the Hurst Exponent ... and financial stuff

motivated by e-mail from Carl P.

Once upon a time, a British government bureaucrat named Harold Edwin Hurst studied 800 years of records of the Nile's flooding.
He noticed that there was a tendency for a high flood year to be followed by another high flood year, and for a low flood year to be followed by another low flood year.

Was that accidental ... or was there *really* some correlation between levels?
Did the height at year 5 have an effect on the height in year 6?

>Are you talking about river levels ... or financial stuff?
Patience.

To analyze, we might do something like this:

1. Note the heights of the n flood levels:
   h(1), **h**(2), ... **h**(n)
2. Let **m** be the **M**ean of these levels:
   M = (1/n) [**h**(1)+**h**(2)+...+**h**(n)]
   Calculate the deviations from the mean:
   x(1) = **h**(1) - **M**
   x(2) = **h**(2) - **M**
   ...
   x(n) = **h**(n) - **M**
   Note that the set of **x**s have zero mean.
   Positive **x**'s indicate that the Nile level was above the average.
3. Now calculate the Sums:
   Y(1) = **x**(1)
   Y(2) = **x**(1) + **x**(2)
   ...
   Y(n) = **x**(1) + **x**(2) + ...+ **x**(n)
   Note that the set of partial sums, the **Y**'s, are sums of zero-mean variables.
   They will be positive if there's a preponderance of positive **x**'s.
   Note, too, that Y(k) = Y(k-1) + x(k).
4. Let **R**(n) = MAX[**Y**(k)] - MIN[**Y**(k)]
   This difference between the maximum and minimum of the n values is called the **Range**
5. Let **s**(n) be the standard deviation of the set of n **h**-values.

As it turns out, the probability theorist **William Feller** proved that if a series of random variables (like the **x**'s) had finite standard deviation and were independent, then the so-called **R/s** statistic (formed over n observations) would increase in proportion to $n^{1/2}$ (for large values of n).

>Huh? The so-called R/s statistic?
Yes. Apparently lots of people are interested in this animal. (See this **PDF stuff**)
This guy, **R/s**, is called the **rescaled range**

Anyway, we now have:
$\mathbf{R}(n) / \mathbf{s}(n) \sim k\mathbf{n}^{1/2}$ ... where k is some constant

If that were true, then we'd expect that:

log(**R**/**s**) ~ log(k) + (1/2) log(**n**)

So, if we were to plot log(**R**/**s**) *vs* log(**n**), we'd expect it to be approximately a straight line with slope (1/2).

>A logarithm to what base?
It doesn't matter.

Anyway, what Hurst apparently found, was that the plot had a slope closer to 0.7 (rather than 0.5).

>So, what's that mean?
I guess it means that the annual Nile levels weren't independent, but this year's level might be expected to affect next year's level.
Indeed, if the slope of the log(**R** / **s**) *vs* log(**n**) "best fit line" is **H**, then we'd expect:

$\mathbf{R} / \mathbf{s} \sim \mathrm{k}\mathbf{n}^{\mathbf{H}}$

>Don't tell me! That **H** is the Hurst Exponent, right?
You got it.

>So what's it got to do with financial stuff?
Patience.
>So where's the spreadsheet?
Patience.
The interesting thing is that many things seem to exhibit this long term patterns or dependence ... seven years of plenty followed by seven years of plenty.
>Sounds like a biblical reference.
Yes. It's called the **Joseph Effect**
>Do you realize that don't have a single picture? A picture is worth a thousand ...

Hurst Examples

Okay, let's look at 300 daily returns for Exxon stock.
We'll call them h(1), h(2), ... h(300).
We calculate the Mean of these 300 returns. We'll call it **M**.
M = (1/300) [h(1) + h(2) + ... + h(300)]
Then we calculate x(1), x(2), ... x(300), the 300 deviations from the Mean:
x(1) = h(1) - M, x(2) = h(2) - M, ... x(300) = h(300) - M.
These devations are shown in green, in Figure 1.
(The average of these deviations is zero!)

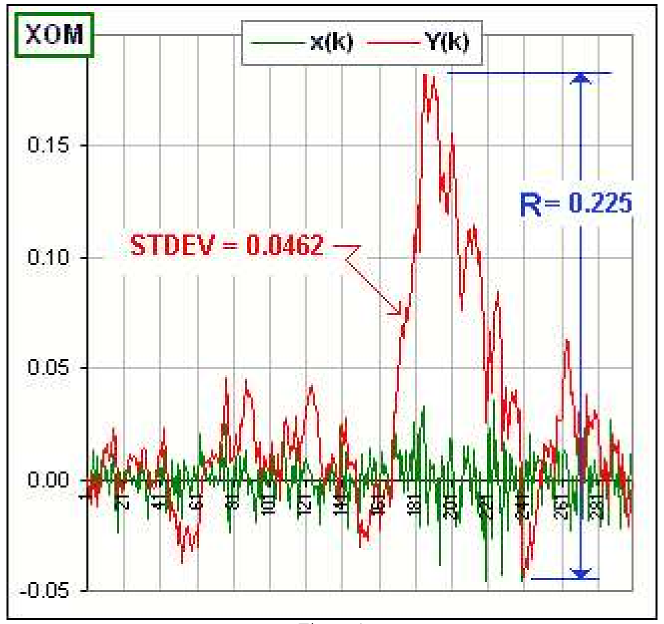

Figure 1

Now we calculate the Y's::
Y(1) = x(1), Y(2) = x(1)+x(2), ... Y(300) = x(1)+x(2)+...x(300).
The Y's are shown in red, in Figure 1.

Now we find the maximum Y and the minimum Y and subtract them.
That's the Range, **R** = Max[Y] - Min[Y] ... in blue.
Finally we calculate the Standard Deviation of the h's:
s = STDEV[h(k)]

>And you get a Hurst exponent ... somehow?

Okay, from the above scheme we note two magic numbers:
n = 300, **R** / **s** = 0.225 / 0.0462 = 4.87.

That'll give us one point on our log(R/s) *vs* log(n) chart, namely
log(300) = **5.70** and log(4.87) = **1.58**.

Now we repeat the above scheme for 310 points, then 320 points *etc. etc.*, each time generating a point on our chart, and ...

>Just give us the chart, okay?
Okay, see Figure 2
You see our first point? We calculated points up to n = 550.

>And the Hurst exponent is ... uh, the slope?
Yes. At least it's an *estimate* of the Hurst Exponent **H** = 0.478.

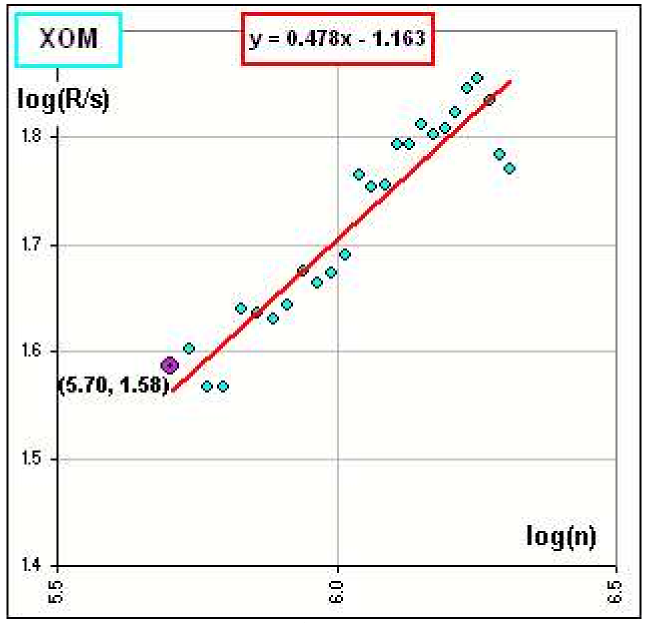

Figure 2

>Pretty close to 1/2, eh?
Yes, and that'd imply that daily returns for XOM are random, uncorrelated, a Brownian motion, independent ...

>Yeah, yeah. Do you always get that?
Patience ...

for Part II

For a great read on Hurst stuff, see **bearcave.com**