

Nonetheless we still want to be able to address the same questions as we did before and we want to ensure that there is a probabilistic framework that grounds our statistical analysis. Our main concern lies with observations that are not identically distributed but we will include a short review for data that exhibit serial dependence.

There are of course many ways to look at dependence in data sets indexed in time or space. Ways of pre-processing data to alleviate dependence issues and possible non-stationary have been considered in [39]. Historically, however, simpler clustering techniques have been employed [40]. We already discussed apropos to the BM, to choose the length of block in such a way that the dependence no longer plays a role or is weak for the extreme value theorem to hold. Ensuring that our observation

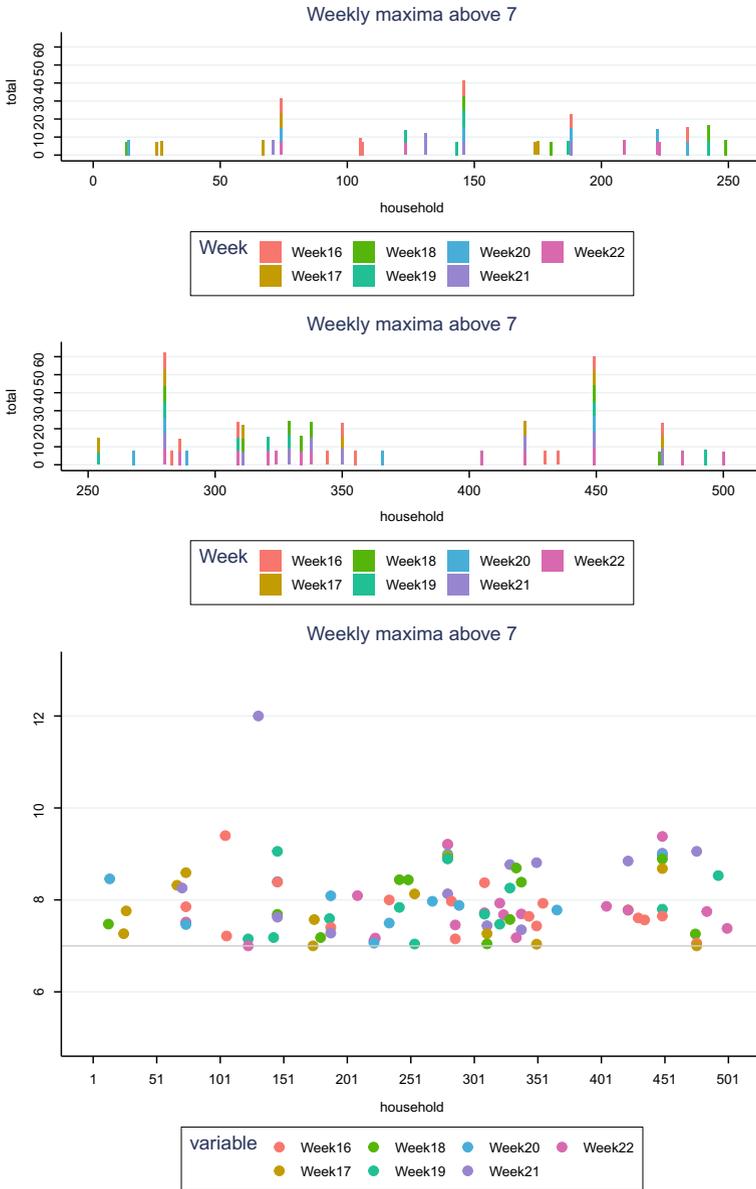


Fig. 4.8 Estimation of the upper endpoint with the POT method

come from separate events is a simple way of ascertaining independence and that the sampled data contains meaningful information on the widely different phenomena emanating from similar hazards (earthquakes, rainfall, storms). For example, if we were considering heatwaves, they may last up to a fortnight thus considering daily

maxima in temperature taken during a heatwave will be part of the same weather system and subsequently dependent and also less representative of the wild. Similarly, if we are considering rainfall, a low pressure system may last two or three days thus maxima taken every 2–3 non-overlapping days may be considered to independent or only weakly dependent. In the same vein, we may look at weekly maxima of the electric load profiles of individual household to weed out the daily and sub-weekly patterns. Thus, the block to be chosen should be application and data specific.

However, sometimes there is also temporal dependence in addition to individual events i.e. profiles change with time. For temperature data, the change comes with the season as well due to the diurnal cycle. Similarly, for electricity data there are many sources of seasonality to consider: the impact from temperature i.e., an annual cycle and the daily cycle as well as human behaviour which may repeat weekly. For example, we may restrict to only taking weekly maxima from the summer, etc. This is where we turn focus to non-stationarity extremes, meaning that the underlying distribution changes over time or across space or both. This aspect will be exploited through the definition of a trend in the frequency of extremes in such a way to maintain integrity as we move across the potentially different distribution functions ascribed to each of the considered time-intervals (or spatial-regions), assumed homogeneous within themselves and heterogeneous between them. The basic structural assumption to the trend on the time-space evolving probability that a certain large outcome is exceeded originates from the concept of comparable tails [41]. There have been estimators accounting for the heteroscedastic (non-stationarity in the scale) setting, first introduced by [42] and further developed by [43] to address challenged arising in the modelling of heavy-tails.

The setup is laid out as follows. Suppose that $X_1^{(n)}, \dots, X_n^{(n)}$ are observations from independent random variables taken at n time points, which are assumed to follow different distribution functions $F_{n,1}, \dots, F_{n,n}$ but sharing a common upper endpoint denoted by x^F . Suppose the following limit relation involving an offset or baseline distribution function F and a continuous positive function c taking values in $[0, 1]$,

$$\lim_{x \rightarrow x^F} \frac{1 - F_{n,i}(x)}{1 - F(x)} = c\left(\frac{i}{n}\right), \quad (4.22)$$

subject to the unifying condition

$$\int_0^1 c(s) d(s) = 1.$$

Doing so allows c to represent the frequency of extremes in the tail. Einmahl et al. [43] advocate a kernel density estimator as ideally suited to tackle the estimation of the scedasis function which can be viewed as a density in itself. Specifically, the estimator for $c(s)$ is given by

$$\hat{c}(s) = \frac{1}{kh} \sum_{i=1}^n I_{\{X_i^{(n)} > X_{n,n-k}\}} G\left(\frac{s - \frac{i}{n}}{h}\right), \quad (4.23)$$

where G is a continuous, symmetric kernel such that $\int_{-1}^1 G(s)ds = 1$, with $G(s) = 0$ for $|s| > 1$. The bandwidth, $h := h_n$ satisfies $h \rightarrow 0$ and $kh \rightarrow \infty$ as $n \rightarrow \infty$, and $I_A = 1$ denotes indicator function which is equal to 1 if A holds true and equal to 0 otherwise. Finally we note that $X_{n,n-k}$ is the global random threshold determined by the k th largest observation. We shall defer the estimation of the scedasis in practice to Chap. 5 using the Thames Valley Vision data.

References

1. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L.: *Statistics of Extremes: Theory and Applications*. Wiley (2004)
2. Castillo, E., Hadi, A., Balakrishnan, N., Sarabia, J.M.: *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, Hoboken, New Jersey (2005)
3. de Haan, L., Ferreira, A.: *Extreme Value Theory: An Introduction*. Springer (2006)
4. Gumbel, E.J.: *Statistics of Extremes*. Columbia University Press, New York and London (1958b)
5. Melinda Gálfi, V., Bódai, T., Lucarini, V.: Convergence of extreme value statistics in a two-layer quasi-geostrophic atmospheric model. *Complexity* (2017)
6. Bingham, N., Goldie, C., Teugels, J.: *Regular Variation*. Cambridge University Press (1987)
7. de Haan, L.: On regular variation and its application to the weak convergence of sample extremes. Ph.D. thesis, Mathematisch Centrum Amsterdam (1970)
8. Ferreira, A., de Haan, L.: On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43**(1), 276–298 (2015)
9. Jeffree, C., Neves, C.: Tilting maximum Lq-likelihood estimation for extreme values drawing on block maxima. Technical report (2018). [arXiv:1810.03319](https://arxiv.org/abs/1810.03319)
10. Ferrari, D., Yang, Y.: Maximum Lq-likelihood estimation. *Ann. Statist.* **38**(2), 753–783 (2010)
11. Ferrari, D., Paterlini, S.: The maximum Lq-likelihood method: an application to extreme quantile estimation in finance. *Methodol. Comput. Appl. Probab.* **11**, 3–19 (2009)
12. Cheng, R.C.H., Amin, N.A.K.: Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Statist. Soc. Ser. B* **45**, 394–403 (1983)
13. Ranneby, B.: The maximum spacing method. An estimation method related to the maximum likelihood method. *Scand. J. Statist.* **11**, 93–112 (1984)
14. Ekström, M.: Consistency of generalized maximum spacing estimates. *Scand. J. Statist.* **28**(2), 343–354 (2001)
15. Huang, C., Lin, J.-G.: Modified maximum spacings estimator for generalized extreme value distribution and applications in real data analysis. *Metrika* **77**, 867–894 (2013)
16. Balkema, A.A., de Haan, L.: Residual life time at great age. *Ann. Prob.* 792–804 (1974)
17. Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* 119–131 (1975)
18. Fraga Alves, I., Neves, C.: Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica* **24**, 1811–1835 (2014)
19. Galambos, J.: A statistical test for extreme value distributions. In: Gnedenko, B.V. et al. (eds.) *Non-parametric Statistical Inference*, North Holland, Amsterdam, pp. 221–230 (1982)
20. Castillo, E., Galambos, J., Sarabia, J.M.: The selection of the domain of attraction of an extreme value distribution from a set of data. In: Hüsler, J., Reiss, R.-D. (eds.) *Extreme Value Theory. Lecture Notes in Statistics*, vol. 51, pp. 181–190 (1989)
21. Hasofer, A.M., Wang, Z.: A test for extreme value domain of attraction. *J. Am. Stat. Assoc.* **87**, 171–177 (1992)
22. Fraga Alves, M.I., Gomes, M.I.: Statistical choice of extreme value domains of attraction—a comparative analysis. *Commun. Stat. Theory Methods* **25**(4), 789–811 (1997)
23. Wang, J.Z., Cooke, P., Li, S.: Determination of domains of attraction based on a sequence of maxima. *Austr. J. Stat.* **38**(2), 173–181 (1996)

24. Marohn, F.: An adaptive test for Gumbel domain of attraction. *Scand. J. Stat.* **25**(2), 311–324 (1998a)
25. Marohn, F.: Testing the gumbel hypothesis via the pot-method. *Extremes* **1**(2), 191–213 (1998b)
26. Segers, J., Teugels, J.: Testing the gumbel hypothesis by Galton's ratio. *Extremes* **3**(3), 291–303 (2000)
27. Neves, C., Picek, J., Alves, M.F.: The contribution of the maximum to the sum of excesses for testing max-domains of attraction. *J. Stat. Plan. Inference* **136**(4), 1281–1301 (2006)
28. Neves, C., Fraga Alves, M.I.: Semi-parametric approach to the hasofer-wang and greenwood statistics in extremes. *Test* **16**(2), 297–313 (2007)
29. Fraga Alves, I., Neves, C., Rosário, P.: A general estimator for the right endpoint with an application to supercentenarian women's records. *Extremes* **20**(1), 199–237 (2017)
30. Neves, C., Pereira, A.: Detecting finiteness in the right endpoint of light-tailed distributions. *Stat. Probab. Lett.* **80**(5), 437–444 (2010)
31. Smith, R.L.: Estimating tails of probability distributions. *Ann. Stat.* **15**, 1174–1207 (1987)
32. Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L.: A moment estimator for the index of an extreme-value distribution. *Ann. Stat.* **17**, 1833–1855 (1989)
33. Fraga Alves, M.I., Gomes, M.I., de Haan, L., Neves, C.: Mixed moment estimator and location invariant alternatives. *Extremes* **12**(2), 149–185 (2009)
34. Drees, H., Ferreira, A., de Haan, L.: On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**(3), 1179–1201 (2004)
35. Qi, Y., Peng, L.: Maximum likelihood estimation of extreme value index for irregular cases. *J. Stat. Plan. Inference* **139**, 3361–3376 (2009)
36. Zhou, C.: The extent of the maximum likelihood estimator for the extreme value index. *J. Multivar. Anal.* **101**(4), 971–983 (2010)
37. Castillo, J., Daoudi, J.: Estimation of generalized Pareto distribution. *Stat. Probabilily Lett.* **79**, 684–688 (2009)
38. Castillo, J., Serra, I.: Likelihood inference for generalized Pareto distribution. *Comput. Stat. Data Anal.* **83**, 116–128 (2015)
39. Eastoe, E.F., Tawn, J.A.: Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **58**(1), 25–45 (2009)
40. Coles, S., Bawa, J., Trenner, L., Dorazio, P.: *An Introduction to Statistical Modeling of Extreme Values*, vol. 208. Springer (2001)
41. Resnick, S.I.: Tail equivalence and its applications. *J. Appl. Prob.* **8**, 135–156 (1971)
42. de Haan, L., Tank, A.K., Neves, C.: On tail trend detection: modeling relative risk. *Extremes* **18**(2), 141–178 (2015)
43. Einmahl, J.H., Haan, L., Zhou, C.: Statistics of heteroscedastic extremes. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **78**(1), 31–51 (2016)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

