Predictability of Time Series

White Paper

Predictability of Time Series

White Paper

July, 2012

www.mu-sigma.com

Predictability of Time Series

A statistical measure used to classify time series and infer the level of difficulty in predicting and choosing an appropriate model for the series at hand.

In 1951 the celebrated British hydrologist H.E. Hurst published a paper titled, "The Long-Term Storage Capacity of Reservoirs." The paper dealt specifically with the modeling of reservoirs, but as it turned out, the results also held valid for a number of other natural systems. Hurst was looking for a way to model the levels of the river Nile so that architects could construct an appropriately sized reservoir system. While his recommendations were not implemented (the 1952 Egyptian revolutions saw to that), he gave life to a statistical methodology for distinguishing random from non-random systems and to identify the persistence of trends, a methodology known as Rescaled Range analysis or R/S analysis.

Many years later, while investigating the fractal nature of financial markets – specifically, the tendency of a time series to regress strongly to its mean or to cluster in a direction – noted mathematician Benoit Mandelbrot happened to stumble across Hurst's work and, recognizing the potential therein, introduced to fractal geometry, in Hurst's honor, the term Generalized Hurst Exponent. Put simply, the Hurst exponent is used as a measure of the long-term memory of a time series.

In addition to the Hurst exponent, Mandelbrot also coined two more terms useful in describing the long-term memory of a time series. He called the first one the Joseph Effect and the second one the Noah Effect. The Joseph Effect tells us whether movements in a time series are part of a long-term trend and refers to the Old Testament where Egypt would experience seven years of rich harvest followed by seven years of famine. The Noah Effect is the tendency of a time series to have abrupt changes and the name is derived from the biblical story of the Great Flood. Both of these effects in a time series can be inferred from the Hurst exponent.

Estimating the Hurst exponent

The Hurst exponent is not so much calculated as it is estimated. A variety of techniques exist for estimating the Hurst exponent (H) and the process detailed here is both simple and highly data intensive. To estimate the Hurst exponent one must regress the rescaled range on the time span of observations. To do this, a time series of full length is divided into a number of shorter time series and the rescaled range is calculated for each of the smaller time series. A minimum length of eight is usually chosen for the length of the smallest time series. So, for example, if a time series has 128 observations it is divided into:

- two chunks of 64 observations each
- four chunks of 32 observations each
- eight chunks of 16 observations each
- 16 chunks of eight observations each

Steps for estimating the Hurst exponent after breaking the time series into chunks:

For each chunk of observations, compute:

- the mean of the time series,
- a mean centered series by subtracting the mean from the series,
- the cumulative deviation of the series from the mean by summing up the mean centered values,
- the Range (R), which is the difference between the maximum value of the cumulative deviation and the minimum value of the cumulative deviation,
- the standard deviation (S) of the mean centered values, and
- the rescaled range by dividing the Range by the standard deviation.

Finally, average the rescaled range over all the chunks.

The rescaled range and chunk size follows a power law, and the Hurst exponent is given by the exponent of this power law. When the frequency of an event varies as the power of some quantity associated with the event, it is said to follow a power law. A wide variety of natural and manmade phenomena follow a power law. For example, the 80/20 rule (20 percent of the population holds 80 percent of wealth), the winner-take-all phenomenon, friend connections in a social network and forest fires all follow power laws.

Interpreting the Hurst Exponent

Using the Hurst exponent we can classify time series into types and gain some insight into their dynamics. Here are some types of time series and the Hurst exponents associated with each of them.

A Brownian time series: In a Brownian time series (also known as a random walk or a drunkard's walk) there is no correlation between the observations and a future observation; being higher or lower than the current observation are equally likely. Series of this kind are hard to predict. Figure 1 provides an example of a Brownian time series and its estimated Hurst exponent. The Hurst exponent for the data plotted above was estimated to be 0.53 - a Hurst exponent close to 0.5 is indicative of a Brownian time series.

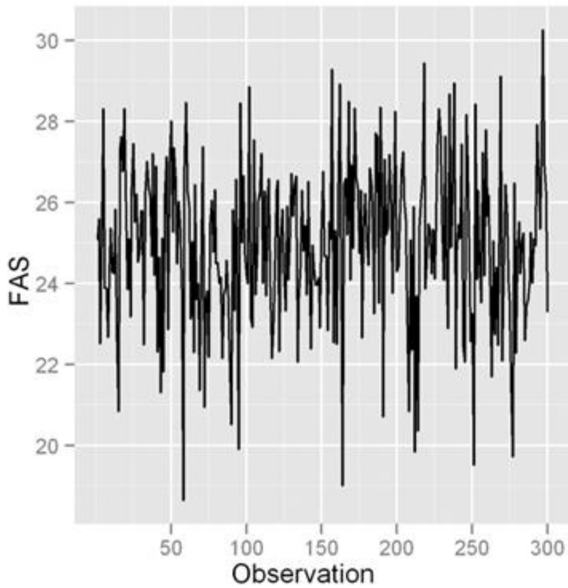

Figure 1: A Brownian time series (H = 0.53).

An anti-persistent time series: In an anti-persistent time series (also known as a mean-reverting series) an increase will most likely be followed by a decrease or vice-versa (i.e., values will tend to revert to a mean). This means that future values have a tendency to return to a long-term mean. Figure 2 provides an example of an anti-persistent time series and its estimated Hurst exponent.

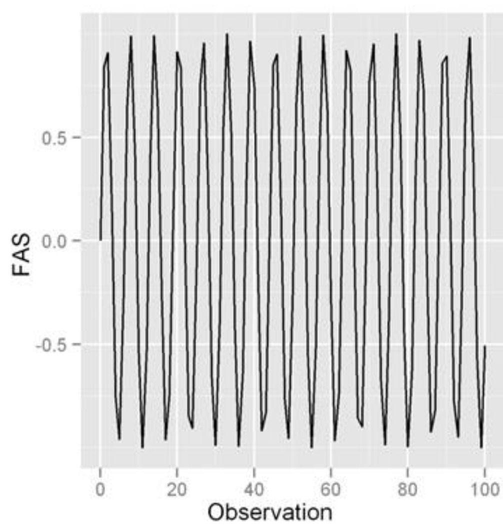

Figure 2: An anti persistent time series (H = 0.043).

The Hurst exponent for the data plotted above was estimated to be 0.043.A Hurst exponent value between 0 and 0.5 is indicative of anti-persistent behavior and the closer the value is to 0, the stronger is the tendency for the time series to revert to its long-term means value.

A persistent time series: In a persistent time series an increase in values will most likely be followed by an increase in the short term and a decrease in values will most likely be followed by another decrease in the short term. Figure 3 provides an example of a persistent time series and its estimated Hurst exponent.

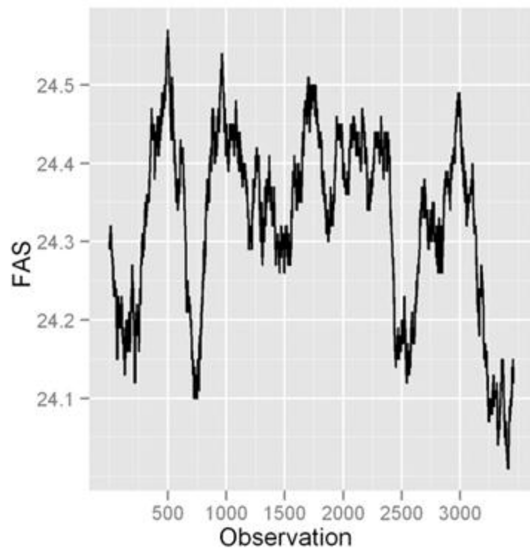

Figure 3: A persistent time series (H = 0.95).

The plot shows the intra-day tick level data for an NYSE traded fund. The Hurst exponent was estimated to be 0.95, which indicates a persistent time series. A Hurst exponent value between 0.5 and 1.0 indicates persistent behavior; the larger the H value the stronger the trend.

Conclusion

The Hurst exponent is a useful statistical method for inferring the properties of a time series without making assumptions about stationarity. It is most useful when used in conjunction with other techniques, and has been applied in a wide range of industries. For example the Hurst exponent is paired with technical indicators to make decisions about trading securities in financial markets; and it is used extensively in the healthcare industry, where it is paired with machine-learning techniques to monitor EEG signals. The Hurst exponent can even be applied in ecology, where it is used to model populations.