Report on *"How far can we go in aluminum extrusion?"*

74th European Study Group with Industry, 26th–30th April 2010

Department of Mathematics, University of Aveiro, Portugal

Problem presented by: Extruverde S.A. – Extrusão de Alumínio

Study group contributors: *Cláudia Neves,*
Elisabete dos Santos Ferreira, Eloísa Catarina Macedo

Report prepared by:
Cláudia Neves (e-mail: *claudia.neves@ua.pt*)
Elisabete dos Santos Ferreira (e-mail: *a31077@ua.pt*)
Eloísa Catarina Macedo (e-mail: *macedo@ua.pt*)

Executive Summary: Extrusion of aluminum is an efficient manufacturing process which allows continuous production. The heated billet (aluminum material) is pushed through a metal die to produce the desired profile. However, a long continuous production increases aging of the die and hampers its capability to yield homogeneously shaped profiles. Because of that, the dies are usually removed before their breaking point and only go back into production after receiving a layer coat for protection of the metal.

This report consists of a preliminary analysis of the extrusion amounts between consecutive maintenance procedures of the dies. A maintenance procedure in its whole encompasses an immersion bath of the die in caustic soda, a polishing operation and possibly a subsequent layering process in a nitriding chamber. The main goal here is to find the optimal life cycle for a die, in the sense that we are looking for a risk level (extrusion amount) above which die-damage occurs with a certain high probability. We shall rely on Extreme Value statistics to answer the question of how far can we go at each continuous operation of aluminum extrusion.

1 Introduction and Preliminary Results

The main production activity of Extruverde is the extrusion of aluminum. Simply put, the manufacturing process consists of squeezing aluminum through a shaped opening (a die) with the aid of a powerful hydraulic press, thus producing an incredible variety of useful products with almost any shape imaginable. Despite considerable efforts by the company Extruverde in providing comprehensive data records to us, we were confronted with sparsity of information in the sense that we did not find enough replicates to establish a feasible tipping point in extrusion, i.e., a value for the total amount produced beyond which defective profiles would follow with high probability. This is the main reason why we shall focus on the data pertaining to the most frequent production available, comprising $N = 111$ data points of the amount of aluminum profile (in Kg) coming out of 6 dies of the same type, referenced by PK40.04. The amounts are displayed in Figure 1 by the date of observation. We can see that the dies alternate in production or extrusion cycles quite often. Every time a die needs to be removed it is immediately replaced

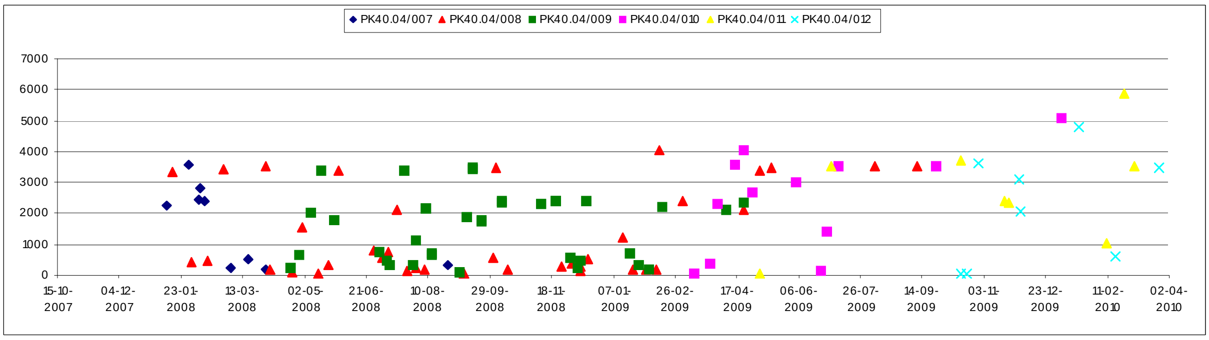

Figure 1: Amount (in Kg) of the aluminum profile yields from 6 dies of the same type PK40.04.

by another die of the same type in order to avoid delays in the aluminum profile manufacture.

A die must be removed from continuous production either because it starts to yield defective (non-homogeneous) aluminum profiles or the die breaks into pieces for some (frequently unknown) reason. Although a die might be in good conditions to carry on with extrusion it is desirable to remove it before any severe damage occurs. We then say that the corresponding aluminum yield has been censored. The dies only go back into production after a two-step maintenance procedure which encompasses: (1) an immersion bath in caustic soda; (2) cleansing and polishing treatments, in order to completely remove the aluminum waste. Sometimes the dies are subsequently submitted to a nitriding layer in order to hardened the die material. The latter is a time consuming process with other costs involved, hence it is only applied when strictly necessary. There are trained and experienced technicians for assessing whether a certain die requires a nitriding layer after the two-step procedure described above. Moreover there are well established upper limits for the total production amount allowed between consecutive maintenance procedures. Crossing an upper limit beyond which operations (1) or (2) are required increases the risk damaging the die. We wish to adjust these values for future guidance in continuous extrusion. Therefore we need some considerations of practical importance.

The data collection at hand consists of amounts of aluminum profile (aluminum yields) produced between consecutive operations of cleansing and/or nitriding. As already mentioned, a die is usually withdrawn from extrusion when it starts to deliver ill-shaped profiles. Because this is a failure event, the amount of aluminum profile produced until this point, i.e., the amount

Table 1: Prescribed maximum for production between nitriding treatments

Die \ Number of Nitridings	1	2	3	4	5	6	∞
PK40.04	2000	4000	5000	6000	10000	10000	10000

before failure occurs corresponds to a uncensored obervation.

Now we emphasize that it is not true that censored values are usually followed by large extrusion yields that could not fit in the prescribed maintenance layout. At any particular instant, a large batch of aluminum profiles needs to be divided in smaller batches that can be extruded without interruption until the next nitriding limit is reached (non-failure). This leads to consecutive large censored observations. Removal from production sometimes occurs simply because no more profiles of the designated shape are needed at that particular instant. Extrusion yields related to the latter are also seen as non-failure or censored data records.

In what follows we shall assume that, once removed from continuous production, a die always enters the two-step procedure that ensures its integrity for the next production period. Regardless of a censored amount of aluminum being pushed through the die, this same die only goes back into production after proper removal of aluminum residue. Since there is no clear cut way of discerning between censored and non-censored observations from the data available, we assume at any arrival, a die should be capable of producing at least 10% of the recommended maximum for nitriding. These prescribed upper limits (in Kg) for a certain die are shown in Table 1. If we consider, for instance, that the probability of a die breaking before delivering 10% of the upper limit, given the die has sustained 2 nitriding treatments, is almost zero then any event yielding less than 400 Kg is very likely to account for a censored yield.

As a matter of fact, the combination of methods described above, nitriding being the ultimate step if necessary, aims to serve a rectifying purpose as it should bring the die back to almost full capacity for the next extrusion cycle. At early stages of extrusion, which entails the die has received a nitride layering only a few number of times - less than 3, say - we find reasonable to assume this purpose is actually achieved. After sustaining a greater number of production cycles and nitridings, the natural aging of the dies hampers its capability of production. The assessment on whether the die should then be used until exhaustion, for a single prolonged extrusion, or for the accumulation of many small production amounts, e.g., by serving as replacements for other

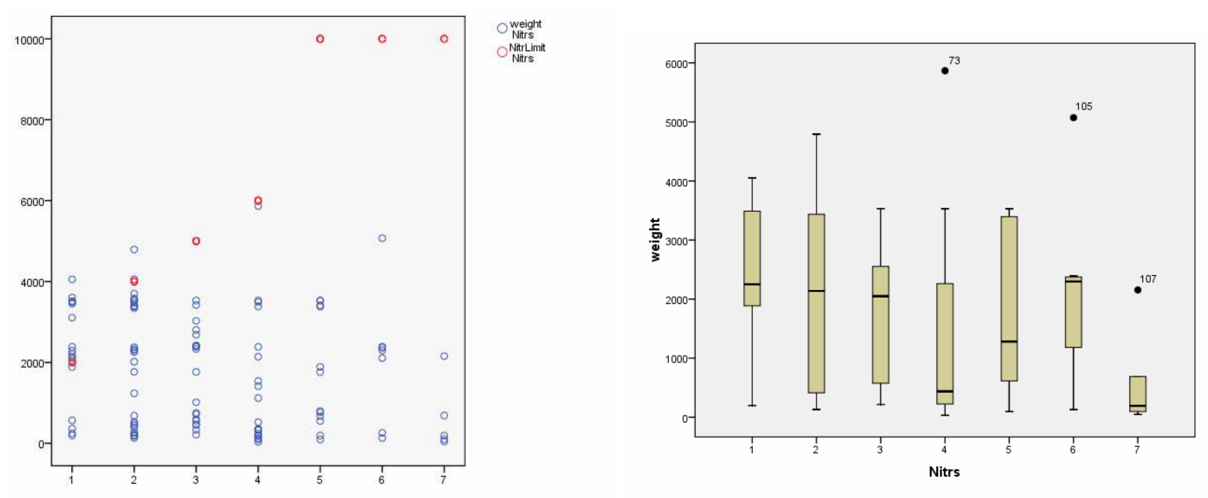

Figure 2: Left: Production amounts *versus* number of nitridings; Right: Rephrasing the plot on the left hand side with a comparative box-plot.

dies of the same type nearly at the end of each production cycle, falls out of scope of the present analysis.

On the left panel of Figure 2, the PK40.04 data is plotted against the number of nitridings. Each blue dot represents a production cycle yielding a certain amount of aluminum profile. The red dots stand for the maximum amount allowed, i.e., the prescribed limits for extrusion yields with respect to the specific nitriding state (cf. Table 1). Particularly at an early state of nitriding, i.e., after a small number of nitriding treatments (less than 3) the extrusion sometimes lingers on, surpassing the upper limits set for continuous production. From this data analysis we hope to be able to adjust these limits for a production cycle in connection with a small number of nitridings. The right hand-side of Figure 2 encloses comparative box-plots of the PK40.04 sampled data of size $N = 111$.

On the basis of the distribution of frequencies of the data in Figure 2, we have found reasonable to proceed by pooling information out of the 3 groups, with respect to nitrindings 1 up to 3. Performing Kruskall-Wallis test upon these 3 groups, regardless of censorship operative on the 3 populations, we found no evidences of significant difference (p.value $p = 0.668$) which corroborates the option for concatenating the three independent samples. Figure 3 depicts the maximum amount yield from each one of the 6 dies of the referenced type PK40.04 against

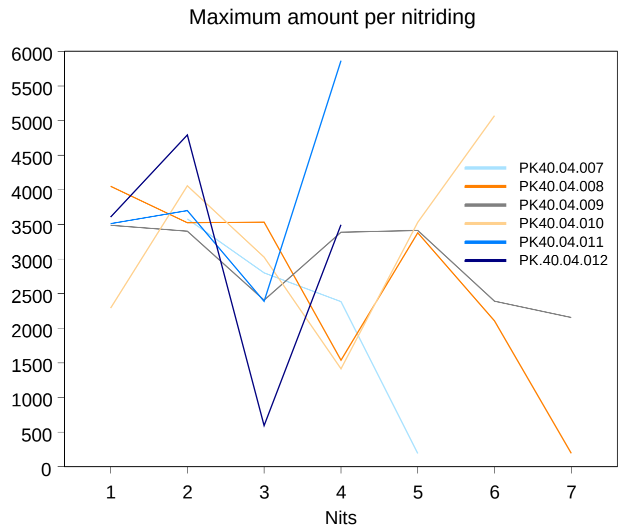

Figure 3: Maximum amount (in Kg) of aluminum profile yields for each individual die PK40.04 after the k-th nitriding, $k = 1, 2, \ldots, 7$.

the number of nitriding treatments. We observe that early staged dies, with at most 3 nitriding treatments, tend to yield extrusion maxima between 2500 Kg and 4000 Kg, while for the remainder stages we observe a mixed behavior in the die usage.

Henceforth we shall assume that the $n = 66$ aluminum amounts comprising groups 1 up to 3 correspond to a realization of a random sample from the same population. We note at this point that the main objective is to calculate a value such that the exceedance probability is small, i.e., we wish to set a risk value at a suitable high percentile. A question of practical importance has just been raised: how to calculate a high percentile (99% or even 95%) if there are less than 100 observations available? Under the present setting we have attained a sample size of $n = 66$ observations, and this is only because we were able to merge three groups. Now, extreme value statistics provides the adequate framework to tackle this problem. Extreme Value Theory (see e.g. [3]) is the theory underpinning our statistical approach. Moreover, since we are facing the possibility of randomly right-censored aluminum yields, the work by Einmahl *et* al. [2] provides adequate statistical tools for extrapolating beyond the range of observed data.

The outline of the present report is as follows. In Section 2 we expound our results and recommendations on future work. Section 3 contains details on the statistical inference methodologies used to tackle this problem submitted by Extruverde to the 74th European Study Group with Industry.

2 Conclusions and Recommendations

Our approach makes the following tangible contributions regarding profile extrusion through dies of type PK40.04:

- Undertaking the usual cleansing and/or nitriding operations, extrusion on recent dies, with at most 3 nitridings, has been conducted with a mild probability of rupture of approximately 30%.
- Given the specific operational conditions, the production data available allow estimation of a value at risk which is up-crossed with a low probability. The values 3917 and 4745 Kg were found with exceedance probability of 0.05 and 0.01, respectively. These values are depicted in Figure 4 and may account for an update in the upper limits for nitriding provided the die has received less than 4 cleansing and/or nitriding treatments.

Average production amounts are given in Table 2 below, with accompanying standard deviation and median production estimates. For instance, we have observed $n_2 = 28$ production yields for the six dies of type PK40.04 subjected to 2 nitriding treatments, which have resulted in the average amount of 1958.32 Kg; the standard error of this estimated mean is $s.e. = 1255.53/\sqrt{28} = 237.27$; on half of the production times, this type of dies was used in order to extrude a maximum of 2137.0 Kg of aluminum profiles, staying far below the imposed upper limit of 4000 Kg (cf. Table 1).

Table 2: Summary statistics

PK40.04 – Number of Nitridings	1	2	3
Mean	2286.33	1958.32	1714.55
Std. Deviation	1255.53	1511.02	1138.64
Median	2250.50	2137.0	2048.5

We have now found the risk values 3900 and 4700 Kg associated with a probability 0.05 and 0.01, respectively, of die damage. These can be regarded as an update of the extrusion limits

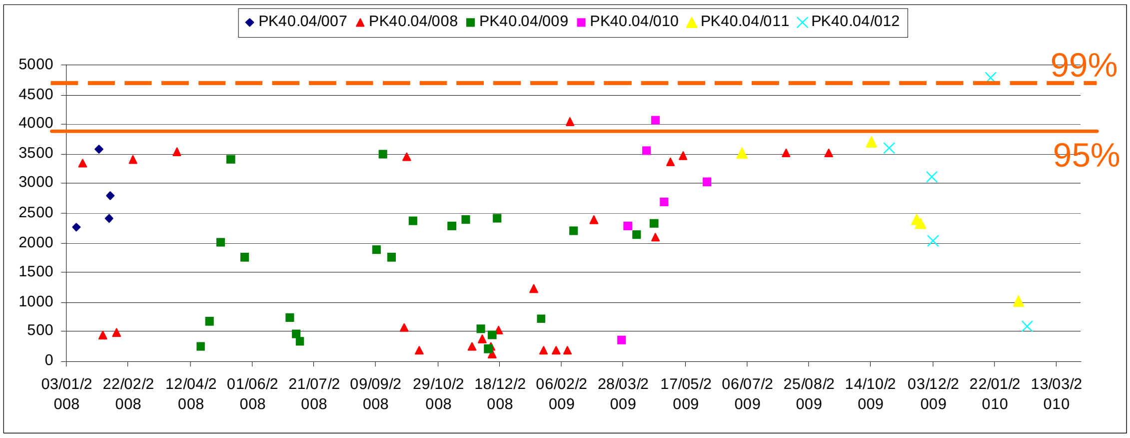

Figure 4: Estimated values at risk with respect to probability $p = 0.01, 0.05$, given a die sustaining up to 3 nitriding treatments.

for a fairly new die, with less than three cleansing and/or nitriding treatments. Nevertheless, these could be improved by following a more systematic scheme of die replacement, since then we would be able give information about a more accurate definition of censored/non-censored data for all nitriding stages.

A subsequent direction for applied research work would naturally emerge if it would be possible to keep track on cleansing/nitriding times and costs involved in order to establish an optimal replacement policy.

Analysis of additional data from ongoing PK40.04 measurements should be undertaken to confirm and expand upon the results reported here. In particular, longer trials should be made possible, albeit under more detailed records as suggested above, thus providing more data and a more realistic assessment.

3 Analysis

Let $X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}$ be a random sample from the same parent distribution function F. Let $X_{1,n}^{(0)} \le X_{2,n}^{(0)} \le \ldots \le X_{n,n}^{(0)}$ be their ascending order statistics. We denote by $\{X_i^{(0)}\}_{i=1}^n$ the true observations for the ith production yield. Since this observation may be censored by a random variable Z_i, independent of $X_i^{(0)}$, sometimes it cannot be observed. Instead we keep records of

$X_i = X_i^{(0)} \wedge Z_i$ coupled with the indicator variable $\delta_i = I(X_i^{(0)} \le Z_i)$, $i = 1, 2, \dots, n$. The latter shows whether or not X_i is in fact censored. The other censoring variables, i.e., those Z_i for which $X_i < Z_i$ contain no information about the underlying distribution function of the true observations would be disregarded even if they were available. In the context of our problem, with respect to reference reference PK40.04, the upper limits from Table 1 are not fixed limits when the number of nitridings comes down to 1 or 2, therefore we find reasonable to address the presence of a random censorship. Hence, given a bivariate random sample $\big\{(X_i, \delta_i)\big\}_{i=1}^{n}$, it is our goal to make inference about the unknown distribution function F at high quantiles. The true distribution function F_Z, underlying the sometimes unobserved i.i.d sample of Z_i, $i = 1, 2, \dots, n$, is thus taken as a nonparametric nuisance parameter. In the framework of extreme value theory (see [3]) we shall assume that both F and F_Z are continuous and belong to some max-domain of attraction, which entails existence of constants $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} P\Big\{\frac{X_{n,n}^{(0)} - b_n}{a_n} \le x\Big\} = G(x),$$

for all x, continuity point of G. Up to location and scale parameters it is possible to redefine constants $a_n > 0$ and b_n so that the only three possible limiting forms of distribution functions G can be nested in a one-parameter family of distributions functions, the Generalized Extreme Value distribution (GEVd), according to the definition:

$$G(x) = G_\gamma(x) := \begin{cases} \exp\big(-(1+\gamma\, x)^{-1/\gamma}\big) & \text{if } \gamma \neq 0, \\ \exp\big(-e^{-x}\big) & \text{if } \gamma = 0. \end{cases}$$

We then say that F is in the domain of attraction of G_γ and use the notation $F \in \mathcal{D}(G_\gamma)$. The shape parameter γ is the so-called extreme value index. The extreme value index $\gamma \in \mathbb{R}$ can be regarded as a gauge of tail heaviness of F: if $\gamma > 0$ then F is heavy-tailed, with a power decaying tail. On the opposite side, the case of $\gamma < 0$ endorses light-tailed distributions with finite right endpoint $x_* := \sup\{x : F(x) < 1\}$. If the extreme value index γ is equal to zero, then we may be in the presence of a variety of underlying behavior in the tail of F. The underlying distribution function F may be a moderately light tail with infinite right endpoint, such as the lognormal distribution, or might be coming down to a light tailed distribution of the exponential type with finite right endpoint.

In the problem at hand, we are interested in looking beyond the prescribed nitriding limits. In particular we wish to estimate a quantile associated with a small exceedance probability. Hence the probabilistic operation of conditioning above a certain high level (yield) of extrusion amounts is quite important. Defining by L_s the extrusion limit at stage $s = 1, 2, ..., m$ (see e.g. Table 1), we wish to investigate the probability of $X_1 - L_s$ being close to zero. This problem shall be addressed in terms of high quantile estimation under the framework provided by extreme value theory. In this respect our main reference is the recent work by [2].

Now let us consider the i.i.d. uncensored observations with tail distribution function given by

$$\overline{F}(x) = P\big\{X_i^{(0)} > x, \delta_i = 1\big\}.$$

We define the expected amount to be produced, per continuous extrusion of aluminum profile, given an extrusion demand greater than t:

$$e(t) := \frac{1}{\overline{F}(t)} \int_t^{\infty} (x - t)\, dF(x) \;=\; E\big(X_1^{(0)} - t \,|X_1^{(0)} > t,\, \delta_1 = 1\big),$$

with empirical counterpart given by

$$e_n(t) := \frac{\sum_{i=1}^{n} X_i\, I(X_i > t)(1 - I^*(X_i))}{\sum_{i=1}^{n} I(X_i > t)(1 - I^*(X_i))} - t, \tag{1}$$

where I^* stands for the empirical distribution function of the censoring random variable Z. The plot in the left panel of Figure 5 displays the estimated mean excess function (1) regarding PK40.04 dies, on the basis of the merged failure data. These $n_1 = 39$ uncensored observations, which result from taking the failure data records at the nitriding stages (numbered) $s = 1, 2, 3$ as one sample realization of the same population, are depicted on the right hand-side of Figure 5. The observed order statistics, after suitable log-transform, have been plotted against the expected Pareto quantiles given by $-\log(1 - p_i)$, $p_i = i/(n_1 + 1)$. The true distribution func-

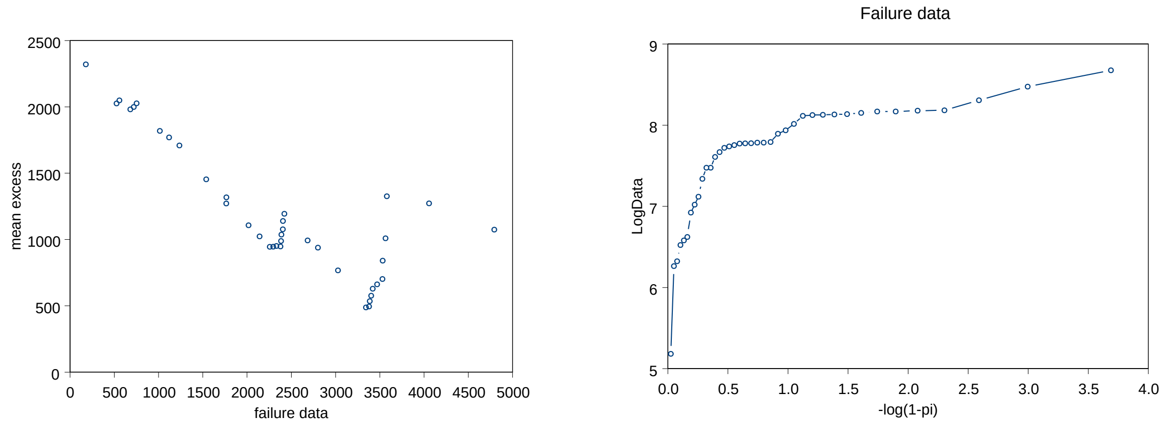

Figure 5: Failure data: mean excess plot (left) and Pareto QQ-plot (right).

tion F underlying the uncensored extrusion amounts seems to detain a heavier tail than the exponential distribution since the estimated mean excess function ultimately increases (cf. [1], p.16). Moreover, the QQ-plot in Figure 5 raises the suspicion that we are in the presence of a not-so-heavy-tailed distribution underlying the uncensored data because of the slow increase in the straight line eventually. Hence we should expect that the distribution function F detains at leat a finite mean and variance, that is, γ_1 less than 1/2. The remainder 27 censored observations suggest similar findings, with a moderately positive tail index $\gamma_2 < 1/2$.

There are several ways to estimate the extreme value index γ, or in the heavy-tailed case, the tail index $\gamma > 0$. Several estimators of widely-spread use are available in the literature (see e.g. Chapter 3 of [3]). The Hill estimator for positive γ, for instance, is connected with the slope of the Pareto QQ-plot like the one shown in Figure 5. We shall make use of the well-known Moment estimator for general γ (see [3], Section 3.5). The Moment estimator is defined as follows, regardless of censorship being operative on the random variables X_i.

$$\hat{\gamma}^M_{n,k} := M_n^{(1)} + 1 - \frac{1}{2}\Big\{1 - \frac{\big(M_n^{(1)}\big)^2}{M_n^{(2)}}\Big\}^{-1}, \tag{2}$$

where, for $r = 1, 2$,

$$M_n^{(r)} := \frac{1}{k}\sum_{i=1}^{k}\Big(\log X_{n-i+1,n} - \log X_{n-k,n}\Big)^r. \tag{3}$$

However, under random censoring, we can only observe X_i in lieu of $X_i^{(0)}$ meaning that the Moment estimator (2) is actually estimating $\gamma \times (\gamma_1 + \gamma_2)/\gamma_2$. We have assumed $\gamma_1 > 0$ and $\gamma_2 > 0$, which means that the true extreme value index γ is such that $\gamma = \gamma_1\gamma_2/(\gamma_1 + \gamma_2)$. Figure 6 contains Moment estimates values of k, i.e., when only a small number $(k+1)$ of upper order statistics is used, the latter providing an adequate setting for application of extreme value statistics. Bearing the above in mind, the Moment estimator involving all the $n = 66$ observations X_i, has to be corrected by $\gamma_2/(\gamma_1+\gamma_2)$. The left panel in Figure 6 also accounts for this correction, which translates into a slightly heavier tail since we obtain γ-estimates around 0.3 quite often. Here we have considered

$$\hat{c}_k := \frac{1}{k}\sum_{i=1}^{k} \delta_{[n-i+1]} \tag{4}$$

as an estimator of the correction factor $\gamma_2/(\gamma_1 + \gamma_2)$, with the notation $\delta_{[n-i+1]}$ for the delta pertaining to $X_{n-i+1,n}$, $i = 1, 2, \ldots, k$.

We are now in a good way to undertake the estimation of high quantiles pertaining to a small exceedance probability. Since we have only $n = 66$ data records available for statistical analysis, it means we are attempting estimation far out the sample range. A good proposal is the estimator for an extreme quantile $x_\varepsilon = F^{\leftarrow}(1-\varepsilon)$ introduced by [2]:

$$\widehat{x}_{\varepsilon,k} = X_{n-k,n} + \widehat{a}\Big(\frac{n}{k}\Big)\frac{\big((1-\widehat{F}_n(X_{n-k,n}))/\varepsilon\big)^{\hat{\gamma}^M_{n,k}/\hat{c}_k} - 1}{\hat{\gamma}^M_{n,k}}. \tag{5}$$

In this respect we have fixed the values $\varepsilon = 0.05$ and $\varepsilon = 0.01$.

Regarding estimation of the scale $a(n/k)$, we have used estimators connected with the so-called moment estimator (see section 4.2 of [3]). More formally, we define

$$\widehat{a}\Big(\frac{n}{k}\Big) = X_{n-k,n}(s_j)\,\hat{\gamma}^+_{n,k}\big(1-\hat{\gamma}^-_{n,k}\big), \tag{6}$$

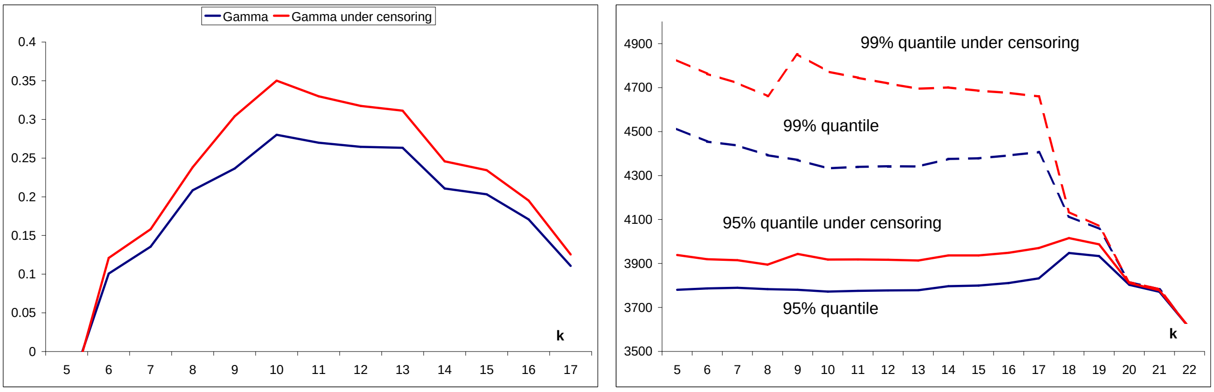

Figure 6: Moment estimates on the basis of the PK40.04 data (left) and estimated quantiles (right), all plotted against the number $(k+1)$ of upper order statistics.

with $\hat{\gamma}^{+}_{n,k}$ coinciding with the Hill estimator given by (3) in the particular case of $r=1$, i.e.,

$$\hat{\gamma}^{+}_{n,k} := M_n^{(1)} = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}, \tag{7}$$

and a suitable estimator for $\gamma_- := \min(\gamma, 0)$ given by

$$\hat{\gamma}^{-}_{n,k} := 1 - \frac{1}{2}\left\{1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right\}^{-1}. \tag{8}$$

The right panel of Figure 6 displays the estimated quantiles, both when censorship is present or not. On the light of the information shared by these plots, we settle with the estimates $\widehat{x}_{0.05,11} = 3917$ Kg and $\widehat{x}_{0.01,12} = 4745$ Kg. We do not provide any confidence intervals because the asymptotic normal limit to be attained by the high quantile estimator (5), after suitable normalization, has not been explicitly deduced in [2].

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2006). *Statistics of Extremes: Theory and Applications*. Wiley, England.

[2] Einmahl, J.H.J., Fils-Villetard, A. and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli,*, **14**, 207-227. Wiley, New York.

[3] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.