Statistics of Extreme Values

L. de Haan
Econometric Institute, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

In the 1950's, after the 1953 flood, the statistical problem of how to determine a safe height of the Dutch sea-dikes on the basis of observed high-tide water levels was studied extensively. Now, thirty years later, both the number of observations and the statistical methodology have grown considerably. This led to a new investigation. I shall report about some theoretical aspects of this work.

1. Introduction: probabilistic theory

The following is mainly a theoretical exposition but I shall often refer to the specific application mentioned above.

The basic problem is the following: suppose $X_1, X_2, \ldots, X_n$ are independent observations (e.g. high-tide water levels observed in the past) from an unknown probability distribution with distribution function F, that is $F(x) = P(X_i \leqslant x)$ for $x \in \mathbb{R}$, $i = 1, 2, \ldots, n$. Let p be a small positive number ($p < 1/n$). Determine a real number x_p (the height of the sea-dike in our problem) such that

$$F(x_p) = 1 - p, \tag{1}$$

i.e. find the $(1-p)$-quantile of F.

Firstly it is clear that we have to *estimate* x_p, i.e., we have to determine a function $\hat{x}_{p,n}(X_1, \ldots, X_n)$ such that $F(\hat{x}_{p,n}) = 1 - p$ is approximately true. Secondly it is clear that since in our problem $p \ll 1/n$ we cannot do anything without some additional assumptions on F (otherwise extrapolation outside the sample is not possible). The additional assumption on F can be parametric, i.e., we assume the function known apart from a few real parameters, or semi-parametric, i.e., we do not assume anything about F except for its behaviour near its upper end point, that is the value $x^* \leqslant \infty$ such that $F(x^*) = 1$, $F(x^* - \epsilon) < 1$ for all $\epsilon > 0$.

Our assumption is of the latter type: Suppose there exist norming constants $a_r > 0$ and b_r $(r = 1, 2, \ldots)$ such that the normed sample extreme converges in distribution, i.e.

$$\lim_{r \to \infty} (F(a_r x + b_r))^r = \lim_{r \to \infty} P\left\{\frac{\max(X_1, X_2, \ldots, X_r) - b_r}{a_r} \leqslant x\right\} \tag{2}$$

$$= G(x)$$

for all x, where G is the distribution function of a non-degenerate probability distribution and $X_1, X_2, \ldots$ a sequence of independent random variables with distribution function F. It is well known (Gnedenko [6]) that for a proper choice of $\{a_r\}$ and $\{b_r\}$ the limiting distributions in (2) are all from the class $\{G_\gamma\}_{\gamma\in\mathbf{R}}$ with

$$G_\gamma(x) := \exp(-(1+\gamma x)^{-1/\gamma}) \tag{3}$$

for those x for which $1+\gamma x>0$. In the special case $\gamma=0$ one should read the right-hand side of (3) as $\exp(-e^{-x})$ for all real x.

One says that F is in the domain of attraction of G_γ (notation $F\in D(G_\gamma)$) for some fixed $\gamma\in\mathbf{R}$ if there are $\{a_r, b_r\}$ such that (2) holds with $G(x)$ replaced by $G_\gamma(x)$. Conditions for $F\in D(G_\gamma)$ and expressions for a_r and b_r in terms of F are well known. What is of interest to us now is that $F\in D(G_\gamma)$ if and only if for some positive function a

$$\lim_{t\uparrow x^*}\frac{1-F(t+xa(t))}{1-F(t)} = -\log G_\gamma(x) = (1+\gamma x)^{-1/\gamma} \tag{4}$$

for all x with $x>0$, $1+\gamma x>0$. This has the following probabilistic interpretation: for a random variable X with distribution function F

$$\lim_{t\uparrow x^*} P\{\frac{X-t}{a(t)}>x \mid X>t\} = (1+\gamma x)^{-1/\gamma} \tag{5}$$

for all $x>0$, $1+\gamma x>0$. The class of distributions whose tails are given in the right-hand side of (5) are called residual life-time distributions.

The relations (2) and (5) have closely related interpretations, as we shall now explain. One can generalize (2) to the limiting distribution of the upper k order statistics ($n\to\infty$, k fixed). Recall that, if the observations $X_1, X_2, \ldots, X_n$ are ordered by size, they are called n-th order statistics and denoted as $X_{(1,n)}\leqslant X_{(2,n)}\leqslant \cdots \leqslant X_{(n,n)}$. So (2) says that the joint distribution of all observations in the sample larger than a fixed order statistic $X_{(n-k,n)}$ is approximately one of a small class of known distributions and (5) says that all observations in the sample exceeding some fixed high level t follow approximately one out of a small class of distribution functions: $1-(1+\gamma x)^{-1/\gamma}$. In the first case the number of observations considered is fixed, $k-1$, in the second case the number is random. Both interpretations can be used for statistical purposes as we shall see.

A third equivalent form of (2) also has a probabilistic interpretation: $F\in D(G_\gamma)$ if and only if for all $x,y>0$, $y\neq 1$

$$\lim_{t\to\infty}\frac{U(tx)-U(t)}{U(ty)-U(t)} = \frac{x^\gamma-1}{y^\gamma-1} \tag{6}$$

where $U:=(\frac{1}{1-F})^{\leftarrow}$ (the arrow denoting the generalized-inverse function). In the case $\gamma=0$ one should read the right-hand side of (6) as $(\log x)/(\log y)$. Moreover (2) holds with $b_r=U(r)$ and $a_r=\{U(2r)-U(r)\}\gamma/(2^\gamma-1)$. An interpretation of (6) is that high quantiles of the distribution ($U(tx)$ with $x>1$)

can be expressed asymptotically in terms of a somewhat lower quantile ($U(t)$). In statistical terms that means that we can extrapolate from a quantile inside the sample to a quantile outside the sample as we shall see.

I shall now sketch a proof of the equivalence of (4) and (6). For simplicity let us assume that F is continuous and strictly increasing on $(-\infty, x^*)$. Replace t in relation (4) by $U(s)$ i.e. $s = 1/(1-F(t))$. Then (4) reads

$$H_s(x) := \frac{1}{s\{1-F(U(s)+xaU(s))\}} \rightarrow (1+\gamma x)^{1/\gamma} \quad (s \rightarrow \infty).$$

Now $H_s(x)$ is a family of non-decreasing functions converging to a continuous function. Then the inverse functions $H_s^{\leftarrow}(z)$ converge to the inverse of the limit function, i.e.

$$\frac{U(sz)-U(s)}{a(U(s))} \rightarrow \frac{z^{\gamma}-1}{\gamma} \quad (s \rightarrow \infty). \tag{7}$$

Now (6) is obtained by using (7) for $z = x$ and $z = y$ and dividing. The proof of the converse is similar. Note that $a_r \sim a(U(r))$ and $(b_r - U(r))/a_r \rightarrow 0$, $r \rightarrow \infty$.

In the application to the height of Dutch sea-dikes we have in mind, the observations $X_1, X_2, \ldots, X_n$ are the observed high-tide water levels during a large number of years. Clearly these observations are not independent. Moreover they do not all come from the same probability distribution. The last difficulty can be resolved (approximately) by using only the observations taken during a few months in the winter (the storm season). The first difficulty is more serious.

The above-mentioned results have been extended to stationary stochastic processes with a weak dependence structure (Leadbetter, Lindgren and Rootzén [9]). In fact these results go through in this case without change. However the dependence in our case is stronger. In order to introduce such a stronger dependence structure from a theoretical point of view I shall now explain briefly the connection between extreme-value theory and Poisson point processes.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with distribution function F for which (2) holds. Consider the random set

$$Q_n := \{(\frac{k}{n}, \frac{X_k - b_n}{a_n}) \mid k = 1, 2, \ldots\}$$

in $\mathbb{R}^2$. For each set of the form $A_{a,b,x} := [a,b] \times (x, x^*)$ let $N_{n,a,b,x}$ be the number of points of Q_n in $A_{a,b,x}$. Then clearly $N_{n,a,b,x}$ has a binomial $(n, (b-a)\{1-F(b_n + xa_n)\})$ distribution. Since by either (2) or (5)

$$\lim_{n \rightarrow \infty} n\{1-F(b_n + xa_n)\} = (1+\gamma x)^{-1/\gamma}$$

(provided $1+\gamma x > 0$), $N_{n,a,b,x}$ has a limiting Poisson $(a-b)\cdot(1+\gamma x)^{-1/\gamma}$ distribution ($n \rightarrow \infty$). Further, it can easily be seen that $N_{n,a,b,x}$ and $N_{n,a',b',x'}$ are asymptotically independent if $A_{a,b,x}$ and $A_{a',b',x'}$ are disjoint. This means that the two-dimensional point processes Q_n of the observations converge in distribution to a Poisson point process on $\mathbb{R}^2$ with mean measure ν given by

$$\nu(A_{a,b,x}) = (b-a)\cdot(1+\gamma x)^{-1/\gamma}$$

(Pickands III [13]). This result remains true under the weak dependence condition referred to above. For the case of stronger dependence we first consider an example.

Let $X_1, X_2, \ldots$ be independent random variables from a distribution function F satisfying (2). Let the sequence $Y_1, Y_2, \ldots$ be equal to

$$X_1, X_1, X_2, X_2, X_3, X_3, \ldots$$

with probability 1/2 and equal to

$$X_1, X_2, X_2, X_3, X_3, \ldots$$

with probability 1/2. Then $Y_1, Y_2, \ldots$ is a stationary sequence, i.e., for all L the distribution of $(Y_{1+r}, Y_{2+r}, \ldots, Y_{L+r})$ does not depend on r. Clearly the point process convergence still holds for the set

$$\{(\frac{k}{n}, \frac{Y_k - b_n}{a_n}) \mid k = 1, 2, \ldots\}$$

with the exception that each point in the limiting point process occurs twice. This is a primitive form of a point process with clustering that in its general form can be described as follows. The set of the projections of all the points on the horizontal axis forms a homogeneous Poisson point process as before. Each projected point on the horizontal line corresponds to a random number of points on the vertical line through that projected point: the cluster. Since we still assume stationarity for the original sequence of random variables, in the limiting point process there is a fixed probability distribution for the number of points on a horizontal line.

An important extra parameter describing the amount of clustering is then the mean number of points $1/\theta$ on a vertical line through a projected point. The parameter θ is called the 'extremal index' by Leadbetter ([10], the phenomenon has been studied before by O'Brien [12]). If $0 < \theta \leq 1$, a satisfactory theory can be developed.

In the application we consider here the above-mentioned clusters are interpreted as the high-tide water levels obtained during one severe windstorm.

The example given above suggests that one can choose between two equivalent ways of performing the computations: applying the theory of the extremal index itself or selecting the largest observation per cluster (windstorm) and treat those as independent observations whereby the total number of observations is reduced by a factor θ. We choose for the latter.

2. Statistical theory: estimation of γ

I now turn to the problem of how to estimate γ using independent observations $X_1, X_2, \ldots, X_n$ from F.

A traditional method uses 'yearly maxima', i.e., breaks the sample into blocks of equal size and uses maximum likelihood estimation under the assumption that the maximum in each block follows *exactly* distribution G_γ (we know by (2) that such a block maximum follows approximately distribution G_γ). Consistency has been proved here under certain conditions (J.P. Cohen [2]). By using this method some information from the sample seems to be lost (the second largest observation in some year may be near the overall maximum).

A less traditional method consists of restricting attention to those observations from $X_1, X_2, \ldots, X_n$ that exceed a certain level $M(n)$ and using the method of maximum likelihood under the assumption that these observations follow *exactly* one of the asymptotic residual life-time distributions from (5). Asymptotic results for this procedure have been obtained by R.L. Smith [16].

I want to discuss two alternative methods in some detail.

a. Pickands' estimator

Relation (6) specializes to

$$\lim_{t\to\infty} \frac{U(2t)-U(t)}{U(t)-U(t/2)} = \frac{2^\gamma-1}{1-2^{-\gamma}} = 2^\gamma. \tag{8}$$

Since the limit relation (6) holds locally uniformly, also

$$\lim_{t\to\infty} \frac{U(q(t)\cdot t)-U(t)}{U(t)-U(t/q(t))} = 2^\gamma \tag{9}$$

provided $\lim_{t\to\infty} q(t)=2$. Our aim is to replace the quantities on the left-hand side of (9) by approximating random quantities so that a consistent estimate of γ is obtained. Let

$$Y_{(1,n)} \leqslant Y_{(2,n)} \leqslant \cdots \leqslant Y_{(n,n)}$$

be n-th order statistics from the distribution with distribution function $1-1/x$ $(x \geqslant 1)$. Then

$$U(Y_{(1,n)}) \leqslant U(Y_{(2,n)}) \leqslant \cdots \leqslant U(Y_{(n,n)})$$

are distributed like the n-th order statistics from the distribution function F. Now it is well known (Smirnov [15]) that, if $k=k(n)\in\mathbb{N}$, $k(n)\to\infty$ and, $k(n)/n\to 0$ $(n\to\infty)$,

$$\lim_{n\to\infty} \frac{k(n)}{n} Y_{(n-k(n)+1,n)} = 1 \tag{10}$$

in probability, so that

$$\lim_{n\to\infty} Y_{(n-k(n)+1,n)} = \infty$$

and

$$\lim_{n\to\infty} Y_{(n-2k(n)+1,n)}/Y_{(n-k(n)+1,n)} = 2 \tag{11}$$

in probability.

Combining (9) and (11) we get

$$\lim_{n\to\infty} \frac{U(Y_{(n-k(n)+1,n)})-U(Y_{(n-2k(n)+1,n)})}{U(Y_{(n-2k(n)+1,n)})-U(Y_{(n-4k(n)+1,n)})} = 2^{\gamma}, \text{ in probability.}$$

We have proved the following. Define $\hat{\gamma}_n^P$ by

$$\hat{\gamma}_n^P := (\log 2)^{-1} \log \frac{X_{(n-k(n)+1,n)}-X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)}-X_{(n-4k(n)+1,n)}}$$

where $X_{(1,n)} \leqslant \cdots \leqslant X_{(n,n)}$ are the n-th order statistics from the distribution F. Then

$$\lim_{n\to\infty} \hat{\gamma}_n^P = \gamma$$

in probability, provided $k(n)\to\infty$, $k(n)/n\to\infty$ $(n\to\infty)$. This estimator has been introduced by J. Pickands III [14]. It can be proved with considerably more effort (Dekkers and de Haan [3]) that under a natural strengthening of condition (2) and a further upper bound on the growth of the sequence $k(n)$ (depending on F)

$$\sqrt{k(n)}(\hat{\gamma}_n^P - \gamma)$$

has asymptotically $(n\to\infty)$ a normal distribution with mean zero and known variance so that a confidence interval for γ can be constructed.

b. A moment estimator

In order to introduce this estimator we first consider the case $\gamma>0$ of (2). Then (4) simplifies to

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}$$

for $x>0$ and this relation is equivalent to

$$\frac{\int_t^\infty (\log x - \log t) dF(x)}{1-F(t)} = \int_1^\infty \frac{1-F(tu)}{1-F(t)} \frac{du}{u} \to \int_1^\infty u^{-1/\gamma} \frac{du}{u} = \gamma, \ t\to\infty \tag{12}$$

(see e.g. Geluk and de Haan [5]). The extra factor $1/u$ in the two integrals on both sides of '$\to$' is necessary since otherwise the last integral may diverge. A sample analogue of the first integral in (12) provides an estimation for γ:

$$M_n^{(1)} := \frac{\int_{X_{(n-k(n),n)}}^\infty (\log x - \log X_{(n-k(n),n)}) dF_n(x)}{1-F_n(X_{(n-k(n),n)})}$$

$$= \frac{1}{k(n)} \sum_{i=0}^{k(n)-1} \log X_{(n-i,n)} - \log X_{(n-k(n),n)}. \tag{13}$$

Here

$$F_n(x) := n^{-1}\sum_{i=1}^{n} i_{\{X_i \leq x\}},$$

with $i_{\{X_i \leq x\}}=1$ if $X_i \leq x$ and 0 otherwise. It is well known (Mason [11]) that, if (2) holds with $\gamma>0$ and $k(n)\to\infty$, $k(n)/n\to 0$ $(n\to\infty)$,

$$\lim_{n\to\infty} M_n^{(1)} = \gamma$$

in probability. In order to extend this estimator to one that is valid for any real γ we note that (2) holds for $\gamma=0$ if and only if (Balkema and de Haan [1])

$$\lim_{t\uparrow x^*}\frac{\{\int_t^{x^*}(x-t)^2 dF(x)\}\{1-F(t)\}}{\{\int_t^{x^*}(x-t)dF(x)\}^2} = 2. \tag{14}$$

In view of the use of (14) as an extension of (12) we first have to make a version of (14) involving logarithms. Indeed (14) implies

$$\lim_{t\uparrow x^*}\frac{\{\int_t^{x^*}(\log x-\log t)^2 dF(x)\}\{1-F(t)\}}{\{\int_t^{x^*}(\log x-\log t)dF(x)\}^2} = 2, \tag{15}$$

provided (of course) $x^*>0$. We shall require this assumption; it does not do any harm in practice. The sample analogue of (15) is now obvious: one can prove that, if (2) is true

$$\lim_{n\to\infty} M_n^{(2)}/(M_n^{(1)})^2 = \begin{cases} 2 & ,\ \gamma\geq 0 \\ (2-2\gamma)/(1-2\gamma) & ,\ \gamma<0 \end{cases} \tag{16}$$

in probability provided $k(n)\to\infty$, $k(n)/n\to 0$ $(n\to\infty)$, with

$$M_n^{(2)} := \frac{1}{k(n)}\sum_{i=0}^{k(n)-1}\{\log X_{(n-i,n)}-\log X_{(n-k(n),n)}\}^2. \tag{17}$$

Moreover one can prove that under (2) for $k(n)\to\infty$, $k(n)/n\to 0$ $(n\to\infty)$

$$\lim_{n\to\infty} M_n^{(1)} = \max(0,\gamma) \tag{18}$$

in probability. Combination of (16) and (18) leads to the estimator

$$\hat{\gamma}_n^M := M_n^{(1)}+1-\tfrac{1}{2}\{1-(M_n^{(1)})^2/M_n^{(2)}\}^{-1},$$

which is clearly weakly consistent. It can be proved (Dekkers, Einmahl and de Haan [4]) that under a natural strengthening of condition (2) and a further upper bound on the growth of the sequence $k(n)$ (depending on F)

$$\sqrt{k(n)}(\hat{\gamma}_n^M-\gamma)$$

has asymptotically ($n \to \infty$) a normal distribution with mean zero and known variance so that a confidence interval for γ can be constructed.

It turns out that the asymptotic properties of the estimators $\hat{\gamma}_n^P$ and $\hat{\gamma}_n^M$ are not much different, but in practice $\hat{\gamma}_n^M$ behaves more nicely, apparently due to the fact that much more observations are used in the definition of $\hat{\gamma}_n^M$.

3. LARGE QUANTILE ESTIMATION

Next we want to solve equation (1), i.e., we want to construct an estimator for a high quantile of the unknown distribution function F. Once again we have to use asymptotic theory and assume (2). Let us proceed in an intuitive way. Relation (2) is still true if we replace n by $n/k(n)$, provided $n/k(n) \to \infty$ ($n \to \infty$). Hence (with $k = k(n)$)

$$\begin{aligned} p &= F(x_p) \approx G_\gamma^{k/n}((x_p - b_{n/k})/a_{n/k}) \\ &= 1 - \exp\left[-kn^{-1}(1+\gamma\frac{x_p - b_{n/k}}{a_{n/k}})^{-1/\gamma}\right] \\ &\approx kn^{-1}(1+\gamma\frac{x_p - b_{n/k}}{a_{n/k}})^{-1/\gamma}. \end{aligned}$$

That is,

$$x_p \approx \frac{(\frac{k}{np})^\gamma - 1}{\gamma} \cdot a_{n/k} + b_{n/k}. \tag{19}$$

We already know how to estimate γ, so it remains to construct estimators for $a_{n/k}$ and $b_{n/k}$. Also, since we are after asymptotic properties for x_p, we must be able to replace $k/(np)$ asymptotically by a constant. In the application we consider here (as in many other ones) we have $p \leqslant 1/n$, so for the asymptotic theory we have to assume that in fact p depends on n ($p := p_n$), $p_n \to 0$ and $np_n \to c$, finite and positive ($n \to \infty$). We then take $k > c$ fixed (not depending on n).

I remark that one can also consider the case $p_n \to 0$, $np_n \to \infty$ ($n \to \infty$). In that case x_p is best approximated by $b_{n/k}$ with $k(n) := [np_n]$, but one still needs to estimate $a_{n/k}$ in order to find the asymptotic properties of the estimator.

Let us now turn to the question how to estimate $a_{n/k}$ and $b_{n/k}$. We recall that (2) holds with $b_r = U(r)$ and $a_r = \{U(2r) - U(r)\}\gamma/(2^\gamma - 1)$. One can also prove that an alternative form for a_n is

$$a_n = nU(n)\{1 - \min(0,\gamma)\} \int_{U(n)}^{x^*} \{\log x - \log U(n)\} dF(x).$$

Replacing these quantities by their sample analogues we get two estimators for x_p, one based on $\hat{\gamma}_n^P$ one based on $\hat{\gamma}_n^M$:

$$\hat{x}_{p_n}^P := \frac{(\frac{k}{np_n})^{\hat{\gamma}_n^P} - 1}{1 - 2^{-\hat{\gamma}_n^P}} \{X_{(n-k+1,n)} - X_{(n-2k+1,n)}\} + X_{(n-k+1,n)}$$

and

$$\hat{x}^{M}_{p_n} := \frac{\left(\frac{k}{np_n}\right)^{\hat{\gamma}^M_n}-1}{\hat{\gamma}^M_n}\cdot X_{(n-k+1,n)}M^{(1)}_n\cdot\{1-\min(0,\hat{\gamma}^M_n)\}+X_{(n-k+1,n)}.$$

It can be proved that, if (2) is true, both

$$\frac{\hat{x}^P_{p_n}-x_{p_n}}{X_{(n-k+1,n)}-X_{(n-2k+1,n)}} \quad\text{and}\quad \frac{\hat{x}^M_{p_n}-x_{p_n}}{X_{(n-k+1,n)}\cdot M^{(1)}_n}$$

have completely specified limiting distributions ($n\to\infty$) so that an asymptotic confidence interval for x_{p_n} can be constructed.

4. Higher-dimensional theory

One can consider the joint distribution of high water levels at several places along the Dutch sea-coast. One then aims at statements concerning e.g. the likelihood of a flood at two different places simultaneously. I shall sketch the theory in a simple but representative case.

Suppose $(X_1,Y_1),(X_2,Y_2),\ldots$ are independent observations from some distribution function $F(x,y)$.

The distribution function of $(\max_{1\leq i\leq n}X_i,\max_{1\leq i\leq n}Y_i)$ is $F^n(x,y)$. Suppose for simplicity that the marginal distributions of X_1 and that of Y_1 are standard exponential (this can be achieved by preliminary transformation). Suppose

$$F^n(x+\log n, y+\log n)\to G(x,y)\quad (n\to\infty), \tag{20}$$

a proper distribution function. Note that $\log n$ is the proper normalization for convergence of the marginal distributions.

The analogue of (4) is here

$$\lim_{t\to\infty}\frac{1-F(t+x,t+y)}{1-F(t,t)} = -\log G(x,y) \tag{21}$$

for all continuity points (x,y) of G. Since the left-hand side represents the measure of a set (note that the upper index c denotes the complement of a set)

$$A_{x,y} := \{(s,t)|s\leq x,\ s\leq y\}^c,$$

this must also be true for the right-hand side, i.e., there is a measure υ such that for all x,y

$$G(x,y) = \exp(-\upsilon(A_{x,y})).$$

Moreover one sees easily that G satisfies

$$G^n(x+\log n, y+\log n) = G(x,y) \text{ for all } n\in\mathbb{N},\ x\in\mathbb{R},$$

hence

$$n\cdot\upsilon(A+\log n) = \upsilon(A)$$

for every Borel set $A \in \mathbb{R}^2$ and $n \in \mathbb{N}$ or, more generally

$$e^u \cdot \nu(A+u) = \nu(A)$$

for every Borel set A and $u \in \mathbb{R}$. It follows that

$$\nu\{(s,t) | s+t > w,\ s-t \in B\} = \pi(B) \cdot e^{-w} \tag{22}$$

where B is any Borel set of $\mathbb{R}$ and π a measure on $[-\infty, +\infty]$ with $\int_{[-\infty,\infty]} e^{|u|}\pi(du) < \infty$. The distribution functions G are thus parametrized by a collection of (say) probability measures:

$$\begin{aligned} -\log G(x,y) &= \nu\{(u,v) \mid u \leq x,\ v \leq y\}^c \\ &= \nu\{(u,v) \mid (u+v)+(u-v) \leq 2x, (u+v)-(u-v) \leq 2y\}^c \\ &= \nu\{(u,v) \mid u+v > \min(2x-(u-v), 2y+(u-v))\} \\ &= -\int_{[-\infty,\infty]} e^{-\min(2x-t,2y+t)}\pi(dt) \\ &= -\int_{[-\infty,\infty]} \max(e^{-2x+t}, e^{-2y-t})\pi(dt) \end{aligned}$$

(de Haan and Resnick, [7]).

The question is how to estimate π. As in the one-dimensional situation one only considers 'high' observations, since it follows from (21) that for all $x,y \in \mathbb{R}$

$$P\{X_1 - t, Y_1 - t) \in A_{x,y} \mid (X_1, Y_1) \in A_{t,t}\} \to$$

$$-\log G(x,y) = \nu(A_{x,y}) \quad (t \to \infty).$$

One can prove that also the following variant holds: for each Borel set C

$$P\{(X_1 - t, Y_1 - t) \in C \mid X_1 + Y_1 > t\} \to \nu(C) \quad (t \to \infty)$$

hence, in particular (cf. (22))

$$P\{X_1 - Y_1 \in B \mid X_1 + Y_1 > t\} \to \nu\{s,t)|s-t \in B\} = \pi(B)$$

for each Borel set B.

This shows how one can estimate π: Consider only those observations $\{(X_{i_k}, Y_{i_k})\}_{k=1}^{N_n}$ for which the sum of the components exceeds a certain level L_n i.e. $X_{i_k} + Y_{i_k} > L_n$ for all k where $\lim_{n\to\infty} L_n = \infty$. The empirical distribution function of $X_{i_1} - Y_{i_1}, \ldots, X_{i_N} - Y_{i_N}$ is the required estimate of the probability measure π (de Haan [8]).

This procedure has not yet been applied to the water-level data.

5. Some results

In this section we give some results of the theory above applied to 1577 extreme high-tide water-levels observed at the Dutch station Hoek van Holland during the winters 1887/88 until 1984/85 (about 100 years).

We want to estimate a quantile with exceedance probability $5 \cdot 10^{-4}$ per year. In the first step of the estimation procedure γ is estimated using both the

Pickands' estimator and the moment estimator for several values of k, the number of upper order statistics involved.

In Figure 1 estimates $\hat{\gamma}_n^P$ are plotted (solid curve) with for each the 95% confidence interval. At the left-hand side (k small), the curve fluctuates very much. This is because only few order statistics are involved here, moreover the extreme order statistics have the highest variance. The confidence intervals are accordingly very large. In the central part of the graph the curve seems to be more or less constant and at the right-hand side, when k is large, the bias takes over (remember that $k(n)$ must be $o(n)$, $n \to \infty$) and the curve decreases.

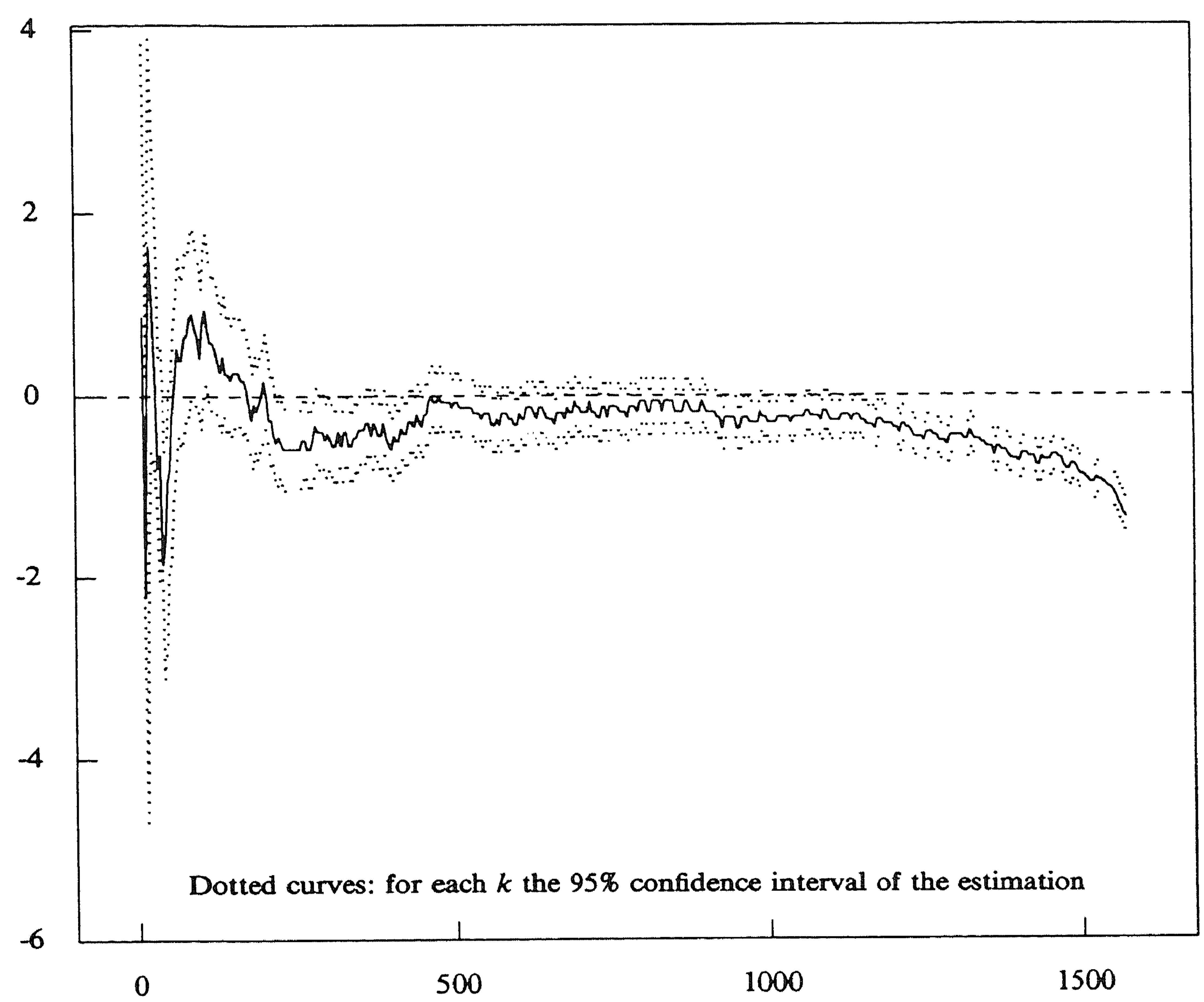

FIGURE 1. Pickands' estimator for γ against number of used upper order statistics

In Figure 2 the moment estimates are plotted with again the 95% confidence intervals. As one would expect the curve doesn't fluctuate as much as the previous one because all k upper order statistics are used for the estimations. The solid curve has many very small fluctuations due to the many ties in the (discrete) observations. At the right-hand end of the curve again the bias takes over.

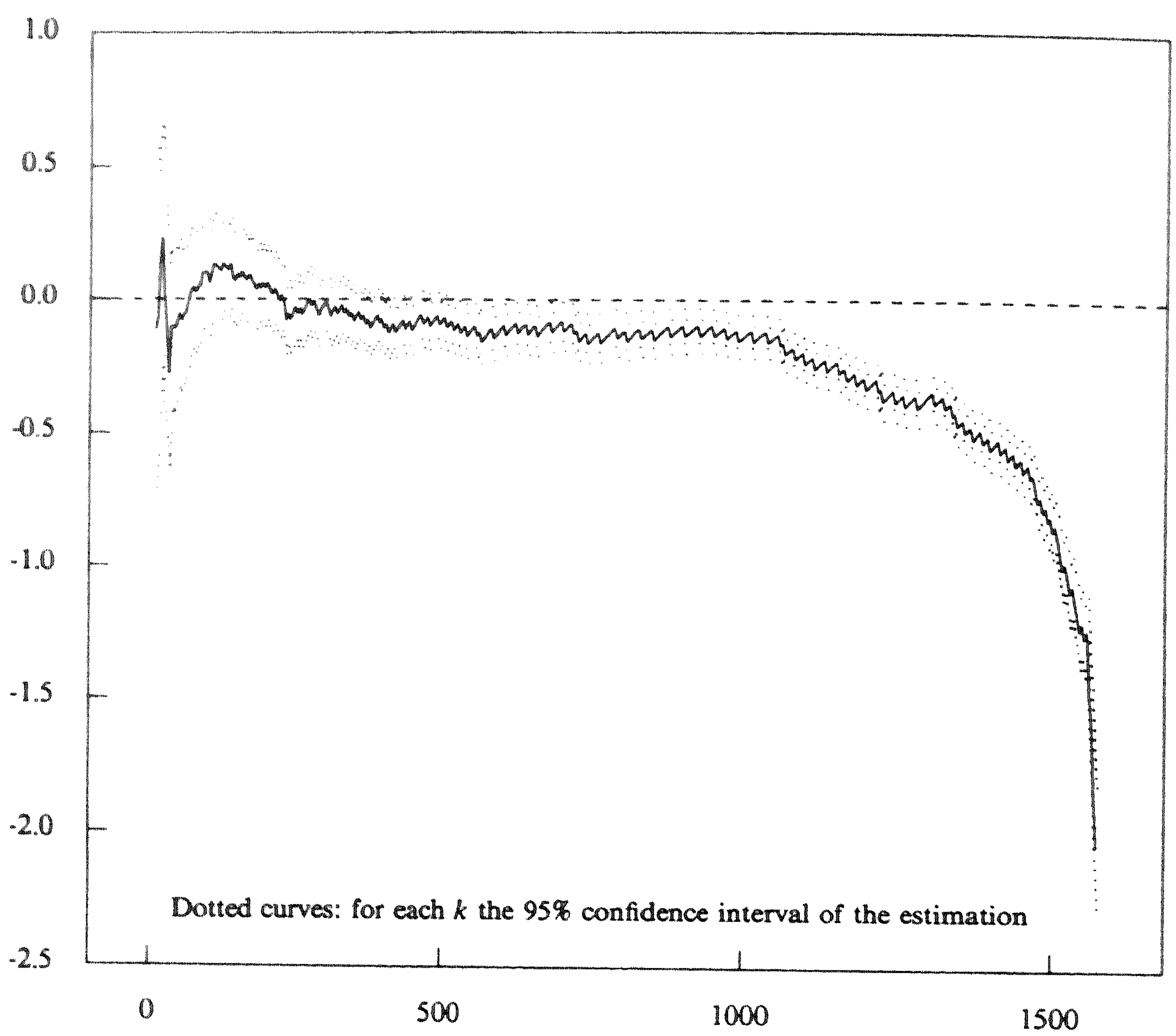

FIGURE 2. Moment estimator for γ against number of used upper order statistics

In Figure 3 the central part of the two mentioned curves has been displayed in more detail (fine vertical scales). It is difficult to decide from this graph what γ should be exactly, i.e., how many order statistics should be used. For k between 500 and 900 the two graphs seem to correspond more or less. The most important conclusion of these figures seems to be that probably γ is negative. This means that the underlying distribution function has a finite right endpoint. Note that any quantile of the extreme-value distributions G_γ is an increasing function of γ. So to play things safe, one can just put $\gamma=0$ and then estimate the quantile of the distribution.

In Figure 4 both the estimates for the $1-5.10^{-4}$ quantile (in cm.) are plotted against the number of used upper order statistics used, with for each k a 97.5% upper confidence bound (dashed curves). Here we assume $\gamma=0$. The solid curve (with a P) corresponds with $\hat{x}^P_{p_n}$ and the dotted curve (with a M) with $\hat{x}^M_{p_n}$. The estimator based on $\hat{\gamma}^M_n$ seems to be more stable than the other one, and they seem to correspond more or less for k between 100 and 200. It is remarkable that for larger k the estimates $\hat{x}^P_{p_n}$ are higher than the 97.5% upperbound confidence for $x^M_{p_n}$.

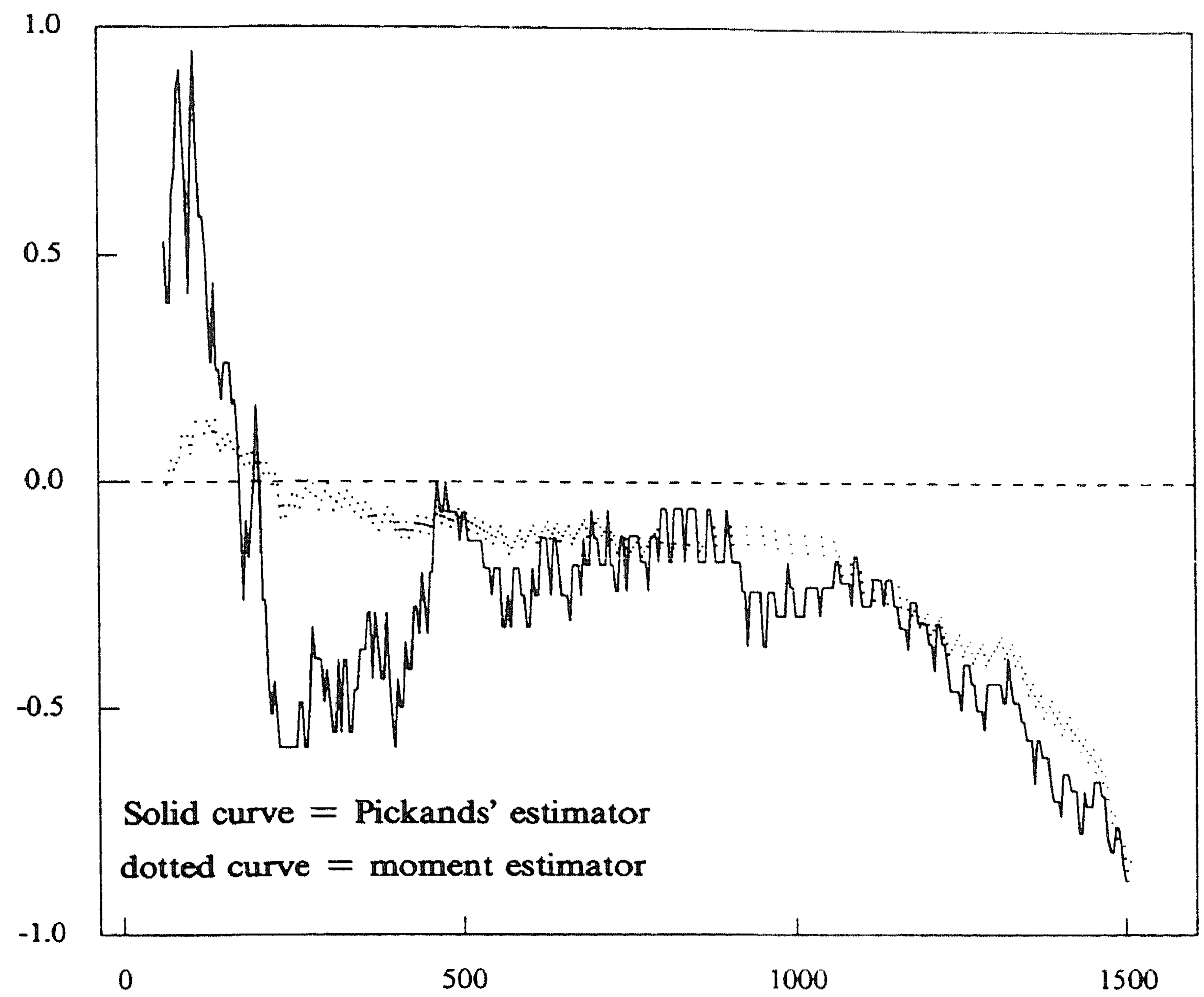

FIGURE 3. Pickands' and moment estimator for γ against number of used upper order statistics

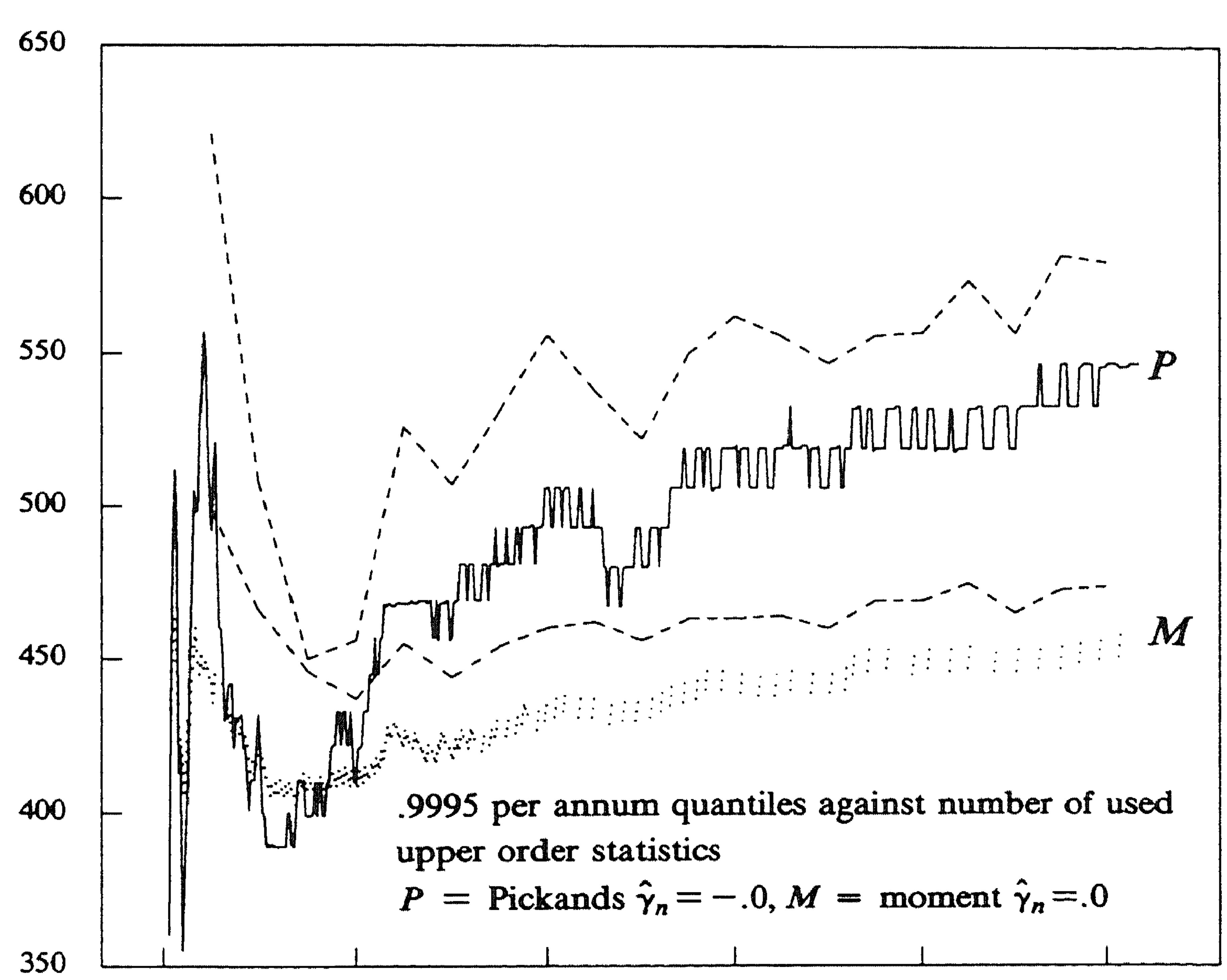

FIGURE 4. Hoek van Holland: high-tide water levels 1887/88...1984/85

In Figure 5 $\hat{x}^P_{p_n}$ and $\hat{x}^M_{p_n}$ are plotted with respectively $\hat{\gamma}^P_n = -.67$ and $\hat{\gamma}^M_n = -.70$ corresponding to $k = 504$ in Figure 3. The estimates are about 70 to 100 cm. lower than the ones in the previous figure and the estimates seem to fluctuate less. Note that both scales are different here.

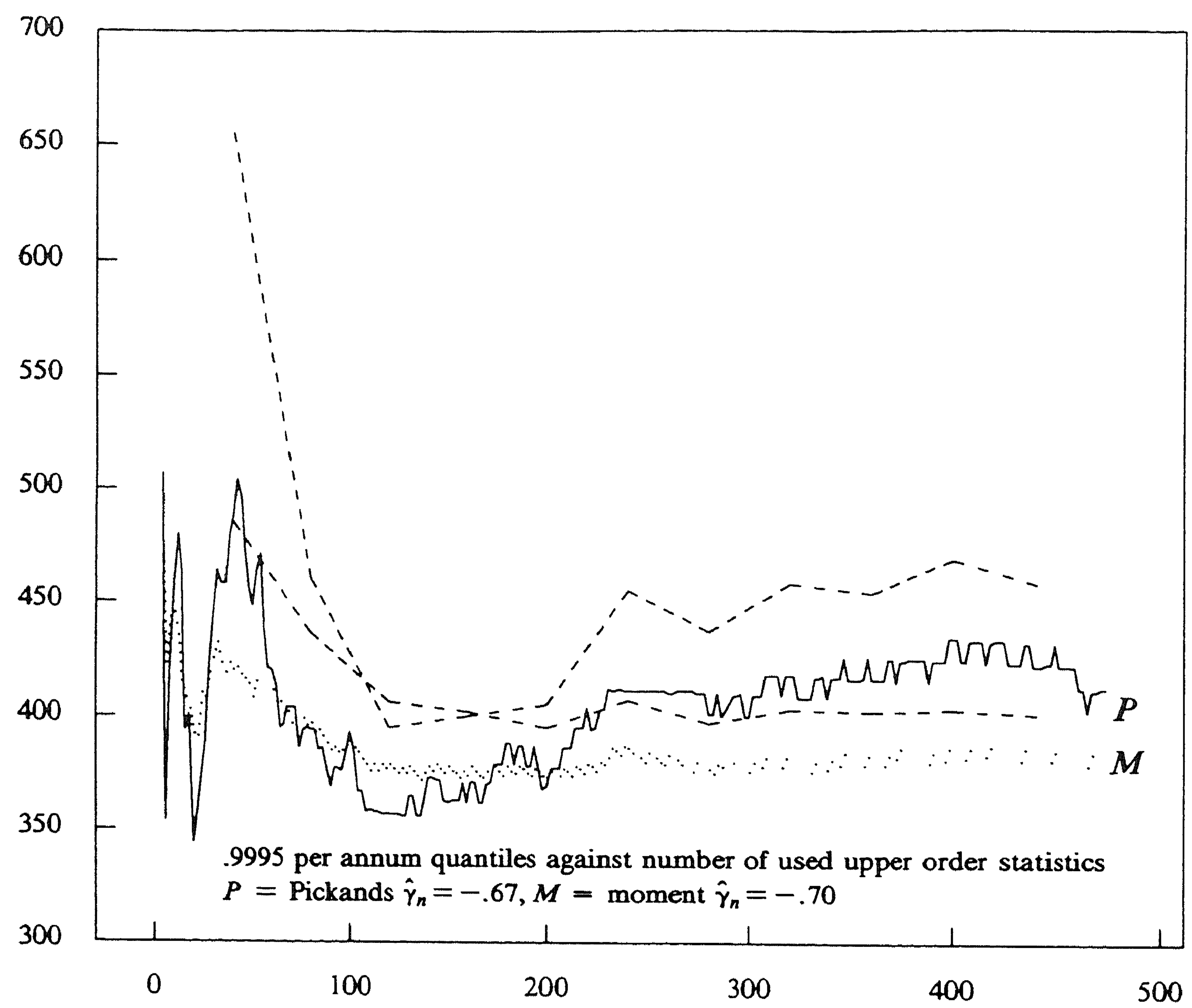

FIGURE 5. Hoek van Holland: high-tide water levels 1887/88...1984/85

ACKNOWLEDGEMENT

Arnold Dekkers wrote Section 5 and provided the numerical illustrations.

REFERENCES

1. A.A. BALKEMA and L. DE HAAN (1974). Residual life time at great age. *Ann. Prob. 2*, 792-804.
2. J.P. COHEN (1988). Fitting extreme value distributions to maxima. *Sankhyā, A, 50-1*, 74-97.
3. A.L.M. DEKKERS and L. DE HAAN (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics* (to appear).

4. A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* (to appear).
5. J.L. Geluk and L. de Haan (1987). *Regular Variation, Extensions and Tauberian Theorems,* CWI, Amsterdam.
6. B.V. Gnedenko (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Math. 44,* 423-453.
7. L. de Haan and S.I. Resnick (1977). Limit theory for multivariate sample extremals. *Z. Wahrsch. verw. Geb. 40,* 317-337.
8. L. de Haan (1985). Extremes in higher dimensions: the model and some statistics. A.C. Atkinson and S.E. Fienberg (eds.). *A Celebration of Statistics. The I.S.I. Centenary Volume, 26.1.* Amsterdam.
9. M.R. Leadbetter, G. Lindgren and H. Rootzén (1982). *Extremes and Related Properties of Random Sequences and Processes,* Springer, New York.
10. M.R. Leadbetter (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. verw. Geb. 65,* 291-306.
11. D. Mason (1982). Laws of large numbers for sums of extreme values. *Ann. Prob. 10,* 754-764.
12. G.L. O'Brien (1974). The maximum term of uniformly mixing stationary processes. *Z. Wahrsch. verw. Geb. 30,* 57-63.
13. J. Pickands III (1971). The two-dimensional Poisson process and external processes. *J. Appl. Probability 8,* 745-756.
14. J. Pickands III (1975). Statistical inference using extreme order statistics. *Ann. Statist. 3,* 119-131.
15. N.V. Smirnov (1949). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl. Ser. 1, no. 67.*
16. R.L. Smith (1987). Estimating tails of probability distributions. *Ann. Statist. 15,* 1174-1207.