17. Long Term Trends and Hurst Phenomena

From ancient times the Nile river region has been known for its peculiar long-term behavior: long periods of dryness followed by long periods of yearly floods. It seems historical records that go back as far as 622 AD also seem to support this trend. There were long periods where the high levels tended to stay high and other periods where low levels remained low[1].

An interesting question for hydrologists in this context is how to devise methods to regularize the flow of a river through reservoir so that the outflow is uniform, there is no overflow at any time, and in particular the capacity of the reservoir is ideally as full at time $t+t_0$ as at *t*. Let $\{y_i\}$ denote the annual inflows, and

$$s_n = y_i + y_2 + \cdots + y_n \qquad (17\text{-}1)$$

---

[1]A reference in the Bible says "*seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them*" (*Genesis*).

their cumulative inflow up to time n so that

$$\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{s_N}{N} \tag{17-2}$$

represents the overall average over a period N. Note that $\{y_i\}$ may as well represent the internet traffic at some specific local area network and $\bar{y}_N$ the average system load in some suitable time frame.

To study the long term behavior in such systems, define the "extermal" parameters

$$u_N = \max_{1\le n\le N}\{s_n - n\bar{y}_N\}, \tag{17-3}$$

$$v_N = \min_{1\le n\le N}\{s_n - n\bar{y}_N\}, \tag{17-4}$$

as well as the sample variance

$$D_N = \frac{1}{N}\sum_{n=1}^{N}(y_n - \bar{y}_N)^2. \tag{17-5}$$

In this case

$$R_N = u_N - v_N \tag{17-6}$$

defines the *adjusted range statistic* over the period N, and the dimensionless quantity

$$\frac{R_N}{\sqrt{D_N}} = \frac{u_N - v_N}{\sqrt{D_N}} \tag{17-7}$$

that represents the *readjusted range statistic* has been used extensively by hydrologists to investigate a variety of natural phenomena.

To understand the long term behavior of $R_N / \sqrt{D_N}$ where $y_i,\ i = 1, 2, \cdots N$ are independent identically distributed random variables with common mean μ and variance σ^2, note that for large N by the strong law of large numbers

$$s_n \xrightarrow{d} N(n\mu, n\sigma^2), \tag{17-8}$$

$$\bar{y}_N \xrightarrow{d} N(\mu, \sigma^2 / N) \to \mu \tag{17-9}$$

and

$$D_N \xrightarrow{d} \sigma^2 \tag{17-10}$$

with probability 1. Further with $n = Nt$, where $0 < t < 1$, we have

$$\lim_{N\to\infty} \frac{s_n - n\mu}{\sqrt{N}\sigma} = \lim_{N\to\infty} \frac{s_{\lfloor Nt \rfloor} - \lfloor Nt \rfloor \mu}{\sqrt{N}\sigma} \xrightarrow{d} B(t) \tag{17-11}$$

where $B(t)$ is the standard Brownian process with auto-correlation function given by $min\,(t_1, t_2)$.To make further progress note that

$$\begin{aligned} s_n - n\bar{y}_N &= s_n - n\mu - n(\bar{y}_N - \mu) \\ &= (s_n - n\mu) - \frac{n}{N}(s_N - N\mu) \end{aligned} \tag{17-12}$$

so that

$$\frac{s_n - n\bar{y}_N}{\sqrt{N}\sigma} = \frac{s_n - n\mu}{\sqrt{N}\sigma} - \frac{n}{N}\frac{s_N - N\mu}{\sqrt{N}\sigma} \xrightarrow{d} B(t) - tB(1), \quad 0 < \mathrm{t} < 1. \tag{17-13}$$

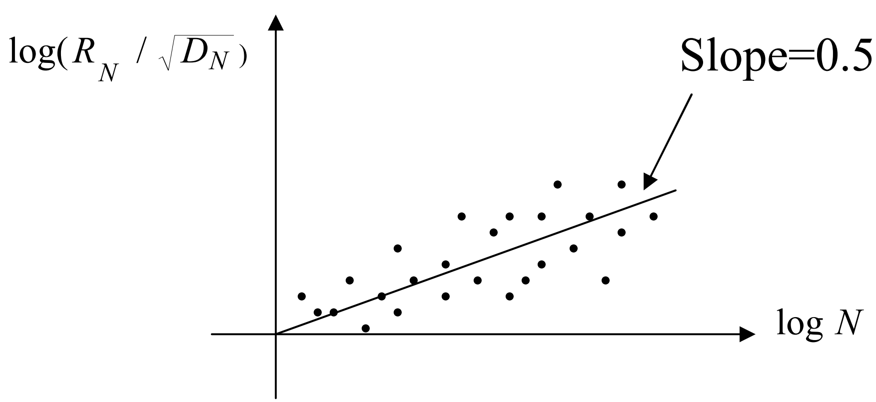
Hence by the functional central limit theorem, using (17-3) and (17-4) we get

$$\frac{u_N - v_N}{\sqrt{N}\sigma} \xrightarrow{d} \max_{0<t<1}\{B(t) - tB(1)\} - \min_{0<t<1}\{B(t) - tB(1)\} \equiv Q, \qquad (17\text{-}14)$$

where Q is a strictly positive random variable with finite variance. Together with (17-10) this gives

$$\frac{R_N}{\sqrt{D_N}} \to \frac{u_N - v_N}{\sigma} \xrightarrow{d} \sqrt{N}Q, \qquad (17\text{-}15)$$

a result due to Feller. Thus in the case of i.i.d. random variables the rescaled range statistic $R_N / \sqrt{D_N}$ is of the order of $O(N^{1/2})$. It follows that the plot of $\log(R_N / \sqrt{D_N})$ versus $\log N$ should be linear with slope $H = 0.5$ for independent and identically distributed observations.



The hydrologist Harold Erwin Hurst (1951) generated tremendous interest when he published results based on water level data that he analyzed for regions of the Nile river which showed that Plots of $\log(R_N / \sqrt{D_N})$ versus log N are linear with slope $H \approx 0.75$. According to Feller's analysis this must be an anomaly if the flows are i.i.d. with finite second moment.

The basic problem raised by Hurst was to identify circumstances under which one may obtain an exponent $H > 1/2$ for N in (17-15). The first positive result in this context was obtained by Mandelbrot and Van Ness (1968) who obtained $H > 1/2$ under a strongly dependent stationary Gaussian model. The Hurst effect appears for independent and non-stationary flows with finite second moment also. In particular, when an appropriate slow-trend is superimposed on a sequence of i.i.d. random variables the Hurst phenomenon reappears. To see this, we define the Hurst exponent fora data set to be H if

$$\frac{R_N}{\sqrt{D_N}\,N^H} \xrightarrow{d} Q, \qquad N \to \infty, \tag{17-16}$$

where Q is a nonzero real valued random variable.

IID with slow Trend

Let $\{X_n\}$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2, and g_n be an arbitrary real valued function on the set of positive integers setting a deterministic trend, so that

$$y_n = x_n + g_n \tag{17-17}$$

represents the actual observations. Then the partial sum in (17-1) becomes

$$\begin{aligned} s_n &= y_1 + y_2 + \cdots + y_n = x_1 + x_2 + \cdots + x_n + \sum_{i=1}^{n} g_i \\ &= n\,(\bar{x}_n + \bar{g}_n) \end{aligned} \tag{17-18}$$

where $\bar{g}_n = 1/n \sum_{i=1}^{n} g_i$ represents the running mean of the slow trend. From (17-5) and (17-17), we obtain

$$
\begin{aligned}
D_N &= \frac{1}{N}\sum_{n=1}^{N}(y_n - \bar{y}_N)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x}_N)^2 + \frac{1}{N}\sum_{n=1}^{N}(g_n - \bar{g}_N)^2 + \frac{2}{N}\sum_{n=1}^{N}(x_n - \bar{x}_N)(g_n - \bar{g}_N) \\
&= \hat{\sigma}_X^2 + \frac{1}{N}\sum_{n=1}^{N}(g_n - \bar{g}_N)^2 + \frac{2}{N}\sum_{n=1}^{N}(x_n - \bar{x}_N)(g_n - \bar{g}_N). \qquad (17\text{-}19)
\end{aligned}
$$

Since $\{x_n\}$ are i.i.d. random variables, from (17-10) we get $\hat{\sigma}_X^2 \xrightarrow{d} \sigma^2$. Further suppose that the deterministic sequence $\{g_n\}$ converges to a finite limit c. Then their Caesaro means $\frac{1}{N}\sum_{n=1}^{N} g_n = \bar{g}_N$ also converges to c. Since

$$
\frac{1}{N}\sum_{n=1}^{N}(g_n - \bar{g}_N)^2 = \frac{1}{N}\sum_{n=1}^{N}(g_n - c)^2 - (\bar{g}_N - c)^2, \qquad (17\text{-}20)
$$

applying the above argument to the sequence $(g_n - c)^2$ and $(\bar{g}_N - c)^2$ we get (17-20) converges to zero. Similarly, since

$$\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x}_N)(g_n-\bar{g}_N)=\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(g_n-c)-(\bar{x}_N-\mu)(\bar{g}_N-c), \tag{17-21}$$

by Schwarz inequality, the first term becomes

$$\left|\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(g_n-c)\right|^2 \le \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2\,\frac{1}{N}\sum_{n=1}^{N}(g_n-c)^2. \tag{17-22}$$

But $\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 \to \sigma^2$ and the Caesaro means $\frac{1}{N}\sum_{n=1}^{N}(g_n-c)^2 \to 0.$ Hence the first term (17-21) tends to zero as $N\to\infty,$ and so does the second term there. Using these results in (17-19), we get

$$g_n \to c \quad \Rightarrow \quad D_N \xrightarrow{d} \sigma^2. \tag{17-23}$$

To make further progress, observe that

$$\begin{aligned} u_N &= \max\{s_n - n\bar{g}_N\} \\ &= \max\{n(\bar{x}_n-\bar{x}_N)+n(\bar{g}_n-\bar{g}_N)\} \\ &\le \max_{0<n<N}\{n(\bar{x}_n-\bar{x}_N)\}+\max_{0<n<N}\{n(\bar{g}_n-\bar{g}_N)\} \end{aligned} \tag{17-24}$$

and

$$
\begin{aligned}
v_N &= \min\{s_n - n\bar{g}_N\} \\
&= \min\{n(\bar{x}_n - \bar{x}_N) + n(\bar{g}_n - \bar{g}_N)\} \\
&\geq \min_{0<n<N}\{n(\bar{x}_n - \bar{x}_N)\} + \min_{0<n<N}\{n(\bar{g}_n - \bar{g}_N)\}.
\end{aligned} \tag{17-25}
$$

Consequently, if we let

$$
r_N = \max_{0<n<N}\{n(\bar{x}_n - \bar{x}_N)\} - \min_{0<n<N}\{n(\bar{x}_n - \bar{x}_N)\} \tag{17-26}
$$

for the i.i.d. random variables, then from (17-6),(17-24) and (17-25) (17-26), we obtain

$$
R_N = u_N - v_N \leq r_N + G_N \tag{17-27}
$$

where

$$
G_N = \max_{0<n<N}\{n(\bar{g}_n - \bar{g}_N)\} - \min_{0<n<N}\{n(\bar{g}_n - \bar{g}_N)\} \tag{17-28}
$$

From (17-24) – (17-25), we also obtain

$$u_N \geq \min_{0<n<N}\{n\,(\bar{x}_n - \bar{x}_N)\} + \max_{0<n<N}\{n\,(\bar{g}_n - \bar{g}_N)\}, \qquad (17\text{-}29)$$

$$v_N \leq \max_{0<n<N}\{n\,(\bar{x}_n - \bar{x}_N)\} + \min_{0<n<N}\{n\,(\bar{g}_n - \bar{g}_N)\}, \qquad (17\text{-}30)$$

[use $\max_i\{(x_i + y_i)\} \geq \max_i\{(\min_i x_i) + y_i\} = \min_i(x_i) + \max_i(y_i)$] and hence

$$R_N \geq G_N - r_N. \qquad (17\text{-}31)$$

From (17-27) and (17-31) we get the useful estimates

$$|R_N - G_N| \leq r_N,$$

and $$(17\text{-}32)$$

$$|R_N - r_N| \leq G_N. \qquad (17\text{-}33)$$

Since $\{x_n\}$ are i.i.d. random variables, using (17-15) in (17-26) we get

$$\frac{r_N}{\hat{\sigma}_X^2\sqrt{N}} \to \frac{r_N}{\sigma\sqrt{N}} \to Q, \quad \textit{in probability} \qquad (17\text{-}34)$$

a positive random variable, so that

$$\frac{r_N}{\sigma} \to \sqrt{N}Q \quad \textit{in probability.} \qquad (17\text{-}35)$$

Consequently for the sequence $\{y_n\}$ in (17-17) using (17-23) in (17-32)-(17-34) we get

$$\frac{|R_N - G_N|}{\sqrt{D_N}\,N^H} \to \frac{r_N}{\sigma\ N^H} \to \frac{Q/\sigma}{N^{H-1/2}} \to 0 \qquad (17\text{-}36)$$

if $H > 1/2$. To summarize, if the slow trend $\{g_n\}$ converges to a finite limit, then for the observed sequence $\{y_n\}$, for every H > 1/2

$$\left| \frac{R_N}{\sqrt{D_N}\,N^H} - \frac{G_N}{\sqrt{D_N}\,N^H} \right| \to \left| \frac{R_N}{\sqrt{D_N}\,N^H} - \frac{G_N}{\sigma\ N^H} \right| \to 0 \qquad (17\text{-}37)$$

in probability as $N \to \infty.$

In particular it follows from (17-16) and (17-36)-(17-37) that the Hurst exponent $H > 1/2$ holds for a sequence $\{y_n\}$ if and only if the slow trend sequence $\{g_n\}$ satisfies

$$\lim_{N\to\infty} \frac{G_N}{N^H} = c_0 > 0, \qquad H > 1/2. \qquad (17\text{-}38)$$

In that case from (17-37), for that H > 1/2 we obtain

$$\frac{R_N}{\sqrt{D_N}\,N^H} \to c_0/\sigma \quad \textit{in probability as} \quad N \to \infty, \tag{17-39}$$

where c_0 is a positive number.

Thus if the slow trend $\{g_n\}$ satisfies (17-38) for some H > 1/2, then from (17-39)

$$\log \frac{R_N}{\sqrt{D_N}} \to H \log N + c, \quad \text{as} \quad N \to \infty. \tag{17-40}$$

Example: Consider the observations

$$y_n = x_n + a + bn^\alpha, \quad n \ge 1 \tag{17-41}$$

where x_n are i.i.d. random variables. Here $g_n = a + bn^\alpha$, and the sequence converges to a for $\alpha < 0$, so that the above result applies. Let

$$M_n = n\,(\bar{g}_n - \bar{g}_N) = b\left(\sum_{k=1}^{n} k^\alpha - \frac{n}{N} \sum_{k=1}^{N} k^\alpha \right). \tag{17-42}$$

To obtain its *max* and *min*, notice that

$$M_n - M_{n-1} = b\left(n^\alpha - \frac{1}{N}\sum_{k=1}^{N} k^\alpha\right) > 0$$

if $n < (\frac{1}{N}\sum_{k=1}^{N} k^\alpha)^{1/\alpha}$, and negative otherwise. Thus max M_N is achieved at

$$n_0 = \left(\frac{1}{N}\sum_{k=1}^{N} k^\alpha\right)^{1/\alpha} \qquad (17\text{-}43)$$

and the minimum of $M_N = 0$ is attained at N=0. Hence from (17-28) and (17-42)-(17-43)

$$G_N = b\left(\sum_{k=1}^{n_0} k^\alpha - \frac{n_0}{N}\sum_{k=1}^{N} k^\alpha\right). \qquad (17\text{-}44)$$

Now using the Reimann sum approximation, we may write

$$\frac{1}{N}\sum_{k=1}^{N} k^{\alpha} \approx \frac{1}{N}\int_0^N x^{\alpha}\,dx = \begin{cases} (1+\alpha)^{-1} N^{\alpha}, & \alpha > -1 \\ \dfrac{\log N}{N}, & \alpha = -1 \\ \dfrac{\sum_{k=1}^{\infty} k^{\alpha}}{N}, & \alpha < -1 \end{cases} \tag{17-45}$$

so that

$$n_0 \approx \begin{cases} (1+\alpha)^{-1/\alpha} N, & \alpha > -1 \\ \dfrac{N}{\log N}, & \alpha = -1 \\ \dfrac{\sum_{k=1}^{\infty} k^{\alpha}}{N^{1/\alpha}}, & \alpha < -1 \end{cases} \tag{17-46}$$

and using (17-45)-(17-46) repeatedly in (17-44) we obtain

$$G_n = bn_0\left(\frac{1}{n_0}\sum_{k=1}^{n_0} k^\alpha - \frac{1}{N}\sum_{k=1}^{N} k^\alpha\right)$$

$$\approx \begin{cases} \dfrac{bn_0}{1+\alpha}(n_0^\alpha - N^\alpha) \approx c_1 N^{1+\alpha}, & \alpha > -1 \\ bn_0\left(\dfrac{1}{n_0}\log n_0 - \dfrac{1}{N}\log N\right) \approx c_2 \log N, & \alpha = -1 \\ b\left(1-\dfrac{n_0}{N}\right)\left(\displaystyle\sum_{k=1}^{\infty} k^\alpha\right) \approx c_3, & \alpha < -1 \end{cases} \qquad (17\text{-}47)$$

where c_1, c_2, c_3 are positive constants independent of N. From (17-47), notice that if $-1/2 < \alpha < 0$, then

$$G_n \sim c_1 N^H,$$

where $1/2 < H < 1$ and hence (17-38) is satisfied. In that case

$$\frac{R_N}{\sqrt{D_N}\,N^{(1+\alpha)}} \to c_1 \quad \textit{in probability as} \quad N \to \infty. \tag{17-48}$$

and the Hurst exponent $H = 1 + \alpha > 1/2.$

Next consider $\alpha < -1/2.$ In that case from the entries in (17-47) we get $G_N = o(N^{1/2}),$ and diving both sides of (17-33) with $\sqrt{D_N}\,N^{1/2},$

$$\frac{|R_N - r_N|}{\sqrt{D_N}\,N^{1/2}} \sim \frac{o(N^{1/2})}{\sigma\ N^{1/2}} \to 0 \quad \textit{in probability}$$

so that

$$\frac{R_N}{\sqrt{D_N}\,N^{1/2}} \sim \frac{r_N}{\sigma\ N^{1/2}} \to Q \tag{17-49}$$

where the last step follows from (17-15) that is valid for i.i.d. observations. Hence using a limiting argument the Hurst exponent

$H = 1/2$ if $\alpha \le -1/2$. Notice that $\alpha = 0$ gives rise to i.i.d. observations, and the Hurst exponent in that case is 1/2. Finally for $\alpha > 0$, the slow trend sequence $\{g_n\}$ does not converge and (17-36)-(17-40) does not apply. However direct calculation shows that D_N in (17-19) is dominated by the second term which for large N can be approximated as $\frac{1}{N}\int_0^N x^{2\alpha} \approx N^{2\alpha}$ so that

$$\sqrt{D_N} \to c_4 N^{\alpha} \qquad as \qquad N \to \infty \tag{17-50}$$

From (17-32)

$$\frac{|R_N - G_N|}{\sqrt{D_N}\,N} \approx \frac{r_N}{\sqrt{D_N}\,N} \to \frac{\sqrt{N}\sigma\ Q}{c_4 N^{1+\alpha}} \to 0$$

where the last step follows from (17-34)-(17-35). Hence for $\alpha > 0$ from (17-47) and (17-50)

$$\frac{R_N}{\sqrt{D_N}\,N} \approx \frac{G_N}{\sqrt{D_N}\,N} \approx \frac{c_1 N^{1+\alpha}}{c_4 N^{1+\alpha}} \to \frac{c_1}{c_4} \tag{17-51}$$

as $N \to \infty$. Hence the Hurst exponent is 1 if $\alpha > 0$. In summary,

$$H(\alpha) = \begin{cases} 1 & \alpha > 0 \\ 1/2 & \alpha = 0 \\ 1+\alpha & 0 > \alpha > -1/2 \\ 1/2 & \alpha < -1/2 \end{cases} \tag{17-52}$$

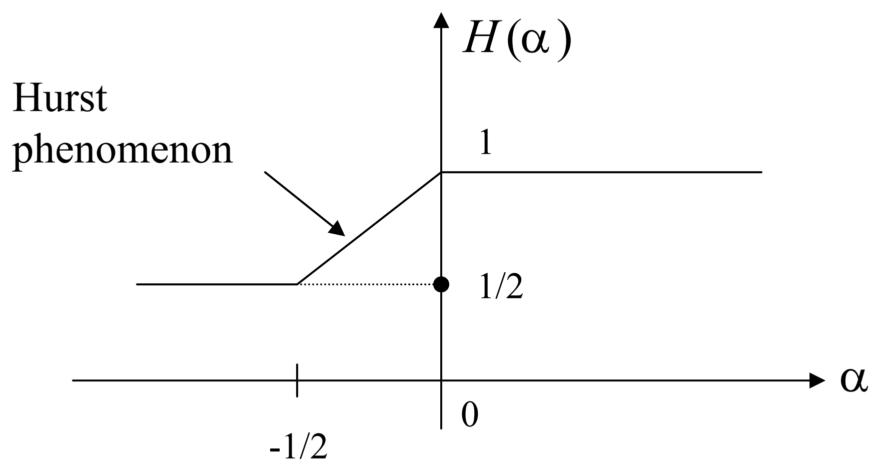


Fig.1 Hurst exponent for a process with superimposed slow trend